ORIGINAL RESEARCH

Rethinking how children individuate objects: spatial indexicals in early development

Frauke Hildebrandt[1] · Ramiro Glauer[1] · Richard Moore[2]

Abstract

The current understanding of cognitive development rests on the premise that infants can individuate objects early on. However, the so-called object-first account faces severe difficulties explaining extant empirical findings in object individuation tasks while alternative, more parsimonious explanations are available. In this paper, we assume that children start as feature-thinkers without being able to individuate objects and show how this ability can be learned by thinkers who do not already implicitly possess the notion of an object. Based on Tugendhat's ideas on the relation between singular terms and object reference, we argue that spatial indexicals comprise the fundamental means of object individuation and describe how feature thinkers might acquire the complex substitutional system of spatial indexicals. In closing, two accounts of object cognition that do not rely on symbolic capacities, namely Pylyshyn's FINST indexes and Burge's perceptual objectivity, are critically discussed.

Keywords Reference · Object individuation · Indexicals · Singular terms · Spatiotemporal frame of reference

✉ Ramiro Glauer
    ramiro.glauer@fh-potsdam.de

    Frauke Hildebrandt
    frauke.hildebrandt@fh-potsdam.de

    Richard Moore
    richard.moore@warwick.ac.uk

1   University of Applied Sciences Potsdam, Potsdam, Germany

2   University of Warwick, Coventry, UK

1 Introduction

In recent years, much progress has been made in our understanding of infant cognitive development. We know more about mechanisms supporting the development of intuitive physics (Baillargeon et al., 2009, 2012; Mascalzoni et al., 2013; Wang & Goldman, 2016), the development of number cognition (Carey, 2004, 2009; Cheung et al., 2017; Sarnecka & Carey, 2008; Schaeffer et al., 1974), the role of cooperation in the development of linguistic and ethical behaviour (Nucci & Gingo, 2010; Tomasello, 2021; Tomasello & Gonzalez-Cabrera, 2017), and how infants come to think about others and their mental states (Blijd-Hoogewys & van Geert, 2017; Sodian & Kristen, 2010; Wellman, 2011). Another central theme of early cognitive development research is infants' ability to individuate objects (Spelke, 1985, 1988; Xu & Carey, 1996; Wilcox & Baillargeon, 1998; Van de Walle et al., 2000; Santos et al., 2002; Xu et al., 2004; Xu & Baker, 2005; Mendes et al., 2008; Yoon et al., 2008; Futó et al., 2010; Stavans et al., 2019). Object individuation is seen as the cornerstone of children's ontologies (Moore et al., 1978; Wynn, 1992; Spelke et al., 1995; Xu & Carey, 1996; Hespos & Rochat, 1997; Wilcox & Baillargeon, 1998; Aguiar & Baillargeon, 1999; Van de Walle et al., 2000; Cacchione & Rakoczy, 2017; Cacchione et al., 2020), plays a central role in empirical research on cognitive development (as evidenced by Lin et al., 2022) and in philosophical debates about propositional thought (Bermúdez, 2007), objectivity (Burge, 2010), and cognitive development (Butterfill, 2020).

In the developmental debate, objects are conceptualised as cohesive entities that move continuously through space and time (Cacchione & Rakoczy, 2017; Carey, 2009).[1] Based on empirical findings, the received view maintains that infants have a notion of objects from around 9 months. The paradigmatic experiments apparently showing that infants individuate objects involve two occluders. From behind the occluders, two objects that look alike appear and disappear subsequently without passing the gap between the occluders. When the occluders are lifted, infants display surprise when only one object is behind one of the occluders. This is interpreted as showing that infants understand that the object could not have passed the gap between the occluders, meaning that they understand that objects move through space continuously. Problematically, this "object-first" account runs into difficulties accounting for the empirical evidence—because infants do not develop appropriate expectations when experimental conditions are only slightly changed. For instance, infants do not expect to see two objects behind one occluder when these objects are balls of different sizes and colours that were not simultaneously perceived. Stavans et al. (2019) attempt to explain infants' individuation failures within the object-first account by invoking two cognitive systems—a physical reasoning system and an object file system—and errors resulting from failed attempts at integrating the output of these systems. However, Stavans et al. overlook that the explanatory burden is almost exclusively carried

---

[1] We focus on concrete objects and leave abstract objects out of our consideration. Human thinkers encounter spatiotemporal objects before learning about abstract objects. Therefore, to understand how we acquire to individuate objects, we need to understand how we learn to individuate spatiotemporal objects. Also, note that the individuation of abstract objects requires quite sophisticated linguistic capacities; see Quine (1974) for a suggestion of how reference to objects tout court (including abstract objects) is acquired.

by only one of the hypothesised cognitive systems, namely the physical reasoning system, which processes featural information. Moreover, an alternative interpretation of the challenging data is available that is simpler, more parsimonious, and theoretically well-grounded (Hildebrandt et al., 2020, 2022).

According to this alternative view, infants cannot refer to selfsame entities that persist over time (i.e., objects) at the outset of their cognitive development. Instead, infants discriminate features and form expectations about feature patterns (e.g., Cohen et al., 2002). Reference to objects is then associated with specific linguistic capabilities—in particular, with the acquisition of terms that refer to one and only one thing, i.e., singular terms (Tugendhat, 2016/1976; Hinzen & Sheehan, 2015; Hildebrandt & Glauer, 2022, 2023; Hinzen & Mattos, 2023). Identifying objects as the same across time and space is mastered only with these terms. Moreover, the first singular terms used referentially are spatial indexicals (Tugendhat, 2016/1976; Glauer & Hildebrandt, 2022; Hildebrandt & Glauer, 2022, 2023).

If this hypothesis is correct, an object-involving ontology develops relatively late in ontogeny and only with mastery of certain natural language forms, marking a fundamental transition from an initially pre-propositional stage to propositional thinking. In many cases, there appear to be pre-propositional capacities that look like their accomplished propositional forms when exhibited in a given situation, but the cognitive system is nonetheless overall confronted with characteristic limitations. For instance, the acquisition of object individuation should guarantee the situation independence of meaning, allow for the distinction between truth and falsehood, and enable human beings to think about possibilities. Singular reference would give rise to the development of predication, the powerful logical tool of quantification, and form the basis for attributing beliefs and desires (for a conceptualisation of these connections, see Hildebrandt & Glauer, 2022). The acquisition of singular terms would transform the initial cognitive capacities into their propositional forms. In a similar vein and from a pragmatic point of view, Rubio-Fernandez (2021) argues that theory of mind development builds on establishing situations of joint attention, which is driven by the linguistic capacity to understand spatial indexicals.

Spatial indexicals comprise several distinguishable semantic features acquired between 3 and 7 years of age (Chu & Minai, 2018; de Villiers & de Villiers, 1973; Webb & Abrahamson, 1976). While the theoretical connection between singular terms and object reference and the alternative interpretation of empirical findings on object individuation are developed in the literature, a detailed account is still lacking of how creatures, if they do not yet possess the notion of an object, can learn to refer to objects via the acquisition of singular terms. Such an account would show that, under the assumption that children start as feature-based thinkers, the acquisition of spatial indexicals is sufficient for object reference. Under the same assumption, an argument that feature-based thinkers could only learn object reference via acquiring spatial indexicals would show the necessity of spatial indexicals for object reference.

In this article, we want to start from the assumption that infants (and other non-linguistic creatures) lack a conception of objects. From there, we want to show how the ability to refer to objects can be learned via the acquisition of a system of inter-defined spatial indexicals by thinkers who do not already possess the notion of an object and give some initial credibility to the claim that spatial indexicals are also

necessary. In Sect. 2, we present the assumption that children begin their cognitive development without grasping selfsame entities, i.e., objects. We aim to make this assumption plausible but will not argue for it here. Section 3 elucidates Tugendhat's (2016/1976) ideas on the relation between singular terms and object reference and argues that spatial indexicals comprise the fundamental means of singular reference. Assuming that children set out as feature-based thinkers, Sect. 4 shows that learning a system of inter-defined spatial indexicals is sufficient for acquiring object reference. Section 5 suggests that it might also be necessary by showing that two prominent attempts to explain object cognition without recurrence to language capacities fail: Pylyshyn's FINST indexes (Pylyshyn, 2001, 2007) and Burge's perceptual objectivity (Burge, 2009b, 2010). Section 6 sums up.

2 Inhabiting a world without objects

Let us assume that young children—and other non-linguistic creatures—are unable to refer to selfsame entities that persist over time (objects) (Strawson, 1959; Quine, 1960, 1974; Tugendhat, 2016/1976, see Cohen et al., 2002, for how perception could be structured without reference to objects). This means that these creatures inhabit a world with an ontology that strongly differs from our own. They can discriminate between features and clusters of features and form expectations about changes in features and clusters of features. However, they lack the idea that these feature clusters belong to entities that are selfsame, that is, have identity criteria. Two ordinary objects, for instance, cannot be in the same place at the same time; and any object must take a continuous path through space.

Note that the claim being made is not about perceptual binding. We do not maintain that infants cannot discriminate clusters of features on several perceptual dimensions and learn which combinations of patterns are to be expected. It is uncontroversial that young infants (not to mention animals) can do this. They also "classify" features, feature changes, and feature interactions according to similarities and differences, attend and react appropriately to changes in their environment, and successfully predict the outcomes of their interactions with things. From the neurotypical human adult's point of view, feature patterns and feature changes are usually correlated with objects. Thus, according to the proposed picture, young infants can adjust their behaviour to *what we call objects* without themselves structuring their sensual input into objects. The claim is that infants do not represent those feature patterns based on identity criteria for selfsame entities: as spatiotemporally individuated objects, identifiable as the same over time—not even briefly and in the immediate surrounding.

Significantly, infants could still use "classification terms" to refer to features and feature clusters in a world without objects. A child would grasp a particular ball, for example, not as a particular individual that persists through space and time but as a ball-like pattern of features. The child would still interact with the ball in ways that she would not interact with other non-ball feature clusters—for example, by using the classification term 'ball.' Nonetheless, while to an adult observer, the child would seem to be referring to an object, the child's interpretation of the world would not

contain objects. The young child would experience only (relatively stable) clusters of features.

The main difference between a feature-based ontology and an object-based ontology is that the former lacks the idea of (token) identity. Agents inhabiting a feature-based world would experience similarities and differences between features. Thus, they could grasp that apples look and taste alike and are different from strawberries. They could also act on (what for us are) objects based on prior experience of how some perceptual patterns predict others. Thus, suppose that an infant wanted to stroke a cat. Based on her prior experience with cats, the infant might predict how an overall feature pattern containing a cat-shaped cluster of features would likely change. It might then successfully anticipate how it would have to behave to stroke "the" cat.

Nonetheless, if it could do this successfully, her considerations would not involve whether the cat-shaped feature cluster is selfsame. In a feature-based ontology, non-objectual thinkers would lack a grasp of token identity—for example, whether the cat I saw some seconds ago is or is not the same cat as the one in front of me now. Since such thoughts are not part of a feature-based thinker's mindset, questions about token identity do not arise, irrespective of whether objects leave the perceptual environment (e.g., in occlusion or containment events) or stay within the perceptual environment. The ability to ask oneself questions about identity (about the selfsameness of objects) would come only later. In Tugendhat's view, which we shall introduce shortly, this could happen only with the mastery of singular terms. In particular, object individuation requires the development of an intersubjective spatiotemporal coordinate system within which objects can be located.

For feature-based thinkers, other questions also do not arise. If we think about a particular apple, we can wonder what that apple was like before we got it, how it might have changed throughout its existence, and what the future has in store for it. However, these thoughts become possible only when we can think about objects—that is, as things located in space and persisting over time. Entities that move in and out of an agent's experience—for example, by leaving the agent's visual field before returning—would not be experienced as coming and going, for the question of whether the returning individual was the same as the one that left would not arise. To paraphrase Strawson (1959), a cat that left the scene and then returned might motivate the feature-based thinker to think "More cat" but not to think "The cat is back." To understand movement, one needs to understand identity and vice versa. Whatever is to be understood as moving must minimally be conceptualised as selfsame. Without a notion of identity, one could notice only processes of feature change. As with individuals who leave and return to a scene, locations inside and outside a feature-based thinker's experiential situation are not seen as selfsame locations. Thus, feature-based thinkers might revisit what for us are the same places and even navigate complex environments confidently to reach them—as when, for example, chimpanzees revisit the same high-yield fruit trees using direct routes indicative of planned travel (Normand & Boesch, 2009; Normand et al., 2009). Nevertheless, these individuals would not conceptualise particular locations. These agents would not conceive of themselves as returning to the same area but would merely return to scenes marked by familiar features. Thus, as in the cat example above, chimpanzees returning to a particular fruit tree might entertain a thought analogous to "More fruit tree" but not "Here is that good tree again". Data

indicating that chimpanzees use Euclidean maps to navigate a forest, approaching the same trees from several different directions (Normand & Boesch, 2009), would then need to be re-interpreted as further evidence that chimpanzees track and respond to stable patterns of features from different vantage points—which themselves need to be conceptualised as feature patterns.

A world without objects would then be marked by very different conceptions of what we think of as movement, objects, time, and space. Very young children might track feature changes, and outcomes of previously experienced feature interactions could be predicted by detecting statistical regularities and sequential dependencies, but such children would not track objects.

Following Tugendhat (2016/1976), Hildebrandt et al. (2022) have argued that the ability to grasp objects develops in ontogeny together with children's acquisition of a spatiotemporal coordinate system which provides identity criteria for objects. Furthermore, this spatiotemporal coordinate system is developmentally associated with learning to use special terms, namely singular terms. These terms permit their users to track objects as tokens over time because the usage rules of basic singular terms (spatial indexicals) evoke the acquisition of a spatiotemporal identification system. If this is right, then if we want to understand how infants and other feature-based thinkers structure their environment and the transition of infant- to adult-like ways of thinking, an account of the foundations and development of object reference should be given.

Overall, there are two general ways in which object reference may depend on the acquisition of singular referring terms: First, it could be the case that by learning to master singular referring terms, assumed feature-based thinking infants can acquire the notions they lack—let us call this the *Sufficiency Claim*. To make a compelling case for this claim, one would need to show (1) how a speaker can come to grasp the existence of objects through the use of singular terms and (2) how a feature-based thinker could learn to use singular terms. The Sufficiency Claim does not preclude that mastering singular referring terms is the only way for assumed feature-based thinking infants to acquire object reference. There may be alternative ways—including non-verbal ones—to acquire an understanding of objects. We will argue for the Sufficiency Claim in the next section.

Second, it could be the case that *only* through the mastery of singular referring terms could assumed feature-based thinkers come to understand the notions of token identity, objects, space, and time—let us call this the *Necessity Claim*. To argue for this claim, one would need to demonstrate (3) that the spatiotemporal coordinates that provide identity criteria necessary for object individuation are missing without the particular system of singular terms. We cannot argue for this claim here. By showing that two prominent non-verbal accounts of object reference fail, we aim to give initial credence to the Necessity Claim.

3 Singular terms and object cognition

Spelling out the relation between object cognition and singular reference was the aim of many twentieth-century analytic philosophers (cf. Russell, 1905, 1956; Searle, 1958; Strawson, 1959; Quine, 1960; Donnellan, 1966; Tugendhat, 2016/1976; Kripke, 1980;

Evans, 1982). Russell, who sought to understand reference by explaining the function of proper names, concluded that the "ambiguous proper name" 'this' is the only logically proper name that refers directly. He concluded that all references to spatiotemporal objects must be demonstratively-perceptually grounded by a demonstrative act, using the demonstrative term "this" to give momentary names (Russell, 1905, 1956). Like Russell, Strawson (1959) pointed out that reference to objects relies on demonstratives, but in contrast, he focused on the specific indexicality of demonstratives: While for Russell, the term-object-relation does not reach beyond the particular use of a demonstrative, Strawson characterised identification as resulting from demonstrative acts which fix referents in the context of speaker-utterances by locating them in the surrounding space. Like Russell and Strawson, Tugendhat argues for the importance of demonstratives when referring to objects. However, he goes further still by elucidating a system of spatiotemporal relations that is not only fixed by using a demonstrative in a particular perceptual situation but by using a system of inter-defined demonstrative terms (Tugendhat, 2016/1976). Following this line of thought, singular referring terms play a pivotal role in thinking about objects (also cf. Perry, 1979).

Tugendhat (2016/1976), in particular, argues that the ability to track individual objects comes only with the mastery of singular referring terms. Moreover, he argues, acquiring these terms is made possible by mastering pairs of indexicals like 'here' and 'there' and 'this' and 'that'. Tugendhat's crucial premise about the origins of object reference derives directly from his understanding of what it takes to refer to objects and to understand the use of an object-referring term. He builds on the distinction between classification expressions and singular terms and notes that the former may have uses that do not require reference to objects. With classification expressions, speakers (including infants) can classify clusters of features—as when an infant reaches for an object and calls out "Ball!" or "Blue!" or points to something and says, "Want that!" Since these language uses are consistent with a merely feature-based ontology, their use is insufficient to demonstrate singular reference.

Singular thought requires being able, in principle, to give a satisfactory answer to the question, "Which one (out of all) is it?". A satisfactory answer to this question forestalls the follow-up question, "Ok, but which one is *that*?". Because non-spatiotemporal definite descriptions would always leave open the possibility that several objects satisfy the description, according to Tugendhat, no non-spatiotemporal definite description alone could provide a satisfactory answer to this question.[2] This means that to individuate an object, one must, in principle, be able to pick it out from all other objects by specifying its location in space and time. A cognitive system needs access to a background system of spatiotemporal coordinates that provides the criteria by which objects can be uniquely identified.

---

[2] Tugendhat identifies four ways in which adults use language to refer to objects:

1. By an indexical demonstrative expression ('this book', 'this mountain')
2. By a description of spatiotemporal relations ('the mountain situated at these coordinates')
3. By other unique relations to something that is already identified ('the murderer of Mr. Maier')
4. By unique properties (e.g. 'the highest mountain').

Tugendhat claims that reference types (2), (3), and (4) ultimately all lead back to (1) because only an indexical demonstrative expression answers the question "Which object is it?" with sufficient clarity that any follow-up questions would be senseless.

It is this conclusion that forms the background for Tugendhat's developmental claim. A mature system of reference and singular thought, which is itself required for object reference, requires the acquisition of an objective spatiotemporal framework for locating objects unambiguously in space. Tugendhat argues that the first stages of such a framework are acquired via developing an intersubjective coordinate system. Such an intersubjective coordinate system is established by mastering pairs of indexical terms. While feature-thinkers use terms like 'that' and 'here', their use of these terms is consistent with a feature-based ontology and thus insufficient for mastering the use of indexical terms. The meaning of indexicals is not exhausted by their referent in a given perceptual situation. The meaning of 'this', for instance, is only fully understood by a speaker who grasps that the object denoted for her by the word 'this' would be referred to by 'that' for her interlocutor. The perceptual situation in which 'this' is used contrasts with another situation in which the *same* object is referred to by saying 'that'. This basic substitutability of demonstratives is part of their meaning. If one does not know that 'this' is systematically substitutable with 'that', one does not grasp the meaning of either. In Tugendhat's words:

> [I]t is the demonstrative expression itself which refers beyond the situation in the requisite manner by being used in such a way that one knows that it can be replaced by *other deictic* expressions if the same thing is referred to from another situation. (Tugendhat, 2016/1976, p. 343, original emphasis)

Tugendhat argues that this kind of term substitution principle incorporates a basic idea of object identity over different situations by providing a very elementary spatiotemporal frame of reference. This frame of reference should not be conceived as an elaborate, say, Cartesian coordinate system. There is no need to fix an allocentric origin nor a need for the axes to be scaled. Only an 'ordinal' space is required in which the overall pattern of relations is preserved by the 'metric' provided by the substitution conditions of 'here' and 'there'—that is, a system of situative points in space. This very simple term-substitution-based coordinate system gives identity criteria for the individuation of objects.

Thus, the critical claim of Tugendhat's argument is that the spatiotemporal frame of reference that adult humans use to refer to objects emerges from the mastery of spatial indexicals. Tugendhat attempts to show that this system of substitutable terms is a constitutive element of an objective spatiotemporal coordinate system[3]; and that it provides the first identity criteria for objects. Such identity criteria are needed for *any reference* to objects—with and without language.

In Tugendhat's view, grasping those term substitutions is only the first step in a two-step developmental process. After establishing an indexical-based intersubjective system of reference grounded in indexical substitution, in the second stage of the developmental process, local and temporal points of this intersubjective coordinate system

---

[3] It might be argued that infants use and understand declarative pointing well before spatial indexicals are learned and that they can thereby refer to objects before learning demonstratives. However, note that declarative pointing gestures accompanying early triangulations cannot fulfil the substitution function of spatial indexicals. While their function to signal directions might be crucial for human communication (and for developing spatial indexicals), they do not mark different distances, nor can pointing gestures be substituted for each other as they are not systematically distinct.

are supplemented by other spatiotemporal descriptions. This culminates in an objective system of coordinates for locating objects in space—for example, in the form of longitude and latitude coordinates—and an objective system for marking time. Thus, to develop a full-fledged object concept, one needs to supplement the intersubjective localisation system with an objective one. To this end, Tugendhat further attributes a significant role in identifying objects to descriptions that function to locate an object. While the first stage is mastered when an agent can use linked pairs of indexicals interchangeably, the second stage is mastered when those can be substituted for objective localising definite descriptions.

4 Learning to use spatiotemporal indexicals

The previous section leaves unexplained how someone who cannot individuate objects could *learn* to use singular terms. However, if infants are feature-based thinkers and if object reference is possible for those who have mastered singular reference, and singular reference requires the grasp of an objective spatiotemporal system of coordinates, then—if Tugendhat's view is to be viable—it must be possible to acquire a grasp of this coordinate system without yet understanding either objects or singular reference. After discussing empirical findings on how children learn spatial indexicals, this section describes a possible developmental trajectory from feature-based term usage to an intersubjective spatiotemporal coordinate system by spatial indexical term use.

Spatial indexicals are among the first words that children use in their early language production, being often the first noncontent words used together with pointing gestures (Clark & Sengul, 1978; Diessel, 2006; Diessel & Monakhov, 2022; González-Peña, 2020; Kita, 2003). They appear in pairs marking a distance contrast ("this"/"that") and are language universal (e.g., Diessel, 2006; but cf. Levinson et al., 2018). Across languages, children's use of demonstratives decreases with age while other types of spatial referring terms become more frequent, suggesting that early demonstrative use expresses an initial frame of reference (Diessel & Monakhov, 2022). Spatial indexicals are seen as providing a conceptual frame of reference emerging prior to all other frames (Tanz, 1980).

While those terms are among the first and most used words in early childhood, studies revealed that children's comprehension of spatial indexicals is not adult-like (Chu & Minai, 2018; Clark & Sengul, 1978; González-Peña, 2020; Webb & Abrahamson, 1976). Clark and Sengul (1978) and González-Peña (2020) agree that the relative distance feature of spatial indexicals—'this' is closer to the speaker than 'that'—is learned around age four or five. However, de Villiers and de Villiers (1973) argue that children are able to understand spatial indexical terms at 3 years of age, while Webb and Abrahamson (1976) found that it is only comprehended by children at the age of seven (also cf. Gonzalez-Peña et al., 2020). In a comparative study with English- and Mandarin-speaking children, Chu and Minai (2018) found that comprehension of demonstratives is above chance around children's fifth birthday but is still non-adult-like in 6-year-olds. Overall, extant empirical findings are inconclusive as to the exact age range at which different semantic features of spatial indexicals are acquired. To

our knowledge, no study has addressed the age at which children can substitute spatial indexicals.

Linguistic research on spatial indexicals and psychological research on joint attention provide evidence that spatial indexicals constitute a universal class of expressions that is of fundamental significance for cognition (Diessel, 2014). In a similar vein, Castañeda (1966, 1968), Kaplan (1989), and Perry (2000) have highlighted the essential role of indexicals in language and thought. According to them, sentences involving indexical expressions are not reducible to sentences without indexicals. Additionally, Martin and Hinzen (2014) and Hinzen and Sheehan (2015) have argued that indexicals have no lexical content but are entirely defined by the "grammar of their use". In particular, when pronouns are used indexically, they express reference without the help of lexical content—something no nominal ever does (Chomsky, 2000; Martin & Hinzen, 2014; Hinzen & Sheehan, 2015, p. 173). Learning those context-specific rules poses challenges that do not arise for learning the meaning of expressions with lexical content. The latter seems easier because what the term refers to is speaker-independent and stays constant. For instance, 'cat' refers to anything that looks cat-like irrespective of who utters it. As a result, it might suffice to associate the sound pattern 'cat' with cat-like feature patterns for primary mastery of the term. On the other hand, learning indexicals requires that one learn that the feature pattern an indexical expression refers to on a given occasion does not fix the future reference of the same term. The meaning of 'here' and 'there' involves a pattern of substitutions that differs from non-indexical expressions. To understand their meaning, one must grasp that the positions of speakers and their interlocutors determine reference and that different terms must be used to refer to the same thing depending on speakers' relative positions. What is *here* for one speaker is *there* for another. Thus, the meaning of spatial indexicals is not exhausted by any object-specific or person-specific feature pattern correlation. The only other kind of feature pattern correlation available for learning holds among symbols.

Here is how feature-based thinkers can begin to learn rules for using indexicals without yet understanding their substitutional and spatiotemporal character: they might start by interpreting spatial indexicals as reachability classifications relative to speakers. Speakers can be grasped in the form of feature patterns, just like other features in one's environment that are associated with what, for us, are objects. And reachability can be grasped as a feature configuration that allows for specific feature interactions. Roughly speaking, reachability might be understood as an expectation about the possible interactions between hand features and features of the nearby environment. Nearby features could thus be understood as "reachable", and a preliminary understanding of the indexical 'here' could be achieved accordingly. For example, when uttered by a speaker, 'here' could be understood as 'reachable-for-the-speaker'. Correspondingly, 'there' would be understood as 'unreachable-for-the-speaker' (Coventry et al., 2008, 2014; Rocca et al., 2019).

Were a feature-based thinker asked, "Please, give me that spoon there", they might disambiguate between different aspects of their sensory scene containing spoon feature patterns by determining spoon feature patterns that are unreachable-for-the-speaker and handing one over to fulfil the request. Thereby, a feature pattern (the spoon) is correlated with a speaker in a manner that could support the interpretation of utterances even without knowledge of the substitution rules governing the use of indexicals. Under

the assumption that young children are feature-based thinkers, their ability to memorise and react to others' expressed goals even when these are not shared—as demonstrated in so-called helping experiments (Warneken & Tomasello, 2007)—shows that associating features and persons is possible for feature-based thinkers.

While the feature pattern associated with 'this here' might be interpreted based on reachability, this initial reachability sense of spatial indexicals is bound to the speakers when uttering a spatial indexical. The speaker feature patterns for whom the reachability of another feature pattern is determined correspond to specific speakers. They are not initially generalised to what would be reachable for anyone from a given position. Positions cannot yet be discerned. However, children observe that the speaker for whom the feature pattern associated with "this here" is reachable might change her position, and another speaker might take her place. As a result, "this here" can neither be interpreted based on object-feature correlations (anything could be here) nor based on person-feature correlations (anyone could be here). This disconnects the use of "here" and "there" from any direct association between features and symbols.

In effect, in combination with the quasi-predicative use of classificatory terms (that are associated with feature patterns), the feature-independent use of "here" and "there" leads to striking difficulties: For a feature thinker, conversations of the following form appear to violate the usage rules for affirmatives. Ava: "Could you hand me the spoon over there?" Karim: "Do you mean this one here?" Ava: "Yes, that one there." While utterances involving indexicals like 'here' and 'there' might be interpreted in a preliminary way using the reachability rules described above, these rules would give rise to puzzling interpretations that seem to diverge from how affirmation is ordinarily used with non-indexical, quasi-predicative terms. For example, the above conversation would be understood as a transition from << Give + spoon-feature + unreachable-for-Ava >> via << spoon-feature + reachable-for-Karim? >> to << Yes + spoon-feature + unreachable-for-Ava >>. For a feature-based thinker, this conversation would have to be interpreted analogously to a non-indexical quasi-predicative sentence similar to: "Could you hand me the water?" "The juice?" "Yes, the water." Tensions that arise from such improbable interpretations should motivate feature-thinkers to search for a better basis for interpreting exchanges that involve indexical substitution.

The lack of stable feature-symbol associations highlights the replacement rules for symbols and might impose a purely language-internal (symbol-replacement) meaning aspect onto 'here' and 'there'. Under certain circumstances, 'here' must be replaced by 'there'. For a feature thinker, 'here' and 'there' might be connected to each other, almost like cat features are connected with the symbol 'cat'. The difference is that, in the case of 'here' and 'there', the correlated features are both symbols. The imposition of such a symbol-replacement meaning aspect partly disconnects 'here' and 'there' from language-external associations.

This language-internal meaning aspect of spatial indexicals leads to the acquisition of an initial spatial frame of reference that can be understood analogously to how physical quantities are made accessible through the scientific development of measurement structures (Mari, 2003, 2005). Measurement consists in assigning elements of a numerical structure to the elements of an empirical structure in a way that preserves certain operations on the elements of the empirical structure. For instance, adding numerical weights corresponds to conjoining masses on a scale. Measurement

promotes our understanding of the measured quantities insofar as characteristics of the numerical structure can be "read back" into the empirical structure. Some of the characteristics of the empirical structure could not be grasped without the development of measurements, especially in the case of more theoretically laden quantities (see Chang, 2007, for a discussion of temperature measurement).

In an analogous way, the symbolic structure consisting of spatial-indexical substitutions comprises the means to "read back" the characteristics of places into the empirical structure of feature patterns that are reachable/unreachable-for-speakers. Places relate in the same way as "here" and "there" are substituted for each other for different speakers at different "reachability distances" from attended-to feature patterns. The structure of symbol substitutions provides a relative ordering of places as nearer and further away. As a result, places are individuated in social interaction by their relative position in a feature-independent, symbol-based spatial frame of reference.

This story of the acquisition of spatial indexicals and their role in developing a spatial frame of reference is a how-possibly story that shows how children, assuming they are feature-based thinkers, *might* acquire spatial indexicals, constituting a fundamental class of singular terms and providing an initial spatial frame of reference in which objects can be individuated. Tugendhat (2016/1976) argues that mastering singular terms suffices for understanding objects. Our how-possibly scenario provides a complementary account of how learning spatial indexicals could suffice to acquire object reference. The following section discusses whether there might also be non-verbal paths to this capability.

5 Is there non-verbal object individuation?—Pylyshyn's FINSTs and Burge's perceptual objectivity

Historically, there have been several attempts at explaining how human thinkers come to understand objects. Like Tugendhat, some researchers argue that the process of coming to structure one's environment into objects is bound to one or the other aspect of linguistic competence (e.g., Davidson, 1963, 2001; Quine, 1960, 1974). Others have stressed the importance of specific high-level cognitive capacities (Strawson, 1959; Evans, 1982, see Burge, 2010 for a discussion). More recently, researchers tend to think that object individuation is a non-verbal process that occurs pre-conceptually in perception (Burge, 2010; Butterfill, 2020; Leslie et al., 1998; Peacocke, 1992; Pylyshyn, 2001, 2007). For reasons of scope, we cannot do justice to all available accounts. A brief discussion of two prominent non-verbal accounts will have to suffice for motivating the Necessity Claim: Pylyshyn's object indexes (FINST) and Burge's perceptual objectivity account.

According to Pylyshyn's (2001, 2007) model of visual object tracking, a small number of objects is individuated by the early visual system by indexing their position in egocentric space. This ability is non-verbal and said to underlie more sophisticated capacities to refer to particular objects. Pylyshyn's account has been widely accepted as a basis for explaining object individuation in recent work in developmental psychology and philosophical discussions of the foundations of object individuation (see, e.g., Butterfill, 2020, Chap. 6).

On the other hand, Burge argues against the mentioned classical accounts on the basis that they over-intellectualise objectivity. For Burge, objectivity results from perceptual constancies, which provide us with the attributives (for colours, shapes, or locations) that play a role in referring to objects.

However, both accounts presuppose what they want to explain. Pylyshyn assumes that the environment is structured into objects and that the mind causally latches onto them. Burge treats spatiotemporal location as if it were akin to any other featural attribute and thereby presupposes identity outright.

5.1 Object indexes

Pylyshyn (2001, 2007) attempts to answer how the mind connects with the world. The overall idea is that this connection results from our sensory contact with the things in our environment. Vision is the investigated case at hand. It is argued that early vision parses the visual input into sensory particulars, also referred to as individuals or things, and represents them independently of any of their properties. These representations are introduced with an analogy. They are said to function like an indefinitely extendable finger that touches a sensory particular in one's environment and sticks to it—hence the illustrative name FINSTs (for FINgers of INSTantiation). There are roughly four or five such FINSTs in the human early visual system. Pylyshyn argued that FINST indexes provide a causal link between the mind and the world. They constitute a foundation for our object individuation and a basis for our object-dependent thoughts. In his words:

> The proposal is that there is in the early visual system a primitive mechanism that accomplishes two tasks: it individuates things in the visual scene and provides a direct reference to a small number of them. In this statement, I mean by "individuates" that *the visual system parses the visual world and segregates things in space and time so they can be treated as enduring individuals*. This entails not only carrying out a figure-ground segregation (which is segregation in space), but also solving the correspondence problem (which is segregation in time). By a "direct reference" I mean essentially a demonstrative reference or an opaque pointer or index (which I have called a FINST) that allows epistemic access to a small number of the spatially and temporally segregated individuals without specifying any of their properties. (Pylyshyn, 2007, p. 206, emphasis added)

Pylyshyn (2007) argues that this kind of demonstrative thought rests on the causal or nomological dependency of the creation (and maintenance) of a FINST index on the appearance of an individual (see Pylyshyn, 2007, p. 82). The visual system respects these individuals' spatial cohesion by visually segregating a figure from its background. And by maintaining a FINST over the presence of a sensory particular, FINSTs respect the temporal endurance of individuals.

Pylyshyn (2007) is cautious to distinguish object individuation and recognition as understood by philosophers like Strawson (1959) or Quine (1974)—which requires

high-level conceptual capacities such as a notion of identity—from the ability to individuate and reidentify sensory particulars perceptually—which is to be nonconceptual (ibid., p. 53). Moreover, he notes that "the notion of individuating has a narrower meaning here than in the more general context where it refers not only to separating a part of the visual world from the rest of the clutter (which is roughly what it means here), but also providing identity criteria for recognitional instances of that individual" (p. 21, fn 6).

The meaning of "individuating" in the sense of "separating a part of the visual world from the rest of the clutter" is narrower than the general notion of individuation insofar as only "under certain conditions (viz., the conditions that allow indexing and tracking) FINSTs do allow us to individuate and even to reidentify certain sensory individuals: They allow us to maintain the identity of tracked objects as enduring individuals" (ibid., p. 53).

According to Pylyshyn, under which conditions the visual system can individuate and track sensory individuals is an empirical question. For instance, the endpoints of lines and "objects that appear to liquefy and "pour" from one place to another or that stretch and slink in wormlike fashion can't be tracked" (ibid., p. 95 f.). At any rate, the conditions enabling tracking are not essential to what is tracked. While certain properties might cause the tokening of an index, these properties are not what the index refers to. FINSTs refer to the *bearers* of the properties that cause them to be tokened (ibid., p. 96).

Importantly, the early visual system "delivers a reference to a selected sensory individual (call it x) to which the argument of a predicate can be bound, so that properties may be subsequently predicated of x—presumably starting with such predicates as *Object(x)* or *Location(x,L)*" (p. 95). Following the philosophical discussion of reference and description (Perry, 1979; Quine, 1960; Strawson, 1959), Pylyshyn (2007, p. jj) argues that reference to individuals cannot be a matter of matching descriptions all the way down. Ultimately, there must be a direct, that is, non-descriptive, reference to particulars that could fill the argument positions of predicates. Giving an account of how this direct reference is grounded in perception is the ultimate goal of Pylyshyn's book.

In itself, the causal dependency of the tokening of a FINST on an object in the visual scene does not ensure that the FINST indexes a particular instead of any other aspect of the visual scene causally involved in creating and maintaining a FINST. The tokening of a FINST index causally depends on various aspects of the environment, notably, all features of an object that are causally relevant for the tokening of a FINST and all intermediate causes. As far as causal dependency is concerned, FINSTs could index any of these. Alluding to causal dependency does not explain how a FINST index fixates on a particular distal sensory individual. As Pylyshyn acknowledges, causal processes by themselves are insufficient to determine referents because "all links [of a causal chain] are equally part of the causal story" (p. 97). And Pylyshyn does not offer an account of how to single out the referent of a FINST from the causal chain that leads to its tokening, because "what determines the particular link in the causal chain that has the predicated property… is one of the "big questions" about how reference is naturalized and is beyond the scope of this monograph" (p. 96).

Not aiming to resolve long-standing philosophical disputes in an empirical monograph about early vision is fair enough. However, deferring to the unresolved philosophical question of reference in a theory that aims to explain how the mind connects to the world referentially is, at best, underwhelming. Deferring to an unresolved philosophical question is especially unfortunate in this case because there is a tendency to employ FINSTs in current philosophical discussions of how the mind represents objects in the first place, suggesting that the issue has been resolved empirically (e.g., Perner et al., 2015; Recanati, 2012).

Critically, Pylyshyn overlooks that a cognitive system need not structure its environment into particulars (of whatever sort) at all, be them objects, proto-objects, individuals, or sensory particulars. The empirical findings and the idea of a causal dependency of FINSTs on what appears on the visual scene are compatible with FINSTs indexing feature patterns without thereby having or creating any sensitivity to the identity criteria of sensory particulars—even under the restricted circumstances enabling indexing and tracking. Without sensitivity to identity criteria, the visual system cannot be said to track particulars of whatever sort. The FINST account could just as well apply to the visual system of thinkers who do not come to structure their environment into particulars at all.

A proponent of the FINST account might bite the bullet and concede that FINSTs do not serve to structure a cognitive system's environment into objects. After all, FINSTs are not presented as explaining individuation in the full-blown sense. However, if FINSTs are to fill the argument positions of predicates, they must refer to particular things. Simply having a FINST tokened does not ensure that it is latched on to a sensory particular or even the same sensory particular for as long as it is maintained. The FINST account does not answer how the perceptual system determines whether a FINST still latches on to the same sensory particular or whether it is to be replaced by another FINST, latching on to another sensory particular. In terms of Pylyshyn's analogy: it must be ensured that the finger sticks to one and the same object for as long as the same FINST is tokened. This requires sensitivity to objects' individuation criteria.

Note that figure-ground segregation and solving the correspondence problem are not sufficient for the individuation of sensory particulars—even if this individuation only is to succeed under the restricted conditions during which a figure is segregated from the background and the visual system determines which aspect of the visual impression corresponds to which aspect of previous impressions. Figure-ground segregation and solutions to the correspondence problem need only be sensitive to perceptual similarities, to-be-expected feature patterns, and changes in sensory features. Thus, segregating a figure from its background does not amount even to the momentary individuation of a sensory particular. No further account is offered for how the early visual system is sensitive to sensory particulars' identity criteria. As a result, the adduced evidence shows that the early visual system is sensitive to feature patterns, but it need not thereby individuate sensory particulars.

Pylyshyn (2007, p. 17 f.) argues that FINSTs index sensory particulars by directly referring to them without relying on any of their properties or their location, much in the way demonstratives in natural languages directly refer to objects. However, the idea that FINSTs directly refer to sensory particulars much in the same way as

demonstratives in natural languages refer to ordinary objects rests on a falsifying simplification of the functioning of demonstratives. Even if we assume that demonstratives refer directly—in the sense that demonstrative reference does not rely on any of the referent's properties—demonstrative reference in natural language cannot be understood simply on the model of an assignment of singular terms to objects similar to a model-theoretic interpretation of a formal language. Structurally, the rules governing the use of demonstratives do not allow for a one–one correspondence between singular terms and referents as the characteristic meaning postulate. The Russellian idea that demonstrative uses create short-lived logical names (this$_1$, this$_2$, this$_3$, …, cf. Pylyshyn, 2007, p. 95) is incorrect. Demonstratives must be substituted for each other in a speaker-position-specific way. Someone uses "this" correctly only if she knows that she has to substitute "this" with "that" in another situation to pick out the same object. This aspect of the meaning of demonstratives is not captured by an assignment of individuals to uses of demonstratives, and it does not appear in the FINST account. FINSTs are not substituted. This alone suggests that there are important differences between FINSTs and demonstratives.

From the Tugendhat-inspired perspective of this article, the central characteristics of demonstratives indeed explain how cognitive systems can refer to objects. But it is their mutual substitutability that explains reference, not the fact that they refer without relying on objects' properties. Substitution is needed to individuate objects because it provides the entrance to an intersubjective spatial coordinate system that allows employing feature-independent identity criteria for objects. FINST indexes do not exhibit the required pattern of substitutions, nor do they provide a spatial coordinate system, or any other means, that would explain a perceptual/cognitive system's sensitivity to objects' individuation criteria. As a result, the FINST account cannot explain how the mind connects to sensory particulars in the first place.

5.2 Perceptual objectivity

Burge sets out to explain empirical objectivity, that is, to give an account of the "minimal constitutive conditions on objective representation of the physical environment" (2010, p. 156). Throughout the book, it is argued that the primary form of objective representation occurs in perception. Traditional views of empirical objectivity by Russell (1905), Strawson (1959), Evans (1982), Quine (1960), and Davidson (1963, 2001) are criticised by Burge as being individual-representationalist. In the analysed accounts, it is assumed that representing an objective physical environment requires the representation of individuation criteria for the objects that inhabit the environment. Individual-representationalist accounts understand objectivity as a conceptually loaded, high-level cognitive capacity. For Burge, however, objectivity results from the exercise of purely perceptual capacities that partly consist "in an ability to single out bodies from a background, locating them in space, to perceive them in relation to other bodies, and to track them over time" (p. 169). Perception "includes the capacity to select individual things in one's field of view, to reidentify each of them under certain conditions as the same individual thing that was seen before, and to keep track of their enduring individuality despite radical changes in their properties" (ibid., p. ix).

Burge cites empirical evidence to the effect that many animals, including human infants, parse their sensory impressions accordingly without possessing any conceptual capacities that would provide them with criteria for objective representation. In Burge's own words:

> What is remarkable is how primitive the origins [of empirical objectivity] are. Perception is the root objectivity and, I believe, the developmental and phylogenetic origin of genuine representation. Perception is shared by humanity with many arthropods, reptiles, birds, and fish, and probably with all other mammals. Perception is constitutively independent of capacities for propositional thought. (2010, p. 548)

Burge's account is based on the insight developed by Strawson (1959) that reference to particular objects cannot entirely rest on descriptions and that demonstrative-like reference has to underlie all forms of empirical objectivity. Following Strawson, Burge concedes that it is plausible that a comprehensive spatial frame of reference provides the identity criteria for physical objects. Burge argues, however, that this plausibility does not carry over to the project of giving minimal conditions for empirical objectivity and that no argument can be found either in Strawson or any of his heirs to the effect that demonstrative-like reference to physical objects requires knowing any set of identity criteria. Identity criteria for the objects referred to do not figure in Burge's minimal conditions for empirical objectivity.

At the same time, Burge acknowledges Strawson's point that spatial relations are critical for empirical objectivity:

> It is, I think, impossible to represent bodies as such without being able to represent specific spatial properties and relations as such. And it is impossible to have a conception of bodies as mind-independent without having some spatial conceptions that one associates with those bodies (2010, p. 172).

However, according to Burge (2010), representing spatial relations as such need not involve a conception of these relations. It is sufficient that the perceptual system operates under principles that we can describe as involving spatial relations. In particular, "[i]n order to use perceptual concepts to distinguish bodies as same or different, [… i]t is enough to be able to track sameness and difference of particular bodies perceptually, and to incorporate this ability into a propositional structure…" (p. 170). Burge thinks that this is shown by the observation that many animals that do not possess concepts—especially concepts of individuation criteria for physical objects—can nonetheless perceive physical objects as such (ibid.).

Burge's account of empirical objectivity is backed by the idea that to explain perceptual or cognitive capacities, it is sufficient to describe these capacities as operating "under principles that *we* can understand and use in explaining them" (p. 169, original emphasis). The principles need not be understood by the being whose cognitive or perceptual capacities we are trying to explain. "No general criteria or principles need be represented, conceptualised, understood, or otherwise grasped, even implicitly" (p. 170).

While one must concede that no such principles need be accessible to the cognitive system, it is nonetheless the case that the principles employed to explain a cognitive

or perceptual capacity must capture the structure of the cognitive or perceptual system itself if that explanation is to be compelling. If the representational content attributed in a psychological explanation is supposed to be explanatorily fruitful, it must make a psychological difference. However, when a cognitive or psychological system is explained in terms that *only we* understand, we run the danger of making an incongruous ascription. To borrow a phrase Glock (2007) employed with a slightly different intent: the risk is "that the rich mental idiom we employ has conceptual connections that go beyond the phenomena to which it is applied". Attributing an ability to perceive physical objects to many animals, including human infants, is potentially one such case.

Let us introduce a distinction made by Hildebrandt et al. (2020) to make this point. Cognitive systems describable by principles that *we* can understand as involving reference to physical objects without committing to an implicit or explicit grasp of these principles by the cognitive system itself can be said to refer *de re* to physical objects.[4] On the other hand, cognitive systems that implicitly or explicitly make the distinctions characteristic of what we call reference to physical objects can be said to refer *de dicto* to physical objects.

The distinction is analogous to the distinction between *de re* and *de dicto* attributions of propositional attitudes (Quine, 1956; Schwitzgebel, 2021) but generalised to any cognitive explanation, including non-propositional cognitive systems. *De dicto* attributions of cognitive capacities are intended to capture how a cognitive system processes its input or structures its environment. The vocabulary used in *de dicto* descriptions of cognitive capacities must capture the distinctions and transitions made by the cognitive system in question. If no terms are yet available to capture a cognitive system's inner workings, new technical terms must be devised. *De re* descriptions of cognitive capacities do not bear this commitment.

The problem with attributions of reference *de re* is that they are not explanatory of *how* a cognitive system comes to behave in a certain way. *De re* descriptions can be useful for a systematisation of behaviour and for making some kinds of predictions. Nevertheless, attributions of *de re* reference do not capture the 'inner workings' of a system. If we are to explain how a cognitive system comes to structure its environment into objects and their properties—even if we are merely after minimal conditions for empirical objectivity—an account is needed that captures the way a cognitive system structures its sensory input. We need an explanation of how *de dicto* reference is possible.

Most importantly, by attributing reference *de dicto* to physical objects, one is committed that the system makes all the distinctions and transitions characteristic of someone who structures her environment into *objects*. Since objects have spatiotemporal identity criteria, a system capable of reference to objects *de dicto* must be sensitive to the spatiotemporal individuation criteria of physical objects. This is not to say that these criteria must be accessible to the cognitive system. It is only to say that

---

[4] Note that this use of "*de re*" differs sharply from Burge's use of the term: "To be a *de re* state or attitude is to bear a peculiarly direct epistemic and representational relation to a particular referent in perception or thought" (Burge, 2009a). As we employ the term, it does not characterise any epistemic or representational relation whatsoever. Our *de dicto/de re* distinction concerns the commitments we make in describing a cognitive system.

the cognitive system must be operating based on all relevant distinctions. It is not sufficient that a cognitive system can distinguish what we call objects by their perceptual features—a duck from a ball, say, or a green cube from a red one. Even sensitivity to the relative stability of such feature patterns does not show that a system can refer *de dicto* to physical objects. As argued above, as long as the system is not sensitive to spatiotemporal identity criteria for physical objects, its perceptual capacities can entirely be explained in terms of generalisations over feature patterns.

Burge seems to endorse that perceptual objectivity is to be understood *de re* (in our sense of the term): it is sufficient to describe these capacities as operating "under principles that *we* can understand and use in explaining them" (p. 169, original emphasis). "No general criteria or principles need be represented, conceptualised, understood, or otherwise grasped, even implicitly" (p. 170). At the same time, Burge holds that for a cognitive system to individuate physical objects, it is enough that the perceptual system can "track the sameness or difference of particular bodies perceptually" (Burge, 2010, p. 170). In our sense of the distinction, Burge attributes tracking sameness or difference *de dicto*.

As noted, Burge thinks that spatial relations are critical for empirical objectivity and, thereby, for the sameness or difference of particular bodies. Sameness or difference of particular bodies can be tracked perceptually without representing or conceptualising the principles governing object individuation because, for Burge, spatiotemporal continuity is an abstract repeatable (a property) akin to other fundamental properties of objects like cohesion, boundedness, and solidity (Burge, 2010, p. 444).

However, Burge overlooks that spatiotemporal continuity is not an abstract repeatable that could be attributed to a particular. We take it that it is one of the central difficulties of Burge's account that he does not consider the fundamental difference between spatiotemporal continuity and other properties of objects. If spatiotemporal location is a property at all, it differs from all other properties of objects in that no more than one object could ever have the same property at the same time. At any moment, all objects can be distinguished by this property. This is what we would usually call identification. Other than all other common properties, spatiotemporal locations do not classify. Objects can unequivocally be identified by their location in space and time but not by any other property.

Moreover, spatiotemporal location is not perceptual. The place of a body (where it is) is determined by the relations among places. Places have their identities relative to a frame of reference and irrespective of the features that happen to be at a place. From an ability to distinguish "places" by the features that happen to be at a place, no sensitivity to the individuation criteria of places results. Deferring empirical objectivity to an ability to track "spatiotemporal continuity" would, again, amount to a *de re* attribution in our sense and would therefore lack explanatory import. As described in Sect. 2, there are alternatives available that explain animals' and infants' abilities to track what, for us, is the spatiotemporal continuity of bodies while respecting the difference between spatiotemporal location and other properties and not making any uncovered commitments concerning the implications of such attributions.

To recall, for feature-based thinkers, tracking "sameness of particular bodies" would consist in attending to certain feature patterns (that we take to belong to bodies) and having certain expectations about the changes such patterns might undergo. Such an

organism would, in effect, track feature patterns in a way that, for us, looks a lot like tracking objects. But in many cases in which bodies interact in unusual ways, such an organism would have different (or perhaps no) expectations about feature changes that would correspond to our expectations about the sameness or difference of bodies. For such an organism, the question could not arise whether a specific feature pattern is the same or not. Being able to consider whether an object is the same or not requires spatiotemporal identity criteria. This is simply because the token identity of objects *is* spatiotemporal.

In effect, token identity cannot be tracked based on similarities and differences in an organism's sensory input—as Burge seems to assume. Because objects' identities consist in their spatiotemporal positions, sameness can only be tracked by a cognitive system sensitive to spatiotemporal identity criteria. Tracking sameness, therefore, requires a feature-independent frame of reference that provides such identity criteria.

To sum up, Burge (2010) recognises the relevance of demonstrative-like reference for empirical objectivity, which he takes to result from basic perceptual capacities. To avoid too-high-level conceptual requirements for objectivity, Burge treats *de re* ascriptions (in the sense of Glauer & Hildebrandt, 2021) of perceptual and cognitive capacities to be explanatory. However, in *de re* ascriptions, no commitment to the inner workings of a cognitive system is made. As a result, no explanation of how a cognitive system comes to behave in a certain way can be given. Moreover, Burge overlooks that spatiotemporal continuity should not be conceived of as an abstract repeatable like other mundane properties of objects. Spatiotemporal continuity provides identity criteria for objects and cannot be perceived based on patterns of similarity between features and expectations about feature changes. It requires a spatiotemporal frame of reference.

6 Summary and outlook

Following Tugendhat (2016/1976) and under the assumption that infants set out as feature-based thinkers, we have argued that the acquisition of an inter-defined system of deictic singular terms (spatial indexicals) is sufficient for the capacity of object reference. The central idea is that the structure of term substitutions (this-for-someone is that-for-someone-else) can be read back onto the feature patterns that are reachable or non-reachable for speakers. By substituting "here" and "there" in the characteristic way, places are ordered as nearer and further away according to different speakers' "reachability distances" from attended-to features. As a result, places are individuated in social interaction by their relative position in a feature-independent, symbol-based spatial frame of reference.

We have then suggested that learning a spatial indexical term substitution system might also be necessary for object individuation. We have discussed two prominent attempts to explain object reference without recurrence to language capacities—Pylyshyn's FINST indexes (Pylyshyn, 2001, 2007) and Burge's perceptual objectivity (Burge, 2009b, 2010)—and argued that neither account succeeds. Pylyshyn presupposes what is to be explained, namely the spatial individuation of particulars,

and Burge illegitimately treats the individuation criteria of objects, namely their spatiotemporal continuity, as perceptual features. The failure of these two prominent accounts is diagnostic of the difficulties faced by non-linguistic accounts of object cognition.

Similar to how the FINST account conceptualises the problem to be solved by the cognitive system, it is common to assume that the main difficulty is figuring out how a mental object representation is latched on to the object it represents. One could say that the task is not to confuse objects. However, if we accept that the perceived environment of a cognitive system is not already pre-structured into particulars, the problem is to explain how the cognitive system structures its perceptual input into particulars in the first place. The task is to figure out that there are selfsame entities that could be confused at all. Because their spatial position individuates (physical) objects, object cognition requires a spatial frame of reference. Burge (2010) saw that the problem of object cognition is one of structuring the perceived environment according to spatiotemporal continuities. He overlooked, however, that a spatial frame of reference cannot be a perceptual feature. Throughout this article, we have argued that sensitivity to the identity criteria of objects could not be obtained based on features alone and that operating with a spatial frame of reference involves being able to pick out objects independently of their features.

The failure of these two accounts does not yet show that acquiring a system of inter-defined indexical terms is necessary for object reference. Realising that spatial continuity is not a feature but depends on a spatial frame of reference, however, suggests that a feature-independent identification of objects requires a *symbolic* presentation of spatial relations—irrespective of whether this requires language or not. It would have to be laid out that only the internal relations between the elements of a symbolic spatial frame of reference allow for the localisation of objects. To argue for the necessity claim, one would have to analyse the symbolic content of any (minimal) spatial frame of reference and show that it corresponds to the structure of the substitution system of inter-defined spatial indexicals. Alternatively, one could argue against the Necessity Claim by showing that there is a non-symbolic route to singular reference. A critical step in this argument would be to formulate a viable account of individuation that does not involve a system of symbols but still provides abstract spatial individuation criteria.

References

Aguiar, A., & Baillargeon, R. (1999). 25-Month-old infants' reasoning about when objects should and should not be occluded. *Cognitive Psychology, 39*(2), 116–157. https://doi.org/10.1006/cogp.1999.0717

Baillargeon, R., Li, J., Ng, W., & Yuan, S. (2009). An account of infants' physical reasoning. In *Learning and the infant mind* (pp. 66–116). Oxford University Press.

Baillargeon, R., Stavans, M., Wu, D., Gertner, Y., Setoh, P., Kittredge, A. K., & Bernard, A. (2012). Object individuation and physical reasoning in infancy: An integrative account. *Language Learning and Development, 8*(1), 4–46. https://doi.org/10.1080/15475441.2012.630610

Bermúdez, J. L. (2007). *Thinking without words* (1. issued as an Oxford Univ. Press paperback). Oxford University Press.

Blijd-Hoogewys, E. M. A., & van Geert, P. L. C. (2017). Non-linearities in theory-of-mind development. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2016.01970

Burge, T. (2009a). Five theses on de re states and attitudes. In J. Almog & P. Leonardi (Eds.), *The philosophy of David Kaplan* (pp. 246–324). Oxford University Press.

Burge, T. (2009b). Perceptual objectivity. *Philosophical Review, 118*(3), 285–324. https://doi.org/10.1215/00318108-2009-001

Burge, T. (2010). *Origins of objectivity*. Oxford University Press.

Butterfill, S. A. (2020). *The developing mind: A philosophical introduction*. Routledge.

Cacchione, T., Abbaspour, S., & Rakoczy, H. (2020). Object individuation in the absence of kind-specific surface features: Evidence for a primordial essentialist stance? *Journal of Cognition and Development, 21*(4), 534–550. https://doi.org/10.1080/15248372.2020.1797746

Cacchione, T., & Rakoczy, H. (2017). Comparative metaphysics: Thinking about objects in space and time. In J. Call, G. M. Burghardt, I. M. Pepperberg, C. T. Snowdon, & T. Zentall (Eds.), *APA handbook of comparative psychology: Perception, learning, and cognition* (pp. 579–599). American Psychological Association. https://doi.org/10.1037/0000012-026

Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus, 133*(1), 59–68. https://doi.org/10.1162/001152604772746701

Carey, S. (2009). *The origin of concepts*. Oxford University Press.

Castañeda, H.-N. (1966). 'He': A study in the logic of self-consciousness. *Ratio, 8*, 130–157.

Castañeda, H.-N. (1968). On the logic of attributions of self-knowledge to others. *The Journal of Philosophy, 65*(15), 439. https://doi.org/10.2307/2024296

Chang, H. (2007). *Inventing temperature: Measurement and scientific progress*. Oxford University Press.

Cheung, P., Rubenson, M., & Barner, D. (2017). To infinity and beyond: Children generalize the successor function to all possible numbers years after learning to count. *Cognitive Psychology, 92*, 22–36. https://doi.org/10.1016/j.cogpsych.2016.11.002

Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge University Press.

Chu, C.-Y., & Minai, U. (2018). Children's demonstrative comprehension and the role of non-linguistic cognitive abilities: A cross-linguistic study. *Journal of Psycholinguistic Research, 47*(6), 1343–1368. https://doi.org/10.1007/s10936-018-9565-8

Clark, E. V., & Sengul, C. J. (1978). Strategies in the acquisition of deixis. *Journal of Child Language*. https://doi.org/10.1017/S0305000900002099

Cohen, L. B., Chaput, H. H., & Cashon, C. H. (2002). A constructivist model of infant cognition. *Cognitive Development, 17*(3–4), 1323–1343. https://doi.org/10.1016/S0885-2014(02)00124-7

Coventry, K. R., Griffiths, D., & Hamilton, C. J. (2014). Spatial demonstratives and perceptual space: Describing and remembering object location. *Cognitive Psychology, 69*, 46–70. https://doi.org/10.1016/j.cogpsych.2013.12.001

Coventry, K. R., Valdés, B., Castillo, A., & Guijarro-Fuentes, P. (2008). Language within your reach: Near–far perceptual space and spatial demonstratives. *Cognition, 108*(3), 889–895. https://doi.org/10.1016/j.cognition.2008.06.010

Davidson, D. (1963). Actions, reasons, and causes. *The Journal of Philosophy, 60*(23), 685. https://doi.org/10.2307/2023177

Davidson, D. (2001). *What is present to the mind*. Oxford University Press.

de Villiers, J. G., & de Villiers, P. A. (1973). A cross-sectional study of the acquisition of grammatical morphemes in child speech. *Journal of Psycholinguistic Research, 2*(3), 267–278. https://doi.org/10.1007/BF01067106

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics*. https://doi.org/10.1515/COG.2006.015

Diessel, H. (2014). Demonstratives, frames of reference, and semantic universals of space: Demonstratives and frames of reference. *Language and Linguistics Compass*. https://doi.org/10.1111/lnc3.12066

Diessel, H., & Monakhov, S. (2022). Acquisition of demonstratives in cross-linguistic perspective. *Journal of Child Language*. https://doi.org/10.1017/S030500092200023X

Donnellan, K. S. (1966). Reference and definite descriptions. *The Philosophical Review, 75*(3), 281. https://doi.org/10.2307/2183143

Evans, G. (1982). In J. H. McDowell (Ed.), *The varieties of reference.* Clarendon Press.

Futó, J., Téglás, E., Csibra, G., & Gergely, G. (2010). Communicative function demonstration induces kind-based artifact representation in preverbal infants. *Cognition, 117*(1), 1–8. https://doi.org/10.1016/j.cognition.2010.06.003

Glauer, R., & Hildebrandt, F. (2021). No facts without perspectives. *Synthese, 199*(1–2), 3825–3851. https://doi.org/10.1007/s11229-020-02958-9

Glauer, R., & Hildebrandt, F. (2022). Sortals, bodies, and variables. A critique of Quine's theory of reference. *Synthese, 200*(4), 307. https://doi.org/10.1007/s11229-022-03767-y

Glock, H. (2007). *Can animals judge?* Retrieved from http://www.philosophie.uzh.ch/institut/lehrstuehle/theoretische2/forschung/concepts/publications.html.

González-Peña, P. (2020). *Spatial deixis in child development*. University of East Anglia. Retrieved from https://ueaeprints.uea.ac.uk/id/eprint/77867/.

González-Peña, P., Doherty, M. J., & Guijarro-Fuentes, P. (2020). Acquisition of demonstratives in English and Spanish. *Frontiers in Psychology, 11*, 1778. https://doi.org/10.3389/fpsyg.2020.01778

Hespos, S. J., & Rochat, P. (1997). Dynamic mental representation in infancy1Portions of this research have been presented at the International Conference on Infant Studies, Society for Research in Child Development, and Association for Research in Vision and Opthamology.1. *Cognition, 64*(2), 153–188. https://doi.org/10.1016/S0010-0277(97)00029-2

Hildebrandt, F., & Glauer, R. (2022). Singular reference. The turn to propositional thinking. *Grazer Philosophische Studien, 99*, 159–185. https://doi.org/10.1163/18756735-00000164

Hildebrandt, F., & Glauer, R. (2023). Becoming episodic: The development of objectivity. *Philosophical Psychology*. https://doi.org/10.1080/09515089.2023.2181152

Hildebrandt, F., Glauer, R., & Kachel, G. (2022). Coming from a world without objects. *Mind & Language, 37*(2), 159–176. https://doi.org/10.1111/mila.12313

Hildebrandt, F., Lonnemann, J., & Glauer, R. (2020). Why not just features? reconsidering infants' behavior in individuation tasks. *Frontiers in Psychology, 11*, 564807. https://doi.org/10.3389/fpsyg.2020.564807

Hinzen, W., & Mattos, O. (2023). Explaining early generics: A linguistic model. *Mind & Language, 38*(1), 256–273. https://doi.org/10.1111/mila.12367

Hinzen, W., & Sheehan, M. (2015). *The philosophy of universal grammar* (First published in paperback).

Kaplan, D. (1989). Demonstratives: An essay on the semantics, logic, metaphysics and epistemology of demonstratives and other indexicals. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481–563). Oxford University Press.

Kita, S. (2003). *Pointing: Where language, culture, and cognition meet*. Psychology Press.

Kripke, S. A. (1980). *Naming and necessity*. Harvard University.

Leslie, A. M., Xu, F., Tremoulet, P. D., & Scholl, B. J. (1998). Indexing and the object concept: Developing 'what' and 'where' systems. *Trends in Cognitive Sciences, 2*(1), 10–18. https://doi.org/10.1016/S1364-6613(97)01113-3

Levinson, S., Cutfield, S., Dunn, M., Enfield, N., Meira, S., & Wilkins, D. (2018). *Demonstratives in cross-linguistic perspective*. Cambridge University Press.

Lin, Y., Stavans, M., & Baillargeon, R. (2022). Infants' physical reasoning and the cognitive architecture that supports it. In O. Houdé & G. Borst (Eds.), *The Cambridge handbook of cognitive development* (1st ed., pp. 168–194). Cambridge University Press. https://doi.org/10.1017/9781108399838.012

Mari, L. (2003). Epistemology of measurement. *Measurement, 34*(1), 17–30. https://doi.org/10.1016/S0263-2241(03)00016-2

Mari, L. (2005). The problem of foundations of measurement. *Measurement, 38*(4), 259–266. https://doi.org/10.1016/j.measurement.2005.09.006

Martin, T., & Hinzen, W. (2014). The grammar of the essential indexical. *Lingua, 148*, 95–117. https://doi.org/10.1016/j.lingua.2014.05.016

Mascalzoni, E., Regolin, L., Vallortigara, G., & Simion, F. (2013). The cradle of causal reasoning: Newborns' preference for physical causality. *Developmental Science, 16*(3), 327–335. https://doi.org/10.1111/desc.12018

Mendes, N., Rakoczy, H., & Call, J. (2008). Ape metaphysics: Object individuation without language. *Cognition, 106*(2), 730–749. https://doi.org/10.1016/j.cognition.2007.04.007

Moore, M. K., Borton, R., & Darby, B. L. (1978). Visual tracking in young infants: Evidence for object identity or object permanence? *Journal of Experimental Child Psychology, 25*(2), 183–198. https://doi.org/10.1016/0022-0965(78)90076-0

Normand, E., Ban, S. D., & Boesch, C. (2009). Forest chimpanzees (Pan troglodytes verus) remember the location of numerous fruit trees. *Animal Cognition, 12*(6), 797–807. https://doi.org/10.1007/s10071-009-0239-7

Normand, E., & Boesch, C. (2009). Sophisticated Euclidean maps in forest chimpanzees. *Animal Behaviour, 77*(5), 1195–1201. https://doi.org/10.1016/j.anbehav.2009.01.025

Nucci, L. P., & Gingo, M. (2010). The Development of moral reasoning. In U. Goswami (Ed.), *The Wiley-Blackwell Handbook of childhood cognitive development* (1st ed., pp. 420–445). Wiley. https://doi.org/10.1002/9781444325485.ch16

Peacocke, C. (1992). *A study of concepts*. The MIT Press.

Perner, J., Huemer, M., & Leahy, B. (2015). Mental files and belief: A cognitive theory of how children represent belief and its intensionality. *Cognition, 145*, 77–88. https://doi.org/10.1016/j.cognition.2015.08.006

Perry, J. (1979). The problem of the essential indexical. *Noûs, 13*(1), 3. https://doi.org/10.2307/2214792

Perry, J. (2000). *The problem of the essential indexical: And other essays (Expanded ed)*. CSLI Publications.

Pylyshyn, Z. (2007). *Things and places: How the mind connects with the world*. MIT Press.

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition*. https://doi.org/10.1016/S0010-0277(00)00156-6

Quine, W. V. (1956). Quantifiers and propositional attitudes. *The Journal of Philosophy, 53*(5), 177. https://doi.org/10.2307/2022451

Quine, W. V. O. (1960). *Word & object*. MIT Press.

Quine, W. V. O. (1974). *The roots of reference*. Open Court.

Recanati, F. (2012). *Mental Files*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199659982.001.0001

Rocca, R., Tylén, K., & Wallentin, M. (2019). This shoe, that tiger: Semantic properties reflecting manual affordances of the referent modulate demonstrative use. *PLoS ONE, 14*(1), e0210333. https://doi.org/10.1371/journal.pone.0210333

Rubio-Fernández, P. (2021). Pragmatic markers: The missing link between language and theory of mind. *Synthese, 199*(1–2), 1125–1158. https://doi.org/10.1007/s11229-020-02768-z

Russell, B. (1905). II.—On denoting. *Mind, XIV*(4), 479–493. https://doi.org/10.1093/mind/XIV.4.479

Russell, B. (1956). In R. C. Marsh (Ed.), *Logic and knowledge: Essays 1901–1950*. The Macmillan Company.

Santos, L. R., Sulkowski, G. M., Spaepen, G. M., & Hauser, M. D. (2002). Object individuation using property/kind information in rhesus macaques (*Macaca mulatta*). *Cognition, 83*(3), 241–264. https://doi.org/10.1016/S0010-0277(02)00006-9

Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition, 108*(3), 662–674. https://doi.org/10.1016/j.cognition.2008.05.007

Schaeffer, B., Eggleston, V. H., & Scott, J. L. (1974). Number development in young children. *Cognitive Psychology, 6*(3), 357–379. https://doi.org/10.1016/0010-0285(74)90017-6

Schwitzgebel, E. (2021). Belief. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy (winter 2021).* Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2021/entries/belief/

Searle, J. R. (1958). II.—Proper names. *Mind, LXVII*(266), 166–173. https://doi.org/10.1093/mind/LXVII.266.166

Sodian, B., & Kristen, S. (2010). Theory of mind. In B. Glatzeder, V. Goel, & A. Müller (Eds.), *Towards a theory of thinking* (pp. 189–201). Springer. https://doi.org/10.1007/978-3-642-03129-8_13

Spelke, E. S. (1985). Perception of unity, persistence, and identity: Thoughts on infants' conceptions of objects. In J. Mehler & R. Fox (Eds.), *Neonate cognition: Beyond the blooming buzzing confusion* (pp. 89–113). Lawrence Erlbaum.

Spelke, E. S. (1988). Where perceiving ends and thinking begins: The apprehension of objects in infancy. In *Perceptual development in infancy* (pp. 197–234). Lawrence Erlbaum Associates Inc.

Spelke, E. S., Kestenbaum, R., Simons, D. J., & Wein, D. (1995). Spatiotemporal continuity, smoothness of motion and object identity in infancy. *British Journal of Developmental Psychology, 13*(2), 113–142. https://doi.org/10.1111/j.2044-835X.1995.tb00669.x

Stavans, M., Lin, Y., Wu, D., & Baillargeon, R. (2019). Catastrophic individuation failures in infancy: A new model and predictions. *Psychological Review, 126*(2), 196–225. https://doi.org/10.1037/rev0000136

Strawson, P. F. (1959). *Individuals. An essay in descriptive metaphysics*. Lowe & Brydone.

Tanz, C. (1980). *Studies in the acquisition of deictic terms*. Cambridge University Press.

Tomasello, M. (2021). *Becoming human: A theory of ontogeny*. Belknap Press of Harvard University Press.

Tomasello, M., & Gonzalez-Cabrera, I. (2017). The role of ontogeny in the evolution of human cooperation. *Human Nature, 28*(3), 274–288. https://doi.org/10.1007/s12110-017-9291-1

Tugendhat, E. (2016/1976). Traditional and analytical philosophy: Lectures on the philosophy of language. Cambridge University Press. German: Vorlesungen zur Einführung in die sprachanalytische Philosophie, Frankfurt am Main: Suhrkamp, 1976. https://doi.org/10.1017/CBO9781316535608.

Van de Walle, G. A., Carey, S., & Prevor, M. (2000). Bases for object individuation in infancy: Evidence from manual search. *Journal of Cognition and Development, 1*(3), 249–280. https://doi.org/10.1207/S15327647JCD0103_1

Wang, S., & Goldman, E. J. (2016). Infants actively construct and update their representations of physical events: Evidence from change detection by 12-month-olds. *Child Development Research, 2016*, 1–11. https://doi.org/10.1155/2016/3102481

Warneken, F., & Tomasello, M. (2007). Helping and cooperation at 14 months of age. *Infancy, 11*(3), 271–294. https://doi.org/10.1111/j.1532-7078.2007.tb00227.x

Webb, P. A., & Abrahamson, A. A. (1976). Stages of egocentrism in children's use of 'this' and 'that': A different point of view. *Journal of Child Language, 3*(3), 349–367. https://doi.org/10.1017/S0305000900007236

Wellman, H. M. (2011). Developing a theory of mind. In *The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp. 258–284). Wiley Blackwell.

Wilcox, T., & Baillargeon, R. (1998). Object individuation in infancy: The use of featural information in reasoning about occlusion events. *Cognitive Psychology, 37*(2), 97–155. https://doi.org/10.1006/cogp.1998.0690

Wynn, K. (1992). Addition and subtraction by human infants. *Nature, 358*(6389), 749–750. https://doi.org/10.1038/358749a0

Xu, F., & Baker, A. (2005). Object individuation in 10-month-old infants using a simplified manual search method. *Journal of Cognition and Development, 6*(3), 307–323. https://doi.org/10.1207/s15327647jcd0603_1

Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive Psychology, 30*(2), 111–153. https://doi.org/10.1006/cogp.1996.0005

Xu, F., Carey, S., & Quint, N. (2004). The emergence of kind-based object individuation in infancy. *Cognitive Psychology, 49*(2), 155–190. https://doi.org/10.1016/j.cogpsych.2004.01.001

Yoon, J. M. D., Johnson, M. H., & Csibra, G. (2008). Communication-induced memory biases in preverbal infants. *Proceedings of the National Academy of Sciences, 105*(36), 13690–13695. https://doi.org/10.1073/pnas.0804388105

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.