INDEXING OBJECTS IN VISION AND COMMUNICATION

By

Gabor Brody

Submitted to

Central European University
Department of Cognitive Science

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Cognitive Science

Supervisors:
Gergely Csibra
Ágnes Melinda Kovács

Budapest, Hungary

2020

DECLARATION OF AUTHORSHIP

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials written by another person, or which have been accepted for the award of any other degree or diploma at Central European University or any other educational institution, except where due acknowledgment is made in the form of bibliographical reference.

__________________________-

Gabor Brody
02/26/2020

ABSTRACT

In this dissertation we explore the representational capacities infants recruit in the process of tokening objects. We argue that infants around the first year of life are equipped with multiple representational systems that serve to uniquely index entities. The main contribution of the present work is the proposal of a novel indexing system, one that is bound to communicative discourses. This system is engaged in referential communication, and creates an index for every entity that is construed as under discussion by the communicative agents. Moreover, we claim that this communicative indexing system is largely independent from visual indexing. Using standard methodologies in the object individuation literature we demonstrate that indexing in a discourse is not based on the spatiotemporal characteristics of the represented entities. Although we found that in isolation infants can use spatiotemporal information to create multiple object representations, we consistently found that if the objects were presented in a referential-communicative context, infants did not take the spatiotemporal separation between objects as an individuating criterion. We also explored the nature of the discourse-bound system by directly assessing the process of index creation within this system. We devised a novel individuation paradigm, where infants did not have any direct perceptual access to the objects and had to derive numerical expectations solely from their interpretation of communicative acts. Within this paradigm, infants could recruit different types of information in order to create novel indices. Our findings suggest that within a discourse context, both distinct communicative agents and distinct referred-to locations generated expectations of multiple objects. Crucially, infants had trouble integrating referential information from disjoint discourse contexts. Together, these results underscore the point that discourse-bound representations are not based on a first-person encoding, but bound to specific communicative contexts. Our proposal and the empirical results have various architectural ramifications for, and raise important questions about, the interface between distinct systems of indexing. Positing a discourse-bound indexing system in infancy is empirically supported and theoretically productive as it can offer a framework to explore the developmental origins of the displacement property of human communication.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to my supervisors, Gergely Csibra and Ágnes Melinda Kovács. Gergő, I am so inspired every day by seeing the way you do science, and how you work for our community. I am extremely grateful for this opportunity to learn from you. Thank you Ági for all your support, discussion, and for your constructive and thoughtful criticism. I always felt that you are ready to take my weirdest ideas seriously, and that forced me to take them seriously also.

I want to thank the external examiners, Justin Halberda and Zsuzsa Kaldy for their suggestions, which helped to improve the manuscript both in clarity and content.

I want to thank all former and current staff of CDC / CEU who provided invaluable help in the experimental part of this dissertation, and in my academic life. These studies would not have not been the same without Dorottya Mészégető, Iulia Savos, Kriszti Andrási, Eszter Körtvélyesi, Ágnes Volein and Mari Tóth. I am extremely grateful for Réka Finta, for providing guidance in any and all issues a Phd student can encounter.

I want to thank Gyuri Gergely and Gergő Csibra for establishing this department. It was a incredible opportunity to be a part of this community, without leaving my beloved home city. I dream that the winds will change once again, so there can always be CEU Babylab in Budapest.

I am grateful to Luca Bonatti, Liz Spelke, and Szilvia Bíró for hosting me for short research visits. These were extremely important phases in my life and in my training. I want to thank Ildikó Király for being an amazing supervisor during my BA and MA studies at ELTE. You always encouraged me, and your trust helped me overcome so many challenges.

I want to thank all my friends and family who have supported me along the way. I am especially indebted to Misi Bródy, Anna Vincze, Kata Bródy, Marci Szigeti, Lili Horváth, Zita Szenteczki, Zsófi Szabó, Máté Kertész.

Special thanks go to Hanna and Misi Kertész for providing company, data, and insight. Your relentless efforts to succeed in my extremely infelicitous variants of the Give-n-task showed persistence, faith, and creativity.

Finally, I want to thank Athulya Aravind, for helping me in every part of the writing process from emotional support to theory development and comma placement. I appreciate so much how you always press me to clearly define my terms (except when I really do not want to). But nothing makes me as excited as us working and playing together every day to define, redefine, and enjoy our happiness.

Table of contents

Chapter 1 — Indexing objects around the first year of life

> *“Identity is utterly simple and unproblematic. Everything is identical to itself; nothing is ever identical to anything except itself. There is never any problem about what makes something identical to itself; nothing can ever fail to be. And there is never any problem about what makes two things identical; two things never can be identical. [...]There might be a problem about how to define identity to someone sufficiently lacking in conceptual resources [...] but since such unfortunates are rare, even among philosophers, we needn't worry much if their condition is incurable.”*
>
> (Lewis 1986, 192–193)

This dissertation aims to explore the nature of human infants' object representations. At first blush, we take a naive realist position (Fodor, 1998, 2006) and define object representations as extensionally as we can: as mental tokens that under everyday circumstances pick out particular objects. While *objects* might be understood extensionally, “mental tokens” are internal notions and thus require further clarification. A token (object representation) is taken to be a specific subcategory of mental representations, one that is *indexable*. In order to pick out particular objects from the environment, any system that represents them have to fulfill some basic functions. It has to (1) maintain the identity of the represented entities, (2) distinguish different entities from each other, (3) provide an address for any represented entity, to which properties might be bound to and queried from, and (4) make this address available for further cognitive processing. Indices, irrespective of their implementation — be it symbols, icons, files, slots, neural activation patterns or something else — are the simplest and to our knowledge only way of adhering to these requirements. A system that assigns an index to each representation it contains has a way of establishing the identity of the represented objects. An object can be tracked as “itself” as long as it has its corresponding index assigned. Objects can be differentiated: two objects are different *iff* they have different indices. Information can be bound to and queried from a specific object *iff* an index can be used to establish which object representation to operate on. And finally, indices might be used by other cognitive processes to refer to the represented objects, even if we assume a high level of cognitive impenetrability of the processes that create and maintain the indices (Pylyshyn, 2003).

What system(s) are present in the infant mind that are able to assign indices to objects? How are indices created and how are they tracked? Are there multiple indexing systems? If yes, how do they interact with one another to establish coreference? Trying to probe these questions, the general strategy of the dissertation is the following. As the process of indexing is not directly observable, the nature of these systems have to be inferred from infants' behavior in well controlled situations. In the present chapter we will detail some of the properties that infants use to establish the identity and non-identity of objects in these circumstances. We will review evidence of the processes of index assignment. We will also show that infants' behavior in specific circumstances can be considered as preliminary evidence for the existence of multiple indexing systems. In Chapters 2 & 3 we will present new empirical evidence that provide partial support for this hypothesis. In Chapter 4 we will focus on infants' ability to represent objects in communicative contexts, and argue for a system of object/referent indexing that is bound to a specific discourse.We start to empirically explore this idea in Chapter 5. In Chapter 6 we will summarize how the present findings contribute to our understanding of the development of object representations, and conclude by pointing out some crucial challenges for future research.

1.1 Spatiotemporal indexing in object based attention

One of the most influential theses in infant cognitive development is that infants' object representations are supported by perceptual mechanisms that encode and track objects in space and time. These views were heavily influenced by advances in vision science: in particular the visual-index theory (Pylyshyn, 1989, 2000, 2001) and the object file theory (Kahneman & Treisman, & Burkell 1983; Kahneman, Treisman & Gibbs, 1992) provided convincing (cf. Alvarez & Cavanagh, 2004, Holcombe & Chen, 2013; Hein & Moore, 2012) support for the so-called *spatiotemporal priority hypothesis* in object directed attention. The general claim is that spatiotemporal properties are privileged compared to any other properties in tracking and encoding visual objects (Pylyshyn 1989, Pylyshyn, 2001; Scholl 2001; Flombaum, Scholl, & Santos, 2009).

Building on these theories, an even stronger claim is prevalent in the developmental literature, with the crucial difference that spatiotemporal priority is not restricted to perceptual processes (e.g. Scholl & Leslie, 1999). More than being a claim about infants' attentional/perceptual capacity to track objects, it is often conceptualized as infants' general capacity to represent objects in the first year of life. This claim would imply that, for young infants, encoding the spatiotemporal characteristics of an object — which roughly translates into tracking its location — could be the single sufficient and necessary criterion for creating and maintaining a corresponding object representation. Any further information (e.g., visual features or kind membership) could only be bound to these representations, thus treating such information as fundamentally secondary, dependent on the index-creating spatiotemporal encoding system.

The way the human visual system can pick out and track objects has been subject to intense research in the last decades. An important insight of this research is that some perceptual mechanisms maintain object representations entirely independent of the attributed featural or conceptual descriptions of the attended objects. Converging evidence comes from studies on ambiguous apparent motion (Ullman, 1979; Nishida & Takeuchi, 1990), visual search (Burkell & Pylyshyn, 1997), subitizing (Trick & Pylyshyn 1993; 1994), and most extensively on multiple object tracking   (Pylyshyn & Storm, 1988; Pylyshyn, 1998, 2001). In the classical versions of this paradigm, participants are witnessing multiple visually indistinguishable objects on the screen moving on continuous but unpredictable paths. Their task is to track multiple target objects in parallel without confusing them with visually indistinguishable distractors. The targets are indicated before the movement starts by flashing for a couple of seconds. The general finding is that adults can track around four objects in these studies under the usual conditions (cf. Franconeri, Alvarez, & Cavanagh, 2013 for an alternative take on the limits). In a particularly illuminating version of the paradigm (Pylyshyn, 2004), participants did not only have to track the targets but also had to remember their features. Objects were revealed to have unique visual or conceptual properties, but only during target presentation, and not during movement. At the end of the trials, participants had to indicate not only which were the targets but also their identity: what properties a particular target had before the movement started. While they still succeeded in tracking the objects, they crucially failed in retaining the binding between the objects and their properties. Furthermore, in another study participants often failed to notice changes in the properties of the objects during tracking (Scholl, Pylysyhyn, & Franconeri, 1999; Bahrami, 2003). These results were interpreted as evidence that these object representations are organized around indexes that take into account the spatiotemporal aspect of the objects, a central claim in visual Visual-Index Theory (Pylyshyn, 1989, 2001,2003). In this framework so called visual indices can pick out objects from the environment without giving them a description[1,2]. This process of picking out is caused by the interplay between the actual physical objects in the world and the architecture of the indexing system. While conceptual or perceptual descriptions might be used to reidentify representations that are created based on these indexes (e.g., after a brief occlusion), they do not play a role in establishing and maintaining the representations. That function achieved by the index's ability to sustain the picking-out relation with the objects (by tracking), even in cases of visual changes (Bahrami, 2003) and brief occlusions (Scholl & Pylyshyn, 1999). The causal aspect of index creation is crucial for Pylyshyn’s theory on visual indexing. It posits a mechanism that creates indices in the visual system without descriptions, and thus provides a paradigm case for a primitive world-to-mind connection.

Extensions and, to a degree, alternatives to the visual-index theory are the object-file theory (Kahneman & Treisman, 1984; Kahneman, et al., 1992) and the object-indexing theory (Leslie, Xu, Tremoulet, & Scholl, 1998; Scholl & Leslie, 1999). These aim to characterize spatiotemporal object representations as mid-level perception: a theorized special layer in the cognitive

[1] In visual-index theory the spatiotemporal properties of the objects are not necessarily encoded in the index either, at least not explicitly. This is relevant for explaining how the indices can be caused by an object rather than describe one.

[2] While we will keep using the term “visual index”, there are reasons to believe that these representations are not modality dependent, as auditory or proprioceptive input might be used create the same type of representations (Pylyshyn, 1989, 2003).

architecture where spatiotemporal, perceptual and conceptual information can be encoded together in so-called object files. Object files are also individuated based on spatiotemporal indexes, but their architecture would allow for much more complex inferences and information encoding compared to the more bottom-up approach the visual-index theory takes (Scholl & Leslie, 1999). This increase in expressive power of the system comes at the cost of having a more opaque, and less predictive theory on cognitive architecture (Pylyshyn, 2003). The differences between these two types of theories are not relevant for our current purposes, and we will just keep invoking visual-index theory because of its clarity with relation to the cognitive architecture. These theories are indistinguishable in their most important thesis: that the spatiotemporal properties of the objects are the sole indexing-relevant properties of object based attention.

1.2 Spatiotemporal indexing in infancy

Infants can individuate[3]objects based on spatiotemporal properties, as it was first demonstrated in a habituation paradigm (Spelke, Kestenbaum, Simons, & Wein, 1995). In its original version, a scene was presented with two separate occluders with a visible empty space (a spatial gap) between them. In the *discontinuous motion* condition, objects emerged from behind both occluders sequentially. As these objects never crossed the gap in the middle, the possibility of only one object being present could be ruled out. First, an object was presented as it moved out from behind one occluder, before returning to its starting position. Then, an object left the other occluder and returned the same way. In the *continuous motion* condition, an object emerged sequentially from behind both occluders in a manner that was compatible with a single object interpretation. First, it left one of the occluders, then it crossed the gap between the two occluders and moved behind the other occluder. By measuring infants' looking time to a one object and two object presentation where the occluders were not present Spelke et al. (1995)

[3] We define individuation as the capacity to create, maintain and differentiate (object) representations. When discussing empirical results that probe infants’ individuation capacities, we will use the term to refer specifically to individuating (representing) two objects simultaneously. Individuating, and by extension representing, two objects can be taken as evidence for treating some object properties relevant for indexing: those that minimally differentiate the two representations.

found that infants had different numerical expectations in the two scenarios. Infants dishabituated to displays of a single object only when the object did not cross the gap; that is, they used the spatial trajectory of the objects as an individuating criterion. Later the study was replicated with 10-months-old infants in a paradigm using familiarization instead of the habituation (Xu and Carey, 1996). The capacity to individuate objects based on spatiotemporal cues is not restricted to looking behavior, as 10-month-old infants can also pass this kind of task that uses a different measure: manual search. If infants are presented with two objects simultaneously and then these objects are hidden in a box, they seem to search for exactly two objects (Van de Walle, Carey & Prevor, 2000). Further evidence for spatiotemporal individuation is infants' ability to enumerate objects in a variety of paradigms where the only available information was spatiotemporal, like encountering serial hiding of objects and transformations on the resulting hidden sets (Wynn, 1992). In these situations infants witness a series of object placements and possibly replacements, and still succeed in building expectations about the resulting set of objects — if the set size does not exceed their working memory limit (Feigenson & Carey, 2003).

When infants individuate objects on a spatiotemporal basis, that property in itself is supporting the corresponding representation (Scholl & Leslie, 1998). In one study (Kibbe & Leslie, 2011) presented 6-month-old infants with two different looking objects, that were hidden at separate locations consecutively. Prior research (Káldy & Leslie 2003, Káldy & Leslie, 2005) established that at this age infants are only able to remember the features of a single object in similar context. Their question was whether infants remember the presence of the object when all its features are forgotten, that is, whether the spatiotemporally grounded object representation prevails when all other information about the object is lost. They compared infants' looking times to three possible outcomes: when the second hidden object was revealed to have the wrong features, when this object vanished, and when no transformation on the object occurred. They found that while the infants were not sensitive to feature changes, they were surprised when the object disappeared. The infants seemingly kept the representation of the object even when the bound featural information was lost. This shows that whatever representational format supports the encoding of the object in cases like this, spatiotemporal information is sufficient for maintaining it.

Multiple authors noticed the analogy between the visual object tracking system and infants' ability to represent objects in space, often arguing that the two literatures are tapping into the same cognitive architecture (Leslie et al., 1998; Scholl & Leslie 1999; Xu, 2005; Carey, 2009; Stavans, et al., 2015). This similarity can be mainly cached out in the following ways. There is a clear set size limitation for a small array of trackable objects, as at bigger quantities performance breaks down (Feigenson, Carey, & Hauser, 2002; Feigenson & Carey, 2005). Both the infant and the adult visual system can track objects through brief periods of occlusion (Scholl & Pylyshyn, 1999; Wynn, 1992) even though infants' ability to maintain object representations is longer by as much as 10-20 seconds than the ones usually tested in the adult vision literature (it is unclear how the visual system would handle these longer occlusions (Pylyshyn, 2003). But crucially both for infants (Cheries, Feigenson, Scholl, & Carey, 2005) and for adults (Scholl & Feigenson, 2004), the ability to track objects breaks down if the disappearance of the objects at occlusion is not presented as a “physically plausible” asymmetrical and gradual deletion by the occluder but as an "implausible shrinking" event, in which the object gradually and symmetrically gets smaller. Similarly, the object representation is lost if the object disintegrates (Kaufman, Csibra, & Johnson, 2005). To make this point more general, if some specific characteristics of objecthood are violated, humans fail to track these entities. In the developmental literature objecthood is usually defined as meeting the criteria of so-called ‘Spelke objects’: "bounded, coherent, three-dimensional objects that move as a whole... " (Spelke, 1990). In a variety of tasks infants fail to track non-Spelke objects, such as a collection of smaller objects (Chiang & Wynn, 2000) or piles of substances, for example, sand (Huntley-Fenner, Carey & Solimando, 2002).

Altogether the parallels between infants' and adults' ability to perceptually track objects are persuasive enough to postulate a shared underlying mechanism: one that encodes objects based on the available spatiotemporal information using visual indices. However, encoding object location might be supported by other mechanisms as well. For example, a belief that "my keys are in my car" is arguably not based on the visual tracking of the keys, but references spatiotemporal information. For the present purposes, we will discuss spatiotemporal information in relation to visual-index based representations, as it is unclear how other types of location information are encoded or used for maintaining object identity in infancy.

The literature on object-based attention helps to clarify some key notions of object indexing. It seems that we can distinguish two kinds of object properties. We define *primary properties* as object properties that are constitutive of the process of object indexing. Based on them infants can individuate objects even when no other information is contained in the representation (Kibbe & Leslie 2011). Secondary properties in contrast are object properties that can be encoded about an object, but not being constitutive of the indexing process. In principle, they could be modified or forgotten without the loss of the object representation. Some properties, like color, can be used for object individuation, but it does not follow that they are also primary properties. For example, we might individuate and expect two pens when we see a red pen and a blue pen, but representing "red" in isolation might not be sufficient to maintain an object representation; we might need an index that this property can be attached to. Thus, the primary/secondary distinction allows for individuating objects based on secondary properties, but with the caveat that the resulting representation would still have to be organized around a primary property: those that are necessary for maintaining object indices. On the other hand, the opposite is not true: the loss of all primary properties should always result in the erasure of the the object representation, and incongruent transformations on a primary property should always create a novel object representation. Thus, encoding at least a single primary property is sufficient and necessary to represent an object, but in their absence there is no representation that secondary properties could be bound to. This argument does not necessitate that encoded primary properties are equivalent to the resulting indices: It is also possible that indices are individuated by system-internal symbols, icons or slots without any reference to the information that necessitated their existence (Pylyshyn, 2003). The primary/secondary distinction only helps to define extensionally what actual object properties are used in the process of indexing, without providing a characterization of how these indices are represented.

1.3 Object representations beyond spatiotemporal indexing

Adults can describe objects in ways that do not take into account spatiotemporal information at all. The “misty garlands of your daydreams” is not something that can be spatiotemporally encoded and it is unclear if a visual index could in principle apply to it. In some fictional worlds teleporting is possible — a concept that explicitly disregards spatiotemporal continuity — and audiences understand them perfectly. Humans can also describe non-coherent objects like a disassembled computer both as as *a computer*, and as *computer pieces* depending on which aspect they think is relevant. But some fundamental cognitive and linguistic phenomena like quantification, counterfactual reasoning and such also requires us to think of real or counterfactual objects, often with set sizes much bigger than what is assumed to be in the domain of the visual-indexing system. While staying agnostic for now on the exact mechanisms, we can minimally say that some thoughts and linguistic expressions pick out and individuate objects based on non-spatiotemporal properties. Although the exact role and nature of descriptive representations is heavily debated in philosophy and linguistics they are not relevant for the simple point that we are trying to make. To accommodate the displacement property of language and thinking, we have to represent objects that do not build on spatiotemporal information. These likely require an indexing system that is not spatiotemporally defined, but by using some descriptive properties of a given object-related thought/expression.

In the following, we will present evidence that infants early in life can represent objects where the primary individuating property of the object is not spatiotemporal in the sense that the representations are not based on visual indices. We will also argue that spatiotemporal properties are not necessarily primary: they can be forgotten in cases where other properties take primacy. Some of these properties (that are mostly referred to as “conceptual” or “kind-based”) have been long identified as being used by a distinct system for object individuation in infancy (Xu, 2002, 2005, 2007), while still maintaining the thesis that their role is secondary compared to spatiotemporal representations both architecturally and in ontogenesis (Xu, 2005, 2007; Carey, 2009). There is considerable debate surrounding perceptually encoded object features being suitable for object individuation in the first year of life (Tremoulet, Leslie, & Hall, 2000; cf. Stavans, Lin, Wu, & Baillargeon, 2015 for a review) but no theory to our knowledge claims it to be a primary property.

1.4 Conceptual object individuation in infancy

The first piece of research that systematically started to assess infants ability to use non-spatiotemporal properties for object individuation is the seminal study by (Xu & Carey, 1996). This study assessed how infants of different ages can individuate objects in the presence of kind relevant cues. In the most important condition infants witnessed two objects that belonged to different kinds (e.g., a truck and a duck) emerging one by one from behind a single occluder repeatedly. After these events, the occluder fell revealing either both objects or just one of them. In this condition only the kind and surface features differentiated the objects but not their spatiotemporal properties. The results showed, that while 12-month-old infants expected two objects to be present, 10-month-olds did not look longer for a single object outcome. Their original conclusion was that conceptual and featural cues are not taken into account until 12 months of age.

Since this first study, much more evidence was uncovered about the development of conceptual object individuation. It turns out that 10-month-olds can already individuate objects in paradigm like Xu and Carey (1996) used, when one of the two presented objects is human-like while the other is not (Bonatti, Frot, Zangl & Mehler, 2002). Ten-month-olds also succeed when the contrast is between a self-propelled agent and an object that is moved by a hand (Surian & Caldi, 2010). It is likely that agents/humans have conceptual descriptions that infants use spontaneously from an early age (Carey, 2009). Also, within the domain of non-agentive objects, infants can succeed under some circumstances that warrant conceptual descriptions. For example, after an ostensive function demonstration, infants individuate artifacts, but only if they are presented to have different functions (Futó, Téglás, Csibra, & Gergely, 2010). Similiar evidence was obtained with much younger infants as well (Stavans & Baillargeon, 2018). Also 9-month-olds succeed in the "standard" Xu & Carey (1996) kind individuation task if the two objects are labelled with two different words during presentation (Xu, 2002), and this effect was also present even when the objects and words were unfamiliar. Furthermore, the perceptual features of these unfamiliar objects are not relevant. Even if there is a single object that is labelled with different labels at different presentations, infants expect two objects (Xu, 2003 (as cited by Xu, 2005)). It is not just labels that have such an effect: if two functions are demonstrated on the same object (at different times), infants also expect two objects (Futó et al., 2010). This evidence suggests that in the first year of life what is required for conceptual individuation is (1) the communicative framing of object presentation (except for agents), and (2) infants' ability to give individuating conceptual or linguistic descriptions to the objects. The role of linguistic information (i.e., different labels) is probably to establish a conceptual distinction between the objects (Xu, 2005, 2007; Dewar & Xu, 2007).

While these studies show that in communicative situations young infants are able to use conceptual descriptions for object individuation, they fail to provide evidence that conceptual properties can take primacy, i.e., that object representations can be created or sustained even in the absence of a visual index. Strong evidence for such a capacity comes from a manual search task by Xu, Cote, & Baker (2005). In this study 12-month-old infants were presented with an opaque box without perceptual access to its content. The experimenter only provided verbal cues to inform the infants of the content of the box. The infants readily individuated objects — searched for two — if the experimenter used two novel labels in the presentation. They did not individuate when the experimenter provided a single label, or provided two different emotional expressions. For this performance infants had to treat the two labelled utterances as mutually exclusive descriptions of objects, and those descriptions were sufficient for creating different object representations. Nevertheless, this result still does not prove that the representations that the infants in this study created were not based one spatiotemporally encoded information. To rephrase, it is not clear whether they created two visual indices with conceptual/linguistic information bound to them, or created two object representations organized around the conceptual/linguistic descriptions to which some spatiotemporal information was bound to.

1.5 Cases that imply that visual-indices are insufficient

If we want to show that the general claim of spatiotemporal priority in infancy is not valid, we have to find contexts where the spatiotemporal properties of objects are not taken into account even when the relevant visual properties are available and infants could use them in principle. Going further, in order to provide evidence that conceptual properties are indexed in these cases, we also need to show that the objects are still represented by the infants, based on some description of these objects. we have identified six types of evidence that speak to these questions.

1.5.1 Forgetting the location of an object while remembering its features.

Yoon, Johnson & Csibra (2008) presented 9-month-old infants with objects in either a communicative or a noncommunicative context to assess what they remembered about the objects after a brief occlusion. In the communicative condition, the protagonist engaged the infants in ostensive communication: greeted them and pointed to the single object present at the scene. In the non-communicative condition the protagonist did not address the infant and performed a reaching action towards the object. After briefly occluding the object, the infants' memory was probed by measuring their looking times and the duration of their first look to three different outcomes: feature change (a different looking object at the same location), location change (the same object at a different location), and no change. In the non-communicative context the infants encoded the spatiotemporal information: they looked longer at the location change, compared to the no change outcome, but they did not encode the object otherwise: the feature change did not elicit different looking patterns from when there was no change. But in the communicative context the infants showed an opposite pattern: they only remembered the features of the object but not its location. This provides evidence for both of the criteria we set up above. In the communicative condition the infants did not encode the spatiotemporal information, while the evidence was available and even used in the non-communicative condition. Furthermore, the object was still represented in that case: the features of the object were not lost. This study shows that infants have trouble with simultaneously encoding the same object as a spatiotemporal entity and also conceptually: as the referent of a communicative act (Csibra, 2010).

1.5.2 Object identity in preference attribution can be based on kinds.

The second case study (Spaepen & Spelke, 2007) is a version of the preferential choice paradigm (Woodward, 1998) aiming to reveal how kind descriptions interact with infants' ability to encode the preferences/goals of an agent. Twelve-month-olds were either habituated to a between-kind preference demonstration (an agent choosing a truck over a doll) or a within-kind one (choosing between different looking trucks, or between different looking dolls). The looking-time results indicated that the infants only encoded between-kind preferences, but not within-kind ones. These results suggest that the perceptual features that are bound to visual-indexes are not used to encode preferences, if conceptual descriptions are available for the objects (12-month-olds are familiar with the relevant object kinds. This is surprising because prior studies show, that already three-month-olds can use these perceptual cues (e.g. Luo, 2011, Choi, Mou, & Luo, 2018). More importantly, this result held even in conditions where spatiotemporal continuity of the objects were not interrupted as they were visible throughout the study. This shows that the way objects are mentally described (DOLL vs. DOLL, or DOLL vs. TRUCK) is causally relevant for whether preference can be attributed. If the objects had been represented on the basis of visual-indexes as primary properties, infants should have noticed the contingency of choosing one kind-member over another, and could have attributed a preference to the protagonists. Any argument to the contrary would have to account for the fact that infants in fact are able to rely on visual-indexes in different versions of the study where the objects are unfamiliar (e.g., Woodward, 1998, Luo, 2011, Luo & Baillargeon, 2005). To see this, assume that infants always take spatiotemporal properties as primary, and consider the standard Woodward (1998) paradigm. It presents objects that are unfamiliar to infants, and we have no reason to think that infants can make a conceptual distinction between the objects. Thus, in order to realize that the agent is behaving systematically (preferring one object), they have to recognize the objects on trial-by-trial basis by re-attaching the corresponding visual indices. The explanandum is the following: If across-trial index re-application is sufficient for preference attribution in such cases, why does this not happen in the otherwise analogous within-kind contrast case?

1.5.3 Spatiotemporal cues might not be necessarily used in agent individuation.

While there is a lot of early evidence for infants' understanding of agents as self-propelled and goal-directed (Gergely, Nádasdy, Csibra, & Bíró, 1995), the conceptual encoding of agents is spontaneous and does not require communication. Ten-month-olds are not only able to discriminate agents from non-agents, but also individuate entities on the basis of this distinction (Bonatti et al., 2002; Surian & Caldi, 2010). We have very little evidence of what primary property infants adopt to the represent an agent. The most relevant study is reported by (Kuhlmeier, Bloom, & Wynn, 2004). They replicated the original spatiotemporal individuation result with objects (Spelke et al., 1995) in 5-month-olds, but found a striking failure of individuation in the condition where the objects were replaced by humans. It seems as if humans were not expected to be spatiotemporally continuous, as continuity violation did not result in human individuation. What might be the explanation? It is possible that 5-month-olds think of agents as beings with special powers, as originally argued (Bloom, 2005). Alternatively, what makes agents special could be that infants do not treat their spatiotemporal properties as a primary, *because* agents are indexed conceptually. This interpretation also leads to a leaner explanation of the findings: the spatiotemporal violation might not have been encoded at all, or it did not warrant individuation because the relevant indexing system just did not take it into account. On this account further assumptions about the exact nature of infants' agency concept are not required.

1.5.4 Infants remember objects even when they lose their visual indices.

One of the most convincing demonstrations that infants use visual-index based representations to encode objects comes from the exploration of infants' working memory capacity (Feigenson, Carey, & Hauser, 2002; Feigenson & Carey, 2003). While infants succeed in representing sets of 1, 2, or 3 objects, they fail with larger arrays in particular ways. In a manual search paradigm, one-year-old infants' search duration was compared across conditions where they had observed different number of objects hidden and retrieved from a box (Feigenson & Carey, 2005). Infants successfully discriminated between arrays where the set size was lower than four. They searched longer after a single object was retrieved in the 2 vs. 1, 2 vs. 3, and 1 vs. 3 comparisons, where items were still present in the box only for the larger array. But they failed in the 4 vs. 1 condition. If 4 items were hidden, of which a single item was retrieved (thus 3 remained), their search behavior did not differ from a condition where initially a single item was hidden and retrieved (thus the box was empty). This finding suggests that the upper limit of encoding spatiotemporally distinct objects is 3, which strengthens the claim that infants encode objects individually as spatiotemporally separate entities. Crucially, further conditions revealed that even if the visual-index system fails to encode the objects as individual entities, infants do not completely lose their representations. In further experiments reported in the same paper, the infants had to choose between two containers in which the experimenter hid different number of crackers. Infants successfully chose the larger set in all the comparisons where both sets were smaller than 4: 1 vs. 2, 1 vs. 3 and 2 vs. 3, but they failed in the 4 vs. 1 condition, showing that spatiotemporal encoding of the objects was impaired. But when they had to choose between a container that had 4 crackers and a container that was empty, they succeeded in choosing the larger set. This shows that whatever properties they used the encode the objects, losing the individual visual indices did not result in complete erasure of their representation. This representation furthermore is an object representation in the sense that properties can be bound to it (that is infants likely searched for a cracker and not something else). In another condition, the infants succeeded in choosing four crackers over a single one if all the individuals in the set of four crackers were substantially larger than the single item at the other location. It is clear that the representation the infants used here contained information about location: infants chose the correct container. But this encoding of the location was most likely not based on visual indices, because the failure in the 4 vs. 1 comparison showed that the different objects were not represented by individual indices attached to the objects.

1.5.5 Infants can create hierarchically structured set representations.

Infants starting from 7 months of age are able to create larger units of representation than a single object, where the existence of each individual object is still encoded (Moher, Tuerk, & Feigenson, 2012; Feigenson & Halberda, 2004). These so called "chunks" are created in response to a large variety of cues. Amongst others, they can be based on spatiotemporal cues, when objects are in close physical proximity (Feigenson & Halberda, 2004), on temporal regularities, when objects systematically co-occur at different time points (Kibbe & Feigenson, 2016), and on perceptual features. In a Xu and Carey (1996) type individuation paradigm Leslie and Chen (2007) sequentially presented 11-month-old infants with object pairs instead singular objects. These pairs in one of the two conditions were identical (two triangles or two discs), while in the other condition mixed (a triangle and disc each). At test, the four objects were revealed. Infants in the identical pair condition fixated on the screen for shorter time compared the mixed condition, implying that they expected four objects (or two pairs) to be present. Leslie and Chen (2007) argued that these objects might be represented using the conceptual description PAIR rather than via four distinct visual indices (as infants supposedly only have 3). Framing the phenomenon of chunking, as driven by conceptual encoding, is also corroborated by findings from older age groups. Evidence from 14 months of age shows that conceptual cues, like shared labels or kind membership, are also available for creating these representations (Feigenson & Halberda, 2008). In this manual search study, the infants' search times indicated that even though they were unable to remember that four identical objects were in the box, when they could chunk these items into two pairs they succeeded. This chunking could be based on prior knowledge of known kinds (like a pair of cars, and a pair of cats), or online, where where pairs of objects were labelled with a distinct nouns. For the current purposes what is most relevant in these findings is that, in order to account for increased memory of individual items, these representations have to be organized hierarchically (Feigenson & Halberda, 2004), and the structure of visual indices are patently non-hierarchical. The system of representation that encodes chunks therefore not only have to create descriptions of sets of objects, but it has to represent the hierarchical relationship between chunks and individual objects.

1.5.6 Infants perseverative errors in search task are influenced by pragmatic factors.

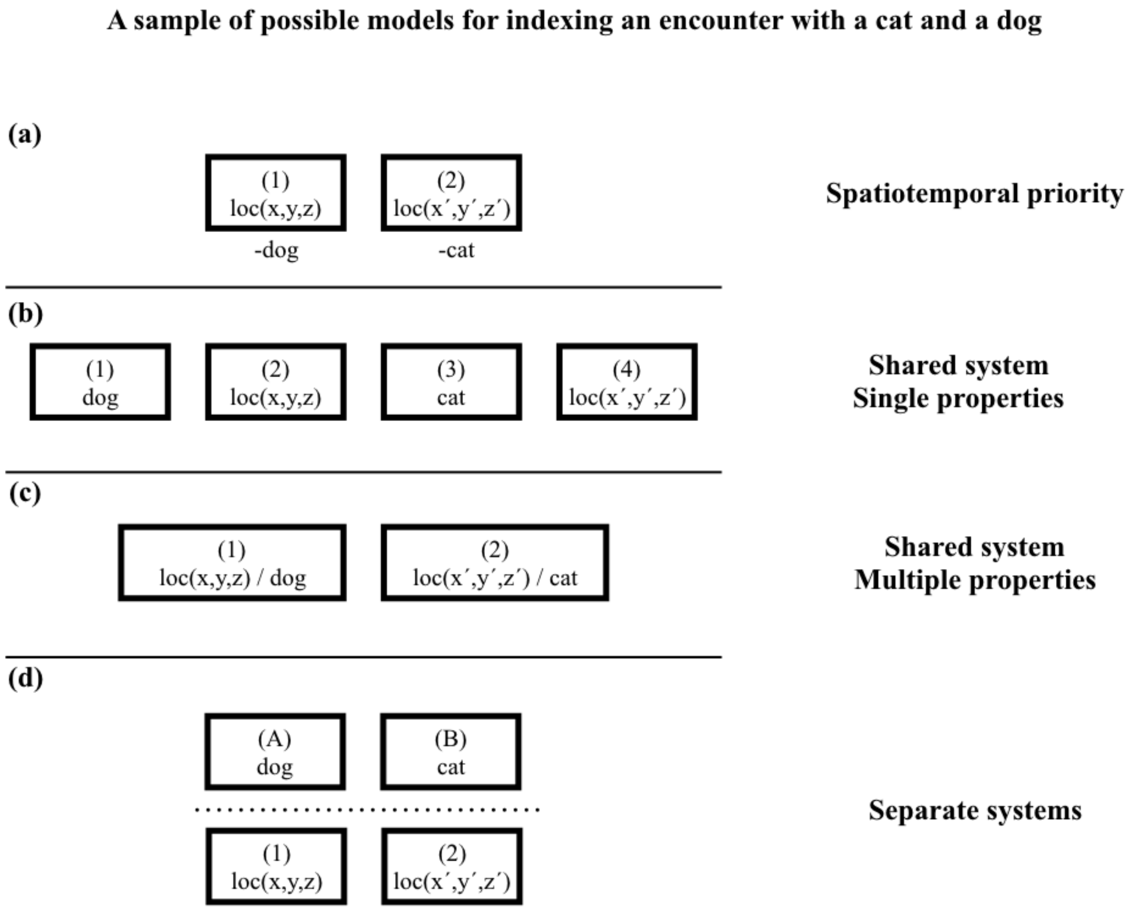
A classic example from the literature on infants' search behavior is the phenomenon of the A-not-B error (Piaget, 1954). After having repeatedly retrieved an object from a location, young infants tend to search at the same place even if they observe the object being hidden at another location. While all spatiotemporal evidence is available for infants to make the proper choice, they perseverate and keep searching at the incorrect location. This is often explained by appealing to inhibition (Diamond, Cruttenden, Neiderman, 1994) and visuo-motor development (Smith, Thelen, Titzer, & Mclin, 1999). But more recent studies show that one of the main driving forces behind the error is the communicative context in which the hiding takes place (Topál, Gergely, Miklósi, Erdőhegyi, & Csibra, 2008). If the amount of communicative cues are reduced for the hiding events, either by changing the behavior of experimenter, or by the infants not seeing the experimenter at all, they commit substantially fewer errors. Even without committing to the rich interpretation of these results that infants interpreted the demonstrations as "conveying generic information," these results minimally show that in a communicative situation infants' encoding of the hiding event changes, and direct spatiotemporal evidence of the location of an object is not taken into account the same way as outside of communicative contexts.

1.6 Possible models of object indexing

To summarize, there is evidence suggesting that infants even before their first birthday can represent objects using indices that did not originate from the visual system[4]. Infants can represent objects with attributed properties even when individual visual indices are lost (Feigenson & Carey, 2005). They do not use spatial information for individuating multiple agents

[4] Visible objects supposedly automatically have a visual-index based encoding (hence Pylyshyn's causal theory). What is clear from these studies, is that infants' expectations and behavior cannot be accounted for solely by these indices and their attributed properties.

(Kuhlmeier et al., 2004). They can forget an object's location while still retaining the representation of its features when it is a referent of a communicative act (Yoon et al., 2008). Their ability to attribute preferences based on visual features (including spatiotemporal properties) is inhibited when they know that the objects belong to the same kind (Spaepen & Spelke, 2007). What possible models can we construe of the indexing system in the light of these results? We offer four (non-exhaustive) options that seem reasonable to think about (cf. Figure 1.1 for illustration). We already provided arguments against option (a), the idea of infants having a single indexing system that relies solely on visual indices to individuate objects. Option (b) is a single indexing mechanism that can deploy different types of indices: ones that depend on descriptive, and others that depend spatiotemporal properties. In option (c) indexes can be complex and composed of both a description and the visual index. Option (d) assumes that there are separate systems responsible for descriptive and visual indexing.



A sample of possible models for indexing an encounter with a cat and a dog

Figure 1.1 Possible models of indexing an encounter with a cat and a dog that are discussed in the text. On this figure visual indices are represented with a set of coordinates, but this is not meant to imply that visual indices track objects this way. Similarly, on the figure all indices are numbered, which is not a necessary precondition for an implementation either. The first model (a) depicts a single system with solely spatiotemporally construed indices with attributed conceptual properties. Below (b) a shared system is depicted where different types of indices coexist. (c) is a variant of (b) where multiple properties can be represented in complex indices. Option (d) proposes that different and incommensurable indices represented by different systems.

We can discard option (a) based on the above presented empirical evidence. Infants represent objects even when there is no good reason to assume that the individual objects are represented via their spatiotemporal properties. Arguments in favor of specific versions of Option (b) are frequent in the literature. Infants' working memory is often argued to be able to index both conceptually and visually encoded entities in the same system (Leslie & Chen, 2007). But without further architectural clarification we do not see how that system could be viable. The main function of an indexing system (irrespective of what it tracks) is to keep track of the identity of all represented entities by providing them with a unique identifier. To offer a tautology, unique identifiers are unique, and any representation with index 1 cannot be the same representation as one with index 2. But if a system can construe a single entity under multiple non-comparable indices, the one-to-one correspondence between the tracked entities and the indexes does not necessarily hold anymore, and the indexing system loses its power of maintaining the identity of the tracked entities. Consider infants' performance in remembering and individuating pairs of objects (let's call the objects "Xs"), and assume infants' conceptual description as: PAIR-OF-X (Leslie & Chen, 2007). Why does working memory have a single index for the pair and not three: One index for the conceptual description: PAIR-OF-X and two visual-indices tracking the individual Xs present? There is nothing about being a PAIR-OF-X *per se* that is mutually exclusive with the visual indices that are tracking the same objects. What is needed for avoiding multiple indexing is a mapping between what is seen and how it is conceptually described: representing the correspondence between the two visual indices that happen to track X and X' and the single conceptual description (PAIR-OF-X). For such mapping to occur, both the domain and the co-domain have to be indexed at least during the encoding and recognizing a pair, and either object *as a member of a pair*. But as the relevant study shows (Feigenson & Halberda, 2008), conceptual descriptions seem to increase rather than decrease working memory capacity. Thus, it seems unlikely that both the domain and co-domain is represented in the same system, as that would predict a decrease in capacity because three indices would be required for construing or recognizing a chunk.

Adult data corroborates the need to disentangle visual and descriptive indexing of items in working memory. Holding verbal items (words) in working memory does not inhibit visual working memory (Luck & Vogel, 1997), or multiple object tracking (Scholl & Xu, 2001), showing that at least some point in ontogenesis indexing of entities in these systems are disjoint. Also, recall the result cited earlier that adults fail to track the initial features of objects in a multiple object tracking paradigm while still succeeding to track each individual object (Pylyshyn, 2004). When adults lose the mappings between the four initially presented features and the four visually tracked objects, they do not assume that there are 8 target objects present (four feature indexed, and four location indexed). They are aware that the features that they encoded at the beginning of a trial are in a one-to-one mapping relationship with the objects that are tracked till the end of the trial. If the representations of the attributed features shared an indexing system with tracking, in the lack of proper mapping, 8 indices would be required. The most straightforward explanation is to assume that the descriptions of the objects' initial features are indexed separately from the corresponding visual indices, so neither of the two individual systems have to retain more than 4 items simultaneously.

Trying to hold onto a more parsimonious single system solution, one can envision an index assignment function that references both spatiotemporal and descriptive properties of an object (option (c), Figure 1.1). This system might result in having complex indices, where the identity of an individual index is defined by multiple properties. This would elegantly explain why both conceptual and spatiotemporal properties can be used for individuation: a new object representation would be automatically warranted in differences due to any of the properties that are constitutive for indexing. But this model would fall prey either to an under-generation problem, similarly to option (a), or to the over-generation problem of option (b). Because, like option (a), this system would always individuate based on spatiotemporal information, it is unable to explain the previously cited cases where infants fail to do so (but still retain some object representation). To account for this data, we might want to allow the system to use non-complex indices at least some of the time. But this would result in the same problems that option

(b) faces. If indices are non-comparable, the main function of identity tracking breaks down, resulting in multiple non-exclusive representations for an object under all possible descriptions.

Contrary to the previously discussed possibilities, option (d) maintains that there is an indexing system (or systems) that is distinct from the visual-indexing system. This model is by no means novel, as it is implicitly or explicitly present in a variety of accounts without explicating the requirements on the architecture of object indexing (Feigenson & Carey, 2005 Feigenson & Halberda, 2004; Carey, 2009; Xu, 2005; Stavans et al., 2015). A paradigm case for such an alternative system could be one that indexes objects based on conceptual properties. This system would allow for identity tracking similarly to option (a), where indices correspond to representations of mutually exclusive entities, but this would only be true within a single system. That is, while the notion of identity might become unproblematic on the level of a specific indexing system, it becomes problematic on the level of the organism. Infants' behavior in variety of tasks might reflect exactly that. With the auxiliary hypothesis that infants prioritize this descriptive system over spatiotemporal indexing under some conditions, we could make sense of the findings that are challenging for visual indices. Thus, an object or an agent can be recognized *as itself* even after violating spatiotemporal continuity (Yoon et al., 2008; Kuhlmeier et al., 2004). Preferences can be attributed solely based on conceptual descriptions, while spatiotemporal and other perceptual characteristics of the objects can be disregarded (Spaepen & Spelke, 2007). This provides a good characterization of how infants remember the existence objects (with attributed properties) when the visual-index system is overloaded, and loses individual indices (Feigenson & Carey, 2005). Furthermore, these multiple indexing systems might provide a characterization of what is hierarchical in chunking. When the relevant concepts are available, a hypothesized conceptual indexing system might be able to take entries for descriptions of multiple objects, like *pair-of-objects, set-of-balls, two-wugs,* and the like. While these representations might require correspondence to visual indexes in order to get encoded, or recognized, they would not take up more than a single index for maintenance in the relevant system (Leslie & Chen, 2007, Feigenson & Halberda, 2004).

Other than providing a model that is sufficient for explaining the infant data, making the systems dealing with object-identity modular (Fodor, 1983; Chomsky, 2018) and in some ways redundant may also have theoretical benefits. In a framework like this, one might be able to keep the local and encapsulated problem of tracking identity within a system, apart from the possibly global, and holistic problem of identity tracking between systems (establishing correspondence between indices of different types). Positing multiple indexing systems might give a clue on why identity, while in a metaphysical sense can be thought of as the basic relation possible (*"Everything is identical to itself; nothing is ever identical to anything except itself"),* can seemingly cause a lot of controversies and even philosophical paradoxes for cognitive systems like humans: The coherence of establishing identity applies only within a system, and different indexing systems might produce different identity judgements.

To restate the goals of the project, in the following chapters we are going try to empirically assess whether infants employ different mechanisms for indexing objects when descriptive encoding is available to them compared to situations where they might use only visual indices to maintain object identity. Chapter 2 provides data from the looking time variant of a spatiotemporal object individuation paradigm (Xu & Carey, 1996). We replicate the original findings that 10-month-old infants are able to use spatiotemporal cues for object individuation. In further conditions we find that when we also provide conceptual/linguistic cues for object identity, infants fail to use the available spatiotemporal cues. In Chapter 3 we conceptually replicate these results using the manual-search object individuation paradigm (Van de Walle, et al., 2000). In Chapter 4 we make the case that one *descriptive* system for object indexing might be linked to communicative understanding, invoking the notion of discourse referents from natural language semantics (Kamp, 1981; Heim, 1982). Chapter 5 empirically assesses how indexing works in the proposed discourse-bound system, and how spatiotemporal information modulates it. Chapter 5 summarizes the contributions of the present work, and aims to provide some directions on what questions are in dire need of further research.

Chapter 2 — Study 1: Individuating objects in a looking time paradigm

Chapter 1 made two separate claims about the mechanisms that support infants' object representations in the first year of life. In accordance with previous literature, it argued that in some cases object representations are supported by visual indices. And contrary to the received view, it also argued that in some situations infants' object representations are supported by different indexing system(s). The paradigm example of such system is the one where indices are organized around some conceptual/linguistic description of the objects.

If, indeed, there are two or more independent indexing systems, we expect there to be situations in which there are discrepancies between them about the number of indexed entities. For example, in some cases there might be a single visual index but two conceptual descriptions present; conversely there might be two visual indices but a single conceptual description. Findings from Xu and Carey (1996) and similar conceptual individuation studies (Xu, 2002; Futó et al., 2010) can be construed as illustrating the former scenario. The latter kind of scenarios, however, remain largely unexplored, although they are the crucial test environments to differentiate the received spatiotemporal-priority hypothesis from alternative views, where conceptual indexing is not parasitic on spatiotemporal encoding. If in such scenarios we find infants behaving as if their numerical expectations were not determined by the number of visually available objects, this would constitute evidence against the spatiotemporal-priority hypothesis. Furthermore, if their numerical expectations are guided by the number of distinct conceptual descriptions, that could provide evidence for the existence of an independent system of object indexing that relies on these descriptions.

In the present study we aimed to create such scenarios. In order to provide evidence for visual-index based encoding, we first attempted to replicate previous results showing that infants can distinguish objects based on their spatiotemporal characteristics. Then we intended to test whether these expectations can be overridden when non-differentiating conceptual encoding is ascribed to the objects. The purpose of these tests was to establish whether spatiotemporal information always serves as a primary property in indexing. Finally, we attempted to show that if there is sufficient evidence to create two distinct conceptual representations, infants can use this information to succeed in individuating objects. We decided to test 10-month-old infants specifically, as at this age there is strong and reliable evidence that they can use both spatiotemporal (Spelke et al., 1995; Xu and Carey, 1996) and conceptual cues for object individuation (Xu, 2002; Xu et al., 2004). Crucially, at this age there is also evidence that in order to use conceptual cues infants need a communicative framing, and without it, they rely solely on spatiotemporal evidence (Futó et al., 2010). Thus at this age we can systematically manipulate conceptual encoding, by providing or not providing a communicative context.

To test these ideas we devised an individuation paradigm similar in logic to the discontinuous motion condition of (Spelke et al., 1995) and the spatiotemporal condition of (Xu & Carey, 1996). In these paradigms infants sequentially and repeatedly got visual access to objects that were located behind two spatially separated occluders. As the objects never crossed the gap in between the two occluders, it was impossible for a single visual index to track both of them (as long as we accept the empirically supported assumption that visual-index-based object representations are not lost during occlusion). Thus, these scenarios required two visual indices and object representations to track both objects. In prior studies, when the occluders were removed, infants looked longer when presented with one-object outcomes compared to two-object outcomes, indicating that indeed they represented multiple objects. Our strategy was to first replicate these findings, and then in specific follow-up conditions provide additional cues that could serve as input for the second (hypothesized) system of conceptual indexing. In one of these conditions, the two objects shared their conceptual descriptions, resulting in a single indexed entity for the conceptual system. In the second such condition, we presented two distinct linguistic descriptions to induce the creation of multiple conceptual indices. The conceptual cues were presented linguistically by labeling the objects during presentation in an ostensive-referential manner. In a variety studies, infants around 10 months of age were shown to create conceptual contrasts between objects based on the differences in their labels (Xu, 2002; Dewar & Xu, 2007, 2009).

In these experiments we planned to test the following hypotheses:

(1) Infants expect multiple objects if the available spatiotemporal evidence implies the presence of multiple objects and there is no interference from linguistic/conceptual information.

(2) Infants' numerical expectations are not based on the number of spatiotemporally available objects, when they can encode the objects under conceptual descriptions:

    A. If a single conceptual description is available, they will not expect multiple objects irrespective of the visually available information.

    B. If distinct conceptual descriptions are available, they will expect multiple objects irrespective of the visually available information.

2.1 Experiment 1 — Continuous Path

2.1.1 Method

To test our three hypothesis we planned to run three conditions. To assess Hypothesis 1 we first aimed to conceptually replicate the original spatiotemporal object individuation paradigm (spatiotemporal condition). Contingent on infants' success in this condition, we planned to add two other conditions, both of them containing communicative cues: pointing and verbal labeling. One of these planned conditions was the same-label condition, where the two objects shared their conceptual descriptions (to test Hypothesis 2A). The second was the different-label condition where we planned to use distinct labels (to test Hypothesis 2B).

The approach of data collection was motivated by the planned Bayesian statistical methods in data analysis. Rather than predefining sample size, we continuously collected and analyzed the results. We could do this as Bayesian statistics is not prone to type I errors; with the disadvantage of not having access to direct significance testing: a standard in the current literature. Thus, data collection was concluded when two conditions were met. The analysis could sufficiently distinguish between possible statistical hypotheses (any hypothesis had to be at least 10 times more likely than the alternative), and the sample had to be counterbalanced. For the analysis we used the fixed effect size variant of the toolkit developed by Csibra, Hernik, Mascaro, Tatone, & Lengyel (2006). Based on the log-transformed looking times this analysis calculates the log-bayes factor for two mutually exclusive statistical hypotheses. It compares the likelihood that infants looking times were longer in response to the incongruent outcome compared to the congruent one (H1), with the likelihood that there is no looking time difference in response to the two outcomes (H0).

We decided to create 3d animated stimuli as it allows for more precise stimulus presentation, and more controlled presentations in general.Infants succeed in a variety of tasks using animated displays that involve goal attribution (Gergely, et al., 1995), social actions (Tatone, Hernik, & Csibra, 2019), word learning (Yin and Csibra 2015) and also object individuation (Surian & Caldi, 2010). As such, we did not expect that this change in the presentation medium to have an effect on our results.

2.1.1.1 Participants

A total of 12 participants were successfully tested (mean age = 10 months 14 days; ranging from 10 months 1 days to 10 months 28 days. An additional 7 infants were excluded from the sample: Experimenter errors concerning the live coding occurred 2 times, 3 infants were fussy and did not attend to the stimulus presentation, and 3 infants looked for the maximum duration for both test trials. We only tested the spatiotemporal condition reported below, as the planned labelling conditions were dependent on results obtained here. The study was approved by the Hungarian United Ethical Review Committee for Research in Psychology (EPKEB). Caregivers were contacted by mail and telephone, and signed informed consent forms before participation. Infants received small toys as gifts after the experiment.

2.1.1.2 Procedure

The infants were seated in a closed experimental room on their caregivers' lap. A hidden camera was used for recording them, located below the 40-inch computer screen that was used for presentation. The screen was distanced approximately 70 centimeters in front of them. The camera fed into a mixer that produced a split-image recording of the infant and the stimuli. The experimenter controlled the stimulus presentation, and coded infants' looking online from outside the testing room. Caregivers were instructed not to talk or point, and were either wearing blinded sunglasses during the stimulus presentation or were asked to close their eyes.

2.1.1.3 Stimuli

The stimuli were computer generated 3d animated displays (Figure 2.1, Exp. 1). The complete experiment consisted of 3 trial types: introductory trials, baseline trials, and test trials. All three trial types shared the same contextual cues: a floor with a wooden pattern, two spatiotemporally separate blue occluders, and a striped background wall. During introductory trials, two objects were used: a short pink polka-dotted cylinder in horizontal orientation, and a long vertically aligned cylinder on three white legs that supported it. During baseline and test trials, 4 pairs of unfamiliar looking objects were used. Two of these pairs were used during baseline, and two during test trials (counterbalanced across infants). The objects were identical within pairs but differed across pairs in shape and color, while being roughly the same in size. Additionally, a green arrow was present in the spatiotemporal condition, and a pointing hand was planned to be used in the labeling conditions. Between consecutive trials an abstract attention-grabbing stimulus appeared at the center of the screen, and if infants were not attending, an experimenter-controlled beeping sound guided them to reorient.

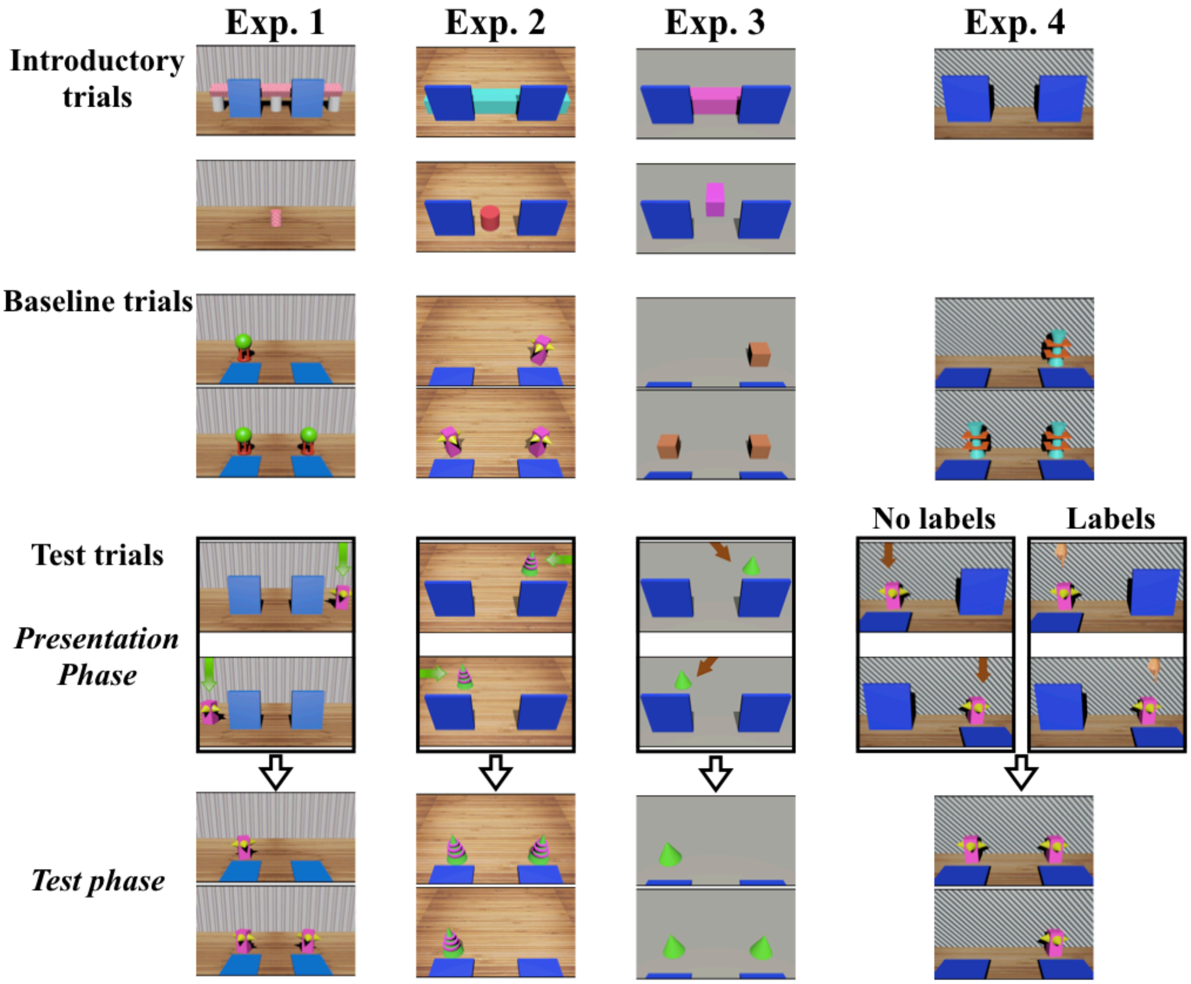


Figure 2.1. Stimulus and design of Experiments 1-4. In one of the introductory trials infants were presented with the movement patterns of the occluder. In the other introductory trial infants witnessed a moving object (not present in Experiment 4). In the 4 baseline trials, one or two objects were revealed after the occluders dropped. In the *presentation phase* of test trials two objects were revealed repeatedly. The depicted frames of these phase were preceded by the objects moving horizontally in Experiment 1, and vertically in Experiment 2 and 3. In Experiment 4, the occluders moved instead of the objects, and in the labeling conditions the arrows were replaced by a pointing hand. After repeated presentations both occluders dropped revealing the outcome (*test phase*).

2.1.1.3.1 Introductory trials

The presentation started with the two introductory trials (14 s each), which were shown in random order to familiarize infants with the setup and the spatiotemporal characteristics of the display. One of them aimed to introduce infants to moving objects by showing them a small cylinder traveling back and forth between the two horizontal endpoints of the screen. In this trial, no occluder was present. The other introductory trial aimed to familiarize infants with how the occluders work in space. The trial started with the large horizontally oriented cylinder placed behind the two occluders so that it was visible both in the spatiotemporal gap in between the two occluders, and both extending from behind each occluder. The occluders repeatedly dropped revealing the whole object and returned to their upright position partially hiding it. Both introductory trials were accompanied by attention-grabbing ringing sound effects.

2.1.1.3.2 Baseline trials

After the introductory trials, in four baseline trials infants' were shown one- or two-object outcomes. The trial started with the occluders in upright position for 1.5 seconds. Then the occluders dropped (1 s) revealing either one or two identical looking objects. The four trials followed either a 1,2,2,1 or a 2,1,1,2 order, counterbalanced across infants. Two of the four object pairs were used in the baseline trials so that either the (a) and (b) pairs or the (c) and (d) pairs were used. Additionally, within participant, we counterbalanced the order of the displayed object pair, alternating every trial. The location of the single object in the one object outcomes was counterbalanced (left, right; or right, left). After the occluders dropped, the length of the trial was contingent on infants' looking behavior. If infants looked away from the screen for 2 seconds or longer, or looked for the predefined maximum looking duration (30 s), the trial ended.

2.1.1.3.3 Test trials

During the two test trials, we presented visual evidence that there were two objects present, one behind each of the two occluders (*presentation phase)*. At the end of these trials, we measured infants' looking times to the one- or two-object outcomes (*test phase*). Every trial started with the occluders in upright position with a single visible object placed laterally (not in the gap). Then a green arrow descended from above the screen pointing to the object (*object presentation*). The arrow stayed still before it returned to its starting location above the display (4 s). During this time music was played to keep the infants engaged. This *object presentation* event was included as a control for our planned labeling conditions, where we intended to switch the arrow to a pointing hand, and the music to verbal labeling. After the disappearance of the arrow, the object moved horizontally behind its corresponding occluder (1 s). Then after 2 seconds an identical looking object emerged from behind the other occluder (1 s), reaching its final location that mirrored that of the first object on the other side of the screen. This *object movement* event, which took 4 seconds in total, was set up in a way that the movement pattern of the two objects was consistent with a single object 'invisibly' crossing the gap in between the occluders. This set of events – object presentation and object movement – was repeated 3 times in total. After the third movement event, an extra presentation event followed in order to present both objects 2 times each, then the object returned behind occlusion (1 s). After a 1.5 seconds delay, both occluders dropped, revealing either only one or both of the objects previously seen (*test phase*). For one-object outcomes, the object revealed was always behind the occluder where the last object disappeared. Whether or not this location was left or right was counterbalanced between infants, just as the trial order (one-object first/ two-object first), and the objects that were used in these trials. After the occluders dropped, the length of the trial was contingent on the infants' behavior, and the last frame of the stimuli was presented until the trial ended. If infants looked away from the screen for longer than 2 seconds, or looked for the maximum 30 seconds according to live coding, the trial was terminated. Infants that looked for the maximum 30 seconds for both test trials were excluded from the sample.

2.1.1.4 Coding

Looking-time data of the test trials were offline coded by the author on a frame-by-frame basis. This measurement started from the first frame that the object(s) became visible in the test trials.

2.1.2 Results and discussion

As shown in Figure 2.2, infants looked at the two outcomes for a similar period of time ($M_{one\text{-}obj}$ = 9.19 s, SD = 6.42 a; $M_{two\text{-}obj}$ = 9.78 s, SD = 5.02 s). The statistical analyses were based on log-transformed data.We calculated the Bayes factor to probe the likelihood that (H1) infants looked longer for the one-object outcome than the two-object outcome versus (H0) that there was no difference between the outcomes – given our data. We were using the fixed effect size variant of the toolkit developed by Csibra et al. (2006) for analysis, and concluded that there was no difference between the two groups ($\log_{10}$BF = -1.10). This result shows that the looking times of the two outcomes were 12.6 times more likely coming from the same distribution, an outcome that exceeds the usual interpretative strength of null results.

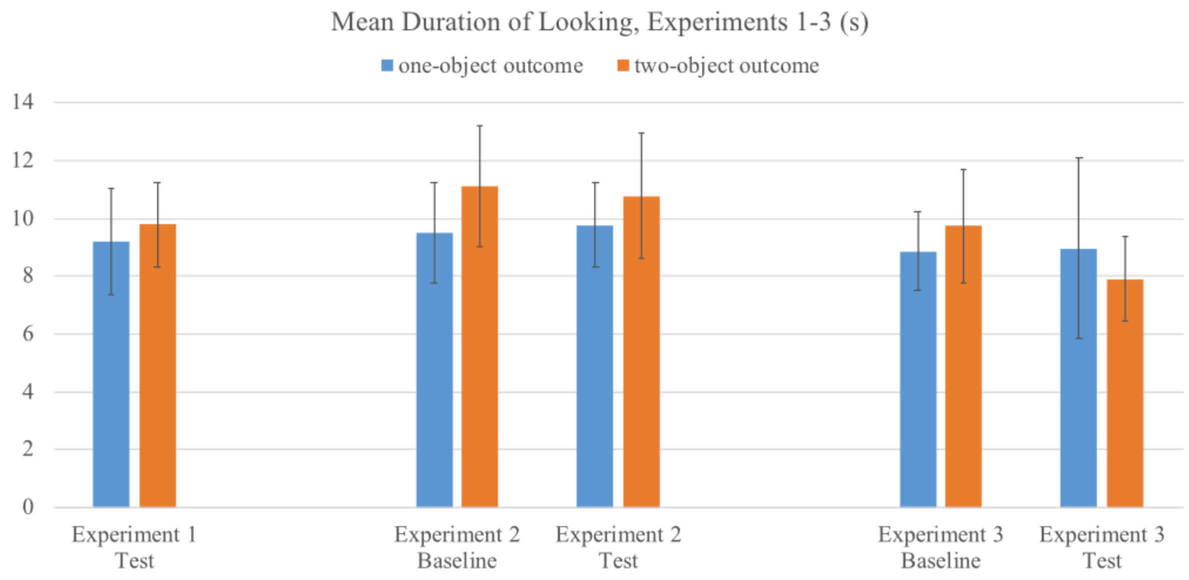


Figure 2.2. Mean duration of looking times in Experiments 1-3 (seconds) in response to one-object versus two-object outcomes. Error bars represent the standard error of the mean.

Experiment 1 failed to replicate the original spatiotemporal individuation finding. The cause of this failure cannot be deduced from the results, but we hypothesized that it was due to surface characteristics of our experimental stimuli. We used a procedure that was very similar to to the original study (Xu & Carey, 1996), but with animated displays. These animations might be different from live stimuli in a number of respects. For example, animations do not produce motion parallax, a depth cue that even young infants are sensitive to (Condry & Yonas, 2013). Thus, proper depth perception might have been harder, possibly resulting in an imperfect appreciation of the spatial relations among animation elements. Compared to previous studies, object movement might have been more perfectly aligned both in speed and direction: possibly increasing the likelihood of the interpretation of the continuous motion of a single object. We decided to run Experiment 2 with the goal of changing the stimuli in ways that might make the individuation easier for infants, even at the cost of making it less similar to the original spatiotemporal individuation studies.

2.2 Experiment 2 — Discontinuous Path

Our main goal in designing experiment 2 was to make sure that infants succeed in the spatiotemporal individuation task. The most important changes were the following. We changed object movement from horizontal to vertical, so that a single continuous object movement couldn't be a viable interpretation anymore. We shortened occlusion times, so those time periods where both objects were occluded became substantially shorter. We changed various surface characteristics of the stimuli to make the objects more salient. We also collected baseline looking times, to assess whether infants have a strong baseline preference for the two objects versus the one object outcome.

2.2.1 Method

The methods used for experiment 2 were the same as in experiment 1 except for the changes discussed below.

2.2.1.1 Participants

In Experiment 2, we stopped data collection at 12 participants (mean age = 10 months 10 days; ranging from 10 months and 0 days to 10 months and 25 days). An additional 10 infants were excluded from the sample: Four due to fussiness, 3 due to technical failures of the presentation screen, 2 due to errors in live coding, and 1 due to parental interference.

2.2.1.2 Procedure and Stimuli

While the general procedure and the apparatus stayed the same as in Experiment 1, we made substantive changes to the surface features of the stimuli (Figure 2.1, Exp. 2). The target objects, the occluders, and the surrounding objects were the same, but we changed the distance and the viewing angle of the camera to the events. This way, the viewer had a steeper perspective, as the camera was placed higher and closer to the occluders. From this perspective, the back wall was no longer visible, and the background consisted only of the textured floor. These changes were implemented in order to increase the perceived size and salience of the objects and the occluders. The modifications also helped to accommodate the altered motion paths that we used in the test trials (discussed below). To make these scenarios simpler, we also introduced small changes to the objects in the introductory trials: The small cylinder's polka dot pattern was changed to a simple red texture, while the complex large cylinder was changed to a simple elongated blue cuboid.

2.2.1.2.1 Introductory trials

The only change we made in the introductory trials was that the occluders were now present in both trials; even in the one that familiarized them with the object movement. This way infants got direct visual access to an object that moved continuously between the two occluders becoming visible in the gap in between them. We hoped that this trial would acquaint infants with how object movement should look like in the test trials if there was only a single object present.

2.2.1.2.2 Baseline trials

We did not make any structural changes the baseline trials. However, the looking times of the infants were not only used for controlling stimulus duration, but also as a dependent measure.

2.2.1.2.3 Test trials

We changed the test trials by presenting vertical, instead of horizontal, object motion. This change was implemented to help infants appreciate the spatial configuration of the setup and make it even easier to discard a continuous motion interpretation of the movement of the objects. The test trials started with only the two occluders visible (0.25 s). Then one of the objects emerged from behind an occluder, and moved towards the top of the screen, orthogonally to the other occluder (1.75 s). After the object stopped, an arrow descended at the corresponding side from the top of the screen, oriented sideways pointing at the object (0.25 s). The arrow stayed stationary for 3.5 seconds before returning to its' starting location out of vision. After an additional 0.25 seconds period of staying stationary, the object returned behind the occluder, using the same motion path as before presentation, but in the opposite direction (1.75 s). This presentation was repeated four times, alternating sides for every repetition. After the last repetition, both objects were occluded for an additional 1 second, after which the occluders fell, and measurement period started the same way as in Experiment 1.

2.2.1.3 Coding

Looking time data of both the baseline and test trials were offline coded by the author on a frame-by-frame basis.

2.2.2 Results and discussion

As in Experiment 1, we compared how long infants looked at the two possible outcomes in the test trials ($M_{\text{one-obj}}$ = 9.78 s, SD = 4.99 s; $M_{\text{two-obj}}$ = 10.78 s, SD = 7.48 s). Again, we calculated the Bayes factor to probe the likelihood that (H1) infants looked longer for the one-object outcome than the two-object outcome versus (H0) that there was no difference between the outcomes. We found no difference between the two conditions ($\log_{10}BF = -1.03$).

We separately analyzed baseline looking data in a similar fashion. After averaging the two trials for each participant, we proceeded with the same analysis ($M_{one\text{-}obj} = 9.49$ s, SD = 6.06 s; $M_{two\text{-}obj} = 11.11$ s, SD = 7.30 s). For the baseline analysis, the statistical hypothesis changed, as in accordance with the literature we expected infants to look longer for two objects compared to one (e.g., Xu & Carey, 1996; Xu, 2002; Surian & Caldi, 2010). Thus, we calculated how much more likely was that (H1) infants looked longer at the two-object outcome compared to (H0) that there was no difference. The null-hypothesis again was the more likely hypothesis, although not statistically a strong effect ($\log_{10}BF = -0.52$).

Once more, there were no clear evidence that infants individuated objects based on their different locations. We did not find a looking-time increase to the one-object outcome in the test trials compared to the baseline either. The repeated failure to find evidence for spatiotemporal object individuation, and the fact that there was no baseline difference made us further consider the explanations for infants' performance. Because none of the three analyses conducted so far identified a systematic difference of looking times between outcomes, it seemed plausible that infants either did not focus on the number of objects present — for example, because they were preoccupied with some other aspect of the presentation —, or that they had trouble with looking away from the 40-inch screen used for stimulus presentation. Note that the size of this screen was larger than what is usually used for presenting stimuli in looking-time studies.

2.3 Experiment 3 — Discontinuous Path, Simplified Version

There were two main goals in designing Experiment 3. We aimed to remove all superfluous details of the stimulus presentation of the previous experiments, and to reduce the size of the stimuli by changing the experimental apparatus. Other than the changes mentioned below, every detail remained the same as in Experiment 2.

2.3.1 Method

2.3.1.1 Participants

In Experiment 3, we stopped data collection at 8 participants (mean age = 10 months 5 days; ranging from 9 months and 18 days to 10 months and 13 days). An additional 4 infants were excluded from the sample: 3 due to fussiness, and one due to looking for the maximum duration in both test trials.

2.3.1.2 Procedure

We changed the location of the experiment. Experiment 3 took place at an experimental room separated by a curtain from the experimenter. The presentation screen was smaller, 24-inch computer monitor, while the distance of the infant to this screen remained the same (70 cm). We reasoned that this change might yield better results, as infants can more easily orient away from stimulus presentation in the baseline and test trials.

2.3.1.3 Stimuli

We also made changes in the stimulus presented (Figure 2.1, Exp. 3). We removed the superfluous details of the stimuli that were intended to make it more interesting but might have contributed to infants attending to aspects of the scene that were irrelevant for object representation. The scene now lacked the wooden texture of the floor as it was changed to a monochrome grey color. The complex test and baseline object pairs were now switched to simple geometrical shapes: pairs of brown cubes, red spheres, yellow toruses, and green cones. In the introductory trials, a big and a small pink cuboid were present. In the test trials, the music that played during object presentations were swapped to various sound effects (different rings, beeps, trumpets etc.) that were presented in a randomized order.

2.3.1.3.1 Introductory trials

The introductory trial presenting the occluder movement did not change. The two occluders repeatedly fell to reveal the parts of the large horizontally placed cuboid they covered. As in Experiment 2, when the occluders were in the upright position, the object was still visible in the spatiotemporal gap. Some more substantive changes were introduced to the trial that presented object movement. We aimed to emphasize even more the spatiotemporal gap in between the two occluders. This was realized with the small cuboid not only moving in a horizontal direction, in between the two occluders, but also stopping when visible at the middle of the spatiotemporal gap. Here it moved back and forth for a short period of time, providing more information about the spatial arrangement of the stimuli.

2.3.1.3.2 Baseline trials

The baseline trials were exactly the same as in Experiment 2 except for the changes in the surface features of stimuli described above.

2.3.1.3.3 Test trials

The test trials were almost the same as in Experiment 2, except for the changes discussed above: objects, background, and the auditory cues during presentation phase. We also changed the behavior of the arrow pointing to the objects. Now the arrow was present for the full duration of the test trials, as it left only after the last object presentation event. The location of the arrow was now changed to the top midpoint of the screen. Its movement pattern changed as well: now it rotated towards the currently presented object rather than targeting the object by a descending motion. Our rationale for this change was to minimize non-relevant motion, and to decrease the load on infants' attention.

2.3.2 Results and discussion

As in the previous experiments, we compared how long infants looked at the two possible outcomes in the test trials ($M_{\text{one-obj}}$ = 8.96 s, SD = 8.81 s; $M_{\text{two-obj}}$ = 7.91 s, SD= 4.10 s). We calculated the Bayes factor contrasting the hypothesis that infants looked longer for the one-object outcome (compared to the two-object outcome) and the null hypothesis that there was no difference between the two outcomes. We concluded that there was no difference between the two looking time distributions ($\log_{10}$BF = -1.50).

We examined the baseline looking data in a similar fashion. Again, after averaging the two trials for each participant, we proceeded with the same analysis ($M_{\text{one-obj}}$ = 8.86 s, SD = 3.89 s; $M_{\text{two-obj}}$ = 9.75 s, SD = 5.56 s). The null-hypothesis was the more likely hypothesis, but not statistically a strong effect ($\log_{10}$BF = -0.60).

The results were in line with our previous experiments. Infants not only failed to individuate objects, but also did not look longer at the two- compared to the one-object outcome in the baseline either. By this point, we had experimented with changing a large variety of surface features of the stimuli (in Experiments 1-3), even if not all possible combinations were tried for obvious combinatorial reasons. We thus decided to analyze the possible reasons for infants' repeated failure, rather than hoping that changing some surface-level visual features could remedy the problem.

The most important difference between the experiments reported above and the ones that reported success of spatiotemporal object individuation in infancy is the presentation medium. Previous studies used real-life presentations (Spelke, et al., 1995; Xu and Carey, 1996) or recorded videos (Kuhlmeier et al., 2004, Experiment 1). As infants can indeed succeed with video-recorded stimuli, it was not the screen-based presentation in general that caused infants' failure in Experiments 1-3. Also, there is ample evidence that infants at this age can understand physical and spatial properties of 3d animated displays when reasoning about agents (e.g., Tatone

& Csibra, 2015; Surian & Caldi, 2010) and their goals (Gergely, et al., 1995). These agents were often completely lacking physical agency cues, like biological motion or facial features, such as eyes.

There is a distinct possibility that would account for infants' failures in a way that is consistent with Hypothesis 1. In principle, it is possible that infants conceptualized the objects in our stimuli as agents. This, in turn, might be responsible for grounding infants' expectations based on conceptual indexing rather than two visual indices. That is, if infants conceptualized the entities presented under a the concept AGENT, spatiotemporal cues for object individuation might have ceased to function similarly to what we predicted in the same-label condition of the study. The validity of this interpretation is backed by prior evidence of 5-month-old infants' failure to use spatiotemporal cues for agent individuation in a similar design (Kuhlmeier et al., 2004). Is it possible that infants construed these objects as agents? Although self-propelled motion in isolation is usually not taken as sufficient cue for agency (Csibra, 2008), there is not much direct evidence on this question using 3d animated stimuli. In our study the movement was not only extremely fluid, but also displayed acceleration and deceleration patterns. These gradual speeding-up and slowing-down behaviors that might be construed as efficiency in action execution. A further possible agency cue that infants might have used is contingent reactivity. In different studies, 10-to 12-month-old infants took contingent reactivity both in a second person scenario (Johnson, 2003) and in third person one (Tauzin & Gergely, 2019) as cue for agency. In our stimuli, the "actions" of the objects were in perfect contingency (disappearance behind one occluder predicted appearance behind the other), infants might have construed these events in such way. But the most relevant cue for attributing agency, *goal-directed action,* might be also inadvertently present: the movement of the objects was consistent with an interpretation that they repeatedly approached the occluders. To sum up, self-propelled motion, goal-directed behavior, contingent reactivity are agency cues that were to some degree present in our experiments. While these cues to a varying degrees were present in other spatiotemporal object individuation studies in which infants did succeed (Xu & Carey, 1996, Spelke et al., 1995), it is possible that our 3d animated presentation highlighted these cues more, resulting in this failure. We will discuss this possibility more detail in the General Discussion. While we have no direct evidence for this interpretation, in order to make sure that this possibility is excluded, we made substantial changes for Experiment 4.

2.4 Experiment 4 — No Object Motion

Experiment 4 aimed to completely remove all presented agency cues of the objects while retaining conceptual equivalency with the previous experiments. We achieved this by changing the general structure of the test trials. Instead of presenting infants with moving objects, we decided to show moving occluders instead. By sequentially and repeatedly revealing what was behind each occluder, the same amount of spatiotemporal information was made available about the objects. This allowed us to expose infants to the objects for the same duration as in previous experiments. As the moving occluders were already present in previous experiments (and in some previous spatiotemporal individuation studies) the general complexity of the stimuli did not increase.

2.4.1 Method

2.4.1.1 Participants

In Experiment 4 we tested all three planned between-subject conditions. These were tested in the following order: spatiotemporal condition, same label condition, different label condition. In the spatiotemporal condition, we stopped data collection at 24 participants (mean age = 9 months 28 days; ranging from 9 months and 14 days to 10 months and 16 days). In the same label condition data collection concluded after 20 infants (mean age = 9 months 29 days; ranging from 9 months and 16 days to 10 months and 14 days). In the different label condition, we tested 24 participants (mean age = 10 months; ranging from 9 months and 14 days to 10 months and 16 days). Other than the 68 successfully tested infants, we had to exclude further 40 infants (exclusion rate 37%). Out of the 40 excluded cases, 12 were caused by technical issues with the apparatus (4 with the recording equipment, and 8 due to script errors), and 9 were caused by errors in the live coding.

A further 14 infants were fussy, and 4 were looking for the maximum duration in both test trials. A single participant was removed because of parental interference.

2.4.1.2 Procedure and Stimuli

We used the same testing booth as in Experiment 3, but the general features of the stimuli were a mixture of the previous experiments (Figure 2.1, Exp. 4). We retained the richer surface features of stimuli of Experiment 1 and 2: the wooden pattern of the floor, a textured wall, and the more complex objects. The arrow presented in the spatiotemporal condition was brown (as in Experiment 3), which in the two labelling conditions was swapped for a photograph of a downward pointing hand of similar size. The sound stimuli of the spatiotemporal condition were the same as in Experiment 3. For the labelling conditions, we used two Hungarian nonwords (“bitye” and “tacok”), and created 6 short Hungarian phrases for each (e.g., “Look! A tacok” or “Wow! A tacok”). We created all possible 12 phrases, which were then used in a randomized order.

2.4.1.2.1 Introductory trials

As there was no object movement present in the test trials, we only presented a single introductory trial showing the occluders’ movement. The trial started with both occluders in the upright position. Then one of the occluders fell (1 s) revealing the empty space behind (1 s). This was followed by it returning to its upright position (1s). This sequence was repeated 6 times, with alternating occluders.

2.4.1.2.2 Baseline trials

The baseline trials remained structurally identical to previous experiments, with the changes explained above.

2.4.1.2.3 Test trials

The test trials in all three conditions had the same event structure as in Experiments 1-3. The presentation phase started with one of the objects visible, while the other one stayed hidden behind the corresponding occluder (4 s). During this time, either the arrow (in the spatiotemporal condition) or the hand (in the two labelling conditions) was pointing to the visible object. In the spatiotemporal condition, a sound effect was played for this duration, while in the labelling conditions one of the 12 phrases was used instead. This was followed by the movement phase: the corresponding occluder was raised, hiding the object; simultaneously the arrow/hand horizontally moved above the opposite occluder (1 s). Then the other occluder fell revealing the other object (1s). These events were repeated on alternating sides for a total of 6 times (3 presentations for each object). In the spatiotemporal condition, a different sound was played during every presentation. In the labelling conditions, a different phrase was played during every presentation. In the same label condition, both objects were labelled with the same label, while in the different label condition, the two objects were labelled with different labels. This way both labelling conditions had the same variance in the carrier phrases, the only difference between the labeling conditions was whether infants heard one or two labels. At the end of the 6th presentation, the arrow/hand moved to the top midpoint of the screen (1 s). During the next 1 second two events simultaneously happened: The arrow left the screen, and the occluder was raised, so that neither objects was visible anymore. After 2 seconds of delay, the occluders simultaneously dropped revealing either the one-object or the two-object outcome. Further counterbalancing and coding was exactly as in the previous experiments.

2.4.2 Results and discussion

We analyzed looking time differences between the one- versus the two-object outcomes separately in all three conditions (Figure 2.3). In all trials, we compared the statistical null-hypotheses (that there is no looking time difference between the outcomes) with specific alternative hypotheses. While in the baseline trials these alternative hypotheses were that infants looked longer during the two-object outcome, it was the opposite in the test trials: that total looking time was longer for the one-object trials.

In the spatiotemporal condition, while we found evidence in the baseline trials that the infants did not look longer for the two-object outcome ($M_{two\text{-}obj}$ = 11.32 s, SD = 6.19 s) than the one-object outcome ($M_{one\text{-}obj}$ = 11.93 s, SD = 6.14 s; $\log_{10}BF_{baseline}$ = -2.04), this pattern changed for the test trials. There, infants looked longer for the one-object ($M_{one\text{-}obj}$ = 13.77 s, SD = 7.47 s) compared to the two-object outcome ($M_{two\text{-}obj}$ = 10.91 s, SD = 7.84 s), providing evidence that infants individuated the objects ($\log_{10}BF_{test}$ = 1.22).

In the same label condition, both in the baseline and in test trials we found that the null-hypotheses were more likely than the corresponding alternative hypotheses ($\log_{10}BF_{baseline}$ = -1.36, $\log_{10}BF_{test}$ = -3.18). Although there was a small numerical difference both in the baseline ($M_{one\text{-}obj}$ = 11.64 s, SD = 5.58 s; $M_{two\text{-}obj}$ = 13.05 s, SD = 6.34 s) and in the test trial ($M_{one\text{-}obj}$ = 13.32 s, SD = 10.16 s; $M_{two\text{-}obj}$ = 13.16 s, SD = 7.43 s) we interpret the results as evidence that infants did not individuate the objects.

The results of the different label condition were quite similar to the ones acquired in the same label condition. The null-hypotheses were more likely in both baseline and test ($\log_{10}BF_{baseline}$ = -1.84, $\log_{10}BF_{test}$ = -1.36). In both cases, infants numerically looked longer for the one-object displays. The mean looking times in the baseline were $M_{one\text{-}obj}$ = 12.49 seconds (SD = 7.75 s) and $M_{two\text{-}obj}$ = 12.36 seconds (SD = 6.54 s), while in the test they were $M_{one\text{-}obj}$ = 12.71 seconds (SD = 10.45 s) and $M_{two\text{-}obj}$ = 11.79 seconds (SD = 9.47 s). As in the same label condition, the evidence points toward infants failing to individuate objects.

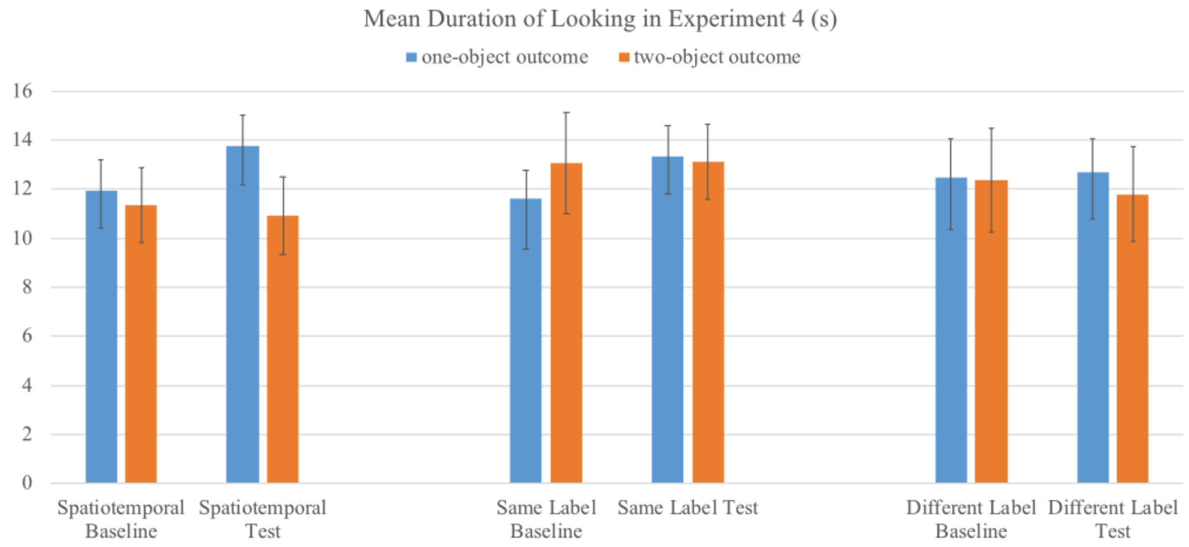


Figure 2.3. Mean duration of looking in the different conditions of Experiment 4. Error bars represent standard error of the mean.

Experiment 4 can be considered a partial success. First, infants (finally) succeeded in the purely spatiotemporal condition. This shows that, even with animated displays, infants can track and remember multiple objects, with their location as their sole differentiator (Hypothesis 1). On the other hand, infants failed to individuate both in the same label condition and in the different label condition. The failure in the same label condition provided evidence for Hypothesis 2A while the different label condition speaks against Hypothesis 2B. The latter result is clearly surprising in light of the fact that infants had the opportunity to use either information source alone or together to individuate objects. This raises the question whether our labelling conditions were adequate, and whether infants conceptualized the objects in the first place. If they did not, it would explain a failure in a study where only conceptual individuation cues were present. Notably, this study also presented redundant spatiotemporal cues. If infants did not conceptualize the objects, they could have simply individuated them based on location information, just as they did in the spatiotemporal condition. On the other hand, if infants successfully conceptualized the objects, what could be the reason for their failure in the different label condition?

A possible explanation is to assume that the infants had no trouble using labelling events to conceptualize the objects, but they failed to build expectations about the contextual presence of these entities. For example, if infants interpreted the labelling conditions as a “word learning game,” the context might not have warranted the evaluation of the immediate presence or absence of the referents. If the physical presence of the referent of labelling is construed not to be relevant in the context of the presentation, infants might have inhibited drawing valid but local inferences about the “here and now”. This hypothesis is indirectly testable. If in such a context infants actually learn the labels, it might imply that the failure of individuation was not due to their lack of conceptualization, but rather to the failure to build expectations about the presence of particular objects. In other words, object conceptualization in some contexts might not only hinder spatiotemporal individuation of objects, but individuation as particulars in general.

2.5 General discussion

In multiple experiments we examined infants' capacity to individuate objects both spatiotemporally and conceptually. Our main goal was to establish whether (1) spatiotemporal properties always take priority in object individuation, and whether (2) conceptual/linguistic information can override expectations based on spatiotemporal cues. With the data gathered in the current study, we can provide a strong negative answer to (1) though we were unable to find definitive support for (2). This conclusion is partially consistent with the architectural stipulations of object indexing that I argued for in Chapter 1. However, the pattern of the obtained results raises a number of questions.

Why is it the case that with a seemingly minor shift to 3d generated animations, previously reported object individuation effects (e.g., Xu & Carey, 1996) did not self-evidently carry over? Out of the four experiments where we expected infants to individuate objects based on spatiotemporal cues, they only succeeded once. After the failed Experiment 1, we tried multiple manipulations both in our stimuli and in our procedure to increase infants' performance in Experiments 2 and 3, without success. As these surface feature level changes did not yield better results, we hypothesized that infants might be more prone to attributing agency when using animated medium. This agency attribution might cause difficulties, as individuating agents using spatiotemporal cues seem to be harder than with objects (Kuhlmeier et al., 2004). When all agency cues were removed – as the objects were stationary – infants succeeded (Experiment 4, spatiotemporal condition). Although we cannot take this as direct evidence that infants' previous failures were due to the attribution of agency, at least we know that it is possible for infants to rely on spatiotemporal cues for object individuation when presented with animated displays. Thus, Experiment 4 provided evidence in favor of Hypothesis 1. However, it is important to consider the possibility that this success is not a reliable finding, given the failures that preceded it. Further, more focused research might be required to understand the interaction of animated displays and spatial cognition.

We found supporting evidence in favor of Hypothesis 2A, as in the same label condition of Experiment 4, infants' individuation performance plummeted. When the two objects were target of referential communication and labelled with the same label, infants seemingly did not expect multiple objects to be present – regardless of the spatiotemporal cues available to them. This implies that the linguistic/conceptual descriptions of the object can interfere with a visual-index based encoding. But in order to show that infants in this condition actually used a different indexing-system, one that can use labels/conceptual descriptions to establish object identity, we need further evidence. When we directly tested this question in the different label condition, infants failed, providing evidence against Hypothesis 2B. This is a striking failure given that this condition contained both spatiotemporal and conceptual cues that provided evidence for the presence of two objects. When trying to interpret these results, we encounter a paradox. If our labelling manipulations succeeded in eliciting conceptual object encoding, why did the infants fail in the different label condition? If the labeling manipulation did not succeed, why did the infants fail to use the available spatiotemporal evidence to individuate the objects?

To resolve the paradox, one possible approach is to question the result of the spatiotemporal condition of Experiment 4. Although we obtained results that indicated that infants succeeded in spatiotemporal individuation, we cannot exclude the possibility that the results we obtained were due to chance (the odds are ~17:1). If that were the case, we would have a consistent set of null-results, which would paint a completely different picture. If infants did not individuate in any of our experiments, then either there was still some specific flaw in our stimuli, or possibly infants are not sensitive to individuation cues when presented with animated displays. There is no theory at the present moment that would explain why animated displays could be different than live presentations or video stimuli. Furthermore, as we cited before, there is a already large literature of looking time studies that successfully used animated presentations. These studies sometimes involved scenarios that required object individuation, or at least the maintenance of object identity. For example in a study by (Hernik & Southgate, 2012), 9-month-old infants were familiarized to repeated presentations of an agent approaching the single object in the scene. In the test trials, the agents had two potential target objects to approach. Infants looked longer when the approach was directed towards the novel object, showing that they maintained the identity of the original target object and differentiated it from the novel one.

If infants have the capacity to individuate objects but can selectively refrain from using this ability depending on presentation medium, it would implicate that either the process of individuation is not automatic, or at least that it does not necessarily result in measurable expectations. Thus, the specific predictions that infants make during the presentation of animated stimuli might be based on context sensitive inferential processes rather than on the automatic and encapsulated processes of indexing objects. In this sense, it could be speculated that infants treat animated displays more like an act of communication rather than an actual real-life scenario.

Even if we disregard this auxiliary hypothesis about animated displays, the results from Experiment 4 raise the possibility that infants' expectations were set up within a broader inferential scheme than just the processes of object-indexing, because they disregarded both conceptual/linguistic and spatiotemporal cues. One such possibility is that the infants understood the labelling conditions as intending to teach them new words. More generally, any interpretation where the referent of the communication is not bound by either the location of the object, or by the time it is observed could result in cessation of local expectations. We will return to this proposal in Chapter 4. But irrespective of this line of thought, previous studies measuring conceptual individuation did work with similar metrics to ours (e.g., Xu, 2002). Thus, it cannot be the case that infants would always fail to build expectations in the "here and now" upon encountering linguistic/conceptual cues.

The results of the current study challenged a mainstream assumption about the relationship between spatiotemporal and conceptual object individuation. This assumption is that, for young infants, objects are always tracked via visual indices. The individuation cues are generally thought of as disjunctive, where either of them in isolation, or both of them in conjunction, could suffice to establish representations of multiple objects. In Chapter 1, we abandoned the assumption that there would be a single indexing system, and argued that when conceptual cues are available, indexing these alone provides the input for identity judgements. This proposal is still consistent with idea that both cues in isolation could be used for object individuation, but predicts specific failures in cases like the same-label condition, where the conceptual information underdetermines the number of objects that are actually present. But the results from the different label condition of Experiment 4 raise the possibility that the integration of these different information sources is even more impaired at 10 months of age than we assumed. Maybe object individuation completely breaks down in cases where both cues are simultaneously present, even if they both provide an identity contrast (evidence for two objects). This counterintuitive proposal might be tenable, as in prior research spatiotemporal and conceptual individuation were never used in conjunction, at least not at this age. This possibility will be discussed in more detail in the general discussion of Chapter 3.

In sum, individuating objects is not straightforward for 10-month-old infants when presented with animated displays of self-propelled objects. After removing such agentive features, infants succeeded in using spatiotemporal information for object individuation. In further conditions, labeling was shown to interfere with this process, but not the way we predicted. Independent of the specific conceptual information presented, infants failed to build expectations that multiple objects were present. It is an open question whether 10-month-old infants can integrate conceptual and spatiotemporal individuation cues at all. In addition, these findings raise the possibility that infants treat animated presentations in a more general explanatory and inferential framework, like communication.

Chapter 3 — Study 2: Individuating objects in the manual-search paradigm

3.1.Lessons learnt from Study 1

In Chapter 2, we set out to systematically assess how 10-month-old infants integrate spatiotemporal and conceptual identity cues for object individuation. We found that even in a scenario where infants could succeed based on spatiotemporal cues, their performance plummeted when conceptual cues (i.e., labels) were added. While we predicted exactly this in the same label condition, where the conceptual/linguistic information did not distinguish between the two objects, this was highly unexpected in the different label condition, where objects were at different locations, and were also labelled with different nouns. Interpreting this result made us question infants' ability to use spatiotemporal and conceptual information simultaneously in object individuation. In particular, we speculated that if both cues are simultaneously present and both independently provide identity contrasts (evidence for two objects), 10-month-old infants fail to integrate the available information, which could result in a lack of numerical expectations altogether. There are a variety of methodological and theoretical issues that we need to address in order to take this interpretation of Study 1 seriously: (1) failed replication attempts raise questions about the validity of the findings, (2) that study might not have presented the strongest possible cues for spatiotemporal object individuation, (3) the violation-of-expectation paradigm may be unable to reveal online processing, and (4) looking time measures cannot differentiate between a lack of expectation and positive expectations that are (not) violated to the same degree across different outcomes. Let us consider each of these in more detail.

(1) It is important to take into account the fact that Study 1 failed to conceptually replicate the spatiotemporal object individuation findings 3 times before we succeeded. Furthermore, we obtained only one positive result as in no other condition did we find looking time differences to the one- versus two-object outcomes. A parsimonious explanation of the results is to disregard

infants' success as due to pure chance, and argue instead that infants failed across the board. In this case, further assumptions about the integration of different types of cues would not be needed. This explanation would still need to be supplemented by a hypothesis on why infants failed in 4 experiments (6 conditions in total) to use the available spatiotemporal information for building expectations about the number objects present. One possibility, already mentioned in Chapter 2, makes reference to issues introduced by the presentation medium of 3d animated displays.

(2) We also have to consider the strength of the spatiotemporal evidence that we provided to the infants. While infants were presented with two spatiotemporally separate occluders, each with a corresponding object behind it, this still cannot be taken as the strongest possible spatiotemporal evidence for the presence of two objects, which is when both are visually available simultaneously. From the viewpoint of the visual-indexing theory, there is no principled reason why this should make a difference. At the same time, our repeated failed replications make an empirical case for providing infants with the best chance of success using spatiotemporal evidence.

(3) A further concern is related to the general logic of the Violation of Expectation (VoE) paradigm. Instead of forcing infants to make online inferences about an event, this methodology allows them to make a post-hoc decision on whether the revealed outcome fits well with information that preceded it. A more direct measure of how infants represent a scene during occlusion would be beneficial. While the distinction between online vs. post-hoc processing is generally unimportant, it might be crucial in cases where redundant cognitive systems (like multiple indexing systems tracking object identity) can point to different outcomes. Suppose that, as we hypothesize, there are two or more independent indexing systems that can track the number of objects in a scene. In cases where the systems encode different number of objects, when an outcome is revealed in a VoE paradigm, it is unclear how the information provided by these different systems contribute to post-hoc inferencing. For example, in our study, even if the infant had expectations on the basis of both systems, any outcome that was encoded by *either*

system could simply require less processing to accommodate, which could result in the lack of measurable looking time difference in response to the two outcomes.

(4) There is another issue related to interpreting infants' looking behavior in a VoE paradigm. When discussing the results in Chapter 2, we raised the possibility that infants had a radically different contextual understanding in the labeling conditions compared to the spatiotemporal condition. Specifically, infants may have understood the labeling events as a kind of teaching scenario, where making predictions about the local presence or absence of the presented objects is not necessary. Given the passive nature of looking time measures in VoE, we cannot tell apart the absence of any prediction from a set of predictions that the two outcomes satisfy to the same degree. Thus, from the results we obtained, it is unclear whether infants expected anything to be present behind the occluders. A solution to this problem is to change this passive measure to one that requires infants to actively engage with the objects based on what they represent.

3.1.2 The design of Study 2

In this chapter, we aimed to design a study which is conceptually equivalent to the study presented in Chapter 2, but is not fallible to the issues raised above. We used the manual search object individuation paradigm developed by Van de Walle et al. (2000). This paradigm has the following advantages. First, by carrying out a conceptually equivalent study that does not use the same presentation medium, we can better assess the validity of the findings in Chapter 2, addressing (1). In this paradigm, we will present the objects simultaneously, thus giving the infant the best shot at succeeding using spatiotemporal cues (2). By moving away from a VoE paradigm, we can make sure that infants' behavior does not just reflect post-hoc processing when presented with an outcome (3). Finally, because manual search is an active and intentional behavior, we are more justified in drawing a link from the dependent measure to the infants object representations (4). This paradigm is also optimal as it is considered a robust and reliable way to test infants' representations (Xu, 2005, 2007; Xu and Baker, 2005).

We conceptually recreated the conditions from Experiment 4 of Study 1. We had three separate experiments: one with only spatiotemporal individuation cues, and two further ones that involved labeling events. One of the labeling experiments involved a single label, while the other involved two separate labels. Given the issues presented above, our core hypotheses from Chapter 2 remained unchanged.

(1) Infants expect multiple objects if the available spatiotemporal evidence implies the presence of multiple objects and there is no interference from linguistic/conceptual information.
(2) Infants' numerical expectations are not based on the number of spatiotemporally available objects, when they can encode the objects under conceptual descriptions:
    A. If a single conceptual description is available they will not expect multiple objects irrespective of the visually available information
    B. If distinct conceptual descriptions are available they will expect multiple objects irrespective of the visually available information

In the manual search paradigm, we make the further assumption that when infants expect an object to be present, they are going to search for it given the opportunity. This should result in longer search durations and/or higher frequency of searches compared to situations where no object is represented.

3.2 Experiment 1 — Spatiotemporal object individuation

In this experiment, we aimed to show that infants are able to use spatiotemporal cues to individuate objects, trying to replicate (Van de Walle et al., 2000). If infants see two spatiotemporally separate objects getting hidden in box, their behavior should reflect that they expect (exactly) two objects in the box. We used two objects that looked the same to make sure individuation would be based on a spatiotemporal basis.

3.2.1 Methods

3.2.1.1 Participants

Twenty-four infants participated in the study (range = 9 months 16 days to 10 months, 15 days; mean age 10 months 5 days). Infants were contacted after they were randomly selected from the CEU Cognitive Development Centre's pool of participants. Most families were contacted via a letter, others were recruited through online advertisements. An additional 21 infants were excluded for various reasons: 12 for passivity (see Inclusion criteria below), 4 for fussiness, 4 for experimenter error, and 1 for finding the hidden compartment. More details on the exclusion criteria are available in the Coding section.

3.2.1.2 Materials and Apparatus

Infants were seated at the shorter side of a table (120x60 cm), in a brightly lit testing room on the lap of their caregiver. The experimenter was sitting orthogonal to infant, to their left. The dimensions of the search-box were 32 cm (length) X 25 cm (width) X and 12.5 cm (height). The box was made out of 0.3 cm thick brown cardboard. On the front, it had a 10-cm high and 12 cm wide opening, covered with a blue textile curtain attached to the top. A hidden compartment was 15 cm deep in the box. Four objects were used that are unfamiliar to infants. They differed in shape and color (Figure 3.1). At the end of the table opposite to the infants an occluder hid all the toys not used in the current trial. Three ceiling-mounted cameras recorded the experimental sessions. One camera recorded the whole event, and other two recorded the box and the reaching events from different angles (Figure 3.2).

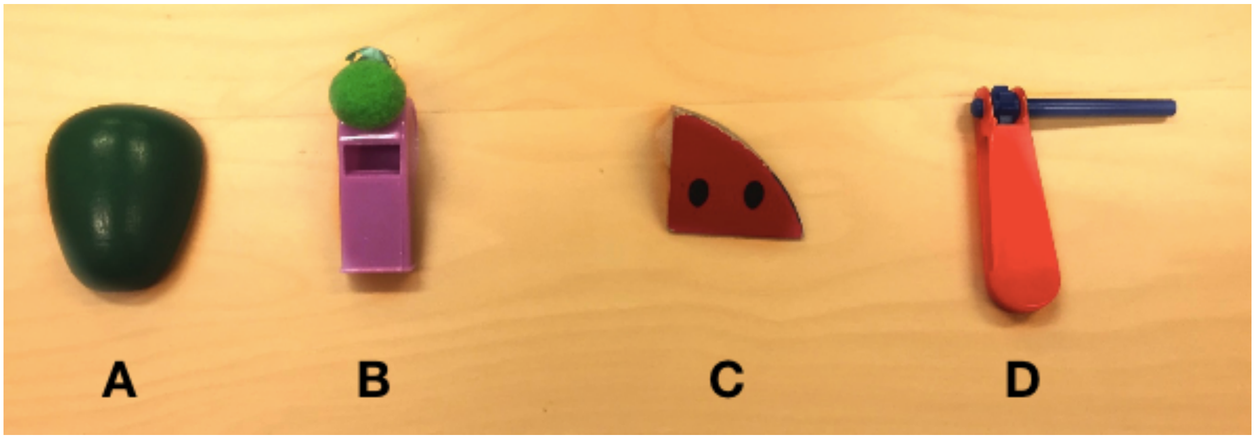


Figure 3.1. Objects used in Study 2. (A) Green oval-shaped object, used in the introductory trial. (B) Pink whistle, with a soft textile attachment, used in training trials. (C) Green wooden object with black dots on it, used in test trials. (D) Red plastic object with a blue attachment used in test trials.

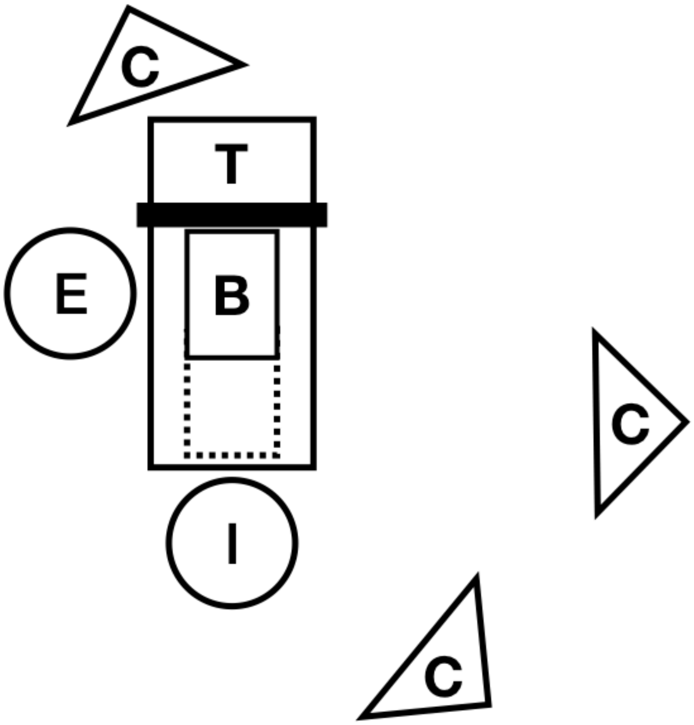


Figure 3.2. Schematic representation of the setup. The infant (I) was sitting on their caregiver's lap. The experimenter (E) was sitting to their left. The toys (T) on the table were hidden behind the opaque occluder. The box (B) was moved closer to infants during the *search phase.* Three ceiling mounted cameras (C) recorded the sessions.

3.2.1.3 Procedure

The experiment consisted of an introductory trial, two training trials, and four test trials in a fixed order. The *introductory trial* always involved a single object which was followed by the one-object and two-object *training trials*, which were presented in a counterbalanced order. Finally, infants were presented with 4 *test trials* (2 one-object and 2 two-object trials), that were presented in either in a 1,2,2,1 or in a 2,1,1,2 order counterbalanced across the participants.

3.2.1.3.1 Introductory trial. The purpose of this trial was to make the infant comfortable with the testing environment, and help them understand the context of our task (a search game). We tried to achieve this with less scripted communication on the part of the experimenter compared to later trials. After the parent and the infant were seated, the experimenter showed the infant the object (Figure 3.1 A), then hid it in the box. The object was hidden only partially, so parts of it still protruded from the box. Initially the object was out of reach, but when it was pushed towards the infants, they had 10 seconds to retrieve the object. During this time the infants were verbally encouraged to search in the box, with the experimenter having complete freedom in her communication. If infants found the object, they were praised. If they did not search, the experimenter gave them the object. Whether or not they retrieved the object by themselves, they had 5 seconds to explore it before the experimenter took it away, while saying "thank you".

3.2.1.3.2 Training trials. The one-object training trial involved a single object, and the two-object training trial involved two separate objects, taken from behind the occluder. The object(s) used in training trials were always the same (Figure 3.1B). In the two-object trial the experimenter serially placed them on the top of the box (at the opposite corners of the closest edge to the infant) in a way that both of them were visible simultaneously. She paused for 4 seconds after both placements. Then she collected both objects into one hand and placed the objects into the box. Both objects were placed into the accessible compartment. After placing the objects, the experimenter pushed the box towards the infants, saying "now it is your turn ". Infants had 10 seconds to find an object, before the experimenter intervened and retrieved it instead. Then during the next 5 seconds infants could explore the object before it was taken away by the experimenter thanking them in the process. These events were repeated with the same timing for the second object. After the second object was taken away, the experimenter pulled the box away from the infants. On the one-object training trials the experimenter first placed the single object to the location that corresponded to the first object location in the two-object trial. Then she moved the object to the other location. This way, the timing and number of actions were kept constant across training trials. For the first of the two training trials (irrespective of the order of the trials) the object(s) were placed similarly to the introductory trial (partly protruding) so infants had continuous visual access to them.

3.2.1.3.3 Test trials. The four test trials either followed a 1,2,2,1 or a 2,1,1,2 order regarding the number of hidden objects. In two-objects trials, the objects were identical. Between test trials, the type of object (Figure 3.1 C and D) varied in either an a C,D,C,D or a D,C,D,C order, counterbalanced between infants. This way both test objects were used in both one-object and two-object trials for each infant. In all test trials the experimenter first grabbed the object(s) from a location out of sight. Then in the *presentation phase* she sequentially placed the objects on the top of the opposite sides of the box, pausing for 4 seconds after both placements if it was a two-object trial. In one-object trials the presentation phase was the same, the only difference being that the object was placed serially to the opposite ends of the top of the box.

During the search phases, the infants had 10 seconds to search in the box for each opportunity. This phase started either at the moment the box was pushed in front of them, or from the moment when the experimenter took the previous toy away and grabbed the box again. If the infant's hand was in the box at the end of this 10-second period, they were allowed to finish their current search. During the search phases the experimenter was looking at the box, while keeping it stable with both hands. If the infants failed to retrieve an object (either because they failed to find it or because the object was in the hidden compartment), the experimenter gave it to them, let them explore it for 5 seconds before she took it away.

During the two-object trials, one of the objects was hidden in the secret compartment so infants could not find it. In these trials, we presented infants with three search phases. First they searched for a ‘real’ first object; an object they could find. Then they searched for a ‘real’ second object, which they could not find as it was in a hidden compartment. Finally, they searched a third time for a ‘phantom’ third object after both objects had been removed. In the one-object trials, infants were only presented with only two search phases. They searched for the real first object and then for a phantom second object, which of course they could not find.

3.2.1.4 Inclusion criteria

Altogether there were 10 search opportunities in the test trials, and in six of these infants normatively had a good reason to search. The criteria for inclusion was that infants had to search in the box at least in 3 trials, no matter which trials these were. This criterion was set up to allow us to exclude completely passive infants without a normative commitment on the trials where infants should search. This decision turned out to be responsible for a high exclusion rate compared to the original study, as we had to exclude 12 infants for this type of passivity.

3.2.1.5 Coding and dependent measures

The coding of the infants’ behavior during the search phases was performed offline on a frame-by-frame basis. Measurements were taken once in the one-object trials (search phase for a phantom second object) and twice in the two-object trials (search phases for a real-but-hidden second object and for a phantom third object). Just as in the original study by Van de Walle et al. (2000), we had two dependent measures: total search duration, and number of reaches into the box. Total search duration was defined as the cumulative duration of all individual searches. The number of reaches measure represents the sum of individual reaching acts, which were separated by periods of not searching. A behavior counted as a searching if all fingers on least one hand of the infant completely disappeared within the box. No quantitative or qualitative judgements were made beyond qualifying such an action as a “search”. The two measures were correlated, and a strong positive relationship was found between them (Spearman's r = .75), which was higher than the moderate relationship reported in the original study (Van de Walle et al., 2000). This correlation is to be expected, as the two measures are not independent.

3.2.2 Results

3.2.2.1 Total Search Duration

To examine whether infants treated the three test scenarios differently, we conducted a one-way repeated-measures ANOVA with test trial type as the independent variable (phantom second, real second, phantom third) and total search duration in milliseconds as the dependent measure (Figure 3.3). We found a significant interaction between trial type ($F(2,46) = 8.619$, $p < .001$, $\eta^2 = .422$). To understand the interaction, a Helmert contrast was computed, as we predicted infants to search more for the real second object when compared to the other two conditions. This replicated the original finding since the infants searched more for the real second object than the average of the other two test trials ($F(1,23) = 15.080$, $p < .001$, $\eta^2 =.396$). Post-hoc paired t-tests revealed that infants reached significantly longer for the real second object (M = 4323, SD = 2678) than the phantom second object (M = 2707, SD = 1917), $t(23) = 2.914$, $p =.008$ or phantom third object (M = 2295, SD = 2385) conditions; $t(23) = 4.35$, $p < .001$). Nonparametric analyses corroborated these results: the infants searched more for the real second object than the phantom second (Wilcoxon $Z = 2.555$; $p = .011$) or the phantom third (Wilcoxon $Z = 3.133$; $p = .002$).

3.2.2.2 Number of reaches

The results of this measure closely matched the ones obtained via the search duration measure (Figure 3.4). We found a significant interaction of trial type in the repeated measures ANOVA ($F(2,46) = 7.313$, $p < .002$, $\eta^2 = .241$). According to the planned t-tests, this effect was driven by significantly more searches for the real second object (M = 1.20, SD = 0.65), than the phantom third (M = 0.68, SD = 0.56), $t(23) = 3.734$, $p < .001$, and a tendency to search more for the real second object than the phantom second one (M = 0.95, SD = 0.56), $t(23) = 1.771$, $p = .090$. The

Helmert contrast was again utilized for replicating the original results. The analysis revealed that the infants searched on more occasions for the real second than on the average of the other two ($F(1,23) = 9.51$, $p < .005$, $\eta^2 = .396$). The nonparametric comparisons indicated significantly more searches for the real second object compared to the phantom third (Wilcoxon $Z = -2.998$; $p = .003$) but failed to show statistical significance when compared to the phantom second (Wilcoxon $Z = -1.639$; $p = .101$).

3.2.3 Discussion

Our study successfully replicated the original finding. Ten-month-old infants were capable of individuating objects based on spatiotemporal cues. This was evident from the fact that they searched more for a second object when two objects were hidden compared to the condition where only one was hidden. The conclusion is also supported by the fact that this search behavior dropped also during the search for the phantom third object, implying that infants represented exactly two objects present in the box.

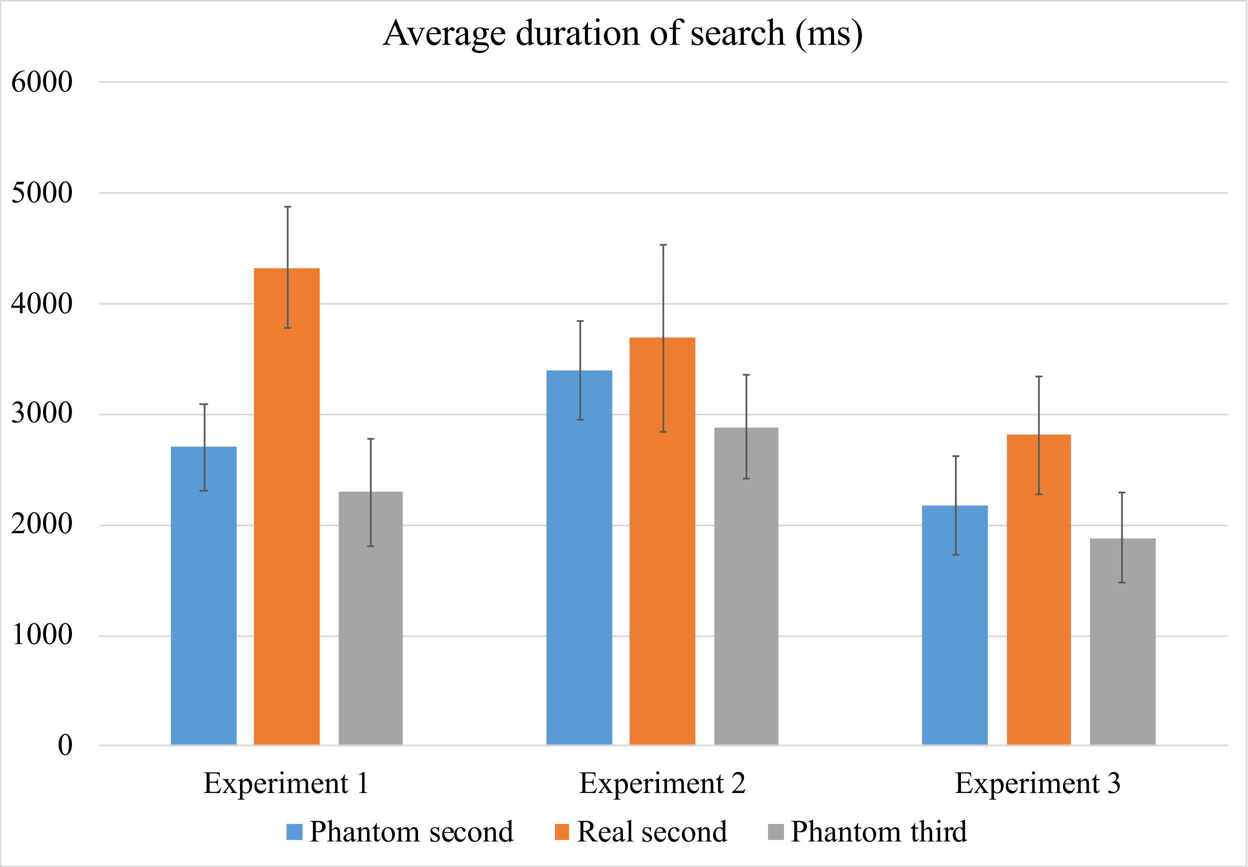


Figure 3.3. Average duration of search (ms) as a function of the trial for each of the three experiments of Study 2. In Experiment 1 the objects were not labeled. In Experiment 2 objects were labelled with the same label. In Experiment 3, when two objects were hidden, the objects were labeled with different labels. We found a significant main effect in Experiment 1, as infants searched longer for the real second object. Error bars represent standard error.

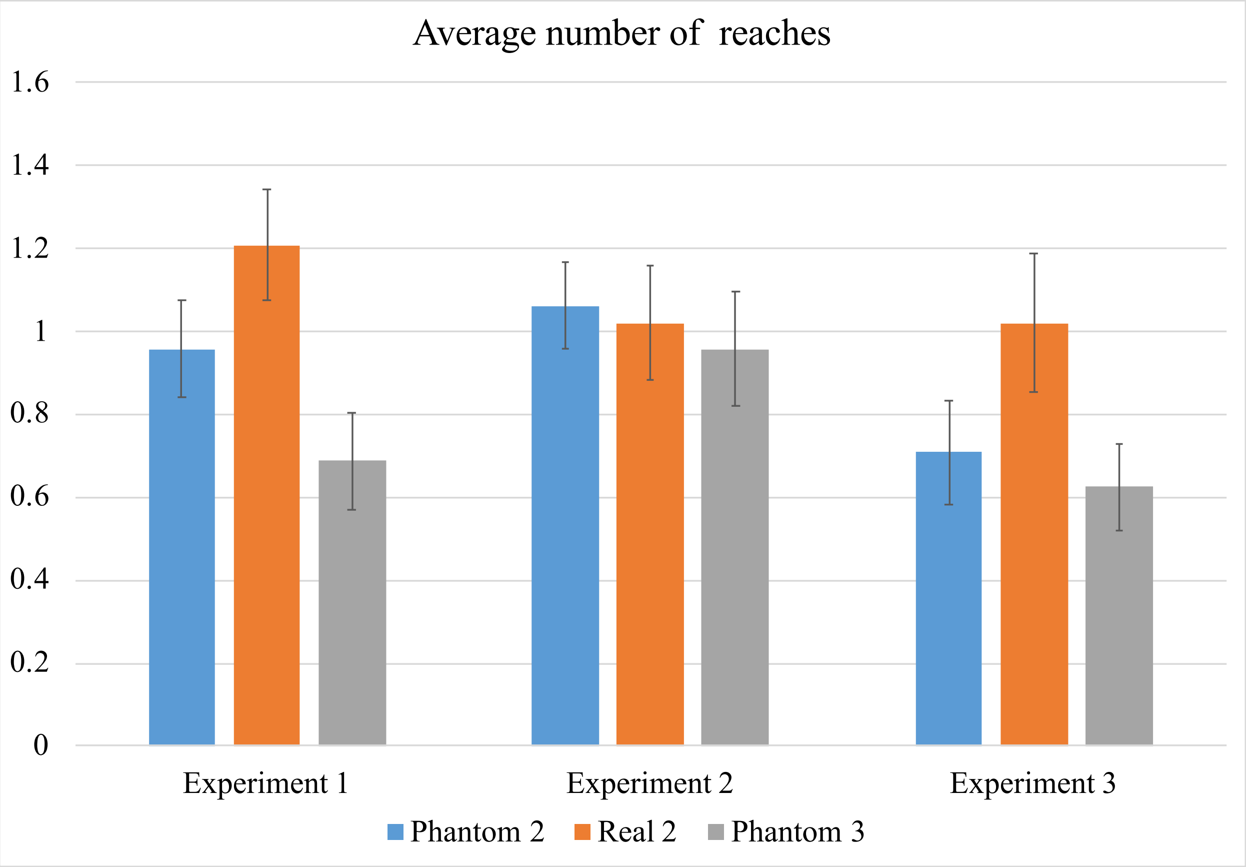


Figure 3.4. Average number of reaches (ms) in a single trial for the three experiments of Study 2. In Experiment 1 the objects were not labeled. In Experiment 2 objects were labelled with the same label. In Experiment 3, when two objects were hidden, the objects were labeled with different labels. We found a significant main effect in Experiment 1, as infants searched longer for the real second object. Error bars represent standard error.

3.3 Experiment 2 — Spatiotemporal evidence with the same label

Experiment 2 aimed to show that the spatiotemporal encoding of the objects is not always privileged: if contradictory conceptual information is available, infants' performance can be disrupted. Our implementation is based on a labelling manipulation (Xu, 2002). In every test trial, the objects were labelled during presentation. Most importantly, in the two-object test trials both objects were labelled with the same label, implying that the objects had the same description, thus not providing evidence for multiple objects. If our hypothesis is false, object individuation should be unimpeded by the labels and infants should still search for a second object longer in the two-object condition compared to the one-object condition.

3.3.1 Methods

3.3.1.1 Participants

In total 24 infants were included in the sample (range = 9 months 16 days to 10 months, 16 days; mean age 10 months 4 days). An additional 20 infants were excluded for various reasons: 12 for passivity, 1 for fussiness, 3 for experimenter error, and 4 because of technical issues with the recording apparatus.

3.3.1.2 Procedure

The key difference between Experiments 1 and 2 was in the *presentation phase* of the test trials. In Experiment 1 the experimenter placed either one or two objects on top of the box and paused 4 seconds after every placement. In Experiment 2 this moment was used for labelling the objects. In two-object trials, the experimenter pointed to the object after each placement and labelled it with pseudo-words. The exact Hungarian phrasing was: "Egy tacok! Nézd! Egy tacok!". This translates to the following English utterance: "A tacok! Look! A tacok!". Importantly, the two objects within a trial had the same label. We used two labels that are frequently used pseudo-words in the literature conducted in Hungarian: "bitye ", and "tacok “. In the one-object trials, the object was pointed to and labelled with the same label at both locations. The labelling events were otherwise identical, so in the one-object trials an object was labelled four times in total.

3.3.2 Results

3.3.2.1 Total Search Duration

We conducted the same analyses as in Experiment 1 (Figure 3.3). We found no effect on trial type on search duration with the repeated measures ANOVA ($F(2,46) = 0.712$, $p = .420$, $\eta^2 = .07$). Search duration in the real second trial (M = 3688, SD = 4163) was contrasted with those of the phantom second trial (M = 3401, SD = 2205), $t(23) = 0.368$, $p = .72$), and the phantom third trial (M = 2884, SD = 2305), $t(23) = 1.07$, $p = .30$, but no differences were found. Nonparametric Wilcoxon tests revealed no significant differences either when comparing the trial types.

3.3.2.2 Number of reaches

The same analysis was conducted for the number reaches, and we found that the results were comparable to the search duration measure. No significant interaction between trial types was found ($F(2,46) = 0.260$, $p = .72$, $\eta^2 = .023$). The number of reaches did not differ significantly for the real second object (M = 1.02, SD = 0.68) compared to the phantom third object (M = 0.95, SD = 0.67), $t(23) = .036$, $p = .72$, or the phantom second object (M = 1.04, SD = 0.55) $t(23) = .647$ $p = .52$. We found no differences between conditions using Wilcoxon tests either.

3.3.2.3 Comparison with Experiment 1

For both of the dependent variables, we conducted a 2x3 mixed type ANOVA to further investigate the effect labelling had on individuation performance, and to directly compare Experiment 1 and Experiment 2. In both cases, we found a main effect of trial type, driven by infants' longer searches for the real second object than the other ones in Experiment 1 $F(2,92) = 5.688$, $p = .005$, $\eta^2 =.110$ for search duration, and $F(2,92) = 5.453$, $p = .006$, $\eta^2 =.106$ for number of reaches. We also found a significant interaction between experiment and trial type for the number of reaches (*F*(2,92) = 5.453, *p* =.039, η^2 =.068), providing some direct evidence that lack of individuation in Experiment 2 was due to the labelling manipulation.

3.3.2.4 Summary

In Experiment 2 we not only consistently failed to find any evidence that infants individuate objects when these objects were labelled with same label, but – via comparing it to Experiment 1 – also learned that the labelling manipulation changed the pattern of search behavior. This shows that, for 10-month-old infants, spatiotemporal properties are not always primary, even if that information is available and could be used.

3.4 Experiment 3 — Spatiotemporal evidence with different labels

We obtained evidence for infants' spatiotemporal individuation in Experiment 1, but we did not find the same pattern of results when infants were presented with conceptual information that in itself did not warrant object individuation (Experiment 2). What exact role did the labeling events play in this failure? Do infants treat the conceptual/linguistic information as a primary property that could maintain an object representation by itself? In order to answer this question, we manipulated the conceptual content available for the infants while keeping all other characteristics of the Experiment 2 unchanged. In Experiment 3, the sole modification we introduced was that, in the two-object trials, the two identical objects were labelled not with the same but with different labels. This way, if infants encode the objects in conceptual terms, they should again individuate them similarly to Experiment 1.

3.4.1 Methods

3.4.1.1 Participants

As in the previous experiments, 24 infants participated in Experiment 3. Their age ranged from 9 months 13 days to 10 months 13 days; their mean age was 9 months 27 days. An additional 22 infants were excluded for various reasons: 18 for passivity, and 4 for experimenter error.

3.4.1.2 Procedure

The only difference between Experiment 2 and Experiment 3 was in the presentation phase of the two-object test trials. In Experiment 3, the two objects got labelled with different, instead of the same labels. In half of the two-object trials, the first labelled object was "bitye " and the second was “tacok", in the other half the order of labels was reversed. The one-object trials were left unchanged; if there was only a single object, it was labelled with a single label.

3.4.2 Results

3.4.2.1 Total Search Duration

We analyzed whether infants searched longer for the real second object compared to the phantom third and phantom second objects using a repeated measures ANOVA (Figure 3.3). We did not find an effect of trial type ($F(2,46) = 1.748$, $p = .185$, $\eta^2 = .71$). We predicted that search duration for the real second object would differ from the other two trials, but the Helmert contrast revealed no significant difference but a trend ($F(1,23) = 3.623$, $p = .070$, $\eta^2 = .14$). Pre-planned paired sample t-tests revealed that this tendency was driven by longer search for the real second object (M = 2815, SD = 2622) compared to the phantom third (M = 1887, SD = 1986), $t(23) = 2.023$, $p = .055$. We found no difference when comparing search duration for the real second object with the phantom second (M = 2178, SD = 2180); $t(23) = 1.220$, $p = .235$). Nonparametric analysis revealed a similar pattern. The comparison between real second to phantom second was non-significant (Wilcoxon Z = -.887; p = .375). But when comparing the real second with the phantom third object, we found a significant difference (Wilcoxon Z = -2.240; p = .025).

3.4.2.2 Number of reaches

Conducting the same analyses on the number of reaches, we found a significant main effect of trial type in the repeated measures ANOVA (F(2,46) = 3.724, p = .032, η^2 = .139). The pre-planned Helmert contrast revealed that infants searched on more occasions for the real second than for the average of the other two objects (F(1,23) = 4.738, p = .040, η^2 = .171). The pre-planned t-tests revealed that infants only searched more for the real second object (M = 1.02, SD=0.81), compared to the phantom third one (M = 0.62, SD = 0.51), t(23) = 2.744, p = .012), but not to the phantom second one (M = 0.71, SD = 0.61), t(23) = 1.606, p = .122. Using non-parametric analyses, we found the same pattern. Infants reached more for a real second object compared to a phantom third (Wilcoxon Z = -2.318; p = .02), but there was no significant difference when with the phantom second object (Wilcoxon Z = -1.414; p = .157).

3.4.2.3 Comparison with Experiment 1 & Experiment 2

When we compared the search durations of Experiment 2 and Experiment 3 (the two experiments with labeling), we found no main effects and no interaction between trial type and experiment (F(2,92) = 0.87, p = .916, η^2 = .002). There was no significant interaction for the other measure, the number of reaches either (F(2,92) = 2.150, p = .122, η^2 = .045). This result indicates that the behavior of the infants in the two labeling conditions were not statistically different. Taken together, infants in the two labeling experiments did not exhibit behavior that would allow us to infer that they represented two objects in the box when there were indeed two of them (real second search trial).

When compared the duration of search with Experiment 1, the ANOVA revealed a strong main effect of trial type (F(2,92) = 9.095, p < .001, η^2 = .165). However, we found no interaction between experiment and trial type (F(2,92) = 1.386, p = .255, η^2 = .029). The pattern of results was the same with the number of reaches: no interaction (F(2,92) = .435, p = .649, η^2 = .009), but a main effect of trial type (F(2,92) = 10.189, p < .001, η^2 =.181). This shows that infants in Experiment 3 did not produce a significantly different search pattern from that of Experiment 1. Nevertheless, the two experiments together still provide evidence for object individuation, mostly driven by the behavior of the infants in Experiment 1.

3.4.3 Summary

We found limited evidence that infants individuated objects in Experiment 3. While they searched longer, and more times for the real second object compared to a phantom third one, the difference compared to phantom second object was not significant. The strongest evidence for them representing two objects inside the box comes from the Helmert contrast, as it revealed that they searched on more occasions for the real second object than in the other two conditions. Crucially though, their performance was not significantly different from either Experiment 1, or from Experiment 2, because the main effects reported in the comparisons with these experiments were mostly driven by the infants' performance in Experiment 1.

Taken together the current dataset does not let us decide on whether or not the infants in Experiment 3 individuated the objects or not. While interpreting these null-results is not easy, the fact that we did not find clear evidence for success in Experiment 3 is unexpected as the infants were presented with both spatiotemporal and conceptual cues about the objects, and these cues were converging. Possible explanations will be reviewed in the General Discussion.

3.5 General Discussion

To date, the usage of spatiotemporal and linguistic/conceptual properties in object individuation was mostly studied in isolation (Xu & Carey, 1996; Xu, 2002). In this and in the previous chapters, we investigated how these two individuating properties could be integrated by infants. In Chapter 2 we found that (1) 10-month-old infants' ability to use spatiotemporal individuation cues was not robust and (2) infants failed to individuate objects in the presence of labeling events. We raised a variety of methodological issues that we tried to tackle in the experiments reported in the current chapter. Otherwise our goal was to create a conceptually equivalent study.

First, we validated our methods via replicating spatiotemporal object individuation in the manual search paradigm (Experiment 1). Infants in this paradigm were provided with the strongest possible spatiotemporal cues for object individuation, simultaneous visual access to both objects. Just as (Van de Walle et al., 2000), we found that 10-month-old infants searched more for a second object in the two-object condition (real second) compared to the one-object condition (phantom second). Also, they searched more only for this second object. After the removal of the second hidden object, they did not keep searching longer for a non-existent third object. This pattern of results shows a robust encoding of the spatiotemporal nature of the presented objects that results in the representation of exactly two objects in the container.

Experiment 2 provided evidence that visually available objects are not always indexed based on their spatiotemporal properties in the first year of life. Infants here failed to use the spatiotemporal cues that were sufficient for individuating objects in Experiment 1. The main manipulation between the studies was the linguistic description of the presented objects (same label for both objects). If infants had taken spatiotemporal properties of the objects as primary, their behavior should not have changed, but the interaction effect between the studies and conditions shows that they treated the two situations differently. This corroborates the findings from Chapter 2, Experiment 4 where we found a similar drop in performance in the corresponding same-label condition. These results are agnostic on how infants' visual system encoded these objects. Without evidence to the contrary, the most plausible assumption is that their visual system created two indices, the same way as the adult visual system does.

In Experiment 3 we failed to find evidence in favor or against hypothesis 2B. Infants did not show an expectation for multiple objects in the box, even though the different labels used here provided distinguishing conceptual information that could have been used for object individuation. Strikingly, despite the two different individuating cues converging in predicting two objects, infants seemingly could not use either to individuate objects in this condition. This is clearly a surprising result in light of the previous literature. Infants of this age have been repeatedly shown to possess the ability to use either cue in isolation for object individuation. Although the pattern of results is less clear in the current study, similar results were obtained in the corresponding different-label condition of Chapter 2, Experiment 4. There, we found statistically strong evidence that infants did not represent multiple objects. Taken together these results raise the following question in relation to Hypothesis 2B. If we accept our interpretation of Experiment 2 as indicating that infants did not take the spatiotemporal evidence into account in the presence of identical labels, why did they fail to individuate objects based on distinct labels in Experiment 3?

A possible reason for the lack of success in Experiment 3 could be that the differently labelled objects shared all their surface features. If infants think that surface features are constitutive of kind membership, they might not accept that objects that look the same can belong to different kinds. There is partial support for this interpretation. Already at 9 months of age, infants expect that objects that have different labels should have different surface features (Dewar & Xu, 2007). But a follow-up study with 10-month-olds also revealed that surface features are not at the core of these expectations (Dewar & Xu, 2009). When identical objects were labelled with different labels, infants expected the objects to have different internal properties - emit different sounds - at a rate similar to a condition where the objects did not share surface features. This shows that while infants may have a bias that objects that have distinct labels also look different on the surface, this bias is not constitutive of making a conceptual distinction. Converging evidence for this conclusion comes from a study on function-based object individuation (Futó et al., 2010) where infants successfully individuated objects based on their functions in a standard Xu & Carey (1996) paradigm, even when the two different functions were associated with surface-identical objects. In the current study, we did not provide any evidence for internal differences between the objects, but our manipulation should have still succeeded in creating a conceptual differentiation. In sum, neither previous literature nor the framework we argue for supports the view that surface-featural distinctions play a crucial role in making a conceptual distinction.

However, we cannot completely discount the possibility that the labelling cues that we provided were not strong enough to override infants' prior biases, and a perceptual differentiation between the objects would have been necessary in this task. In light of this possibility, it seems important to reconsider what drives infants' expectations in labeling based individuation studies in the first place.

Conceptual or label-based object individuation has mostly been discussed in the infancy literature by invoking the notion of sortal concepts. What is assumed to distinguish sortals from other concepts is their ability to provide principles of individuation and identity. We can count the number of objects, chairs, or trucks, but we cannot count the number of "red", without specifying what is the thing that "redness" applies to (Xu, 1997, cf. Wiggins, 1997). The main underlying assumption in the literature is that, for infants, nouns of natural language map onto basic level (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) sortal concepts. It is further assumed that basic-level sortal concepts are mutually exclusive, resulting in a one-to-one correspondence between distinct nouns and sortals (Xu, 1997, 2005, 2007). What follows from this is that noun labels are *also* mutually exclusive, each one mapping onto a different sortal concept.

Assuming that these theoretical assumptions are valid, infants should not have construed the distinct labels of our study as applicable to a single entity, as they should pick out non-overlapping sortals. However, there might be language specific reasons to seriously consider the option that infants did not treat the labels as nouns in the first place. Hungarian allows for dropping (not pronouncing) nouns more liberally than English (the language that the cited word-based individuation studies used). Most relevantly, the Hungarian sentence “Ez egy kék _ ” (“This is a blue _ ”) is grammatical. If an addressee is unaware of the syntactic classification of the word “blue”, this environment will not help to disambiguate whether it is a noun or an adjective inside a noun phrase where the head noun is omitted. Consequently it might be harder in Hungarian to determine whether an unknown label is a noun. If infants misconstrued the syntactic class of our novel labels, it might have resulted in non-sortal and potentially non mutually exclusive mappings or just in a general failure to map the words to concepts.

We also have to consider the possibility that the standard analysis of the relationship between nouns and basic level sortal concepts is not entirely valid, giving us a a wider range of interpretation for infants' failure. Suppose when infants hear nouns, they not only consider basic-level sortals, but also ones that might belong to a different hierarchical level (e.g. toy/ball, animal/moose). In this case it would be possible for the two objects to have shared a single conceptual description that could satisfy both sortals. For instance a single object can be felicitously labeled as both a ball and a toy. More radically, suppose when infants hear nouns they do not only consider sortal concepts. If instead infants mapped the labels onto some other property of the objects (e.g. redness, smallness). This would also allow for multiple labels to apply to a single object especially since within a condition the objects shared all features.

That said, there is considerable empirical support for the noun to basic-level sortal hypothesis. Infants do seem to expect one-to-one correspondence between the number of labels used and the number of objects expected to be present (Xu, 2002). Amongst others, this result is the backbone of the hypothesis that different labels are mapped onto mutually exclusive sortal concepts (Xu, 1997, 2005, 2007; Carey, 2009). There still remains a possibility of adopting a weaker variant of this hypothesis. Specifically, rather than being a principle, reflective of the conceptual/linguistic architecture, mapping nouns to mutually exclusive sortals could be interpreted as a bias that can be overcome based on contextual factors. This is analogous to infants' expectation that objects that have different labels should also have different surface features (Dewar & Xu, 2007). While infants expect this to be the case, they have no trouble accepting scenarios in which it is not true (Dewar and Xu, 2007, 2009).

Altogether, while it is important to consider infants' lack of success in Experiment 3 as evidence against the hypothesis that they treat label-based conceptual descriptions as primary, the data we collected does not support this conclusion in its strongest form. In fact, Experiment 3 strengthens the conclusions of Experiment 2: spatiotemporal properties are not always privileged in the first year of life. Could there be alternative explanations that account for the entire set of data (Studies 1 & 2) without jeopardizing the spatiotemporal priority hypothesis? We see two possible explanatory strategies of this sort: (1) Claiming that our labeling manipulation simply disrupted object tracking due to attentional or working memory limitations, or (2) assuming that the infants successfully individuated in all experiments, but their search and looking behavior did not reflect this underlying knowledge.

(1) The simplest explanation of the results — infants' successes in the spatiotemporal conditions and failures in the labeling conditions — invokes a hypothesized disruptive effect of labeling. This disruption could simply be derived from the fact that there are more things to attend to in the scene: linguistic communication increases information processing load, and the novel labels demand more working memory. While there are no a priori reasons to dismiss this explanation, it is notably incompatible with the architectural stipulations of visual indexing. Visual indices thought to be automatically assigned by a tracking system that operates with a high degree of encapsulation (Pylyshyn, 2003; Scholl & Leslie, 1999). This stipulation also has empirical support. In adults, the number of objects tracked in a multiple object tracking paradigm is unaffected even while maintaining multiple linguistic items in memory (Scholl & Xu, 2011). Moreover, visual working memory of objects is generally unaffected by verbal loads (Luck & Vogel, 1997). The disruption hypothesis is also incompatible with the literature on conceptual object individuation in general. In previous studies where infants had to individuate objects based on labeling information, they succeeded without attentional disruption. The only structural difference in our study was that the spatiotemporal information in itself was sufficient for individuating two objects. In one study (Robinson & Sloutsky, 2008) auditory input negatively impacted 8-month-old infants' individuation performance, but crucially irrespective of the kind of auditory input. Labeling and other sounds both had the same sort of negative input, while we only found a disruption that was caused by the labeling. Taken together, the idea that object indexing breaks down from labelling due to attentional load or memory constraints, while

possible, does not fit well with the architecture of object indexing and it is not directly corroborated by previous data.

(2) Assume that infants successfully individuated the objects in all of our experiments. Is it possible that their interpretation of the context changed across conditions in a way that is reflected in our data? We elaborate on two intuitive possibilities. The first possibility is that the infants' failure to search longer in Experiment 2 (same label) was due to motivational factors related to the way infants treat kinds. Maybe they were less interested in retrieving the second object because the second exemplar was not estimated to increase their net benefit (since the objects belonged to the same category). This would still be compatible with a success in Experiment 1: although the objects shared surface features, there was no evidence of them belonging to the same conceptual category, so the differential search patterns between these two experiments could be warranted. The stipulation that infants might treat members of a category as equivalent is justified. In studies with older children, participants often treat members of the same kind as interchangeable and their differences unimportant. But the equivalence of kind members, to our knowledge, never manifested as expectations of diminishing returns, but rather as expectations of how these objects would behave (Gelman, 2003; Butler & Markman, 2012). In a variety of studies, infants of the same retrieve objects that belong to the same kind, and motivational control conditions do not make a difference in their performance (Feigenson et al., 2002). Moreover, from the fact that two objects are equivalent in value and function it does not follow conceptually that the utility of a second object has to drastically decrease. Most importantly, this analysis fails to account for the results of Experiment 3. In that condition infants had explicit evidence of the objects belonging to different categories, which should imply an increase of their net benefit of getting both objects. Thus, the 'diminishing returns' account predicts that the infants should have best performed in this case, but in actuality, their performance was not statistically better than in Experiment 2. It is unclear whether the motivational account can say anything at all about looking behaviors (Chapter 2). Even if it can, we found a similar pattern of results in Study 1, and thus this account should fail for the same reasons that we outlined above.

Could one explain the results by appealing to infants' differential understanding of the context in the different conditions? In Chapter 2 we considered the possibility that labeling led infants to conceive of the events as a "word learning game". In a similar vein, maybe the labeling manipulations in Experiment 2 and Experiment 3 in this chapter caused infants to consider the context to be a "word learning game" rather than a "searching game". Our choice of the manual search paradigm was partially intended to rule out this type of alternative explanation by using an active measure. In fact, infants kept searching across conditions in all of our studies. This is hard to explain if there was a radical change in their understanding of the context, since it is not obvious why they should have searched in the box at all. The fact that they did search indicates that they represented *something* in the "here and now". The challenge for this task-intelligence based explanation is to explain why infants' knowledge of the contents of the box did not influence this search behavior, in contrast to their contextual understanding of the task, whatever it may be. Previous empirical data also speak against this account. In other studies that involved similar labeling manipulations, infants successfully searched longer for two objects (Xu, 2002), showing that labeling alone does not induce such radical reinterpretations of the context. More generally, while this sort of "task intelligence" based explanation might make intuitive sense, it is not easy to spell out a principled version that can accommodate both our results and previous data.

On a broader conceptual note, if we accept an alternative explanation in which the infants *always knew* the correct number of objects, but still failed in the labeling conditions because of pragmatic or motivational factors, this interpretation should be applied across the board to all previous studies where infants did not individuate objects. Doing so would undermine the logic of the individuation paradigm as a whole. Consider the study by (Xu, 2002) where the infants searched more for a second object in the condition where they heard two different labels, but searched less for it in the condition where they heard a single label twice. Adopting the motivation based alternative explanation outlined above, we could very well interpret these results as consistent with infants expecting two objects in both conditions, but just being less motivated to search in the same label condition. If we take this type of explanation seriously, some of our most basic assumptions and prior understanding of conceptual object individuation would be in jeopardy. In turn, the current study and its hypotheses would also be ungrounded. For these reasons, we are weary of accepting such alternative explanations unless (1) there are principled reasons to believe that they only apply to our studies or (2) there are reasons to believe that they can be applied to all individuation studies and yield a consistent interpretation across the board.

3.6 Conclusions

In two sets of studies (Studies 1 & 2 in Chapters 2 & 3), we tested the integration of spatiotemporal and conceptual cues in object individuation. Our goal was to probe which of these cues (if any) takes primacy in generating numerical expectations in 10-month-old infants. Primary properties are defined as lying at the core of indexing processes that track object identity. Thus, identifying which property takes priority could inform us about the underlying architecture of the indexing systems.

We found that 10-month-old infants took spatiotemporal properties as primary as long as there were no conceptual/labeling information present. On the other hand, we consistently found that when objects were labelled, this spatiotemporal primacy went away. We did not, however, find strong evidence to support the hypothesis that in such cases conceptual properties gained priority, leaving open key questions about the nature of conceptual object individuation. In the following chapter (Chapter 4), we will put forward a novel proposal about the architecture of a conceptual indexing system that makes a different set of predictions about what drives conceptual/linguistic object individuation.

Chapter 4 — Indexing objects in a discourse

In Chapters 2 & 3 we found evidence that a single visual indexing system is insufficient to account for 10-month-old infants' object representations. When the objects were labeled, infants failed to take into account the available spatiotemporal evidence for two objects. Surprisingly, we also found that infants failed even when they were presented with a conceptual contrast (different labels conditions). Taken together, the leanest valid generalization is that infants' expectations were modulated by the labeling manipulations in general rather than by the specific conceptual information that was presented. That is, the conditions that included labeling events succeeded in engaging an alternative system of indexing, but that system might not have assigned separate indices to the objects in response to different labels. The most straightforward treatment of this generalization is that conceptual descriptions provided by labels are not primary but secondary properties in the relevant conceptual indexing system, whatever it may be. This means that while separate object representations may be established using these properties, in themselves they are not providing indices.

On the other hand, the previous chapters also demonstrated that labeling manipulations succeeded in engaging an alternative system of indexing. A goal of this chapter is to gain a better understanding of what this system might be. We first show that sortal concepts, while potentially sufficient for object individuation, can not by themselves create indices. We then suggest that when this alternative system for indexing objects is engaged, referential communication plays a crucial role. Taking this link seriously and relying on existing theories within natural language semantics about discourse representation (Kamp, 1981, Heim, 1982), we argue for the existence of a discourse bound indexing system in infancy. We present evidence that the existence of such a system is supported both by the literature on object representation and the literature on referential communication. Finally, we formulate some of the challenges that the notion of an independent discourse-based indexing system faces.

4.1 The role of kind sortals in object individuation

It has been proposed that sortal based object individuation is supported by a system that is independent from spatiotemporal object representations (Xu, 2005, 2007). Could this proposed individuation system also be an indexing system? The formulation of the sortal theory in the infancy research is quite different from the formulations in the philosophical/linguistic literature (Wiggings, 1997; Hirch, 1997; Rips Blok, & Newman, 2006; Blok, Newman, & Rips, 2007), and we will keep discussing the more relevant psychological variant (Xu, 1997, 2005). The main property of sortal concepts is that they provide principles of individuation and identity. A contrast frequently used in the literature is one between count nouns and other types of syntactic constituents, like adjectives. The main idea of sortal based individuation is that we can count the number of objects, chairs, or trucks, but we cannot count the number of "red" without specifying what is the entity that redness can apply to. While concepts like CHAIR, TRUCK, and maybe even OBJECT are providing us with a specification of what *particulars* fall under it, RED crucially does not. This mapping between count nouns and sortal concepts has some explanatory power. If infants map count nouns of natural language onto basic level sortal concepts and, moreover, they treat basic-level sortal concepts as mutually exclusive, it follows that objects that are labelled with different count nouns have to be different objects. This prediction is supported by a wide range of evidence from early conceptual object individuation studies (see Chapter 1 for a review).

Crucially, however, the question of whether these sortal level conceptual descriptions could provide an autonomous indexing system is different from the question of whether they are used in object individuation. Recall that we defined primary properties as constitutive of the indexing process within a system, so they have to be necessarily encoded in order to maintain an object representation. Properties that are secondary in a system are ones that might be used in object individuation, but they are not part and parcel of the indexing process per se. Could we treat sortal concepts as primary properties under these terms, i.e., as an indexing system? The answer is negative. A positive answer would be a category mistake because sortal concepts are type level descriptions and do not directly pick out objects. An already represented entity can fall under a sortal description, but without representing an entity in the first place, sortal concepts have no way of providing indices that they could then be applied to.

To illustrate, consider the following passage explaining how individuation using the sortal based individuation system is supposed to work: *"[...]if an object seen at time 1 falls under one sortal concept and an object seen at time 2 falls under another sortal concept, then they must be two objects."* (Xu, 2007, p. 401). This formulation highlights the distinction that we want to capitalize on between primary and secondary properties. Infants indeed individuate the objects based on differences in their conceptual description. But the indexing of these objects is not done by sortal concepts, but by the visual system. The sortal concept is merely describing the indices, but the actual indices are "object seen at time 1" and "object seen at time 2". We know that these are not based on sortal concepts, but on "seeing" the objects, because in the lack of visual indices there wouldn't be anything that the descriptions could apply to. Thus, object individuation can be based on represented entities falling under different sortal descriptions, but sortal descriptions can only be secondary properties. The same issue arises with the natural language analogy that the sortal theory also relies on. The word "red" is an adjective and we cannot count the number of red. In contrast, we can count the number of trucks and "truck" is a count noun. But just as "red" is insufficient to establish a countable object, the word "truck" alone is also insufficient to talk about any particular truck. Count nouns in English require further linguistic elements in order to invoke particulars or token level descriptions of that given count noun. Noun phrases can contain a variety of constituents that contribute to invoking representations of particular entities, like articles, ("the truck") numerals, ("three trucks") or quantifiers ("both trucks"). But without the elements that do the actual counting or aid tokening, the word "truck" is insufficient to express thoughts of particular trucks.

Taken together, without a way of tokening particulars, sortal concepts can only be accommodated into this framework as secondary properties. Object representations are described rather than established by sortal-type conceptual content. If these concepts are not primary properties, then using them in individuation might be possible, but is not always necessary. This view is compatible with both the results obtained in Studies 1 & 2 where infants failed to use different labels to individuate objects, and all previous data where infants succeeded (e.g Xu, 2002).

4.2 Discourse referents

We will propose an alternative indexing system that is able to create and track representations that are bound to communicative agents or to a communicator-specific discourse. Our proposal builds on theories of discourse representation from the field of linguistics. In natural language semantics, the notion of discourse reference was first introduced by Karttunen (1968) to distinguish reference within a discourse from *reference* proper. Since then, various theories, e.g., discourse representation theory (Kamp, 1981) and dynamic semantics (Heim, 1983) have aimed to capture how natural language creates referents in a communicative context and keeps track of them, accounting for a variety of linguistic phenomena like the interpretation of anaphoric expressions. One influential and psychologically relevant way of analyzing the linguistic means of these processes is Heim's (1982) earlier theory called file change semantics (FCS). Building on Karttunen's notion of discourse referents, FCS posited an indexing system using file-like representations, not unlike psychological theories of object files (Kahnemann, Treissman, & Gibbs 1992), or mental files (Recanati, 2012; Perner & Leahy, 2016). What makes files in FCS and mental/object files different from each other is that FCS focuses on indexing entities not from a first-person referential viewpoint but from the point of a communicative discourse. The existence of a file is not granted by being anchored to visually available referents in the world, but having a file implies an implicit belief in the existence of the referent in the discourse (Heim, 1982; Partee, 2008).

Consider indefinite and definite noun phrases in FCS. They are differentiated based on their indexing properties. Indefinites carry a novelty requirement: used felicitously, they introduce a new file to the discourse, that is they create a novel index. In contrast, definites carry a familiarity requirement, they pick out files/indices that are already represented in the discourse. For example, the utterance “The cat is on the mat” requires a familiar cat and familiar mat known to all participating communicators. In FCS terms, it requires an already represented cat file/index and a mat file/index. If we change one of the definite articles to an indefinite, yielding a sentence such as “a cat is on the mat,” the interpretation changes. Now the cat is no longer familiar though we still require a familiar mat. In FCS terms the phrase “a cat” creates a new cat file/index to the discourse. Importantly, novelty and familiarity are local to the common ground (Stalnaker, 1974, 2002) of the discourse. The same utterances might be about different cats and mats in conversations with different interlocutors. Conversely, the same cat and mat might varyingly demand using the indefinite versus the definite depending on the interlocutor.

The main insight we aim to incorporate from this literature is the idea that there are socio-cognitive mechanisms that can track referents in relation to a discourse. While infants might not be in full command of all the linguistic apparatus that are routinely employed by adult language users (e.g., the first direct evidence for the distinction of indefinite/definite articles is from 18-month-olds; see Choi, Song, & Luo, 2018), we assume that the underlying apparatus to encode discourse bound representation is present from an early age. To account for the strong relationship between conceptual object individuation and communication, we propose that early in infancy particulars of a kind are represented in this discourse-bound system. That is, other than a system that represents an object at a location (visually-indexed object), infants can also represent an object under discussion (a discourse referent) [5].

4.3 Evidence from the literature on object representation

Positing a discourse-bound indexing system in infancy has architectural consequences. If, for infants, communicative acts can have a primary index creating function, it has to be

[5] When it comes to talking about kinds qua kinds (e.g., “dinosaurs are extinct”) we propose that discourse referents are always required. Kind concepts are not things one can directly perceive because concepts are abstract notions. Thus, visual indexing cannot pick kind concepts out directly (cf. Csibra & Shamsudheen 2015) but communicative reference to kind concepts can (Carlson, 1977).

demonstrably independent from the processes that bind properties to indices (e.g., individuating objects based on sortal concepts). In practice, this means that *within this system,* (1) infants' knowledge of a relevant sortal concept is insufficient for creating an index without a communicative context[6] and (2) a communicative context even without the necessary sortal knowledge should be sufficient for creating an index.

4.3.1 (1) Communication is necessary for index creation

Word recognition studies by show that by infants in the first year of life possess some word-to-concept mappings that could in principle let them succeed in the standard object individuation paradigm (Bergelson and Swingley, 2012; Parise and Csibra, 2012). Still, infants fail in kind-based object individuation until their first birthday in cases where the objects are not labelled during presentation (Xu & Carey, 1996 Xu, 2002). From this pattern alone, it is unclear what the labeling events contribute to infants' success. Are labels required to make the conceptual distinction between the objects salient? Or (as our model posits) is the facilitation also due to the communicative context, which engages an alternate indexing system?

There are recent studies that point to the latter possibility. In a recent study, referential communication was dissociated from the labels in the standard object individuation paradigm and infants still succeeded without explicit labeling (Shamsudheen & Csibra 2016). During object presentation, the objects were pointed to, but not labeled. This alone was enough to help the infants succeed in building an expectation for two objects when these objects belonged to familiar kinds. This is convergent with earlier results obtained from an individuation study by Futó, et al. (2010). In this study ten-month-old infants were presented with two objects that had different functions (one played music, the other had blinking lights). Even though the objects differed both in function and in surface features, infants only expected two objects in the presence of ostensive-referential communicative presentations of these functions. Again a conceptual contrast (here, a function) was insufficient.

[6] Importantly it does not exclude the possibility that other types of indices can use sortal-information to individuate objects.

This relationship between communication and conceptual encoding is supported by evidence from non-individuation studies as well. In a recent study, Pomiechowska, Brody, Csibra, and Gliga (in prep.) presented presented 12-month-old infants with two objects: one familiar, one novel. They measured infants' looking behavior in response to hearing either the label for the familiar word or a novel word. In the baseline condition, upon hearing a familiar label, infants looked at the familiar object, showing that they have the necessary conceptual/lexical knowledge to recognize the kind that familiar object belonged. When they heard a novel label, they did not look at the novel object, replicating previous findings that infants at that age are not abiding by the mutual exclusivity bias (cf. Halberda, 2006). In the critical condition, before hearing the labels, infants saw a pointing action directed at the familiar object. In this case 12-month-old infants not only looked at the familiar object when they heard the corresponding label, but now they *also* succeeded in making a mutual exclusivity inference: they looked at the novel object when they heard the novel label. In order to derive the inference that the novel label cannot refer to a familiar object, infants had to realize what the familiar actually object was. It seems that at least one of the two communicative acts were necessary to subsume the familiar object under a description. Either the pointing act alone, in the critical condition, or the corresponding label, during testing in the baseline condition. The main difference between the conditions is that in pointing condition infants could also use this encoding in their premise for making a mutual exclusivity inference *because* the pointing act preceded the novel label.

4.3.2 (2) Communication is sufficient for index creation

If object representations can be indexed in the discourse and such discourse-binding is a primary property, then infants should be able to create object representations even when no further information is available. This means that within this system, creating a novel object representation should not depend on the availability of relevant kind concepts or any information that is not a part of a referential communicative act. In Chapter 1, we already described the study by Yoon et al. (2008), where in a communicative context 9-month-old infants remembered the features of the object but forgot its location. As these were novel objects, there is no reason to believe that the way infants encoded the objects referenced a basic level sortal concept. At the same time, there are reasons to believe that this representation was not visually-indexed either: infants forgot the objects' location. Thus, the index must have been provided by the referential communicative act itself.

Going further, there is evidence that sortal-free indices can be created even without any direct visual access to the object, showing that index-creation can be solely dependent on referential communicative acts. Around one year of age, infants not only understand pointing gestures as referring to visually available objects (Behne, Carpenter, Liszkowski, & Tomasello, 2012), but can also take it refer to objects that are occluded or currently absent from the physical context (Liszkowski, Carpenter, & Tomasello, 2007). Most relevantly, Pätzold & Liszkowski (2019) provided evidence in a pupil dilation study that 12-month-old infants posit the presence of an object solely based on referential pointing acts. An agent's communicative pointing action was sufficient to create an object representation, even in the absence of any direct visual evidence or conceptual description of the object. In a different paradigm similiar results were obtained by Csibra & Volein (2008) showing that already at 8 and 12 month infants expect an object's presence in response to communicative gaze (see also Moll & Tomasello, 2004). Further evidence comes from a search task where infants inferred the existence of an object in one of two boxes in response to communicative gaze and pointing, but failed to do so if similar behavior was presented in a non-communicative setting (Behne, Carpenter & Tomasello, 2005).

While these results are strong evidence against the idea that sortal concepts are required for indexing, in most of these cases infants expected the object *at a location.* How can we know that location information is not necessary to establish discourse referents? In a study by Moll, Koring, Carpenter & Tomasello (2006), 14- and 18-month-old infants were playing with an object with the experimenter. When a new experimenter entered the scene and pointed to *the side* of the object (an unconventional pointing act) with excitement, infants tended to understand this gesture as referring to the object that they played with. But if the same behavior (excitement, unconventional pointing) was displayed by the initial experimenter, who had already communicated to the infant about the object, infants' behavior changed: they started to look around in the room, as if searching for a novel referent. This can be construed as evidence for infants positing the presence of a further object, one that they did not know the location of, based on their interpretation of the communicative act. Notice how well this result can be accommodated in a framework like FCS. FCS posits that the interpretation of some communicative acts can create novel indices and other communicative acts can refer back to already represented ones. Novelty and familiarity is assessed in specific discourses. It seems reasonable to assume that the unconventional pointing action expressed by the experimenters was understood by infants as expressing novelty. But the interpretation of these otherwise behaviorally indistinguishable actions depended on what was already part of a specific discourse. For the experimenter with whom the object was not yet part of the discourse, the object could be construed as novel entity. On the other hand, for the initial experimenter this object was already part of the discourse, and the same exact action could not meet the novelty requirement referring to the familiar object. From the infants' first person perspective the two behaviors express novelty to the same degree. But from the point of the two distinct communicative discourses, novelty has different consequences.

4.4 Evidence from the literature on communication

The object representation literature indicates that infants can index entities in relation to discourse contexts, without referencing any other type of information. Such an indexing system assumes infants to be equipped with a variety of cognitive capacities. At a minimum, they have to be able to: (1) identify possible communicators to set up discourse contexts, (2) monitor others' mental states to establish (or approximate) common ground, and (3) differentiate between communicative acts performed by different agents. In the following, we argue that there is evidence for these capacities in infants by the first year of life.

4.4.1 (1) Picking out communicators is a crucial precondition of having discourse-bound object indexing. It is well established that even at an age when infants fail in making use of other object-kind distinctions, they can spontaneously treat humanness (Bonatti et al., 2002) or more generally agency (Surian & Caldi, 2010) as a strong conceptual boundary relevant for object individuation. Infants are quite expert in picking out communicators also. They are sensitive to a variety of interconnected ostensive signals, like eye contact (Farroni, Johnson, Menon, Zulian, Farugna & Csibra, 2005), infant-directed speech (Cooper & Aslin, 1990), or contingent responsivity (Murray & Trevarthen, 1985). These signals have been argued to help infants recognize the presence of a "communicative intention" even if they can not recognize the informative intent (the content) of the message (Csibra, 2010). Under various theories of human communication (Grice, 1989; Sperber & Wilson, 1995), a communicative intention, indicated by ostensive signals, is generally understood as letting the addressee(s) know the intention to communicate some information. If this notion is on the right track, an infant who recognizes a communicative intention must have at least an implicit grasp of the necessary concepts such as addressee, information, intent and so forth. There are empirical reason to think that infants' interpretation of ostensive signals reflects more than just an attentional sensitivity to these cues. These signals change infants' behavior in ways that reflect a genuine understanding of communicative intentions. As we cited previously, infants' understanding of objects radically shift in communicative contexts (Xu, 2002; Futo et al., 2010; Yoon et al., 2008). Prior ostensive signals also change infants gaze-following behavior both overtly (Senju & Csibra, 2008) and covertly (Farroni, Johnson, & Csibra, 2004). Finally, and as discussed before, ostensive signals result in changes in referential expectations. Infants expect to find an object at the location where communicative gaze (Csibra & Volein, 2008) or pointing (Behne et al., 2005) is directed.

4.4.2 (2) Indexing an object in a discourse establishes the common ground with the discourse participants where these indices can be maintained. Common ground can be defined as the set of beliefs and information that are shared by discourse participants (Stalnaker 1974, 2002). To establish or approximate common ground, infants thus have to be able to infer mental states of other agents. Mental state attribution, or theory of mind, is traditionally linked to the ability to attribute false beliefs (Wimmer & Perner, 1983). While there is a range of evidence from 7 months of age that infants attribute false beliefs to other agents in some tasks (Kovács, Téglás, &

Endress, 2010; Onishi & Baillargeon, 2005), the evidence and its implications are heavily debated (Jacob, 2012; Kovács, 2016; Dörrenberg, Rakoczy, & Liszkowski, 2018). The debate centers around the fact that these tasks do not match up well with the traditional verbal false-belief tasks on which children fail before they turn 4 years (Wimmer and Perner, 1983; Wellman, Cross, & Watson, 2001). There are architectural and terminological questions related to infants' representational capacities for attributing "belief proper" versus some notion of a "pre-belief" (Apperly & Butterfill, 2009, Rakoczy, 2014). Here, we appeal to a different literature to argue for infants' capacity to attribute mental content to others, at least to the degree that a discourse based representational system would require.

Infants start to produce pointing gestures around 12 month of age (Tomasello, Carpenter, & Liszkowski, 2007). They use it flexibly: to request an object (imperatively), to share information (declaratively) and to request information (interrogatively) (Liszkowski, 2005; Southgate and Begus, 2012). There are good reasons to believe that these actions are intentional (i.e., non-reflexive and motivated) and informative (Tomasello et al., 2007, Harris & Lane, 2014; Southgate, van Maanen, & Csibra, 2007). The way infants produce these communicative acts are discourse specific. For example, Begus and Southgate (2012) found that infants modulate the amount of interrogative pointing behavior in response to the differences in knowledgeability of their interlocutor. Sixteen-month-old infants produced more interrogative points in the presence of the reliable partner compared to an unreliable one.

These behaviors show that infants can differentially reason about communicative partners to establish what information they can provide. Doing so requires the ability to entertain a state of affairs where the other person has access to different information content from the infant. Insofar as this mental content enters into intentional behavior, giving rise to actions like the just discussed cases of interrogative and declarative pointing, it has to be represented explicitly. After all, it makes no sense to interrogate information that is not assumed to be (possibly) represented by the conversational partner. Similarly, it makes no sense to share information without representing a state of affairs where the addressee comes to learn this information. This follows from the premise that the goal of an action is logically and thus psychologically prior to representing how one performs it (Fodor, 2008). If we believe that communicative behaviors like interrogative pointing are goal directed acts, then it must be that the goal of that action must be represented in order to execute it.

The main takeaway of the above discussion is that if we accept that infants' communicative acts are informative and intentional it follows that they can attribute different information content to the participants of a discourse. That is, they can represent both themselves and their communicative partner as asymmetrically having or lacking some particular information. While this argument doesn't show that the information is represented as "full blown propositional attitudes" it shows that both the entertained and attributed content have to be explicit and in a compatible format, at least to a degree that makes intentional communication possible[7].

4.4.3 (3) In order to index objects relative to the discourse infants are required to distinguish different discourse contexts. This, at a minimum requires infants to bind communicative acts to specific agents, and by extension bind referents to the respective communicators. Indeed, thirteen-month-old infants expect that a pointing gesture and a simultaneously provided label are co-referential only if they come from the same communicator but not if they are produced by two persons (Gliga & Csibra, 2009). Around 12 and 14 months of age infants, after interactions with two distinct adults, understand an ambiguous request for an object as referring to the one that was already present in the corresponding discourse (Saylor & Ganea, 2007; Saylor, Ganea & Vázquez, 2011). There are also a variety of studies in the joint attention literature that show that

[7] This argument doesn't address false belief attribution. As common ground consists of beliefs shared by the both participants, remembering attributed false beliefs are not required for maintaining it, at least on a purely definitional basis. Thus, removing some information from the common ground in some cases might create equivalent communicative outcomes as attributing a false belief. Removing the relevant belief from common ground in the Sally-Ann task (as opposed to attributing a false belief) could result in the same desire of communicating the correct location as not attributing any belief. On the other hand, when it comes to action prediction — the gold standard for measuring Theory of Mind — false belief attribution is required. If the main function of Theory of Mind is to help approximate common ground and make cooperative communication possible, it might not be surprising that infants and young children, while are able to attribute mental content, do not necessarily invest in attributing beliefs that happen to be false (cf. Leslie 2000).

infants can assess the familiarity or novelty of a given object relative to previous shared experience with a communicative partner (Tomasello & Haberl, 2003; Moll & Tomasello, 2006). For example Moll et al. (2006) found that 14-month-old infants interpret an excited ambiguous request as referring for to a novel object compared to the objects that were already present in their discourse.

These results do not show that infants cannot have a kind-referring interpretation of a given communicative act (cf. Csibra & Gergely, 2011; Egyed, Kiraly, & Gergely, 2013). On the contrary, it seems to be the case quite often (Csibra & Gergely, 2009; Csibra & Shamshudeen, 2015). A simple way to accommodate generic information to the current framework is to define it as information that can be assumed to be common ground across different discourse contexts: e.g., object valence (Egyed, Kiraly, & Gergely 2013) or labels for kinds (Gelman, 2009). It also seems possible that infants can encode discourse specific and discourse-generalizable information from a single event. In a study by Buresh and Woodward (2007), 13-month-old infants restricted their interpretation of a communicative reaching action to a particular agent while generalizing the novel label that was provided during the action across agents (cf. Kampis, Somogyi, Itakura & Kiraly, 2013).

4.5 Challenges for an independent discourse based indexing system

We cited evidence showing that infants can pick out communicative agents, and maintain object representations that are indexed in relation to the corresponding discourse context. We argued that there are empirical reasons to assume that infants construe or approximate common ground, as infants attribute asymmetric informational content to themselves and to other agents, demonstrated by their understanding and production of declarative and interrogative communicative acts. The behaviors that infants express generally reveal a sophisticated understanding of communication (Csibra, 2010; Tomasello, 2009). There are expansive research traditions probing adult semantic and pragmatic systems (for an overview see Heim & Kratzer, 1998), but less on the computational mechanisms that are already in place in infancy (cf. Bohn &

Frank, 2019; Goodman & Frank, 2016). Postulating a discourse based indexing system in infancy seems beneficial from both an empirical and a theoretical perspective, and it results in new questions and puzzles. We want to briefly describe two challenges for future research. We should have a better understanding how infants construct and track discourse referents (1) and what the space of entities is that they can index (2).

4.5.1 (1) The first explanatory challenge is to characterize how discourse referents are created and maintained. Under what conditions do infants create a novel discourse referent, and under what conditions can infants construe a communicative act as referring anaphorically, i.e., picking out something that is already present in the common ground? In the case of natural language as the medium, theories like FCS have rich linguistic data to hypothesize over. For example, the already mentioned definite/indefinite contrast was analyzed in various complex environments that resulted in better understanding of how they create and maintain indices within a linguistic discourse (Heim, 1982). Testing environments like these are not viable for our purposes, as infants lack the necessary verbal proficiency. But in order to acquire such linguistic apparatus, infants might already have to entertain the relevant notions conceptually, since without a system that keeps track of the referents that are present in the discourse, they would not be able to update their beliefs in relation to communicative acts in the first place. So how do infants do this? The few pieces of this puzzle that are available indicate that infants in some cases treat different noun phrases with different (head) nouns as referring disjointly (Xu, 2002), but not in others (Studies 2 & 3). Fitting well with the FCS, there is also evidence that under some conditions, novelty expressed by the interlocutor is construed as introducing a novel referent to the discourse (Haberl & Tomasello, 2003; Moll & Tomasello, 2004, Moll et al., 2006). In the next chapter we will try to make progress on this question by empirically assessing the relationship between spatial information and discourse information.

4.5.2 (2) A second, closely related task is to characterize the conceptual space of the entities that can be construed as a possible discourse referent. That is, what types of secondary information can discourse referents bind? At a bare minimum, it seems that infants can treat communicative acts as referring to particulars[8] (e.g. objects with or without kind description) and as referring to kind concepts, expressing generalizable information. The kind relevant interpretations might help to explain why infants and children can take objects as exemplars of kinds in nonverbal communication (Csibra & Shamsudheen, 2015) or map words to concepts (e.g. Waxman & Gelman, 2009; 2010; Gelman, 2004, Yin and Csibra, 2015). Reference to kinds might also be fruitful in thinking of a variety of other phenomena, like treating object function as kind relevant in a communicative setting (Futó et al., 2010). At the same time when infants request an object by pointing to it, they probably do not request a kind concept but a particular, as concepts — as opposed to particulars — cannot be touched (mutatis mutandis for informing someone of the location of an object (Liszkowski et al., 2008)). How do infants know whether to take a communicative act as being about kinds or about entities in the absence of identifying cues from natural language?

As a first stab at this problem, let us imagine that infants represent a taxonomy of communicative acts with attributed schemata that restricts the range of possible interpretations for a given discourse referent. This could be analogous to how the grammar of natural languages can constrain possible interpretations of referential expressions. The English utterance "dogs like treats" is constrained to a kind referring interpretation, where the discourse referent introduced by *dogs* does not invoke any particular dog, only the kind concept DOG. In contrast, the utterance "my dogs like treats" makes reference to particular dogs that one owns. The simple point here is that analyzing the whole utterance, rather than just the noun *dog*, can constrain the hypothesis space for picking the correct discourse referent[9]. It remains to be worked out what such a taxonomy of non-verbal communicative acts look like, and what information can be used by infants to identify the relevant taxonomical entry. But if this taxonomy includes categories like "labeling act" or "function demonstration", these might help to constrain to possible

[8] Particulars here are understood as entities in the world. Reference to particulars does not necessarily involve picking out a unique entity. For example an imperative pointing request directed at a duck might be interpreted as request for the unique duck that was pointed to, or just for any duck. But under both notions it have to be understood as request for a particular.

[9] For an extended discussion of kind reference in natural language cf. Carson (1977) and Chierchia (1998)

interpretation of the discourse referent to kind concepts. Analogously, “requesting act" might also be a taxonomical entry, with the attributed rule that the request cannot be of a kind concept, helping to restrict the interpretation of the discourse referent to particulars.

The learned or innate cues that infants use for understanding communicative acts is an open empirical question on any account of communication[10]. A system that would put interpretative constraints on non-verbal communicative acts is descriptive rather than explanatory, but the idea might help to create a more unified notion of early communicative understanding, as it provides a framework to analyze early biases for interpreting communicative acts as having kind-referring or episodic content.

[10] A paradigm case for innate sensitivity to communicative acts might be ostension (Csibra, 2010). Another taxonomical entity that might be available around the first year of life is possibly pretense (Leslie, 1987)

Chapter 5 — Study 3: Individuating discourse referents

In Chapter 4 we developed arguments why communicative acts can provide infants with indices that represent objects. We argued that these indices are grounded in the discourse rather than in the visual-encoding of the entities. In the current chapter we investigate how this discourse based indexing system functions in one-year-old infants.

Our main focus is to assess how objects are individuated in distinct discourses and what properties contribute to the individuation process. The main issue is how referents from different discourse contexts can get unified. Because the indexing process is not based on infants first person knowledge but relative to particular discourses, it is an open question whether expectations of multiple contexts can get integrated. Can discourse bound indices get treated in a unified way, so that infants' expectations are based on multiple contexts simultaneously? If there are two hidden objects in a scene, and evidence for the existence for each object is present only in one discourse, do infants' expectations reflect information from both context (expecting two) objects? If we can answer this question affirmatively, that would mean that the relevant system of representation can individuate objects between discourse contexts and not only within a discourse. We test this question by manipulating the number of separate communicators (to operationalize separate discourses) that engage in referential communication with the infant to provide information about the presence of objects in the scene.

Our second question relates to the encoding discourse-referents in physical space. Following the literature reviewed in Chapter 1, we argued that spatiotemporal object individuation is based on processes that visually index the objects (Spelke et al., 1995; Xu & Carey, 1996; Scholl & Leslie, 1999). On the other hand, visual evidence is not necessary to encode objects at a location. Evidence from the literature on communicative gaze and pointing suggests that infants can represent discourse referents at a particular place (Csibra & Volein, 2008; Pätzold & Liszkowski, 2019; Behne et al., 2005). We also cited evidence that implies that representing the location of an object is not a necessary precondition for creating a discourse referent (Moll et al., 2006). The best treatment of this pattern is to think of location information as a secondary property of discourse referents. We aim to test whether infants can use this secondary information not only to attach it to the objects (i.e., to encode their location), but also to individuate them without any direct visual evidence. That is, if referential communication is directed at different locations, do infants interpret them as index-creating acts? In order to test this question, we manipulate the number of cued locations that are present in the referential communicative setting. This manipulation also serves as a natural control to visual-index based encoding. If, contrary to our arguments in Chapters 1 & 4, infants create visual indices rather than discourse-bound ones in response to referential communication, then their expectations about the number of objects should be only influenced by the distinct locations that referential-communicative acts highlight.

To test these two interrelated questions we devised 3 hypotheses:

1. If infants only treat object location as an individuating (primary) property, even without direct visual access to objects, then the number of cued locations should have an effect on infants' expectations. Thus, according to Hypothesis 1, the number of represented objects should be determined by the number of cued locations in the scene, resulting in an expectation of 2 objects if there are 2 cued locations.

2. In our model, referents in different discourses are indexed disjointly. If infants successfully integrate these indices, they should expect multiple objects if they encounter referential acts in distinct discourse contexts. The corresponding Hypothesis 2 predicts that the number of represented objects should be determined by the number of communicators in the scene. If there are 2 communicators, infants should expect 2 objects in the scene.

3. According to the mixed Hypothesis 3, both of the above hypotheses are correct. Thus, the number of expected objects should be determined both by the number of cued locations and

the number of communicators. Infants should expect multiple objects if either the number of cued locations or the number of communicators are higher than 1.

5.2 Experiment 1

In Experiment 1 we provided information about objects solely through referential communication, more precisely through gaze direction, pointing action and verbal cueing, without labels or direct visual access to the objects. We systematically manipulated the number of communicators (1 or 2) and the number of cued locations (1 or 2) and measured infants' looking time to assess their expectations about the number of objects (1 or 2) in the scene. This let us assess how information is integrated from different discourse contexts, and how spatiotemporal information modulates this process.

5.2.1 Methods

5.2.1.1 Participants

We included 48 infants in the final sample. Their age ranged from 11 month and 13 days to 12 month and 16 days. The mean age was 12 months exactly (SD = 8 days). A further 18 babies were tested but excluded from the sample: 1 for parental interference, 10 for fussiness, 7 for experimenter error. Of these 7 experimenter errors, 4 were errors were in the counterbalancing script, 3 in live coding. We did not have to exclude any infants due to other pre-specified exclusion criteria (see Coding below). We recruited participants through the database of the Cognitive Devleopment Center at Central European University. All personal data were handled in accordance to the GDPR regulations. The study and the data protection measures were approved by the Hungarian United Ethical Review Committee for Research in Psychology (EPKEP). The parents did not receive any compensation for the participation, but infants were gifted small toys after the session. All parents signed a consent form before the study.

5.2.1.2 Procedure

After infants arrived to the lab, they were familiarized with with the environment while the caregiver was introduced to the details of the study and signed the consent form. After this period, the testing took place in a dimly lit room, where the infants were seated in their parents' lap roughly 80 cm away from a 24 inch flat screen monitor used for stimulus presentation. Parents were instructed to close their eyes during the study, and not to talk or interfere with the infant's behavior other than keeping them in a sitting position. The experimenter controlled stimulus presentation and live coded looking behavior during test trials, where the duration of presentation was contingent on the infants' gaze.

In between consecutive trials, an attention-grabbing animation was presented to help infants re-orient towards the screen. This animation consisted of colorful arrows presented centrally on a black background. In case infants did not look at the screen, the trials did not start but a short auditory stimulus (a ringing sound) was presented until they reoriented towards the screen.

5.2.1.3 Design

We adopted a 2X2X2 design with two within-subject and one between-subject independent variables (Figure 1A). The between-subject manipulation was the number of possible hiding locations (occluders) present. One group of infants were assigned to the two-location conditions, the other group to the one-location condition. The groups were assigned 24 infants each. The two within-subject independent variables were the number of communicators, and the number of objects at outcome. The number of communicators refers to the number of humans present and communicating to the infant in the demonstration phase of the test trials. In half of the trials two people were present, while in the other half only a single person was present. In the two-communicator trials both people were communicating to the infant. At test, we presented infants with both one-object and two-object outcomes. Altogether, every infant was assigned to either the one-location or the two-location context, and was presented 4 test trials: 1 communicator and 1 object; 1 communicator and 2 objects; 2 communicators and 1 object; and 2 communicators 2 objects. These 4 test trials were presented in every possible order to achieve a fully counterbalanced design. Our dependent measure was log-looking time (Csibra et al., 2016). Looking time measurement started at the moment the objects were revealed in the test trials.

Before the test trials, infants also were presented with two introductory and two familiarization trials. In the two introductory trials, they were presented with each of the two communicators alone without the objects. In the two familiarization trials infants watched one- or two-object outcomes without communicators. The introductory trials always preceded the familiarization trials, but for both trial types the order of presentation was locally counterbalanced (independent of the counterbalancing of the experimental conditions).

5.2.1.4 Materials

The stimuli were presented as short video clips. In each clip, a desk with black barrier at the far end was the scene of the events. The occluders were placed on the desk in front of the barrier, while the communicators were sitting behind it. The communicators were trained actors who differed in gender and in a variety of other perceptible features. In the two-communicator conditions they were sitting next to each other, while in the one-communicator conditions they were seated at the center of the frame (Figure 5.1).

In the two-location conditions, the occluders were placed at the opposite sides of the desk, while in the one location condition a larger occluder was placed centrally. The occluders moved by themselves. In the two-location conditions, they revealed the object by leaving the scene in a horizontal direction, while in the one-location conditions the occluder moved vertically, leaving towards the bottom of the screen. The occluders were orange colored, and had a brick-like pattern. They were created (and animated) during the post processing of the video clips.

A red fabric-textured curtain, which hid the communicators was also animated and added in post-processing phase. It never completely disappeared from the screen. It was either visible only at the very top of the screen, revealing the communicator(s), or it was lowered, so that in conjunction with the barrier it hid them. When the curtain was lowered its bottom edge was considerably higher than the occluder(s) to make sure that they do not create a unitary occluding surface, which would allow for a single-occluder construal of the two-locations conditions. This also ensured that the communicators had no physically plausible ways of manipulating the objects, when hidden.

During the outcome phase of the familiarization and test trials, 4 unfamiliar objects were presented, organized into two pairs (Figure 5.2). For every infant one of the two pairs was used for the familiarization trials, while the other pair was used in the test trials. The 4 objects differed substantially in shape, pattern and color, but roughly had the same size.

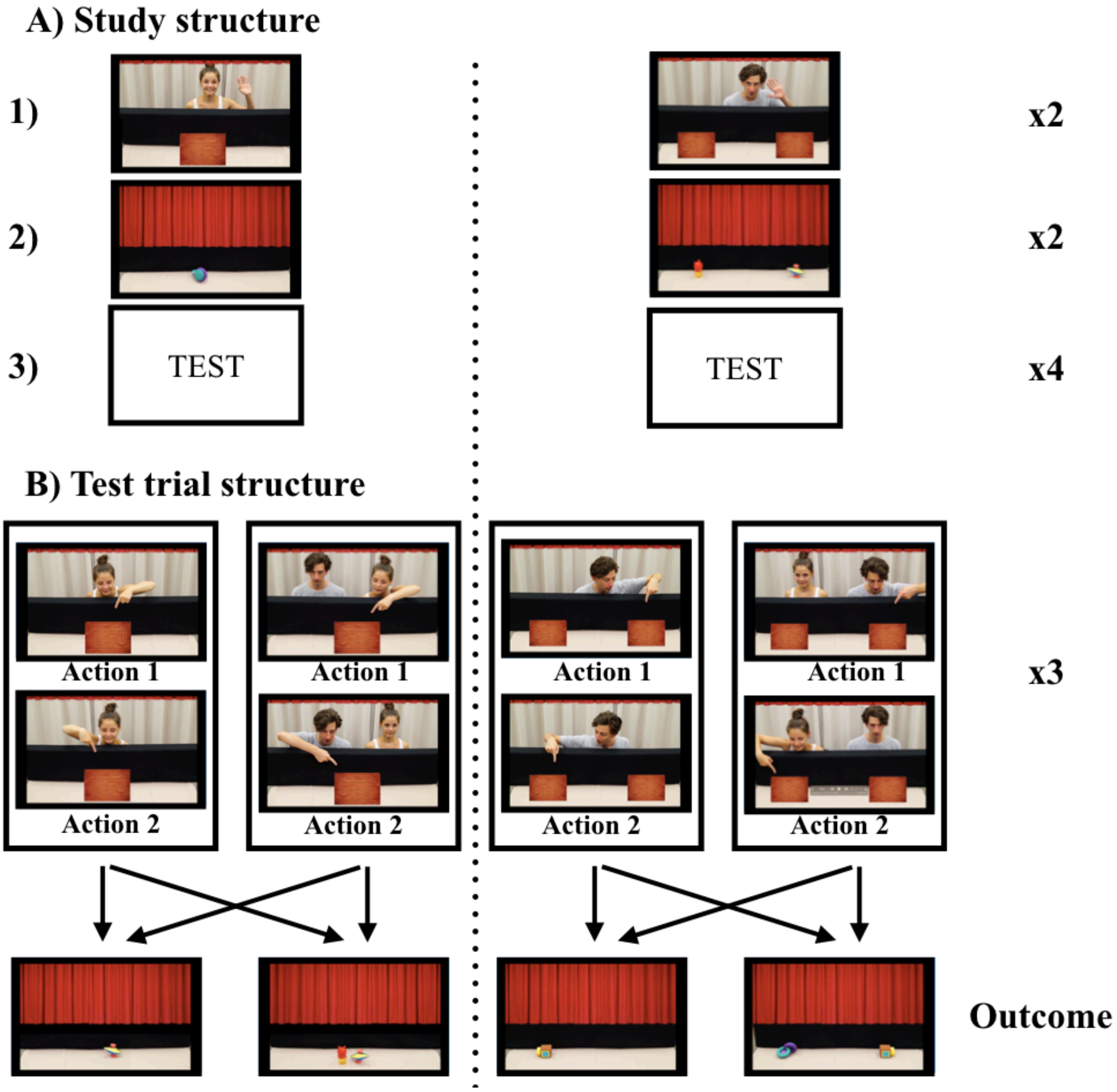


Figure 5.1 (**A**) Every infant was assigned to either the one or the two locations condition. The study started with 2 *introductory trials* (1) where both communicators greeted the infant. It was followed by 2 *familiarization trials (*2) revealing the possible number of objects at outcome. Finally 4 *test trials* were presented. *(***B**) The test trials started with a *demonstration* that included 2 referential actions each. These demonstrations were repeated quasi identically 3 times. A single participant saw all four possible test trials, where the number of communicators and the number of objects at outcome were varied orthogonally.

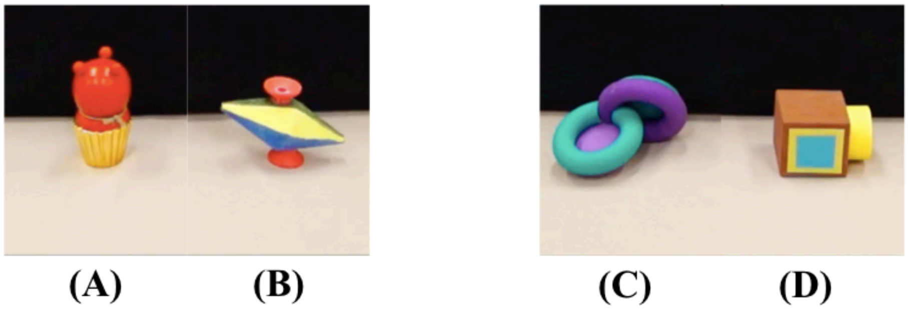


Figure 5.2. Object pairs in the study. Either (A) and (B) were used for familiarization and (C) and (D) for the test trials, or the other way around.

5.2.1.5 Introductory trials

The purpose of these trials was to introduce both communicators to the participants before the test trials. The trials started by raising of the curtain, which revealed one of the two communicators. They greeted the infant by waving to them, and with uttering “Szia Baba!” (the Hungarian equivalent of “Hi Baby!”). After this greeting, the curtain was lowered again, to hide the communicator. The occluders were present, but they stayed stationery for the whole duration of the introductory trials. Trial length was 4.5 s and was not contingent on infants’ looking behavior. A second introductory trial was presented with the other communicator.

5.2.1.6 Familiarization trials

The familiarization trials had two main purposes: (1) to introduce the infants to the behavior and nature of the occluders, and (2) to familiarize them with the fact that either one or two objects are present at the scene. Both familiarization trials lasted a total of 12 seconds. At the starting frame of these trials the occluder(s) were present and the curtain was lowered. After 1 second, the occluder(s) left the screen revealing the object(s) for ~3 seconds, before returning to their starting position. These events were repeated 3 times, revealing the same outcome every time. The second familiarization trial was the same, but revealed the other outcome type (1 vs. 2 objects). In the two-locations/one-object outcome, the object’s location was counterbalanced across infants. The object pair that was used in familiarization trials were not used in the test trials.

5.2.1.7 Test trials

In the test trials we aimed to assess infants’ numerical expectations — as measured by their looking times to one or two object outcomes — in response to referential communication. In a particular trial, these communicative acts were performed by either one or two communicators, highlighting either one or two physical locations. Every trial started with the curtain raised, and the communicator(s) present. Their gaze was downward directed, not looking at any objects in the scene (3 s). A communicator then looked in the camera greeted the infant by waving, and uttering “Szia Baba” (“Hi Baby!”).

Then they continued by saying "Nézd csak!" ("Look!") and pointing behind an occluder (4 s). In the *one–location* conditions the pointing was always directed to single occluder present on the scene. In the *two-communicators / two-locations* condition, the pointing was directed towards the occluder closest to the communicator. In the *one-communicator / two-locations* conditions, the pointing was directed towards either one of the two possible occluders. The pointing lasted for 5 seconds, with the communicator switching her/his gaze between the location and the viewer. After this, the communicator returned into their initial position directing their gaze away from the camera. This event was repeated, by the same person in the *one-communicator* conditions, and by the other person in the *two-communicators* conditions. Crucially for both the *one-communicator / two-locations* and the *two-communicators / two-locations* conditions the second pointing act highlighted the second location. As a result, irrespective of the condition, every occluder-hidden location was referred to by a communicator, and every communicator present did engage in referential communication. In the *one-communicator* conditions, the communicator alternated between using their right and left hands for pointing, so that number of different actions are equalized with the *two-communicators* conditions at a perceptual level.

These demonstrations (which included two instances of referential communication each) were repeated an additional two times, with minor modifications. In the repetitions the communicator(s) did not greet the infant but uttered new verbal frames. For the second presentation they used the phrase "Látod?" ("Can you see (it)?"). For the third presentation the utterance was "Figyelj!" ("Watch!"). Otherwise the structure of the repetitions were the same as the first demonstration.

The demonstrations within the test trials took 54 seconds to complete. After this period the curtain was lowered (1 s) hiding the communicators (with a sound effect accompanying the event). After an additional 0.5 s the occluder(s) also left the screen revealing one or two objects as the outcome. In the one-object / two-location conditions the object was always present at the location last pointed to. After revealing the outcome the experimenter coded the infants' looking time live. Trials ended when the infant looked away for 2 seconds, or the maximum looking time (50 s) was reached.

5.2.1.8 Coding

All looking times were off-line coded by a hypothesis-blind trained research assistant on a frame-by-frame basis. We also coded how many of the 6 communicative acts the infants saw during the presentation. We planned to exclude infants who would see fewer than 4 acts in a given trial, but no infant had to be excluded for this reason.

5.2.2 Results

We analyzed the log-transformed looking times using a mixed-model 2x2x2 ANOVA with the number of locations being a between-subject factor, while the number of communicators and the outcome being within-subject variables. We found a strong three-way interaction between these factors ($F(1,46) = 7.190$, $p = .010$, $\eta^2 = .135$). We found no two-way interactions, but a marginally significant main effect of outcome ($F(1,46) = 3.889$, $p = .055$, $\eta^2 = .078$).

Looking closer at the pattern of results of the untransformed looking times (Figure 5.3) we can see that in the two-communicator/one-location (one object: M = 8.56 s, SD = 9.64 s; two objects: M = 7.11 s, SD = 3.70 s) and the one-communicator/two-location condition (one object: M = 10.96 s, SD = 7.79 s; two objects: M = 9.67 s, SD = 10.18 s) infants' looking times were roughly equal for the two outcomes, while in the one-communicator/one-location (one object: M = 6.46 s, SD = 3.98 s; two objects: M = 10.95 s, SD = 8.79 s) and two-communicator/two-location condition (one object: M = 7.12 s, SD = 5.79 s; two objects: M = 10.75 s, SD = 5.96 s) infants looked longer at the two-object outcome.

To further explore this pattern, we conducted the two available 2x2 ANOVAs within the two groups of infants. We found that in the one-location group there was a marginal main effect of outcome, as infants looked longer at the two-object outcomes ($F(1,23) = 4.164$, $p = .053$, $\eta^2 =$

.135), but there was no interaction between the factors. In the two-locations group there was a significant interaction between the number communicators and the outcomes (*F*(1,23) = 6.643, *p* = .017, η^2 = .224), without any main effects.

Comparing across groups (which differed in the number of occluders present) in the one-communicator conditions, we found an interaction (*F*(1,46) = 5.405, *p* = .025, η^2 = .105), as the infants looked longer at the two objects when only a single occluder were present. A similar analysis within the two-communicator conditions found a main effect of outcomes (*F*(1,46) = 4.560, *p* = .038, η^2 = .090), but the interaction did not to reach statistical significance (*F*(1,46) = 2.799, *p* = .101, η^2 = .057).

Altogether, this complex pattern of results indicates that neither the number of communicators or the number of locations were uniquely predictive of the infants' looking times for the different outcomes. But two distinct patterns of looking behavior can be identified. In pattern (a) conditions, the infants looked longer for the two-object outcomes than the one-object outcomes (one communicator/one location; two communicators/two locations), while in pattern (b) conditions the infants looked roughly equally at the two outcomes (one communicator/two locations; two communicators/one location). Pre-planned t-tests confirmed this analysis. In the pattern (a) conditions, the infants looked longer at two objects than one (*t*(23) = 2.417, *p* = .024 in the one-communicator/one-location condition; *t*(23) = -2.099, *p* = .047 in the two-communicator/two-location condition), while in pattern (b) conditions there was no difference (*t*(23) = 1.237, *p* = .229 in the one-communicator/two-location condition; *t*(23) = 0.375, *p* = .771 in the two communicator/onelocation condition).

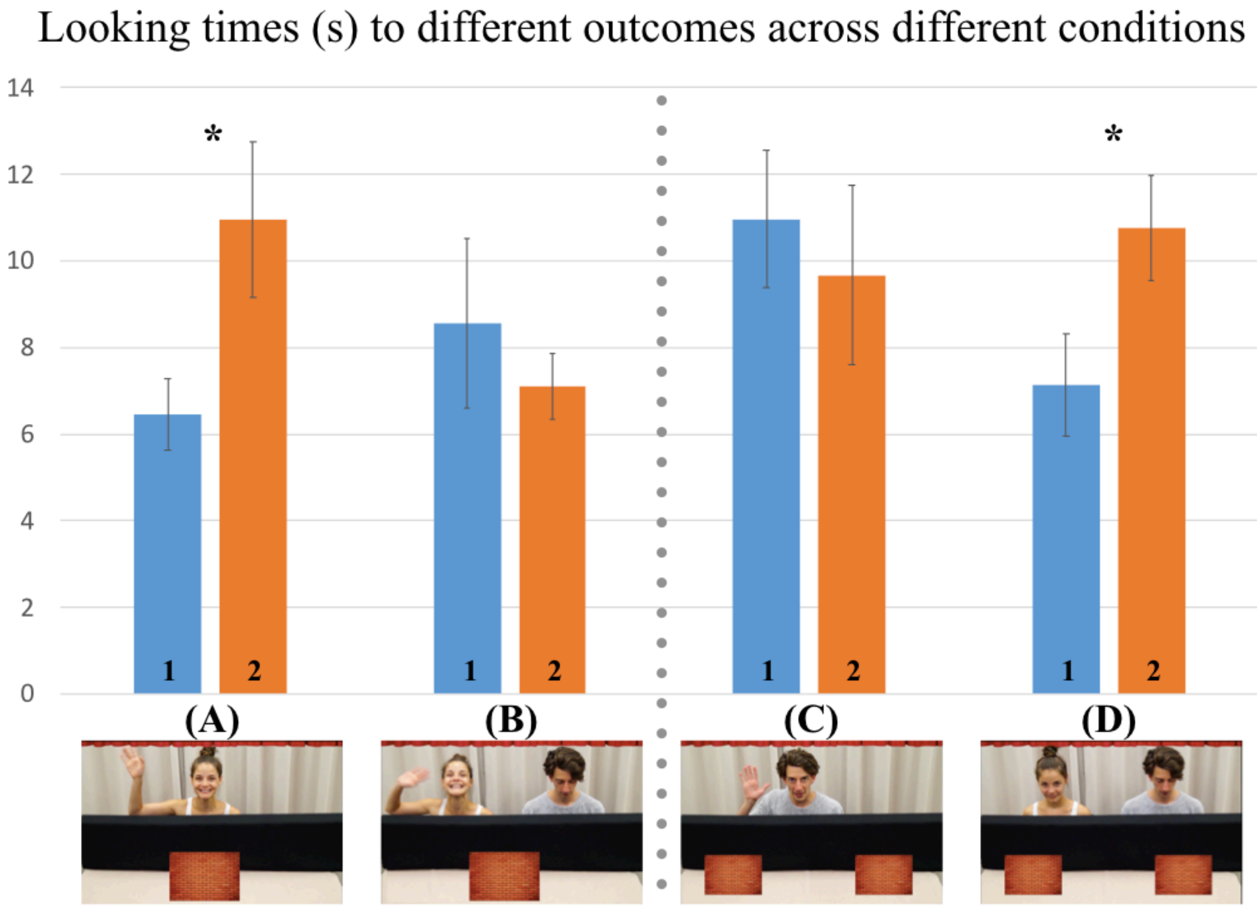


Figure 5.3. Infants' looking times (s) to the different outcomes in Experiment 1. Bars marked with (1) represent one-object outcomes, and (2) represent two-object outcomes. Conditions (A) and (B) are from the one location group: (A) one communicator, (B) two communicators. Conditions (C) and (D) are from the two locations group: (C) one communicator, (D) two communicators. We found a three-way interaction between outcome, number of communicators, and number of cued locations (p = .010). Stars represent significant differences within conditions (p < .05). Error bars represent standard error of the mean. Statistical analyses were performed on log-transformed data.

5.2.3 Discussion

Neither a purely location based (Hypothesis 1) or purely communicator based (Hypothesis 2) explanation of looking behavior will suffice to explain the data that we collected. The evidence also rules out Hypothesis 3: the one-location/one-communicator condition was not the only condition where infants were less surprised at a single object. On the other hand, the data shows that infants' numerical expectations are dependent on both of our manipulations. In order to account for the obtained looking time pattern, we start by offering two possible explanations.

Out of the four pre-planned t-tests that analyzed looking time differences within conditions, we found a significant difference in two: in the one-communicator / one-location condition and in the two-communicators/two-locations condition. The infants in these pattern (a) conditions looked longer at the two-object outcome compared to the one-object outcome. This difference changed in the pattern (b) conditions, where the infants looked roughly equally at the two outcomes. While it seems intuitive to account for these results as infants expecting one object in pattern (a) conditions, but not in pattern (b) conditions, we have to consider the possible baseline biases of looking behavior.

There are theoretical reasons to believe that the one-communicator/one-location condition reflects infants' baseline expectations. In that condition, infants did not have either spatiotemporal or communicative cues to generate numerical expectations of multiple objects. Methodological considerations also supports this view. It is frequently found in individuation studies that infants look longer at two objects than one in baseline conditions, arguably because of the increased perceptual complexity of the scene. If we take this condition as baseline, then we found successful individuation – as operationally defined by expecting two objects – in both pattern (b) conditions, but not in the other pattern (a) condition (two-communicators/two-locations). This would imply that infants can use either of the presented individuation cues in isolation but fail to use them in conjunction. In plain English, infants would be able to individuate an "object here" versus an "object there", and they would also be able to individuate an “object she points to” versus an “object he points to”, but fail when they would need to use “the object here that she points to” versus “the object there that he points to”. This failure to use both cues in conjunction still requires an explanation.

There is an alternative way of looking at infants’ behavior in Experiment 1. It is also possible that the similar looking times for one- and two-object outcomes are closer to infants’ baseline preferences (pattern (b)). Thus, these two conditions could be interpreted as infants’ failure of changing their numerical expectations based on the cues presented. If pattern (b) reflects baseline looking times, we would still need to account for the results in pattern (a) conditions. In these scenarios infants looked longer for two-object outcomes compared to one-object outcomes. This would mean that in these cases infants expect a (contextually) unique object. This pattern, while yet to be observed in individuation studies, is something that could fit well into the linguistics and philosophical literature on discourse referents. In different research traditions it is argued that pointing and natural language demonstratives pick out a unique referent (Kaplan, 1989). If in our experiment the infants expected that pointing implied a contextually unique referent, surprise at the two-object outcome in the one-communicator/one-location condition was justified. This explanation again fails to elegantly account for the two-communicators/two-location condition.

We provided two possible explanations to account for the collected data. Based on infants’ baseline biases, our data can imply that they successfully individuated *iff* a single individuation cue was present, or that they expected a unique individual *iff* neither or both individuation cues were present. It is possible that both of these statements are true. Infants might have expected a unique individual in pattern (a) conditions, and expected two objects in pattern (b) conditions. It is unlikely that neither of these statements are true as infants numerical expectations changed depending on the presented cues. But, crucially, under either of these descriptions we lack an explanation for infants’ behavior in the two-communicators/two-locations condition.

To probe which of our two interpretations of the data is more likely to be correct, we ran a baseline experiment with the goal of removing all referential cues from the test trials. This could help us decide whether pattern (a) or pattern (b) looking time deviates from the baseline.

5.3 Experiment 2 — Baseline

This experiment intended to capture infants' looking behavior to the displays that were presented in the main experiment, but without providing cues of referential communication. The study used the same procedure and materials as Experiment 1, with some changes in design and stimuli.

5.3.1 Methods

5.3.1.1 Participants

Altogether 24 infants participated a study. Their mean age at testing was 12 months and 2 days (SD = 8 days). A further 5 infants were excluded from the sample (4 for fussiness, 1 for parental interference).

5.3.1.2 Design

We used a 2x2 within-subject design as no communicators were present in the baseline test trials. The two within-subject factors were outcome (1 or 2 objects revealed) and number of occluders (1 or 2 occluders present). In the two occluder/one object condition we counterbalanced the occluder that hid the object. We used the same object pairs as in the main experiment, revealing the same objects in the same counterbalancing orders to achieve as high perceptual similarity with the outcomes as possible.

By not having the communicators present at the test trials, we did not only remove referential communication from the design, but communication altogether. To address this issue, we inserted one of our “introductory trial” type videos between consecutive test trials, flanked by the usual attention grabbers before and after.

5.3.1.3 Introductory trials

The introductory trials remained the same as in the main experiment, with the exception that no occluders were present anymore.

5.3.1.4 Familiarization trials

We changed the familiarization trials because we used the number of occluders as a within-subject manipulation. We increased the number of familiarization trials to four, so that we could present infants with all four possible states: one occluder/one object; one occluder/two objects; two occluders/one object; two occluders/two objects. In order to achieve this, we also shortened these trials to 6.5 seconds each. The trials started with the occluder(s) present for 1 second. Then the occluder fell revealing the object(s) (2 s). The occluder(s) then were raised to the starting position hiding the object(s) for an additional 1.5 seconds. The object(s) was then revealed till the end of the trial (2 s). The quick pace of these trials ensured that infants keep attending through the four familiarization trials.

5.3.1.5 Test trials

These trials were almost identical to the last seconds of the test trials in Experiment 1. The curtain was lowered at the already at the first frame, and after 2 seconds delay, the occluder(s) left the screen, revealing the outcome. We added a sound effect that was used in the test trials of the main experiment (when the curtain fell) to the beginning of the trial. This way in the test trial there was no referential communication (or even agents) and we could measure infants’ baseline looking times to the same outcomes as in Experiment 1.

5.3.2 Results and Discussion

We analyzed the log-transformed looking times with a 2x2 repeated measures ANOVA with outcome and number of locations as independent variables. We found no interaction and no main effects (Figure 5.4). This shows that infants in this experiment did not have a strong bias to look longer at two objects compared to one. Obviously, our baseline experiment differed in a variety of ways from the Experiment 1. Trials were considerably shorter as no demonstrations were included. No agents were present in the scene at all. A single infant witnessed both one- and two-location outcomes. Though even as infants usually look at two objects longer in the baseline of individuation studies, it seems like that our stimuli did not induce such difference.

This result suggests that in Experiment 1 the changes from baseline expectations were induced by the one communicator/one location and the two communicators/two locations conditions. In particular, infants in these conditions were prone to look longer for two objects rather than one, implying an expectation of a unique object

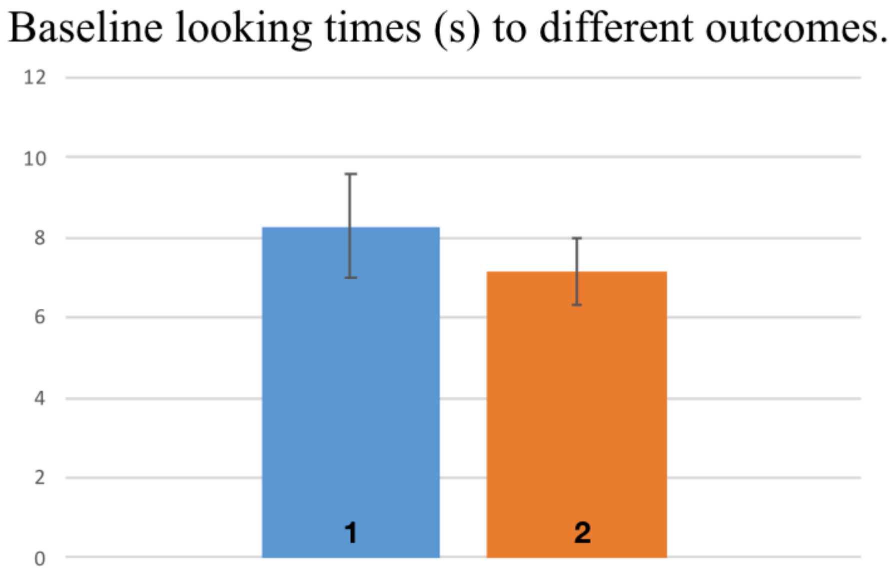


Figure 5.4. Baseline looking times (s) to one object (1) and two objects (2) outcomes. Error bars represent the standard error of the mean.

5.4 General discussion

We started this chapter by asking how infants' numerical expectations are modulated by different cues in ostensive-referential situations. We tested two possibly individuating information types: location and discourse, the latter one operationalized by separate communicative agents. We argued in previous chapters that location information, or more generally spatiotemporal properties are one of the earliest emerging individuation cues, probably processed in the visual system, but in the current study we only provided communicative information to encode such information. Empirical and theoretical considerations that unify a range of phenomenon in the infant literature on object individuation and referential communication led us to hypothesize about a further, communicator-based object individuation process (Chapter 4).

We contrasted these two types of information in a paradigm that manipulated the number of communicators and the number of spatially separate locations. We put forward three hypotheses: object individuation based on location (H1), on the number of communicators (H2), or on either one (H3). We found that infants were sensitive to both spatiotemporal and communicator based cues, but not in the way predicted by the third hypothesis. After claiming that a low-level account would fail to explain these results, we offer three possible explanations that are compatible with the notion of discourse-bound indexing. The first two maintains that infants can integrate indices from disjoint contexts, but they differ in how they interpret infants' looking time pattern. One claims that infants interpreted some communicative acts to build expectations of multiple objects, while the other claims that in some conditions infants built expectations of a unique individual. Finally we offer a third explanation that posits that infants fail to integrate indices from disjoint discourse contexts.

5.4.1 Low-level explanations

Notice that there seems to be no straightforward explanation that could account for all the results without appealing to discourse-relevant phenomena. There was no single perceptual feature present (e.g., the number of occluders, the number of communicators, directionality of pointing acts) that were uniquely predictive of infants' looking time pattern. We also do not have any reason to assume that infants employed visual-index based representations during familiarization. Before revealing the outcome, no objects were visible, and visual-indices require objects to track. On a richer interpretation of the concept of location-based indices (e.g., object files), some top-down mechanisms might create object representations inferentially that are still spatiotemporally indexed (cf. Scholl & Leslie, 1999; Xu, 2005; Carey, 2009). Still, infants did not always expect two objects when referential communication was directed to two locations, and object files by definition should always individuate based on spatiotemporal cues. There is perhaps a further version of object file theory, in which spatiotemporal continuity violations do not trigger individuation. But it is unclear whether that version of the theory could keep all the explanatory power over the phenomenon that it explains very well: the variety of tasks where infants do use the location of visually available objects for individuation.

5.4.2 Explanation 1: Individuating objects between discourses

A possible interpretation of our data is that infants build expectations of two objects based on either, but not both, types of information. This interpretation fits well with earlier individuation studies methodologically, and it is mostly consistent with the idea that infants can integrate information from different discourse contexts. A change from the one-communicator /one-location condition to the one-communicator/two-location condition could imply that infants take into account the location of the object within a discourse context: if there is a novel referential act to a new location, a novel discourse referent is created. The two-communicator/one-location condition could be taken as evidence that, in case of separate discourses, individuation happens in each discourse context independently, but the two pieces of information are integrated. This would imply that, solely based on the the distinctness of the communicators, infants build expectations of two objects, even if the spatiotemporal evidence is insufficient for such a conclusion. The only result that would require further assessment is the two-communicator/two-location condition where the infants in Experiment 1 did not expect two objects, even though both cues were present.

This failure might be due to a limit in working memory capacity. The infants in the two-communicator/two-location condition of Experiment 1 might have tried to remember more than the 3 items in their working memory, which is the working memory limit of this age. One can argue that even if the infants had expected two objects, they might have been forced to keep four items in working memory: two agents and two objects. This is especially likely as, at some point of the trial, both agents and objects were simultaneously occluded. It is well established that at this age infants are prone to a catastrophic memory failure when they are trying to remember more than three items (Feigenson & Carey, 2005). While they do not completely forget about the existence of the items, they are not individually maintained anymore. The breakdown of performance might be attributed to a breakdown of working memory. This interpretation is testable by re-running the study without the curtain hiding the agents during test trials.

5.4.3 Explanation 2: Expecting a unique object

However, the baseline experiment (Experiment 2) points to an interpretation of the data where infants' expectations were driven by other factors. It seems that infants expected a unique object if there was a single communicator and a single location, and if there were two communicators and two locations, while lacked this expectation in the other two cases. While we did not predict this outcome, uniqueness can be conceptualized well in a theory of communication. For example the idea that something is unique can be asserted in natural language: (eg. "there is only one ball behind the occluder") and furthermore sometimes argued to be contentful part of definite noun phrases (Russel, 1905; Strawson, 1950). While it is unclear whether pointing acts can have a definite interpretation in infancy, later in life this interpretation is available, possibly even the default. Some natural languages like Russian and Polish do not necessitate the marking the definite/indefinite nature of a noun phrase, and it has been argued that in these ambiguous or unmarked cases speakers of these languages default to the definite interpretation (Dayal, 2004;

Simik & Demian, 2020). Also in sign languages pointing can be used to signal definite anaphoric reference, when signers point back to the same location where they introduced a novel discourse referent. Thus a pointing gesture might lose its discourse independent referentiality, but still maintain an anaphoric (definite) discourse referential reading (Schlenker, 2011). Altogether this explanation is theoretically more demanding, but has some support from the baseline experiment. Still, just as before it fails to account for infants' behavior in the two-communicators / two-locations condition.

5.4.4 Explanation 3: Discourse-bound indices are not integrated

Instead of the above explanations, we can try to describe the pattern of results from a solely discourse referential viewpoint, focusing on the assumption that index-creating communicative acts have a novelty requirement, and anaphoric communicative acts have a familiarity requirement (reviewed in Chapter 4)[11]. If we take the results of the two-communicator/two-location condition at face value, we would conclude that infants are unable to integrate indices *across* discourse contexts. Theoretically this is easy to accommodate: any represented index has to be local to the specific discourse, as discourse-bound indexing is defined as indexing *within* the common ground of the given discourse. The two different communicative acts displayed by the two communicators could be construed as distinct from each other and thus novel from the infants' first-person perspective. But if instead we analyze it from the perspective of either of the two discourses that infants supposedly maintained, in neither of the two contexts two novel index-creating action was presented. In each discourse a single action was repeated multiple times. Putting the pieces together, we can argue that because there was only a single (novel) communicative act *within* each discourse, there was no reason to create two indices in *either*. This would mean that infants can not integrate discourse referents from different contexts. We

[11] Notice that for this interpretation we assume that the pattern was due to infants' expectations of two referents in the corresponding condition, rather than expectations of a unique referent in the other conditions. One benefit of this explanation is that the conclusions are not dependent on which of the two stances we take on the data. One could rephrase the same argument by appealing to the fact that a familiarity requirement for anaphoric reference to a unique entity was only met in in the one-communicator/one-location and the two-communicator/two-location conditions.

can easily extend this analysis to both of the one-communicator conditions. Infants in the one-communicator/two-location condition successfully interpreted the second pointing act, directed at a *novel* location, to establish the existence of the *novel* index. In contrast in the one-communicator/one-location condition the repeated pointing actions to the *familiar* location were not novel actions, and thus these did not result in novel indices. Finally, how did infants establish the presence of a two indices in the two-communicators/one location condition? According to this analysis, infants should have not expected two objects in this condition because different referential acts were presented *between* discourse contexts rather than *within* a single discourse. Still, they did.

At this point we are forced to say that the successful index creation can only happen *within* a discourse. So to account for infants' expectation of two objects in the two-commincators/one-location condition, we are required to make an auxiliary assumption that infants established a single discourse context rather than two. This would mean, that our manipulation in this condition failed to succeed in creating separate discourses. But if we grant this assumption, we could then argue that the novelty requirement was met when the second communicator engaged in a novel index-creating communicative act *within* the shared discourse context, one that included both communicators. There are arguments in favor of this auxiliary assumption. In this condition (as opposed to the two-communicator/two-location condition) the communicators shared gaze direction as they were looking and pointing to the same location. This is cue that infants take into account in implicit ToM tasks (e.g. Onishi & Baillargeon, 2005). As beliefs are the main currency of establishing a common ground, having overlapping visual perspectives or beliefs might be highly relevant in establishing discourse boundaries. More generally, it is likely that establishing a discourse boundary is dependent on how third parties communicate with each other, and also on their attentional states (cf. Vouloumanos, Onishi, & Pogue, 2012; Vouloumanos, Martin & Onishi, 2014; Krehm, Onishi & Vouloumanos, 2014).

This interpretation is also appealing as it is directly testable. We can derive the clear prediction that if infants construed of a single discourse in a scenario with two-communicators and two-locations, they would expect two objects. This could be operationalized by the two communicators talking to each other, and attending to the others' gaze and pointing direction. In the same vein, if we could ensure that infants construed the two-communicator/one-location situation as two separate discourses, they should expect a single object. This is a harder to operationalize because it is unclear what cues infants use to establish separate discourses instead of a single discourse. One possibility is that if the two experimenters repeatedly swapped locations, only one of them being present for each communicative act, that would be sufficient to achieve this.

5.4.5 Conclusions

To sum up, the solely discourse referential description of the data fits well with the theoretical notion of discourse based indexing system, it is not inconsistent with available data, and provides novel predictions. Crucially, it assumes that our manipulation of eliciting multiple discourses failed in one condition. But establishing whether communicative agents belonged to the same discourse required further inferences about the relations between third parties and the context. The processes that might be responsible in establishing who is participating in a specific discourse context might be part of the complex and holistic problems of belief fixation (Fodor, 2008) rather than the possibly encapsulated problems of indexing entities.

Infants have a sophisticated understanding of communication. Our data in conjunction with the literature described in Chapter 4 demonstrate that interpretation of communicative reference go beyond simple mappings to and visual indices.

Chapter 6 – Discussion

This dissertation aimed to examine how infants represent objects. More specifically, we were interested in how objects get tokenized *as particulars*. A key starting assumption was that representing objects as particulars requires a specific computational apparatus: an indexing system. While it is well-established that infants have an indexing system that utilizes visual evidence of objects, we argued throughout that there are empirical and theoretical reasons to believe that they might have further indexing systems. This hypothesis set the stage for our empirical studies. We began by probing for the potential interaction of distinct indexing systems in two individuation studies (Chapter 2 & 3). Having obtained partial support for our hypothesis, Chapter 4 proposed a novel theory of a second indexing system, linking it to communicative understanding. Chapter 5 was aimed to test predictions derived from this theory. In this chapter, we will summarize the main theoretical and empirical consequences of the findings from previous chapters, discuss an important limitation of the present investigation, and provide a set of open questions for further research.

6.1 Representing particulars in vision and in communication

We started out in Chapter 1 by following Pylyshyn's insight (2006) that indexing entities is a precondition for entertaining structured representations of particulars. What this means is that in order to see, think, or talk about a particular entity, some internal mechanism has to represent it separately from all other entities. This mechanism, a system of indexing, can act as the basis for a variety of functions, including recognizing and creating representations of particulars, and providing an address to those particulars to which information can be bound. In order to fulfill these functions, the representational system has to be constrained in a way that the resulting indices are comparable and uniquely identifiable. We argued that to meet these requirements within a single indexing system, only indices of the same type and format should coexist.

Throughout, our main interest has been in the nature and architecture of this indexing process. Some well-known explanatory frameworks for how representations of a given type can be individuated are the visual-indexing theory (Pylyshyn, 2003), the object-file theory (Kahnemann et al., 1992), and the object-indexing theory (Scholl & Leslie, 1989). Though the technical details and implementations diverge, a core shared feature of these theories is that they all identify the spatiotemporal nature of the input as the main information source relevant for indexing. In general, they have explanatory power over a wide range of phenomena in infants' behavior where perceptual and spatiotemporal information is available (Leslie et al., 1998; Scholl & Leslie, 1999; Carey, 2009). At the same time, we also reviewed a set of phenomena where visual-index based representations do not seem to be sufficient to explain infants' behavior: infants seem to be able to represent objects even in cases where they lose the visual index. This is what led to our hypothesis regarding the existence of further indexing systems beyond visual-indexing. In particular, a variety of results show that in communicative contexts, infants seemingly disregard spatiotemporal information. We proposed that in these contexts, objects are indexed independently, and not in perceptual systems.

Studies 1 & 2 tried to empirically assess this idea. In two individuation studies, we replicated previous results showing that infants can use spatiotemporal cues to posit the existence of multiple objects. In both studies, we ran further conditions where the objects were labeled in referential-communicative contexts. In these conditions, infants' expectations radically changed, as they seemingly did not use the same spatiotemporal cues to build expectations of the number of objects. Importantly, this performance did not improve when the labeling events included different nouns. This was markedly different from previous findings (Xu, 2002, Xu 2004). We interpreted these results as supporting the general idea of multiple individuation systems, and as suggestive evidence against the notion that sortal concepts necessarily mediate the process of individuation in communicative contexts.

What then mediates indexing in these communicative contexts? In Chapter 4 we took the adult communicative system as a starting point. Natural language makes available ways of discussing entities that are not dependent on perceptual evidence. We can talk about entities that exist only in the past or the future, or in some possible but not actual state of affairs, just to name a few of the relevant phenomena. Understanding and producing these communicative acts require a system where the relevant entities are represented and indexed. Similarly to visual indices, the system has to be able to fulfill a variety of functions, like recognition of already represented entities, creating novel representations, binding properties to the representations and reasoning about the entities. In natural language semantics, research on discourse representation explains these capacities by appealing to a discourse-bound system of indexing (Kamp. 1981; Heim, 1982, 1983).

Returning to infancy, we tried to assess whether this system of indexing could be applicable early in development. One-year-old infants have a sophisticated understanding of communicative acts (Tomasello, 2009), and we argued in Chapter 4 that there are variety of theoretical and empirical reasons to assume that infants possess the necessary preconditions to entertain a discourse-based indexing system. Infants recognize communicative partners, bind communicative acts to specific discourse participants, and establish/approximate a common ground in these discourse contexts. We also argued that the available evidence implies that communicative acts are sufficient to induce an expectation of an object, even when sortal or spatiotemporal information is not available. We took this to show that referential communication has an index creating function for infants. As a starting point on how indexing works in this system, we invoked File Change Semantics (Heim, 1982), a theory that uses notions of familiarity and novelty in order to distinguish between index-creating versus anaphoric communicative acts.

In Chapter 5 we experimentally probed the idea that infants are equipped with such a discourse-based indexing system. To do so, we tested whether infants can build referential expectations of multiple objects from multiple discourse contexts. We manipulated the number of distinct communicators (as a means to operationalize different discourse contexts) and the number of locations that referential communication was directed at. The pattern of results that we obtained was complex and did not neatly fit any of our hypotheses. Our best interpretation of the findings is that infants failed to integrate indices from two disjoint communicative contexts. This underscores the idea that indices are bound to the discourse: if two communicators independently referred to distinct spatial locations, infants did not expect multiple objects. On the other hand, we also found evidence that within a discourse context, different sources of information like location, or the identity of speaker can modify infants' numerical expectations. As an important caveat, note that this interpretation relies on a post-hoc auxiliary assumption. We needed to posit that the shared attention of two communicators to a single location can result in infants encoding them as part of a unitary discourse.

6.2 Index creation in the discourse-bound system

Altogether the reviewed empirical evidence and the studies we conducted support the idea that infants are equipped with multiple systems that can represent objects as particulars. More specifically, our proposal was that beyond the visual-indexing system, infants, like adults, entertain a discourse-bound system of indexing. Here we discuss some properties of that system.

As we discussed before, indexing in a discourse is solely dependent on the interpretation of communicative acts. As a direct result of this, the perceptible properties of objects are not deterministically responsible for the creation of novel indices. This makes the resulting system capable of representing particulars without any perceptual access to them and even allows for the representation of ones that do not exist in the actual world (e.g., in counterfactual statements). On the flipside, the resulting system does not have any primary properties that could be defined extensionally. The discourse referent of "the cake" in the sentence "The cake that Mary never baked" does not exist in the actual world, and as such has no actual properties that could be used to separate it from the discourse referent of "the dog" in the sentence "The dog that Peter never had". In the discourse, they still are indexed separately, allowing us to selectively attribute information to them (for instance one could believe that the dog would bark (if it existed) but not the cake).

If there are no extensionally defined primary properties, that leaves open the question of what information infants use to separate distinct indices. What is the minimal differentiating property between any two represented entities? Given there are no extensional properties that infants could rely on, the simplest solution is to assert that indices are solely separated by *them being separate indices*. Implementing this system using symbolic representations, we would require a list of symbols that can be in one-to-one correspondence with the discourse referents. Differences amongst the symbols would suffice as a basis for minimal differentiation between tokened entities. This crucially does not help to elucidate the nature of the information that compels the system to create indices in the first place. In natural language, the way the discourse referents are invoked can be marked, for example, with the definite / indefinite articles. These words can help the listener distinguish whether an entity under discussion is novel or familiar in the discourse, which in turn result determines whether or not a new index is created (Heim, 1982, 2011).

There is no straightforward way to apply the same reasoning to infancy, where it is less clear how non-verbal communicative acts could provide deterministic cues that mark the necessity of index creation. It seems likely that the specific interpretations that infants make are context dependent. For instance, if in some context infants cannot interpret a communicative act as referring to an already represented discourse referent, they create a new index. This is not particularly insightful. If there are no external properties that always distinguish entities that belong under indices *and* there are no communicative acts that are *necessarily* index creating, the claim that infants create a new index when they interpret a communicative act as introducing a novel entity, just puts all the exploratory burden on the process of interpretation.

While we are not sure on how to make progress on the theoretical background of interpretative processes behind index creation in infancy, there is room for exploratory empirical research to probe some characteristics of that system. In fact, the studies we presented can be analyzed in this framework. In studies 1 & 2 we found that the 10-month-old infants likely did not interpret the communicative acts (pointing and labeling to two visible objects) as two separate index creating actions, even in conditions where two different labels were used. Under this interpretation neither the kind information (indicated by the label) nor the direction of the pointing (pointing to different spatial locations) compelled infants to treat these actions as employing different indices. In contrast, we know from studies by Xu (2002) that 9-month-old infants can interpret two referential-communicate acts as referring to two objects if the two acts involve different labels. And the one-communicator/two-location condition of Study 3 can be described as evidence that 12-month-old infants can interpret two pointing actions directed at different locations, as referring to two distinct entities. Taken together, it seems that under some conditions infants use either location and kind information to interpret referential acts as novel and index-creating, while in other conditions they do not; an analysis that again provides limited insight, other than showing that these cues are not deterministic, but subject to contextual interpretation.

The two studies where we found that infants did not expect multiple objects (Studies 1 & 2) differ from studies where they do. In these two studies infants had multiple sources of information at their disposal, as spatiotemporal information for the visual-indexing system was also available to individuate the objects. Is it possible that this visual evidence for two objects interfered with the discourse-based indexing of two entities? While we have no way to directly address this question without further experimentation, it highlights some underlying assumptions behind the present work that we have yet to explicate.

6.3 Interface between indexing systems

We have argued throughout the dissertation that, to account for infants' behavior in a set of studies, the visual-indexing system (as understood by the literature) is insufficient, and that it is better described in a framework for discourse-based indexing. But these studies were not directly measuring the number of discourse-based indices that infants employed. For obvious reasons, what we measured in all experiments, were responses to outcomes that presented information to the visual-indexing system: spatiotemporally distinct objects. Thus, in order to connect our measurements (e.g., looking times to different numbers of spatiotemporally distinct objects) to the theoretical construct of discourse-bound indices, we need further assumptions of how discourse referents are translated into expectations of perceptually available, spatiotemporally distinct entities.

This is not only a methodological problem but a general one. Note that if we accept that the mechanisms for indexing objects/entities in perception and in discourse are largely independent, we need a further theory to account for the indisputable fact that quite frequently we perceive the objects that we talk about. This means that we need some specification of the interface between the systems. How is between-systems coreference established? And what does establishing coreference computationally entail? These problems are far from trivial, as the relationship between visual indices and discourse referents in adulthood is not bijective.

Visual indices operate on singular objects, while the domain of entities that discourse referents can take (at least for adults) is infinite and only constrained by our conceptual repertoire. Thus, the simplest and most desirable solution of positing that "every visual index has to be mapped to a single discourse referent and the other way around" is hopeless as a general purpose solution. Even if we only consider the overlapping domain of physical objects, a desirable bijective relationship does not hold. Discourse referents can pick out pluralities ("the ducks"), thus being mapped by multiple visual indices (surjective mapping). The displacement properties of communication also allow discourse referents not to be mapped by visual entities (like "Mom" when she is not around) resulting in injective mappings. Even more profoundly, the different discourse referents can be mapped by the same visual index ("a ball", "a round object", "a toy") resulting in non-injective, non-surjective mappings. The fact that there could be no simple mapping functions to establish correspondence highlights the complexity of this interface problem. The required systemacity cannot be achieved without a rich archive of background knowledge and contextual understanding. Mapping a visual object that has the perceptual features of a ball to a discourse referent like “the toy” prima facie requires the knowledge that balls are toys, implying that this process again is inferential.

Taken together, there are no simple solutions for establishing correspondence between entities that visually indexed and entities that are indexed in the discourse. But until further progress is made on uncovering the computational processes that drive communicative interpretation in infancy, we still need a set of some assumptions on how these systems exchange information in order to make progress. We propose that this starting assumption can be characterized as a bias along the following lines: *if not otherwise specified* a discourse referent is mapped by a visual index. We can interpret some of the results we obtained in the view of this bias. Take the one-communicator/one-location condition of Study 3. Infants looked longer at two-object outcomes compared to a single object, although in the familiarization no information specified whether the discourse referent picks out a plurality of spatiotemporally separate objects or not. Similarly, in the study by Xu (2002), where infants built expectations of two spatiotemporally distinct objects in response to utterances that included different nouns, positing this bias of mapping discourse referents to spatiotemporally separate entities is a necessary assumption in the current framework.

The application of this bias would imply that infants try to establish an injective relationship between the two systems at hand. While they might be biased to map all discourse referents to distinct visual indices, they are not biased to map distinct visual indices to distinct discourse-referents. Maintaining a visual index does not compel infants to create a corresponding discourse representation, as shown by the studies in Chapter 2 & 3 where, in the labeling conditions, infants witnessed multiple visual objects but it did not generate expectations of multiple objects.

On the other hand, we have good reasons to believe that the ceteris paribus clause we included when characterizing the bias is necessary. Infants can entertain discourse referents where the injective mapping to spatiotemporally separate entities does not hold. In a recent study we found that 12-month-old infants can successfully learn that a complex noun phrase (including a head noun and a modifier/determiner) can refer to a plurality of objects rather than a single object (Pomiechowska, Brody, Teglas, & Kovacs, 2018). This again shows that positing an injective bias carries no explanatory power: It simply redescribes the way infants behaved in most previous studies, and provides an assumption that empirical research can build on.

6.4 Conclusions

We explored how infants around one year of life create and maintain representations of objects. We argued for a cognitive architecture where multiple indexing systems fulfill these functions with different internal structures. We provided converging evidence to the well-established idea that infants have a system that uses perceptual evidence for object indexing, where the creation of an object representation is based on the spatiotemporal information. We found that in communicative contexts, where the objects were labelled, the same visual information did not seem to guide infants' expectations. Building on research in natural language semantics we proposed a novel system for indexing, one that represents entities in relation to a specific discourse. This dissertation leaves a variety questions open. In particular, we are in dire need of a predictive theory that could describe the interpretative system that infants use for encoding a referential communicative act as index-creating or anaphoric. Another open question is related to the interface between distinct systems of indexing, and how the mind establishes coreference between entities that are tokened in multiple systems. On the other hand, our notion of discourse-bound indexing can not only increase our understanding of how objects are represented in infancy, but it can also provide a new framework to investigate fundamental issues in development. Characterizing a system that can represent entities that are not perceptually available is an important step towards understanding the displacement property of human thought and language.

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological science*, 15(2), 106-111.

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states?. *Psychological review*, 116(4), 953.

Bahrami, B. (2003). Object property encoding and change blindness in multiple object tracking. *Visual cognition*, 10(8), 949-963.

Begus, K., & Southgate, V. (2012). Infant pointing serves an interrogative function. *Developmental science*, 15(5), 611-617.

Behne, T., Carpenter, M., & Tomasello, M. (2005). One_year_olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental science*, 8(6), 492-499.

Behne, T., Liszkowski, U., Carpenter, M., & Tomasello, M. (2012). Twelve_month_olds' comprehension and production of pointing. *British Journal of Developmental Psychology*, 30(3), 359-375.

Blok, S. V., Newman, G. E., & Rips, L. J. (2007). Out of sorts? Some remedies for theories of object concepts: A reply to Rhemtulla and Xu (2007).

Bloom, P. (2005). *Descartes' baby: How the science of child development explains what makes us human*. Random House.

Bohn, M., & Frank, M. C. (2019). The pervasive role of pragmatics in early language. *Annual Review of Developmental Psychology*, 1, 223-249.

Bonatti, L., Frot, E., Zangl, R., & Mehler, J. (2002). The human first hypothesis: Identification of conspecifics and individuation of objects in the young infant. *Cognitive psychology*, 44(4), 388-426.

Buresh, J. S., & Woodward, A. L. (2007). Infants track action goals within and across agents. *Cognition*, 104(2), 287-314.

Burkell, J. A., & Pylyshyn, Z. W. (1997). Searching through subsets: A test of the visual indexing hypothesis. *Spatial Vision*, 11(2), 225.

Butler, L. P., & Markman, E. M. (2012). Preschoolers use intentional and pedagogical cues to guide inductive inferences and exploration. *Child development*, 83(4), 1416-1428.

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

Carlson, G. N. (1980). *Reference to kinds in English*. New York and London.

Cheries, E., Feigenson, L., Scholl, B. J., & Carey, S. (2005). Cues to object persistence in infancy: Tracking objects through occlusion vs. implosion. Poster presented at the annual meeting of the Vision Sciences Society, 5/7/05, Sarasota, FL. [Abstract published in Journal of Vision, 5(8), 352a

Chiang, W. C., & Wynn, K. (2000). Infants' tracking of objects and collections. *Cognition*, 77(3), 169-195.

Chierchia, G. (1998). Reference to kinds across language. *Natural language semantics*, 6(4), 339-405.

Choi, Y., Mou, Y., & Luo, Y. (2018). How do 3-month-old infants attribute preferences to a human agent? *The Journal of Experimental Child Psychology*, 172, 96-106.

Choi, Y., Song, H., & Luo, Y. (2018). Infants' understanding of the definite/indefinite article in a third-party communicative situation. *Cognition*, 175, 69-76.

Chomsky, N. (2018). Two notions of modularity. inOn Concepts, Modules, and Language: Cognitive Science at its Core, 25-40.

Condry, K., & Yonas, A. (2013). Six-month-old infants use motion parallax to direct reaching in depth. *Infant Behavior and Development*, 36(2), 238-244.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant_directed speech in the first month after birth. *Child development*, 61(5), 1584-1595.

Csibra, G. (2008). Goal attribution to inanimate agents by 6.5-month-old infants. *Cognition*, 107(2), 705-717.

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language*, 25(2), 141-168.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in cognitive sciences*, 13(4), 148-153.

Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1567), 1149-1157.

Csibra, G., Hernik, M., Mascaro, O., Tatone, D., & Lengyel, M. (2016). Statistical treatment of looking-time data. *Developmental psychology*, 52(4), 521.

Csibra, G., & Shamsudheen, R. (2015). Nonverbal generics: Human infants interpret objects as symbols of object kinds. *Annual review of psychology*, 66, 689-710.

Csibra, G., & Volein, A. (2008). Infants can infer the presence of hidden objects from referential gaze information. *British Journal of Developmental Psychology*, 26(1), 1-11.

Csibra, G., Hernik, M., Mascaro, O., Tatone, D., & Lengyel, M. (2016). Statistical treatment of looking-time data. *Developmental psychology*, 52(4), 521.

Dayal, V. (2004). Number marking and (in) definiteness in kind terms. *Linguistics and philosophy*, 27(4), 393-450.

Dewar, K., & Xu, F. (2007). Do 9-month-old infants expect distinct words to refer to kinds?. *Developmental psychology*, 43(5), 1227.

Dewar, K., & Xu, F. (2009). Do early nouns refer to kinds or distinct shapes? Evidence from 10-month-old infants. *Psychological Science*, 20(2), 252-257.

Diamond, A., Cruttenden, L., & Neiderman, D. (1994). AB with multiple wells: I. Why are multiple wells sometimes easier than two wells? II. Memory or memory+ inhibition?. *Developmental Psychology*, 30(2), 192.

Dörrenberg, S., Rakoczy, H., & Liszkowski, U. (2018). How (not) to measure infant Theory of Mind: Testing the replicability and validity of four non-verbal measures. *Cognitive Development*, 46, 12-30.

Egyed, K., Király, I., & Gergely, G. (2013). Communicating shared knowledge in infancy. *Psychological science*, 24(7), 1348-1353.

Farroni, T., Johnson, M. H., & Csibra, G. (2004). Mechanisms of eye gaze perception during infancy. *Journal of cognitive neuroscience*, 16(8), 1320-1326.

Farroni, T., Johnson, M. H., Menon, E., Zulian, L., Faraguna, D., & Csibra, G. (2005). Newborns' preference for face-relevant stimuli: Effects of contrast polarity. *Proceedings of the National Academy of Sciences*, 102(47), 17245-17250.

Feigenson, L., & Carey, S. (2003). Tracking individuals via object_files: evidence from infants' manual search. *Developmental Science*, 6(5), 568-584.

Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition*, 97(3), 295-313.

Feigenson, L., & Halberda, J. (2004). Infants chunk object arrays into sets of individuals. *Cognition*, 91(2), 173-190.

Feigenson, L., & Halberda, J. (2008). Conceptual knowledge increases infants' memory capacity. *Proceedings of the National Academy of Sciences*, 105(29), 9926-9930.

Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. *Psychological Science*, 13(2), 150-156.

Flombaum, J. I., Scholl, B. J., & Santos, L. R. (2009). Spatiotemporal priority as a fundamental principle of object persistence. In B. Hood & L. Santos (Eds.), *The origins of object knowledge* (pp. 135-164). Oxford University Press.

Fodor, J. A. (1983). *The modularity of mind.* MIT press.

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Oxford University Press.

Fodor, J. A. (2006). Who ate the salted peanuts? *London Review of Books*, 28, 18.

Fodor, J. A. (2008). *LOT 2: The language of thought revisited*. Oxford University Press

Franconeri, S. L., Alvarez, G. A., & Cavanagh, P. (2013). Flexible cognitive resources: competitive content maps for attention and memory. *Trends in cognitive sciences*, 17(3), 134-141.

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.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford University Press, USA.

Gelman, S. A. (2004). Psychological essentialism in children. *Trends in cognitive sciences*, 8(9), 404-409.

Gelman, S. A. (2009). Learning from others: Children's construction of concepts. *Annual review of psychology*, 60, 115-140.

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, 56(2), 165-193.

Gliga, T., & Csibra, G. (2009). One-year-old infants appreciate the referential nature of deictic gestures and words. *Psychological Science*, 20(3), 347-353.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in cognitive sciences*, 20(11), 818-829.

Grice, H. P. (1989). *Studies in the Way of Words*. Harvard University Press.

Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive psychology*, 53(4), 310-344.

Harris, P. L., & Lane, J. D. (2014). Infants understand how testimony works. *Topoi*, 33(2), 443-458.

Heim, I. (1982). The semantics of definite and indefinite noun phrases. *Doctoral Dissertations Available from Proquest*

Heim, I. (1983). File change semantics and the familiarity theory of definiteness. *Semantics Critical Concepts in Linguistics*, 108-135.

Heim, I. (2011). Definiteness and indefiniteness. *Unpublished Lecture Notes*

Hein, E., & Moore, C. M. (2012). Spatio-temporal priority revisited: The role of feature identity and similarity for object correspondence in apparent motion. *Journal of Experimental Psychology: Human Perception and Performance*, 38(4), 975.

Hernik, M., & Southgate, V. (2012). Nine_months_old infants do not need to know what the agent prefers in order to reason about its goals: On the role of preference and persistence in infants' goal_attribution. *Developmental science*, 15(5), 714-722.

Holcombe, A. O., & Chen, W. Y. (2013). Splitting attention reduces temporal resolution from 7 Hz for tracking one object to< 3 Hz when tracking three. *Journal of vision*, 13(1), 12-12.

Huntley-Fenner, G., Carey, S., & Solimando, A. (2002). Objects are individuals but stuff doesn't count: Perceived rigidity and cohesiveness influence infants' representations of small groups of discrete entities. *Cognition*, 85(3), 203-221.

Jacob, P. (2013). A puzzle about belief-ascription. Mind and society: Cognitive science meets the philosophy of the social sciences. Synthese Philosophy Library, Berlin: Springers.

Johnson, S. C. (2003). Detecting agents. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1431), 549-559.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive psychology*, 24(2), 175-219.

Káldy, Z., & Leslie, A. M. (2003). Identification of objects in 9_month_old infants: integrating 'what'and 'where'information. *Developmental Science*, 6(3), 360-373.

Káldy, Z., & Leslie, A. M. (2005). A memory span of one? Object identification in 6.5-month-old infants. *Cognition*, 97(2), 153-177.

Kamp, H. (1981). A theory of truth and semantic representation. Formal semantics-the essential readings, 189-222.

Kampis, D., Somogyi, E., Itakura, S., & Király, I. (2013). Do infants bind mental states to agents?. *Cognition*, 129(2), 232-240.

Karttunen, L. (1968). *What do referential indices refer to?* Rand Corporation.

Kaufman, J., Csibra, G., & Johnson, M. H. (2005). Oscillatory activity in the infant brain reflects object maintenance. *Proceedings of the National Academy of Sciences*, 102(42), 15271-15274.

Kibbe, M. M., & Feigenson, L. (2016). Infants use temporal regularities to chunk objects in memory. *Cognition*, 146, 251-263.

Kibbe, M. M., & Leslie, A. M. (2011). What do infants remember when they forget? Location and identity in 6-month-olds' memory for objects. *Psychological Science*, 22(12), 1500-1505.

Kovács, Á. M. (2016). Belief files in theory of mind reasoning. *Review of Philosophy and Psychology*, 7(2), 509-527.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, 330(6012), 1830-1834.

Kratzer, A., & Heim, I. (1998). *Semantics in generative grammar* (Vol. 1185). Oxford: Blackwell.

Krehm, M., Onishi, K. H., & Vouloumanos, A. (2014). I see your point: Infants under 12 months understand that pointing is communicative. *Journal of Cognition and Development*, 15(4), 527-538.

Kuhlmeier, V. A., Bloom, P., & Wynn, K. (2004). Do 5-month-old infants see humans as material objects?. *Cognition*, 94(1), 95-103.

Leslie, A. M. (1987). Pretense and representation: The origins of theory of mind. *Psychological Review*, 94(4), 412–426.

Leslie, A. M. (2000). How to Acquire a Representational Theory of Mind. *Metarepresentations: A Multidisciplinary Perspective*, (10), 197.

Leslie, A. M., & Chen, M. L. (2007). Individuation of pairs of objects in infancy. *Developmental Science*, 10(4), 423-430.

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.

Lewis, D. (1986). *On the plurality of worlds* (Vol. 322). Blackwell: Oxford.

Liszkowski, U. (2005). Human twelve-month-olds point cooperatively to share interest with and helpfully provide information for a communicative partner. *Gesture*, 5(1-2), 135-154.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657), 279-281.

Luo, Y. (2011). Three_month_old infants attribute goals to a non_human agent. *Developmental science*, 14(2), 453-460.

Luo, Y., & Baillargeon, R. (2005). Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychological Science*, 16(8), 601-608.

Moher, M., Tuerk, A. S., & Feigenson, L. (2012). Seven-month-old infants chunk items in memory. *Journal of experimental child psychology*, 112(4), 361-377.

Moll, H., Koring, C., Carpenter, M., & Tomasello, M. (2006). Infants Determine Others' Focus of Attention by Pragmatics and Exclusion. Journal of Cognition and Development, 7(3), 411-430.

Moll, H., & Tomasello, M. (2004). 12 and 18 month old infants follow gaze to spaces behind barriers. *Developmental science*, 7(1), F1-F9.

Moll, H., & Tomasello, M. (2006). Level 1 perspective_taking at 24 months of age. *British Journal of Developmental Psychology*, 24(3), 603-613.

Moll, H., & Tomasello, M. (2007). How 14-and 18-month-olds know what others have experienced. *Developmental psychology*, 43(2), 309.

Murray, L., & Trevarthen, C. (1986). The infant's role in mother–infant communications. *Journal of child language*, 13(1), 15-29.

Nishida, S. Y., & Takeuchi, T. (1990). The effects of luminance on affinity of apparent motion. *Vision research*, 30(5), 709-721.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs?. *Science*, 308(5719), 255-258.

Partee, B. H. (2008). *Compositionality in formal semantics: Selected papers*. John Wiley & Sons.

Pätzold, W., & Liszkowski, U. (2019). Pupillometry reveals communication_induced object expectations in 12_but not 8_month_old infants. *Developmental science,* 22(6), e12832.

Perner, J., & Leahy, B. (2016). Mental files in development: Dual naming, false belief, identity and intensionality. *Review of Philosophy and Psychology*, 7(2), 491-508.

Pomiechowska, B., Brody, G., Csibra, G., & Gliga T. (in prep.) Twelve-month-olds deploy logical inference to determine referents of new words

Pomiechowska, B., Brody, G., Teglas, E., & Kovacs, A.M. (2018) Do 12-month-olds use the principle of compositionality to interpret complex noun phrases? *Poster presented at the VIII. Budapest Conference on Cognitive Development*, Budapest, Hungary.

Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, 32(1), 65-97.

Pylyshyn, Z. (2004). Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. *Visual cognition*, 11(7), 801-822.

Pylyshyn, Z. W. (2000). Situating vision in the world. *Trends in cognitive sciences*, 4(5), 197-207.

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition*, 80(1-2), 127-158.

Pylyshyn, Z. W. (2003). *Seeing and visualizing: It's not what you think*. MIT press.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial vision,* 3(3), 179-197.

Rakoczy, H. (2017). In defense of a developmental dogma: Children acquire propositional attitude folk psychology around age 4. *Synthese*, 194(3), 689-707.

Recanati, F. (2012). *Mental files.* Oxford University Press.

Rips, L. J., Blok, S., & Newman, G. (2006). Tracing the identity of objects. *Psychological Review,* 113(1), 1.

Robinson, C. W., & Sloutsky, V. M. (2008). Effects of auditory input in individuation tasks. *Developmental Science*, 11(6), 869-881.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive psychology*, 8(3), 382-439.

Russel, B. (1905). *On Denoting*. Reprinted in Marsh, RC (ed.). 1956. Bertrand Russel. Logic and Knowledge. Essays 1901-1950, 39-56.

Saylor, M. M., & Ganea, P. (2007). Infants interpret ambiguous requests for absent objects. *Developmental Psychology,* 43(3), 696.

Saylor, M. M., Ganea, P. A., & Vázquez, M. D. (2011). What’s mine is mine: Twelve_month_olds use possessive pronouns to identify referents. *Developmental Science,* 14(4), 859-864.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, 80(1-2), 1-46.

Scholl, B. J., & Feigenson, L. (2004). When out of sight is out of mind: Perceiving object persistence through occlusion vs. implosion. *Talk given at the annual meeting of the Vision Sciences Society, 5/1/04, Sarasota, FL*. [Abstract published in Journal of Vision, 4(8), 26a

Scholl, B. J., & Leslie, A. M. (1999). Explaining the infant's object concept: Beyond the perception/cognition dichotomy. In E. Lepore & Z. Pylyshyn (Eds.), *What is cognitive science?* (pp. 26-73). Oxford: Blackwell.

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology,* 38, 259-290.

Scholl, B. J., Pylyshyn, Z. W., & Franconeri, S. L. (1999, March). When are featural and spatiotemporal properties encoded as a result of attentional allocation?. *Investigative Ophthalmology & Visual Science* (Vol. 40, No. 4, pp. S797-S797).

Scholl, B. J., & Xu, Y. (2001). The magical number 4 in vision. *Behavioral and Brain Sciences*, 24(1), 145-146.

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology,* 18(9), 668-671.

Shamsudheen, R. & Csibra, G. (2016). Communicative reference to familiar objects induces kind-based individuation in 9-month-old infants. *Poster presented at the Budapest CEU Conference on Cognitive Development, January 2016, Budapest, Hungary*.

Simik, R., & Demian, C. (2020) Definiteness, uniqueness, and maximality in languages with and without articles. *Journal of Semantics*

Smith, L. B., Thelen, E., Titzer, R., & McLin, D. (1999). Knowing in the context of acting: the task dynamics of the A-not-B error. *Psychological review*, 106(2), 235.

Southgate, V., Van Maanen, C., & Csibra, G. (2007). Infant pointing: Communication to cooperate or communication to learn?. *Child development*, 78(3), 735-740.

Spaepen, E., & Spelke, E. (2007). Will any doll do? 12-month-olds' reasoning about goal objects. *Cognitive Psychology*, 54(2), 133-154.

Spelke, E. S. (1990). Principles of object perception. *Cognitive science*, 14(1), 29-56.

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.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and Cognition*. Wiley-Blackwell.

Stalnaker, R. (1974). *Pragmatic presupposition, Semantics and Philosophy*. New York, 197-214.

Stalnaker, R. (2002). Common ground. *Linguistics and philosophy*, 25(5/6), 701-721.

Stavans, M., Lin, Y., Wu, D., & Baillargeon, R. (2019). Catastrophic individuation failures in infancy: A new model and predictions. *Psychological review*, 126(2), 196.

Strawson, P. F. (1950). On referring. *Mind*, 59(235), 320-344.

Surian, L., & Caldi, S. (2010). Infants' individuation of agents and inert objects. *Developmental Science,* 13(1), 143-150.

Tatone, D., Hernik, M., & Csibra, G. (2019). Minimal cues of possession transfer compel infants to ascribe the goal of giving. *Open Mind*, 3, 31-40.

Tauzin, T., & Gergely, G. (2019). Variability of signal sequences in turn-taking exchanges induces agency attribution in 10.5-mo-olds. *Proceedings of the National Academy of Sciences*, 116(31), 15441-15446.

Tomasello, M., & Haberl, K. (2003). Understanding attention: 12-and 18-month-olds know what is new for other persons. *Developmental psychology*, 39(5), 906.

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child development*, 78(3), 705-722.

Tomasello, Michael. *The cultural origins of human cognition*. Harvard University Press, 2009.

Topál, J., Gergely, G., Miklósi, Á., Erd_hegyi, Á., & Csibra, G. (2008). Infants' perseverative search errors are induced by pragmatic misinterpretation. *Science*, 321(5897), 1831-1834.

Treisman, A., Kahneman, D., & Burkell, J. (1983). Perceptual objects and the cost of filtering. *Perception & psychophysics*, 33(6), 527-532.

Trick, L. M., & Pylyshyn, Z. W. (1993). What enumeration studies can show us about spatial attention: evidence for limited capacity preattentive processing. *Journal of Experimental Psychology: Human Perception and Performance*, 19(2), 331.

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological review*, 101(1), 80.

Ullman, S. (1979). The interpretation of structure from motion. *Proceedings of the Royal Society of London. Series B. Biological Sciences,* 203(1153), 405-426.

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.

Vouloumanos, A., Martin, A., & Onishi, K. H. (2014). Do 6_month_olds understand that speech can communicate? *Developmental Science*, 17(6), 872-879.

Vouloumanos, A., Onishi, K. H., & Pogue, A. (2012). Twelve-month-old infants recognize that speech can communicate unobservable intentions. *Proceedings of the National Academy of Sciences*, 109(32), 12933-12937.

Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in cognitive sciences*, 13(6), 258-263.

Waxman, S. R., & Gelman, S. A. (2010). Different kinds of concepts and different kinds of words: What words do for human cognition. *The making of human concepts*, 101-130.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta_analysis of theory_of_mind development: The truth about false belief. *Child development*, 72(3), 655-684.

Wiggins, D. (1997). Sortal Concepts: A Reply To Xu. *Mind & Language*, 12(3−4), 413-421..

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1), 103-128.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, 69(1), 1-34.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, 358(6389), 749-750.

Xu, F. (2003) The effects of labeling on object individuation in 9-month-old infants. *Manuscript under review*.

Xu, F. (1997). From Lot's wife to a pillar of salt: Evidence that physical object is a sortal concept. *Mind & Language,* 12(3_4), 365-392.

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition*, 85(3), 223-250.

Xu, F. (2005). Categories, kinds, and object individuation in infancy. In *Building object categories in developmental time*(pp. 81-108). Psychology Press.

Xu, F. (2007). Sortal concepts, object individuation, and language. *Trends in Cognitive Sciences*, 11(9), 400-406.

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.

Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive psychology*, 30(2), 111-153.

Xu, F., Cote, M., & Baker, A. (2005). Labeling guides object individuation in 12-month-old infants. *Psychological Science,* 16(5), 372-377.

Yin, J., & Csibra, G. (2015). Concept-based word learning in human infants. *Psychological Science*, 26(8), 1316-1324.

Yoon, J. M., Johnson, M. H., & Csibra, G. (2008). Communication-induced memory biases in preverbal infants. *Proceedings of the National Academy of Sciences*, 105(36), 13690-13695.

ACKNOWLEDGEMENT

TO EXTERNAL FUNDING AGENCIES

CONTRIBUTING TO PHD DISSERTATIONS

Name of doctoral candidate: *Gábor Bródy*

Title of dissertation: **Indexing Objects in Vision and Communication**

Name of supervisor(s): Gergely Csibra, Ágnes Melinda Kovács

External funding agency: **European Research Council**

Acknowledgement:

This research has received funding from the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2013) under **grant agreement No 249519 – OSTREFCOM** and Horizon 2020 programme (E/R/H20/2014-2020) under **grant agreement No 742231 – PARTNERS.**