ICONS IN ACTION: REDEFINING ICONICITY FOR THE COGNITIVE SCIENCES

Wim Pouw
Donders Institute for Brain, Cognition, and Behaviour
Radboud University Nijmegen

Forschungskolleg Humanwissenschaften
Bad Homburg
wim.pouw@donders.ru.nl

September 16, 2024

ABSTRACT

Iconicity is a term used in cognitive science and gesture studies to denote an informative relation between the form of an utterance and the meaning of that utterance. With good iconic design, the form of an utterance can directly invite a suitable perceiver with a certain degree of initiation, to grasp a meaning in the right direction. Despite the now increasingly touted importance of iconicity for understanding human languages, it proves difficult to define more formally. When the term is defined, researchers tend to base iconicity on resemblances, such that A is iconic of B, if A resembles B in some relevant respect. In the philosophy of depiction fundamental issues have been raised against resemblance-based accounts. Even when barring such metaphysical issues, it has recently been argued that for all practical research purposes, a 'state-of-the-art' definition of iconicity should also do away with resemblances. Instead iconicity is in the eye of the beholder (mind-to-world relation) as opposed to a property of the environment (world-to-mind relation). In this paper I suggest for all practical purposes that there is an alternative explanatory route available to us, which is paved more broadly by 4E approaches (embodied, embedded, enactive, extended) and the philosophy of depiction. Taking this road should lead to a "distributed" view, where iconicity arises in a niche-constructed organism-environment system. This paper provides the bare bones for such a view, broadening the discussion, and hopefully drawing other perspectives in so that a distributed view of iconicity can become fully fleshed out.

***K*eywords** Iconicity · Cognitive science · 4E cognition · Philosophy of depiction · Representation

1 Introduction

In cognitive psychology, anthropology, linguistics, and semiotics, the term iconicity is often used to describe a relation between the form and the meaning of a signed, spoken, and/or gestured utterance. These utterances can be whole-body and gestural, such as conveying the concept of a tiger by showing one's teeth and making claw-like fingers. Or iconicity may be acoustic in nature - growling our way to the meaning. Iconicity includes tracing gestures where one conveys being home by outlining a house with two pointing fingers. Or, it may include enactive and almost pre-tense-like utterances, such as opening a visibly absent front door and expressing an enacted happiness for being home. Even in highly conventionalized systems of expression, such as lexical items in signed languages or spoken languages, iconicity is often present to do semantic work (Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015). Iconicity as such is a mixed bag of semiotic modes of expression (Ortega & Özyürek, 2020; Streeck, 2008).

How to define such a mixed bag? The iconic relation is not an arbitrary relation. It is not just any utterance that can convey or constrain a meaning iconically. To a matter of degree, the form of the iconic utterance is constrained to relate to its meaning. Importantly, it is not simply that there is some systematic co-occurrence of form and meaning

(Winter, Woodin, & Perlman, 2023), but the form *qua form* is designed to relate to its meaning. With good design, a suitable perceiver, with a certain degree of initiation, will already be grasping the meaning in the right direction - just by perceiving the form. Humans are all initiated in some way in being part of humankind, but not all humans are similarly initiated in a particular social-cultural practice. Thus, depending on ones different initiations, different meanings can arise, which is to say that different organism-environment relations can become possible and stabilized.

How can form relate to its meaning? It is not uncommon for researchers studying iconicity in spoken and signed languages to invoke a definition of iconicity that is purely based on resemblance, such that some utterance is iconic of something if it resembles that something in some relevant way (Moita, Abreu, & Mineiro, 2023; Ortega, 2017) [1]. This can be problematic in some contexts for a number of reasons.

In the area of the philosophy of depiction for example, resemblance on its own has been shown to be a weak basis for understanding representations, such as pictures, drawings, and other modeling behaviors that would include iconic utterances like gestures or ideophones too (Goodman, 1976; Kulvicki, 2013). For example, a resemblance is not enough to make something a representation, for two things (e.g., twin sisters) can resemble each other but they need not represent each other (Goodman, 1976). Also, a painting resembles some landscape, and the landscape also resembles the painting, but we would only say that the painting represents the landscape and not the other way around (Goodman, 1976). The same goes, even more obviously, for the tiger and the growling, whole-body gesture of a tiger. Thus, resemblance cannot be the sufficient characteristic of iconicity (Goodman, 1976). Something more is needed, see e.g., Abell (2009).

The issues associated with pure-and-simple resemblance definitions, like the ones Nelson Goodman famously addressed, cut as deep as the metaphysics of reference and aboutness (Hutto & Myin, 2013; Millikan, 1984; Ramsey, 2007). Perhaps an empirically-oriented cognitive science may simply bracket such issues when it comes to studying iconicity: for all practical purposes an emphasis on resemblance will do.

However, barring any such metaphysical quandries, it has been recently suggested by researchers themselves that holding resemblance as a necessary and sufficient characteristic of iconicity is problematic for all practical purposes for the cognitive sciences too (Winter et al., 2023). According to these authors the research on iconicity has been riddled with conceptual inconsistencies, one of which is to assume that real resemblances are doing iconic work, while in fact iconicity is in "the eye of the beholder" (p. 6) (Winter et al., 2023). Specifically, because "iconicity is always mediated by perceptual and cognitive processes" (p. 6) and because research shows a variation in the ways iconic processes unfold for different users of icons a 'state of the art' (p. 25) definition should actually be: "A signal in any modality or medium exhibits iconicity when its production and/or interpretation involves a sense of resemblance between at least some aspect of its form and at least some aspect of its meaning". As such this 'subjectivist' definition aims to mitigate the wrongful emphasis on real resemblances by construing iconicity as a "sense" or an imagination in the mind (mind-to-world), and not a property of the environment (world-to-mind relation) (Bayne & Williams, 2023). This recent account by Winter and colleagues (2023) seems to go quite some steps further in pushing iconicity inward than other well-known accounts which helpfully emphasizes that iconicity requires taking into account structural mappings, which can be understood in terms of features of an utterance being mapped onto mentally represented image schemas (Emmorey, 2014). All these previous accounts are - in my reading - similar however in nature, given that they emphasize on internal goings-on of the perceiver in explaining how iconicity can possibly come about. This overemphasis on internal representations is at issue in the current paper, and instead the current definition aims to distribute out the iconic process.

And so the pendulum swings from iconicity being based on real resemblances in the world, to being in the eye of the beholder. Fundamentally, this pendulum follows oscillations in age-old debates in philosophy of mind about idealism and realism (Rolla & Figueiredo, 2021), where classic cognitive scientists favor a mind-to-world relation and more embodied accounts like ecological psychology try to motivate a type realism based on a coordinated organism-environment system (Turvey, 2018). In this paper I argue that the renewed interest in practically oriented quarters of cognitive science for defining iconicity in gestures, signs, and spoken utterances as mind-internal properties, may be calibrated towards a more distributed view by important work in theoretical psychology (Gibson, 1971; Lande, 2023), philosophy of mind (Facchin & Rucińska, 2024; Millikan, 1984; Noë, 2012; Ramsey, 2007), and the philosophy of depiction (Kulvicki, 2013, 2020). This calibration is needed for all practical research purposes too, not just philosophical ones, as I will try to argue throughout.

In the end, every cognitive scientist will need to work with a certain ontology and set of pre-empirical commitments that will help ground a particular research program like the study of iconicity. There is no single state-of-the art or theory-neutral way to go about that. Different pre-empirical commitments, or ignoring, comes at significantly different

[1] For example, Moita et al., (p1) define iconicity as "The similarity between certain properties of the linguistic form (such as spoken and signed phonology forms) and certain sensory-motor properties of the corresponding referent". Ortega et al., defines iconicity as "linguistic forms emulate perceptual, sensori-motor characteristics of a referent" p. 12

point of departures in the study of iconicity. Definitions can in this way solidify wrong or biased answers, or worse, wrong or biased questions. This paper aims to provide some reasons for why iconicity whould not be understood as a purely mind-internal affair, and is aimed to relate to a broader literature in radically embodied cognitive science to help define iconic activity. I draw particularly on 4E (embodied, enactive, embedded, extended) types of approaches to inform the current definitional work (Cuffari, 2012; Facchin & Rucińska, 2024; Hutto, 2022; Maddalena, Ferrucci, Bella, & Santarelli, 2024; Malafouris, 2016; Noë, 2012, 2015; Rączaszek-Leonardi & Zubek, 2022; Szokolszky & Read, 2022; Thibault, 2021; Turvey, 2018), which is a perspective that should help us think about iconicity in more distributed and relational way.

1.1 Overview and caveats

In section 2 key theoretical issues are discussed that are important for formulating a definition of iconicity. In section 3, I offer a more distributed definition of iconicity that is informed by previous observations in section 2. I also motivate some supposed benefits of this definition, for example how this definition can be naturally related to the concept of arbitrariness. In section 4, I offer concluding remarks and emphasize that the study of iconicity could benefit from a better integration with 4E approaches in cognitive science.

1.2 Key issues inviting a distributed understanding of iconicity

The following issues seem to arise with a pure-and-simple resemblance definition, but also a "subjectivist" based definition emphasizing a mind-internal process. This list of conceptual breakpoints are not exhaustive, but should make apparent what is at stake when preferring one definition over another.

1.2.1 Representationhood and representational actions

The first issue is that icons are representations, by which I mean physical, external utterances, or things, like maps. If our definition does not account for the simple assumption that iconic utterances function as physical representations in the context of communication, then we fail to have a coherent definition of iconicity. A key feature of representations is that they can misrepresent or be in error (Haugeland, 1998; Ramsey, 2007). They have truth or accuracy conditions; iconic representations can be used incorrectly, to a matter of degree (Haugeland, 1998). The job description of an iconic representation is to act as a substitute - a stand-in - for something (within a wider set of practices and concerns). This alone should make us skeptical about pure-and-simple resemblance definitions of iconicity, as it is not clear from the definition how something can be wrong in resembling something. We need someone to (potentially) *use* the resemblance in a certain way for a particular purpose - a consideration shared among many accounts of representation (Millikan, 1984; Ramsey, 2007), and can be traced back to Peirce's views on reference (Peirce, 1991). Fulfilling the 'use' criterion ensures that there are accuracy conditions of the icon as something can then serve its use or not [2].

A similar kind of problem potentially arises for the subjectivist definition. In what way can we say that your *sense* of resemblance - which is you experiencing an icon - is wrong? Well, we could say that, there is no real resemblance in the world. In this particular reading, it seems you can be wrong in your sense of a resemblance. But importantly, since in the subjectivist definition it is explicitly stated that there is no need for a definition to interface with externally verifiable resemblances, it opens up a reading where there are no corregiability conditions. A sense of resemblance in this reading is therefore like the experience of pain, it is simply true in virtue of experiencing it - it is an incorrigible state of affairs. The sense of iconicity, like pain, is there, even when it does not reflect 'real' resemblances or 'real' damage to the body. Of course, we are perhaps permitted to say that our experience of pain is in error if it is *taken to indicate* damage to the body (when there is none). But if your definition of pain is that it is a sense of damage to the body, or if your definition of iconicity is that it is a sense of resemblance, you cannot be wrong about it if you experience it to be the case. In understanding pain that might not be a problem, after all, pain is a kind of disturbance that can be understood as damaging in virtue of having it, and therefore pain does not require to be representationally understood. In the case of iconicity, it seems obvious to me, this type of incorrigibility poses deep worries, for if there is no verifiable basis for iconicity the iconic utterance fails to function as an external representation. To resolve this, it seems that any account of iconicity needs to relate verifiable resemblances or properties in the environment into their definition to have a more coherent account of iconicity (e.g., Emmorey 2014).

[2] In the case of a map, we can say the map misrepresents, if a user of the map encounters critical misalignments with that of the territory with that of the map. After all, the map becomes unusable in that case for the practice of way finding. We can, similarly, understand an iconic utterance to be incorrectly used as representative of something, when no other (similarly initiated) agents misalign on the representativeness of the utterance. After all, the iconic becomes unusable for the practice of communication.

That iconic utterances are to be understood as external representations does not automatically invoke sophisticated internal mental representations or schemas (cf. Emmorey 2014). As Noë (2012, p. 99) has commented, an external model does not need to produce a mental duplicate model at all, in order to fulfill its representational duties. Applied to gesture, I can perfectly use my gestures as a stand-in for mental rotation or actual rotation of an object (Cappuccio, 2013), but this does not necessitate for this gestural representation to produce a mental representation of actual rotations for the gesture to fulfill its representational function. In fact, gestures are representational actions and their representationhood is not inherited by mental representations of the thing the gesture represents (indeed, if we hold this, an infinite regress looms, for what grounds my representation of the representation; Bickhard 2007; Dennett 1971). The gestures literally stand-in for fully internally constructed mental models, obfuscating them, and they are in this sense representative rather than representational (a distinction advocated by Noë 2012). Indeed, while representational actions are sophisticated in some sense that require an explanation that includes particular mental capacities, we should not overintellectualize them, as the purpose of social-materially constituted representations do the modeling work for us when we use them (Emmorey, 2014; Winter et al., 2023). Thus we should be careful overintellectualizing iconicity: Chimpanzees can learn to use a scale model to locate a hidden object (Kuhlmeier & Boysen, 2002); Children as early as 1 to 2 years old can use a banana as a phone, way before any complex meta-representational abilities come online.

1.2.2 The phenomenology of iconicity and the utterance specification of iconicity

We further have to do justice to the special type of perceptual experience with iconic/depictive things. Depiction is often construed as having a dual nature in experience (Kulvicki, 2013). Most clearly this is evident in the perception of a painting. We see the painting *as a painting*, which can be perceived as a flat 2D surface with colored patches that also presents some 3D landscape. If this were not the case, we either would see no landscape at all - just the flat color-patched surface -, or we would be fooled, and hallucinate the 3D landscape. While there is discussion about whether we are aware of both aspects of a depiction at once or in alternation depending on visual attention (like Jastrow's duck rabbit illusion), and how we would go about discerning this scientifically, there is a general agreement that there is this dual or enmeshed nature to depictive representations (Kulvicki, 2013). As Noë puts it "[Hillary Clinton] does show up for you when you see her in a picture, but she shows up precisely as - obviously, palpably, manifestly - *not* present" (Noë, 2012, p. 84, original emphasis). Depictive representations, according to Noe, thus allow us to experience something present that is "*visibly* absent" (p. 87, original emphasis). This has an important implication. The representation does not present its subject because it simply invokes a sense of resemblance. Seeing something in a picture, or seeing something in a gesture, Noe would insist, is manifestly not just recreating the experience or the sense of that something. Rather depiction is a special mode of perception - a special aspectual mode of relating oneself to the environment, akin to Gibson would call 'second-hand perception' (Gibson, 1971).

We also dont have to be very mystic or overly intellectualistic about this special sense of iconic perception (Facchin & Rucińska, 2024). In gesture it can be seen that a throwing gesture is a gesture rather than an action because, for starters, their is often no ball that is thrown. Relatedly the dynamics of the gesture kinematics are not constrained by throwing an external object with a certain inertia (Facchin & Rucińska, 2024; Runeson & Frykholm, 1983) and there are potential demonstrative exaggerations or modulations of context-relevant features in the iconic utterance (Trujillo, Simanova, Bekkering, & Özyürek, 2018), much like how a caricature functions in the pictorial domain (Gibson, 1971). Furthermore the overall context, including what is said, also contributes to the perception of a gesture being a gesture rather than being an instrumental action (Facchin & Rucińska, 2024). Thus there is no poverty of the stimulus that when it comes to perceiving depictive content.

What does this mean for our characterization of iconic utterances? Firstly, utterances qua their dynamic unfolding can contribute to iconicity (Facchin & Rucińska, 2024). It is not only constituted in the mind of the beholder, the iconicity is literally imprinted in the utterance. Similar to how there is a kinematic specificity to bowling with a heavy versus lighter bowling ball (Runeson & Frykholm, 1983), there is also kinematic specificity of iconically gesturing to bowl a strike. Interestingly, *dissimilarities* rather than similarities of the utterance with the iconic stand-in (e.g., gesturing throwing a ball vs. really throwing a ball) can thus also contribute to iconic perception, going against pure-and-simple similarity accounts, as well as going against an account based on purely subjective or objective similarities. Similarities are only *part* of the story, and structural dissimilarities do much of the iconic work too.

1.2.3 The embodied and social-material constitution of icons

While we should acknowledge that iconicity is not fully 'read of' from the things themselves (Noë 2012, ch. 5), whether a thing or expression serves its use as an icon *is*(!) intrinsic to that thing. Simply consider that icons/depictions are designed by us precisely to fulfill a depictive role with certain purposes - they are designed in a certain way, to have the intrinsic properties that they have, to do the function they perform. Much like hammers, that function well as hammers because they have intrinsic properties typical of hammers, depictive tools, too, have properties that allow us to use them

effectively within our practices and concerns. And equally, hammering cannot not be understood just by contemplating the properties of hammers, but only within a wider set of concerns (e.g., shelter), objects (e.g., nails), and skills (hands that can can grasp (Heidegger, 2010). Hammers, and iconic utterances such as gestures and ideophones, are quite similar in this way, they are tools (Noë, 2015).

Importantly, tools, like iconic modes of communication, are social-cultural practices that have evolved with in our niche-constructed evolutionary trajectory (Rolla & Figueiredo, 2021). We are not simply bound by a pre-given environment out of which certain affordances arise that were and are fixed (Rolla & Figueiredo, 2021), but we actively change the environment, which in turn changes us (Deacon, 1998; Heyes, 2018), where new affordances arise and are stabilized within social-cultural practices (Malafouris, 2016; Reed, 1996). This reciprocal process includes the social institutions and practices in place that guide our differientated iconic perception of utterances. Consider for example that drawings to communicate objects might become less iconic over repeated references but they retain the iconic qualities that can be traced back to early utterances (Hawkins, Sano, Goodman, & Fan, 2023). What this means is that we - iconic toolmakers - are continuously actively shaping iconic utterance structures in a way that allows us to make optimal use of them through aligning with our physiology and social-cultural practices and institutions. Malafouris (2016) similarly shows that this temporally extended affair occurs over long timescales by analyzing the archaeological record in the evolution of bookkeeping tools, where the type of materials shaped the practices of bookkeeping which then further shaped how the materials were optimized over time. To say that iconicity is therefore only in the mind, or is only in the world, does not do justice to the reciprocal causation of *icons in action*.

Another reason that is offered about why senses of resemblances are needed to explain iconicity is that it is often not clear how real resemblances can possibly do the necessary work to support iconicity (Winter et al., 2023). Especially when it comes to cross-modal associations from sound to things. For example, the "bouba" utterance is to some degree iconic of a rounded shape and "kiki" to some degree iconic of a spiked utterance, which seems only explainable by some contrived mental mapping. After all, how can things have similarity if they are crossing a modality? However, a proper understanding of our embodiment could actually provide a bridge here. In ecological acoustics it has for example been found that there is information about the length of an object by attuning to what sound it makes when it hits the ground (Carello, Wagman, & Turvey, 2001). As such there is direct information about object properties in acoustics. Similarly physical properties have sound qualities and these can explain why a rounded object is more similar to Bouba than Kiki (Fort & Schwartz, 2022). There are many such continuities that underline that there is no fundamental schism between modalities that requires a translation process through mental mappings, but the world offers structured patterns in the form of a multidimensional ambient array of patterned energy (Stoffregen, 2003). Similarly there may be sounds, when produced by biological agents, that would require increased subglottal pressures, which in turn requires movements of the rib cage and the surrounding myofascial-skeletal system (Pouw & Fuchs, 2022). Particular high amplitude vocal sounds or emphasized parts of speech are therefore bound to be representative of increased and pulse-like movement that recruit whole-body activations such as postural adjustments (Pouw & Dixon, 2022). In this type of iconicity, where a sound "sounds like" an effortful multimodal utterance, the expressive nature of effortful sounds should be seen as a "constitutive symbol" (Malafouris, 2016) (p. 97): The iconic relation ("this sounds effortful" or "this sounds big") is thus not a contrived similarity mapping between A and B, but a mapping that only exists because of the phenomenon of biological sound production making A and B often contingent and inseparable (a structural similarity).

In sum the social-physical-material constitution of iconicity renders false the idea "that iconicity relies on only a sense of resemblance, and not objective similarities between form and meaning" (Winter et al. 2023, p. 13). Similarly, by an emphasis of the niche-constructed evolution of iconic utterances we should also reject the idea that there are only resemblances out there that are independent from our social-cultural practices.

1.2.4 The structure of iconic representations

Once we acknowledge that iconic utterances are social-culturally scaffolded tools, we naturally emphasize the need to study how we learn to use and shape the structural features of icons in productive ways. We can for example make an intuitive distinction between an *enacting* gesture depicting RUNNING versus a *hand-shape* gesture depicting RUNNING. The representational technique that is used may be more semantically transparent than others - at least to the uninitiated (see e.g., Ortega and Özyürek 2020). How to understand these differences? There are some helpful concepts from the philosophy of depiction that can be readily applied.

Syntactic repleteness (Goodman, 1976, cited in Kulvicki, 2013) provides one level of analyses where the syntactic features of the sign may correspond more or less with the syntactic features of the referent. Kulvicki (2013) considers an example of a black and white picture of a real-world scene, which is less replete than a color picture of that scene. This is because the colors in the color picture are syntactically employed to represent color in the scene, while in black and white picture, color is not employed in the way that the scene is partitioned on the dimension of color (in the black-and-white picture it is contrast that matters, not color). Similarly, the RUNNING gesture is more replete

in its representational technique, as the different features that make a difference for this gesture as standing out as a RUNNING, is much like the differences that make a difference for a running action to stand out from other actions. The hand-shape running gesture can be said to have structural similarities in the form of a isomorphism between legs alternating and fingers alternating, and perhaps between a translational movement of the body and a translational movement of the hand. Note, that the running enacting vs. hand-shape gesture is much more directly resembling the running action and functions much more like images do in depicting their referents.

It seems therefore we need to conclude that enactive gesture is potentially more transparently depictive (Kulvicki, 2013). However, syntactic repleteness does necessitate that the referent is more easily picked out however - it does not ensure semantic repleteness. A caricature of some person might be more likely to pick out a person among a group of similar looking persons than a realistic picture would, though the caricuture is syntactically less replete than a realistic picture. Syntactic repleteness does mean that the representational technique draws on differences and similarities that might be more close to the type of differences and similarities picked out in the environment. They may be syntactically more natural or familiar, in this sense. The familiarity should be assessed in relation to evolved capacities and social-cultural practices. Indeed, some representation might be more syntactically replete for a chimpanzee, than for a human, given that syntactic features of a representation might be closer to the features that matter in dealing with the real world, or learned skills.

Kulvicki (Kulvicki, p.96) offers another aspect to the syntax of depiction, suggesting that some representational techniques are more syntactically *sensitive* than others. One representational technique may tolerate subtle differences in the syntactic features without changing its content, while for another system, subtle differences do change its content. For example, if I change the cadence of my running in the enacted gesture, where I run in place, versus in the hand-shape gesture, where my fingers "run" on an imaginary surface, it might be that subtle differences in cadence are more likely to be seen to represent FAST running in the enacted gesture as compared to the case of the hand-shape gesture. Under this understanding the representational technique of enacted gestures tends to be more *syntactically* sensitive than hand-shape gestures. In general, Kulvicki (2013) notes that the syntactic sensitivity of images is very high relative to say a map, and written text is much less syntactically sensitive (e.g., change the font of the text, and the content stays stable). Indeed, syntactic sensitivity is precisely an important aspect that may allow us to differentiate between one representational technique over another (Ortega & Özyürek, 2020; Streeck, 2008). As far as we know, these conceptual tools have not been applied to the analyses of iconicity in multimodal language (see also Lande 2023 for in-depth treatment of the possible syntax of images), though they can be helpful for analyzing different types of representational techniques in iconic expressions.

What this section is aiming to show, is that one can go about studying iconicity by understanding how users might differ in structural ways in how they use icons. This means, that instead of retreating into the skull by suggesting that iconic perception cannot be understood by objective similarities between form and meaning, or by externalizing everything to pre-given singular similarities, we should be analyzing how users embedded in particular social-cultural practices tend to use iconic utterances in structural ways. For example, by starting to study how syntactically replete and sensitive a particular iconic mode of communication is. These are objectively study-able relations between form and meaning (to which only a certain initiated set of organisms may be sensitive to). The study of syntactic repleteness and sensitivity, is just an example of a conceptual framework for objective study of structural similarities in form and meaning that can be part of a cognitive science of iconicity.

1.2.5 Redefining iconicity for the cognitive sciences: A distributed definition

In the above selective overview we have given some considerations that problematize similarity-based or subjectivist definitions of iconicity in cognitive science. On the basis of the previous section I think we need a merger of views in some way, where having a particular initiation allows one to structurally attune to environmental patterns for iconic use. Only an organism with a particular "Umwelt" (Munz, 1993; von Uexküll, 2001) that has been initiated (evolved, enculturated) into a niche to learn to selectively attune to regularities in the environment can the world show up in a certain way. So can we design a practical definition that allows us to mark the type of processes that together, i.e., co-constitutively, make up iconic activity as a distributed practice? so not mind internal, not mind-independent, but distributed? The following definition aims to address this:

An utterance B' is more iconic of B when
1. B' is acted upon as a representative of B
2. by attuning to structural (dis-)similarities K(B, B')
3. in a context C for an agent A possessing embodied and social-cultural knowledge ES(A)
3. with a higher accuracy

Though I think the definition has a simple list of considerations, I fully acknowledge that this is not a "lean" definition. But we have to keep in mind that the current definition is aimed to serve practically oriented cognitive scientists to help identify the key elements of an iconic process. This definition is not designed to capture some kind of fundamental essence of iconicity.

To exemplify each part of the definition: B' can be a tracing gesture of a house. The context C might be the speech that goes along with the gesture or the previous discourse. Embodied knowledge (Munz, 1993) might be constituted in the fact that one has a certain body belonging to a species evolved to tap into certain invariants of the world relative to the agent, and the social-cultural knowledge might consist in the use of tracing as a technique of presenting outlines of an object. The structural (dis)similarity K(B, B') might be the relation between the contours of a typical house and the traces of the gesture. The accuracy, in this case might be the degree that the tracing gesture is indeed able to be reliably representative for the a house. Note we are emphasizing the dissimilarity as much as the similarity, but henceforth we will just use the term structural similarity to imply both structural relations between form and meaning.

With this definition what differentiates iconic activity from arbitary symbolic activity? It is only when the agent uses the structural resemblances to use B' as representative of B. So in other words, it is only if the form - qua form - contributes to B as being used as a representative of B', that iconicity arises. If it is by mere association or higher-order systematic association, where the agent uses B' as B without attuning to the similarity, it is not an iconic perception (for that agent).

Importantly, this "objective" structural similarity is not what is perceived however. What is perceived are the structural similarities that are "just right to offer the possibility of" (p. 3) iconic activity (Rączaszek-Leonardi & Zubek, 2022), i.e., to use something as a stand-in. Iconic perception is a process of suitable agent attuning to structural similarities that are embedded within a context. We can learn from Ecological Psychology (Turvey, 2018), 4E more generally (Rączaszek-Leonardi & Zubek, 2022), and views within New Materialism (Barad, 2007), in that they teach us how to be realists about iconicity without denying a role for the perceiver. Consider for example the definition of an affordance by (Turvey, 2018) "... X affords activity Y for organism Z on occasion O if and only if X and Z are mutually compatible on dimensions of relevance to Y." (p. 330). Our definition is similar in nature in the sense of trying to reconcile some stable, reproducible, and coherent property (iconicity) to emerge between person and environment. Taking from Karen Barad (Barad, 2007), we can call this stable and reproducible iconic process a type of "material-discursive activity". A temporally extended affair, where object and subject become continuously differentiated as sign and mind through "intra-activity". As such it makes no sense of thinking about iconicity as an objective property of the environment, or a subjective property of the mind (Rączaszek-Leonardi & Zubek, 2022). Iconicity does not exist *before* or *apriori to* contact between entities we can only sensibly call an 'iconic object' and a 'mind poised for iconic perception' after some interactions have stabilized (Barad, 2007). Yet, this distributed iconic process, while relational and based on continuous reciprocal interventions (Hacking, 1983), is as real and as objective as any other perception-action activity an organism-environment system might stably reproduce.

The realism, thus obtained, is not obtained by assuming that we subjectively represent objective features correctly or not (these terms do not make sense in the case of iconic activity). Indeed, we are actually in error to take for granted that that iconic signs are concrete objects that humans have a special "sense" for, as they are rather entities and sensitivities that "must instead be stabilized in the flow of experience with effort, which is often accomplished through routinized actions, coactions and the structures of language" (p. 6, Rączaszek-Leonardi and Zubek 2022) and other syntactic-depictive practices (see section 1.2.4). Rather it is attained by assuming that a type of interaction or "intervention" is stably and coherently reproduced (Hacking, 1983) by the organism-environment system.

Our definition naturally emphasizes embodied or social-cultural knowledge that allow for iconicity to unfold. Given that iconicity "is accessible to us via our engagements with it, its dynamical and social nature is highlighted; the environment can be characterized in terms of processes, changes and stabilities rather than in terms of sets of static objects and their features" (p4, Rączaszek-Leonardi and Zubek 2022). This term is aimed to explicate that a perceiver is always already initiated in a certain way. Most fundamentally, two agents may be different because they are part of a different species, which makes them differently suitable as an perceiving agent of a particular iconicity: An elephant might have a problem with interpreting a 2D picture while a human infant with minimal social-cultural initiation might not. Even within the same species differences in suitability arise. It might be that enacting gesture of RUNNING might be relatively easy to learn to perceive for human infants as they are syntactically replete with its source, while other types of gestures like the hand-shape gesture of RUNNING might not. However, an infant might still need to learn further nuances of the gesture as opposed other gestures, e.g., the infant might need to learn the syntactic sensitivity that is typically employed with a gesture in this community. Thus competent iconic perception is not an either-or phenomenon.

To summarize, being part of the same community, species, genus, or perhaps class or phylum, might sometimes be enough to be a suitable perceiver for some iconic utterances to be representative of something with a certain degree of accuracy. Many types of initiations can be considered, including social-cultural initiations that for example promote the use of particular structural similarities with a certain pictorial or postural-gestural syntax. This emphasis on embodied

and social-cultural knowledge clearly thereby answers to the subjectivist emphasis on the need to acknowledge that there is a important and indeniable role for the perceiver (Winter et al., 2023). Importantly, here it is emphasized that the particular suitability of a perceiver should be sought in social-cultural practices and embodied sensori-motor knowledge which together constrain our possible engagements with the (social-cultural) environment. It thereby directs the cognitive scientists to a wider set of practices, rather then something happening inside the perceiver.

1.2.6 Context, degree of accuracy, arbitrariness

The current section will focus on the idea of accuracy (component 4 of the definition). This is a special element in the distributed definition as it can do important conceptual work that relates arbitrariness directly to iconicity, but only because we explicitly have a term that is about the accuracy conditions.

It is important to acknowledge, that sometimes the reason why a gesture B' is representative of say an action B, can only be understood relative to the context. So a vague gesture B' that is not reliably a good substitution for B, may become an optimal gesture reliably substituting for B given some context that helps constrain what structural similarities should be attuned to over others. As such the context can contribute to the semantic transparency of some iconic utterance. This is nothing controversial of course, though it is non-trivial to work out in what ways context like co-ocurrent lexical signs or speech help to increase semantic transparency.

Lopes (p.16) has an interesting discussion on sign languages in that he distinguishes between lexical utterances that have no resemblance and no iconicity, those that have resemblance such that is supposedly immediately understood as iconic (e.g., the ASL sign for duck), and those that have resemblances but do not have immediate iconicity. For the latter 'no immediate iconicity' category Lopes argued that the ASL sign for Rabbit, resembles a rabbit, but it does not depict it, because this resemblance becomes apparent only when you know the referent of the sign (Lopes, 1996, p 16). I would disagree that the utterance is not iconic, and our distributed definition would require us to provide a further specification: iconicity for whom, in what context? Under the distributed definition, it is not iconic for everyone, nor iconic in every situation, but it sure can be iconic for an agent with a certain perspective. For example, when the semantic field ('animals') has been already been introduced in the discourse through a different iconic lexicalized utterance. Iconicity should then be understood within a context. We can follow Facchin and Rucińska (2024) the Praxeological Enactive account of pretense here (Rucinska, 2019; Weichold & Rucińska, 2022), which has done a lot to explicate the dynamic context that allows for representative actions: "Ongoing creation of pretense meanings, shaped by materiality of the objects one plays with and interactive situation one is in, trumps the view that pretense contents need to be created and represented in advance of action" (p.3).

In a common way of speaking, we may thus say that utterance is more or less iconic. When we say it is highly iconic, it seems that we implicitly set the agent as some kind of null value, which is the perspective of the minimally initiated - of someone sharing a basic set of sensori-motor capacities and having some general social-cultural initiations, but no privileged type of knowledge. I think this understanding comes close to what Kulvicki calls "bare-bones content" as opposed to "fleshed out content" (Kulvicki, 2006), or what semioticians will call "primary" and "secondary" iconicity (Felix & Zlatev, 2010). On such a view, the ASL sign of duck, has strong fleshed-out iconicity for a particular agent with a certain initiation (e.g., in ASL), while its bare-bones iconicity, may not reliably include the concept duck for most minimally initiated perceivers.

Now lets assume for a moment that a null value - a minimally initiated perceiver - for the following reasoning. If we have utterances with high iconicity it is easily used as representative for something. Simply because the utterance is not exactly the same as the thing it is representative of, we can almost always assume that the utterance is representative to a certain degree. Now consider an utterance that does not have immediate iconicity like this, following Lopes categorization of the stylized rabbit ASL sign. This utterance is underspecified, such that there is a very low level of accuracy with respect to acting as a representative of a rabbit. Such a vaguely iconic utterance also can be used as representative for a whole range of other things - at least from the perspective of an uninitiated agent. This under-specification makes it stand in a more arbitrary relation to its referent, because the notion of arbitrariness can be fully understood as the degree of underspecification. A is more arbitrarily related to B, when A is more underspecified relative to B.

There is an interesting realization concerning underspecification that relates to *over*specification. Namely, when a certain utterance is more underspecified, the accuracy of using something B' as representative for B is low. As such, given that there are many vague utterances that have a low level of accuracy, it becomes more likely that there are other utterances Z that can perform at that level of accuracy to act as a representative (Hazen, Griffin, Carothers, & Szostak, 2007). To say this in more common terms then, when an utterance becomes more vague, it is true that other utterances could become equally suitable at that level of accuracy. This thereby relates to synonymy, such that two iconic utterances may be synonymous in acting as a representative at a certain degree of accuracy.

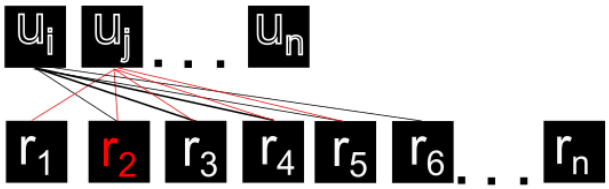

Figure 1: Note that when an utterance becomes underspecified relative to some target referent (r_2) there will be more utterances that have a similar level of underspecification. There are more synonyms at that level of accuracy (Hazen et al., 2007) Note, indeed that the competing utterances with respect to r_2 may not necessarily share all the elements within their set of potential referents, though they must share r_2, and the number of elements of the set is the same as this sets the specification with respect to r_2.

Thus, more under-specification begets more similar-performing alternatives. There are more possible alternatives that could function at that level of specification. Since there other possible utterances that could have done the job equally well there is more pressure to select.

To exemplify, there are many other objects used in representative actions that could be as accurately representative for some person as a salt-shaker. We could use something else instead. To understand why the salt-shaker is selected as an icon, one needs to understand the context or particular social-cultural initiation. Indeed, it is precisely when something is "not obviously iconic" in the ways we would commonly speak about - which just means not iconic for someone minimally initiated - that we need to invoke notions of convention and other constraints to understand how that minimal structural similarity can be tolerated while *functioning accurately* - i.e., representatively.

Without such historical-embodied-social-cultural constraints, we can assume that not very iconic objects wither in their rate of use within a community. Taking into account accuracy conditions, either for the uninitiated (bare-bones iconicity), or with the constraints in place that characterizes the ecological use of the iconic sign (fleshed-out iconicity), means addressing the issue of maintaining realism head on: "Living systems... enact stances on the world that can be and are corrected by the responses (the feedback) that the world, including other persons, gives back" and inappropriate perceptions and differentiations, i.e., low accuracy signs, "will tend to be de-selected" (Thibault 2021 p. 28)". The current definition thereby also closes an open loop that is implicit in the subjectivist view. It closes the loop by suggesting that iconic perception is an active socially-culturally constrained activity that is in constant informational contact with the environment, and as soon such contact is lost, either due to constraints falling away or because the informational specificity without such constraints was too feeble to begin with, then we can assume the iconic sign will not survive as a tool used for that particular representative purpose.

Thus the degree of accuracy naturally invites more emphasis on context and convention (or initiation as I have called it). By including level of accuracy in our definition we thereby also allow - under the right context - for very under-specified iconic utterances to potentially serve as representative for anything else. Gestures may still be vague on their own, but this is because they are not vague in the context of speech, or in the context of discourse. On the other hand, under the current definition we can characterize that a caricatural gesture can be in potential more iconic than a more realistic gesture (Gibson, 1971). This could be because a caricature could in principle exaggerate features that are maximally representative of something, thereby increasing its degree of accuracy relative to other structural similarities that are less representative. Under the current definition this does likely require some initiation or context to support the iconic process, as the structural resemblances that are attuned to deviate from the 'usual perceptual skills' that we ascribe to the minimally initiated. Exaggerated gesture may thus be more iconic, then more faithful gestures, if your used to such ways of representing. This underlines that (structural) resemblances, or senses of resemblances, can only be part and not sufficient for the iconic process to unfold.

Next to context-dependence, by adding degree of accuracy in our definition we further invite a more gradient understanding of iconicity. This is very important because co-speech gestures can be very and deliberately vague in what they convey, and they often do not seem to function to single out a very particular interpretation. This is precisely because they are designed for their meanings to be fleshed out by the linguistic context, on the fly. For example, co-speech gestures are on this view often designed to be vague on their own, while it is quite clear in the spoken or signed context why they are iconic off. Iconic utterances can thus be designed to have flexible interpretations, and they are more art-like then language-like in that respect.

1.3 Discussion

Definitional disputes about iconicity can be couched in a age-old debates about realism and idealism. Pure-and-simple similarity based definitions - where iconicity arises when there is some relevant resemblance between form and meaning - are realist in that iconicity is held to exist prior to our cognition (it is pre-given). Instead the recent subjectivist definition (Winter et al., 2023) can be couched in idealistic terms, where the apparent reality of the world is a projection of internal laws of the mind (Rolla & Figueiredo, 2021) - iconicity is in the eye of the beholder. Based on some critical reflections on a wider literature in the philosophy of depiction, philosophy of mind, and 4E approaches to cognition, I have put forward a distributed definition, which aims to underline that "meaning is the temporally emergent property of [physical]-material engagement, the ongoing blending between the mental and the physical" (Malafouris, 2016) (p. 117).

Concretely, the distributed definition invites a division of explanatory labor in the study of iconicity. Iconicity is not in the eye of the beholder, it is for example also in the kinematic signatures typical of iconic utterances that invoke certain structural ways of resembling (Rucinska, 2019; Runeson & Frykholm, 1983). As such, it invites human movement scientists to study kinematic signatures of iconicity that can be attuned to in iconic perception. Moreover, while for the non-initiated, salt shakers are not very iconic for particular persons, sometimes the salt shaker is quite suitable to stand in for a person. Different definitions invite different potential answers and emphases to how the salt shaker can be representative of a person. If one operates under the assumption that iconicity is mind-internal, it is in some sense trivial to account for the idea that a particular person can be represented by a salt shaker. One just invokes a sense of resemblance. According to a distributional definition a mind-internal representation cannot really explain why the salt shaker can show up as a this or that person. Why the person shows up like this, requires the wider set of practices that set up a language game. So even if we assume tongue-in-cheek, that the subjectivist-inspired neuroscience can find a sense of resemblance of that person in he brain, we really need to ask how the neural correlates of the person engram got there. Which requires recourse to embodied social-cultural initiation that allow for exploiting certain minimal structural resemblances. As such the distributed definition invites research on how certain artefacts come to reliably act as representatives for things and what discursive blending of the physical and the mental needs to take place (Hawkins et al., 2023; Malafouris, 2016). Such blending is a temporally extended activity, where conventional ways of structural resemblance can arise on the fly in multimodal conversational interaction (Hodge & Ferrara, 2022; Rasenberg, Özyürek, & Dingemanse, 2020; Thibault, 2021).

The distributed definition invites researchers to think about the different elements of the iconic process. If we parametrize, or explicitly bracket, these elements in our experiments, I think we come further with our understanding of how iconic activity unfolds, as compared to understanding iconicity as based on subjective senses, or as based on objective resemblances, or as based on subjective representations of objective resemblances.

Of course, there is much more work needed to further explicate the current account. More work is needed what structural similarities could possibly consist of. I also did not discuss or commit myself to differing 4E perspectives, which might prove incompatible for helping define iconicity (Bickhard, 2007; Hutto, 2022; Malafouris, 2016; Rucinska, 2019; Szokolszky & Read, 2022). In that sense, theorists and researchers within the 4E approach, semiotics more generally (Felix & Zlatev, 2010), or the philosophy of depiction, might find the current account too general or repetitive too add much to the debates already. Note however, that the current paper actually tries to invite a more interdiscipinary group of scholars to engage with defining iconicity for the cognitive sciences, a timely challenge posed by Winter et al. (2023). Thus, the main aim of this paper was to broaden the discussion and invite more alternatives to how to think about iconicity.

More generally, I think whatever position one takes about the particular mind-world relation that iconicity entails, ultimately boils down to a much deeper philosophical assumption of truth and reality. It boils down to your philosophy of natural science (Hacking, 1983). It is interesting to note here that Peirce, known in linguistics for coining the term icon, is known in philosophy for his pragmatism about truth, which he called a type of "objective idealism" (Hacking 1983, p. 60). Peirce defended the idea that truth is not a correspondence relation. To say that something is true of the world, is merely a community more regularly coming to the same conclusions. So is this a rabbit or a duck? If one holds a correspondence theory of truth, the drawing can be said to correspond (is similar) in some way to a duck or rabbit. But in a Peircian logic, it is true that it is a rabbit and a bunny, given that a commmunity tends to regularly settle on either, and it very much true, that a set of suitable perceivers entrenched in this and that context where rabbits are the relevant animal, this bunny can undeniably be a rabbit and nothing else. The correspondence theory seems to be less flexible to deal with context-sensitive perception (Barad, 2007). Under such a ill-posed view, to settle the truth, a mental representation must make it true. I think a more pragmatic understanding of how suitably initiated perceivers tend to regularly come to similar conclusions, based on continuously stabilizing organism-environment interactions will suit a practically minded cognitive scientist better than any definition that focuses on internal representations as

the ultimate mark of iconicity. This is because the cognitive scientist is invited to look at the whole process of iconic activity, which is not confined to mental representations, but more distributed.

When broadening out the discussion to include other perspectives, I think it will become apparent that there are other phenomena that are much like iconicity, in that they amount to a nich-constructud environmental property used in some perceiver-relative way. For example, the phenomenon of iconicity could be compared to hearing the beat in music. For hearing a beat, it is true that different people can perceive different beats in a musical piece depending on ones' initiation (Phillips-Silver & Trainor, 2005), and only some species can perceive a beat at all (Patel, 2021). It is also clear that the beat is not always derivable from the metrical structure of the signal alone (e.g., in the case of missing beats); i.e., the explanation requires a blend between the physical and the mental. Therefore, in the case of beat perception it is similar to iconic perception very tempting to to hold that beats are not physically there to be perceived, as a "genuine perceptual engagement with the world requires latching onto some" perceiver-independent property of the world. Thereby it is tempting to arrive at the conclusion that the beat is in the eye (or ear) of the beholder. Recently, Bayne and Williams (2023) have argued that this temptation should be countered with a deeper reasoning, and offer their own strategies how one can maintain a certain realism when it comes to perceiving the beat (Bayne & Williams, 2023). In the case of the perceiving the beat, it is neither an illusion nor simply a pattern in the environment, but it is more appropriately understood as a brain oscillatory dynamics being so that the brain resonates to the external structured energy patterns (i.e., meter) in the music (Raja, 2018). We could argue that this is not so different from iconic utterances, where signs are shaped to resonate to initiated minds.

References

Abell, C. (2009). Canny Resemblance. *The Philosophical Review*, *118*(2), 183–223.

Barad, K. (2007). *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Duke University Press.

Bayne, T., & Williams, I. (2023). In search of the beat. *Mind & Language*, *38*(3), 907–924. doi: 10.1111/mila.12439

Bickhard, M. H. (2007, August). Language as an interaction system. *New Ideas in Psychology*, *25*(2), 171–187. doi: 10.1016/j.newideapsych.2007.02.006

Cappuccio, M. L. (2013). Pointing as an Instrumental Gesture: Gaze Representation Through Indication. *Humana Mente Journal of Philosophical Studies*(24).

Carello, C., Wagman, J. B., & Turvey, M. T. (2001). Ecological acoustics 1 Acoustic Specification of Object Properties..

Cuffari, E. (2012, December). Gestural sense-making: Hand gestures as intersubjective linguistic enactments. *Phenomenology and the Cognitive Sciences*, *11*(4), 599–622. doi: 10.1007/s11097-011-9244-9

Deacon, T. W. (1998). *The symbolic species: The co-evolution of language and the brain*. W.W. Norton.

Dennett, D. C. (1971, January). Intentional Systems. *The Journal of Philosophy*, *68*(4), 87–106. doi: 10.2307/2025382

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015, October). Arbitrariness, Iconicity, and Systematicity in Language. *Trends in Cognitive Sciences*, *19*(10), 603–615. doi: 10.1016/j.tics.2015.07.013

Emmorey, K. (2014, September). Iconicity as structure mapping. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130301. doi: 10.1098/rstb.2013.0301

Facchin, M., & Rucińska, Z. (2024, March). Public Charades, or How the Enactivist Can Tell Apart Pretense from Non-pretense. *Erkenntnis*. doi: 10.1007/s10670-024-00787-7

Felix, A., & Zlatev, J. Z. (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Σημειωτκή - Sign Systems Studies*, *38*(1-4), 298–348.

Fort, M., & Schwartz, J.-L. (2022, November). Resolving the bouba-kiki effect enigma by rooting iconic sound symbolism in physical properties of round and spiky objects. *Scientific Reports*, *12*(1), 19172. doi: 10.1038/s41598-022-23623-w

Gibson, J. J. (1971). The Information Available in Pictures. *Leonardo*, *4*(1), 27–35. doi: 10.2307/1572228

Goodman, N. (1976). *Languages of Art: An Approach to a Theory of Symbols*. Hackett Publishing.

Hacking, I. (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511814563

Haugeland, J. (1998). Representational Genera. In J. Haugeland (Ed.), *Having Thought: Essays in the metaphysics of mind* (pp. 171–206). Harvard.

Hawkins, R. D., Sano, M., Goodman, N. D., & Fan, J. E. (2023, April). Visual resemblance and interaction history jointly constrain pictorial meaning. *Nature Communications*, *14*(1), 2199. doi: 10.1038/s41467-023-37737-w

Hazen, R. M., Griffin, P. L., Carothers, J. M., & Szostak, J. W. (2007, May). Functional information and the emergence of biocomplexity. *Proceedings of the National Academy of Sciences*, *104*(suppl_1), 8574–8581. doi: 10.1073/pnas.0701744104

Heidegger, M. (2010). *Being and Time*. SUNY Press.

Heyes, C. (2018). *Cognitive Gadgets: The Cultural Evolution of Thinking*. Cambridge: Harvard University Press.
Hodge, G., & Ferrara, L. (2022, June). Iconicity as Multimodal, Polysemiotic, and Plurifunctional. *Frontiers in Psychology*, *13*. doi: 10.3389/fpsyg.2022.808896
Hutto, D. D. (2022, December). Getting real about pretense. *Phenomenology and the Cognitive Sciences*, *21*(5), 1157–1175. doi: 10.1007/s11097-022-09826-6
Hutto, D. D., & Myin, E. (2013). *Radicalizing Enactivism: Basic Minds Without Content*. MIT Press.
Kuhlmeier, V. A., & Boysen, S. T. (2002). Chimpanzees (Pan troglodytes) recognize spatial and object correspondences between a scale model and its referent. *Psychological Science*, *13*(1), 60–63. doi: 10.1111/1467-9280.00410
Kulvicki, J. V. (2006, August). Bare Bones Content. In J. V. Kulvicki (Ed.), *On Images: Their Structure and Content* (p. 0). Oxford University Press. doi: 10.1093/019929075X.003.0007
Kulvicki, J. V. (2013). *Images*. London: Routledge. doi: 10.4324/9781315884578
Kulvicki, J. V. (2020). *Modeling the Meanings of Pictures: Depiction and the philosophy of language*. Oxford University Press. doi: 10.1093/oso/9780198847472.001.0001
Lande, K. J. (2023). Pictorial syntax. *Mind & Language*, *n/a*(n/a). doi: 10.1111/mila.12497
Maddalena, G., Ferrucci, F., Bella, M., & Santarelli, M. (2024, September). Introduction. In G. Maddalena, F. Ferrucci, M. Bella, & M. Santarelli (Eds.), *Gestures: Approaches, Uses, and Developments* (pp. 1–12). De Gruyter. doi: 10.1515/9783110785845-001
Malafouris, L. (2016). *How Things Shape the Mind*.
Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories: New Foundations for Realism*. MIT Press.
Moita, M., Abreu, A. M., & Mineiro, A. (2023, January). Iconicity in the emergence of a phonological system? *Journal of Language Evolution*, *8*(1), 1–17. doi: 10.1093/jole/lzad009
Munz, P. (1993). *Philosophical Darwinism: On the Origin of Knowledge by Means of Natural Selection*. London ; New York: Routledge.
Noë, A. (2012). *Varieties of Presence*. Harvard University Press.
Noë, A. (2015). *Strange Tools: Art and Human Nature*. Farrar, Straus and Giroux.
Ortega, G. (2017). Iconicity and Sign Lexical Acquisition: A Review. *Frontiers in Psychology*, *8*.
Ortega, G., & Özyürek, A. (2020, February). Systematic mappings between semantic categories and types of iconic representations in the manual modality: A normed database of silent gesture. *Behavior Research Methods*, *52*(1), 51–67. doi: 10.3758/s13428-019-01204-6
Patel, A. D. (2021, August). Vocal learning as a preadaptation for the evolution of human beat perception and synchronization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1835), 20200326. doi: 10.1098/rstb.2020.0326
Peirce, C. S. (1991). *Peirce on Signs: Writings on Semiotic*. UNC Press Books.
Phillips-Silver, J., & Trainor, L. J. (2005, June). Feeling the beat: Movement influences infant rhythm perception. *Science (New York, N.Y.)*, *308*(5727), 1430. doi: 10.1126/science.1110922
Pouw, W., & Dixon, J. A. (2022, July). What you hear and see specifies the perception of a limb-respiratory-vocal act. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1979), 20221026. doi: 10.1098/rspb.2022.1026
Pouw, W., & Fuchs, S. (2022, October). Origins of vocal-entangled gesture. *Neuroscience & Biobehavioral Reviews*, *141*, 104836. doi: 10.1016/j.neubiorev.2022.104836
Raja, V. (2018, March). A Theory of Resonance: Towards an Ecological Cognitive Architecture. *Minds and Machines*, *28*(1), 29–51. doi: 10.1007/s11023-017-9431-8
Ramsey, W. M. (2007). *Representation Reconsidered*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511597954
Rasenberg, M., Özyürek, A., & Dingemanse, M. (2020). Alignment in Multimodal Interaction: An Integrative Framework. *Cognitive Science*, *44*(11), e12911. doi: 10.1111/cogs.12911
Reed, E. S. (1996). *Encountering the World: Toward an Ecological Psychology*. Oxford University Press.
Rączaszek-Leonardi, J., & Zubek, J. (2022, December). Is love an abstract concept? A view of concepts from an interaction-based perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *378*(1870), 20210356. doi: 10.1098/rstb.2021.0356
Rolla, G., & Figueiredo, N. (2021). Bringing Forth a World, Literally. *Phenomenology and the Cognitive Sciences*, 1–23. doi: 10.1007/s11097-021-09760-z
Rucinska, Z. (2019). Social and Enactive Perspectives on Pretending. *Avant: Trends in Interdisciplinary Studies*, *10*(3).
Runeson, S., & Frykholm, G. (1983). Kinematic Specification of Dynamics as an Informational Basis for Person-and-Action Perception: Expectation, Gender Recognition, and Deceptive Intention. *Journal of Experimental Psychology: General*, 31.
Stoffregen, T. A. (2003, April). Affordances as Properties of the Animal-Environment System. *Ecological Psychology*, *15*(2), 115–134. doi: 10.1207/S15326969ECO1502_2
Streeck, J. (2008). Depicting by gesture. *Gesture*, *8*(3), 285–301. doi: 10.1075/gest.8.3.02str

Szokolszky, A., & Read, C. (2022, December). Pretend play with objects: An ecological approach. *Phenomenology and the Cognitive Sciences*, *21*(5), 1043–1068. doi: 10.1007/s11097-021-09755-w

Thibault, P. J. (2021). *Distributed languaging, affective dynamics, and the human ecology: The sense-making body, Vol. 1*. New York, NY, US: Routledge/Taylor & Francis Group.

Trujillo, J. P., Simanova, I., Bekkering, H., & Özyürek, A. (2018, November). Communicative intent modulates production and comprehension of actions and gestures: A Kinect study. *Cognition*, *180*, 38–51. doi: 10.1016/j.cognition.2018.04.003

Turvey, M. T. (2018). *Lectures on Perception: An Ecological Perspective* (1edition ed.). Routledge.

von Uexküll, J. (2001, July). An introduction to Umwelt. *Semiotica*, *2001*(134), 107–110. doi: 10.1515/semi.2001.017

Weichold, M., & Rucińska, Z. (2022, December). Pretense as alternative sense-making: A praxeological enactivist account. *Phenomenology and the Cognitive Sciences*, *21*(5), 1131–1156. doi: 10.1007/s11097-021-09770-x

Winter, B., Woodin, G., & Perlman, M. (2023, November). Defining iconicity for the cognitive sciences. In *Handbook of Iconicity in Language*. Oxford University Press.

2 Author Note

I am looking for feedback on this preliminary version of the paper (feel free to reach out, wim.pouw@donders.ru.nl).

3 Funding

This work is funded by a personal career 'VENI' grant awarded to Wim Pouw from the NWO ((VI.Veni.201G.047). The work is further supported by the DFG priority programme ViCom (grant PO 2841/1-1).

4 Acknowledgements

I want to thank Cornelia Ebert for inviting me for a month stay at the center for advanced study of the Goethe University in Frankfurt (Forschungskolleg Humanwissenschaften) and for helpful discussions and encouragements to take a close look at iconicity. I want to thank Aleksandra Cwiek for her comments on an early half-baked version of this paper. I want to thank Susanne Fuchs for commenting on the first version of this paper. This paper was improved based on a presentation given at the VICOM workshop on demonstratives, with special thanks to comments from Cecile Meier, Patrick Trettenbrein, Rui Liu. These acknowledgements do not signal endorsements.