CrossMark

From Sensations to Concepts: a Proposal for Two Learning Processes

Peter Gärdenfors [1,2]

Abstract This article presents two learning processes in order to explain how children at an early age can transform a complex sensory input to concepts and categories. The first process constructs the perceptual structures that emerge in children's cognitive development by detecting invariants in the sensory input. The invariant structures involve a reduction in dimensionality of the sensory information. It is argued that this process generates the primary domains of space, objects and actions and that these domains can be represented as conceptual spaces. Once the primary domains have been established, the second process utilizes covariances between different dimensions of the domains in order to identify natural clusters of entities. The clusters are then are used to determine concepts as regions in the spaces. As an application, the processes are used to resolve the so-called 'complex first paradox' that emerges from the fact that children, in general, learn nouns earlier than adjectives, even though nouns are semantically more complex than adjectives.

1 The Blooming, Buzzing Confusion

Our sensory influx is extremely rich. For example, whenever we open our eyes, there is a constantly fluctuating wash of light captured on our retinas. It is something of a miracle that our brains manage to sort up the sensory information and immediately identify and categorize a vast number of entities in our surroundings. The miracle becomes even larger when it is considered that these categories must be *learned* from experience. The learning process is rapid, which is witnessed, among other things, by the fact that children start communicating about the categories after about a year. The following famous quote from William James' *Principles of Psychology* (James 1890, 462) expresses the problem elegantly:

✉ Peter Gärdenfors
    Peter.Gardenfors@lucs.lu.se; http://www.lucs.lu.se/Peter.Gardenfors

[1]   Lund University Cognitive Science, Lund, Sweden

[2]   University of Technology Sydney, Ultimo, Australia

"The baby, assailed by eyes, ears, nose, skin, and entrails at once, feels it all as one great blooming, buzzing confusion; and to the very end of life, our location of all things in one space is due to the fact that the original extents or bignesses of all the sensations which came to our notice at once, coalesced together into one and the same space."

The problem I want to address in this article is how children can create categories and concepts out of such a "blooming, buzzing confusion". I argue that two learning process are involved. The first constructs the underlying primary perceptual structures that emerge in children's cognitive development. These structures will be modelled in terms of conceptual spaces (Gärdenfors 2000, 2014) that are presented in Section 2. My thesis concerning this process is that it detects various *invariants* in the sensory input. To some extent, my analysis follows the program of Gibson (Gibson 1966, Gibson 1979) although my approach is more cognitively focussed. My aim in Section 3 is to show that at least space, object, and action domains are very natural outcomes of a reduction of sensory information in terms of invariants. I argue that these primary domains correspond to separate sets of invariants. In other words, relying on invariants makes it possible to present the domains as conceptual spaces that are considerable reduced in complexity when compared to the sensory input. This process transforms the quickly changing sensations into a relatively invariant representation of the environment. Since I take the perceptual structures to be learned, my position is an empiricist one, in contrast to the nativist view of, for example, Carey (2009) and Spelke (Spelke 2000, 2004).

Several philosophers and psychologists make a distinction between sensations and perceptions (e. g. Humphrey 1993 and Gärdenfors 2003). Sensations are what is received by our senses and perceptions are 'interpreted' sense data. In the present context, the distinction can be described as that sensations are turned into perceptions by mapping them into the conceptual spaces that are constructed from different kinds of invariants. Harnad (1990) makes a related distinction between iconic and categorical representations. The iconic representations are "internal analog transformations of the projections of distal objects on our sensory surfaces" and categorical representations contain those invariant features that "distinguish a member of a category from any non-members". However, Harnad does not specify what the invariants are or how they are determined, but only mentions that they can be picked up by artificial neuron networks.

The second learning process consists of the mechanism that utilizes the primary domains for *concept formation*. For this task (Section 4), I rely on *covariances* between different dimensions (features) of what is perceived in order to identify natural clusters of entities. These clusters are then used to construct *regions* of the underlying conceptual spaces. The regions are interpreted as the intensions of concepts. In Section 5, I then argue that during children's development there is a continued dimensionalization of the conceptual spaces that makes it possible for children to attend to particular features of the perceptual input, for example, colour and size.

Obviously, I will not be able to provide the details of these two learning processes, but my proposal should rather be seen as a research program.[1] As an application, I show

---

[1] The proposed two processes are not the same as in the 'complementary learning systems theory' (Kumaran et al. 2016). The two processes proposed here could rather been seen as parts of the cortical system of that theory.

in Section 6 that using the two processes I propose one can explain some of the intriguing phenomena of concept learning and the corresponding language development, in particular the so called 'complex first paradox' (Werning 2010) that emerges from the fact that children, in general, learn nouns earlier than adjectives in spite of adjectives being semantically less complex than nouns.

A note on terminology: I use the word 'category' as referring to a class of entities and the word 'concept' as referring to the mental representation of such a class that can be used to categorize entities.

2 Background: Conceptual Spaces

A central idea of the conceptual spaces framework is that concepts can be represented geometrically (Gärdenfors 1990, 2000, 2014). Conceptual spaces are mathematical entities in the form of dimensional structures, often (but not always) with a metric defined on them. More exactly, the dimensions of these spaces are interpreted as representing fundamental properties (qualities) that objects may possess to different degrees, so that objects can be mapped onto points in the space in accordance with the degree to which they instantiate a property. The quality dimensions correspond to the different ways stimuli can be judged similar or different. For example, one can judge tones by their pitch, and that will generate a similarity ordering of the auditory perceptions. Distances between representations of objects are then supposed to measure how similar the objects are to each other, where the similarity is not overall similarity but similarity in the property – for example colour, weight, taste, shape – that the space is supposed to model. The coordinates of a point within a conceptual space represent particular instances along each dimension: for example, a particular temperature, a particular weight, etc.

Conceptual spaces that have been discussed in the literature include colour space, taste space, olfactory space, various auditory spaces, as well as shape spaces, musical spaces, spaces to represent actions, events, emotions, moral concepts, scientific concepts, and epistemic concepts.

As a paradigmatic example, consider human perceptual colour space (see Figure 1). This space is three-dimensional, with one dimension – the vertical axis – standing for brightness, which goes from white to black through various shades of grey; the second dimension is the hue circle; and the third dimension is saturation, which is the intensity or depth of a colour.

The primary function of the dimensions of a conceptual space is to represent various qualities of objects in different *domains,* where a domain represents a particular set of properties, for example colours. Since the notion of a domain is central to the analysis, I should give it a more precise meaning. One way to do this is to rely on the notions of *separable* and *integral* dimensions, which I take from cognitive psychology (Maddox 1992; Melara 1992). Certain quality dimensions are integral: one cannot assign an object a value on one dimension without giving it a value on the other(s). For example, an object cannot be given a hue without also assigning it a brightness (and a saturation). Likewise the pitch of a sound always goes with a particular loudness. Dimensions that are not integral are separable: for example, the size and hue dimensions. Using this distinction, a domain can now be defined as a set of integral dimensions that are separable from all other dimensions.
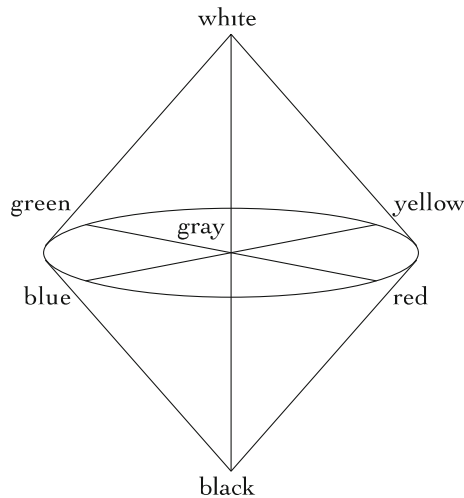
Fig. 1 A geometric representation of human perceptual colour space

In earlier works on conceptual spaces (Gärdenfors 2000, 2014; Gärdenfors and Löhndorf 2013), the problem of the *origins* of the domains has barely been discussed. The problem presented in the introduction can be formulated as follows: How do children obtain their perceptual domains? In particular the problem pertains to the domains of space, actions and object properties that form the basic ontology of our perceived world. Traditionally, there are two answers to this type of question: (1) the domains are *innate* (nativism); and (2) the domains are *learned* (empiricism). My solution will be of the second type, although I will argue that the organisation of the brain generates constraints on the learning processes.

3 Primary Domains

3.1 Extracting Structure: Invariants in Perception

The first learning process to be analysed thus concerns the origin of the fundamental domains that build up the perceptual structures of an infant. My thesis concerning this process is that the sensory input, at an early stage of development, becomes sorted into a number of general ontological domains. In this section I outline how a theory of *invariants* in the perceptual input can be exploited to generate such domains. My approach is to some extent inspired by Gibson's (1966, 1979) 'ecological approach' to perception, more precisely, his notion of information invariance. He writes: "The individual does not have to construct an awareness of the world from bare intensities and frequencies of energy; he has to detect the world from invariant properties in the flux of energy" (Gibson 1966, 319). The brain does this by *resonating* with what the senses receive. Gibson (1966, 201) defines an invariant as a 'non-change' that persists during change. In particular, the most important information for perception is what remains invariant as an agent moves through the environment

(see also Cutting 1986).[2] Gibson's definition is not very precise and not very useful for identifying invariants, so in my analysis, I will mainly rely on well-known types of invariants.

Given that the brain has a strong capacity to detect invariants, a fundamental question is for which perceptual domains these mechanisms work the best. It is natural to assume that the domains are the ones that infants learn first. To develop this idea, I take inspiration from the works of Spelke and others (Spelke 2000, 2004; Spelke and Kinzler 2007; Carey 2009) who have proposed four 'core knowledge domains' that are embedded in perceptual processing: objects, action, number, and space. For example, Spelke and Kintzler (Spelke and Kinzler 2007, 89) write:

"These systems serve to represent inanimate objects and their mechanical interactions, agents and their goal-directed actions, sets and their numerical relationships of ordering, addition and subtraction, and places in the spatial layout and their geometric relationships."

My first objective is to argue that an analysis of perceptual invariants can explain why space, objects and actions should form the basis for the first domains that children develop.[3] In contrast to Spelke and Carey, my position is empiricist. Even if Spelke does not explicitly use the word 'innate' in her characterization of the core knowledge systems, it is clear that her basic position is nativist. And Carey (2009, 11) writes "The claim that core cognition exists is a nativist claim".[4] Carey (2009, Ch.2) argues against the empiricist accounts proposed by Piaget and Quine as a support for her nativist position. As regards these positions, I find her arguments convincing. She admits in passing that it might be possible to develop an empiricist model of concept learning based on artificial networks (Carey 2009, 60). What I am proposing in this paper is a new kind of empiricist model of the development of primary domains, using conceptual spaces based on learning perceptual invariances as a modelling framework.[5] My account will provide some arguments, albeit not conclusive, for *why* these domains are primary.

3.2 Space

A central idea in Gibson's approach is that the visual field is determined from information that generates invariants such as texture gradients, occlusions and visual flow. The brain tunes in to such invariants at a very early stage. For example, when we

---

[2] It is interesting to note already Kaila (Kaila 1939/2014) introduced invariances as a way of sorting up perceptual experience: "Another important class of invariances is constituted by so-called physical objects, or material objects. After all, every object contains a regularity, for objects are constituted by distinct properties that hold together in a regular manner. Space is a system of invariances, it is part of our conception of space that it possesses a structure described by a certain geometry, a structure that remains the same everywhere [...]".

[3] Here I will not develop the argument for the domain of numbers. I believe, however, that the technique of studying invariants can be applied also to the child's developing understanding of numbers. For example, the number of a collection of objects is invariant with respect to the location and the identity of the objects. For similar arguments see Harbour (2014) and Johansson (2015).

[4] But she adds: "Notice also that 'innate' does no mean 'present at birth'. Many representational capacities arise from maturational processes" (Carey 2009, p. 12).

[5] I avoid Spelke's use of 'core' knowledge structures (and Carey's (Carey 2009) 'core' cognition) since it is connected with an nativist position, and instead speak of primary domains.

turn our heads and let our eyes follow along, the image that reaches the retina changes very rapidly. But, just as quick, our brain calculates a representation of the room that remains still in relation to the direction of our body.

During her first months, a child learns to coordinate her sensory input–vision, hearing, and touch–with her motor activities (Thelen and Smith 1994). One outcome of this motor babbling is an *egocentric* representation of space that is used to coordinate seeing with acting. As Gibson (1979: 2) wrote, "the environment to be perceived […] is not the world of physics but the world at the level of ecology". The egocentric space allows an individual to see its *field of action*. As long as only the head is moved and not the rest of the body, there is no change in an individual's possibilities to act. Since it is primarily the hands that are to be guided, it's more efficient if the brain creates a room that is constant in relation to their possibilities.

The egocentric representation of space is invariant of eye, head and body direction. The representation thus maintains a constant relation between the body location and the surrounding objects. The constructed space is basically a three dimensional Euclidean space with the body location as its origo.

The visual domain then expands throughout the child's development. In particular, by coordinating auditory information with visual, the represented space extends beyond the child's current visual field to cover the entire physical space. The child can then direct its attention outside its immediate visual field. It should be emphasized that the resulting representation is not just an extension of the visual domain but an *amodal* abstraction from visual, auditory, tactile, and perhaps even olfactory experiences.

A more advanced invariant of the representing space comes with the ability to represent an *allocentric* space, that is, a space that is independent of the location of the individual. Such a representation allows an individual to shift the perspective (Piaget 1954).[6] Consequently, the allocentric representation of space is not only invariant of eye, head and body orientation but also of *body location*. A concrete example of the use of allocentric space is the ability to give road directions where one has to imagine the route and movements along it.

The adult visuo-spatial domain should be seen as a combination of an allocentric representation and an egocentric representation. The two representations are connected to two different types of functions: The egocentric for reaching and interacting with objects, the allocentric for navigating through the environment (Gallistel 1990). The double aspect of our spatial representation is revealed by the two linguistic codes we have established for referring to positions: egocentric *left* and *right*, and allocentric *west* and *east* (or *north* and *south*). Similarly, what is *behind* the house from my egocentric perspective may be *in front of* the house from an allocentric perspective.

There are strong arguments for that the experience of space is not innate but must be learned through *interaction* with the world around us (e. g. Held and Hein 1963; Agrawal et al. 2015).[7] The process that creates our three-dimensional perception space

---

[6] The distinction beween egocentric and allocentric corrsponds to Gibson's (1966) distinction between 'perspective structure' and 'invariant structure'.

[7] In contrast to this position, Carey (Carey 2009, p. 12) writes: "Even though stereoscopic depth perception is not present at birth, I would want to say that it is innate, for the child does not have to learn to compute depth from the discrepancies between the two images on the two retinas". I don't agree. Apart from the two references given in the text, there is further evidence that we must learn to see. For example, some blind people who have regained vision, have problems perceiving depth (Gregory 1970).

– partly on the basis of the two-dimensional images provided by our eyes – must learn how the sensory impressions can be used to create meaningful fields of action. When one gets a new pair of glasses, for example, the conditions for this process are altered, and it takes a while before the brain has adjusted its construction of space to the new invariants and can provide the perceptions one needs for carrying out precise actions, for example walking down stairs without stumbling.

It is important to note that the egocentric and allocentric spaces that are generated by extracting the various forms of invariants considerably *reduce* the complexity of the information compared to what is transmitted from the retina to the brain. To the extent that the constructed allocentric space is invariant under Galilean transformations (that is, rotations and translations), it follows that what is conserved in visual perception is that space is three-dimensional Euclidean. One aspect of the Galilean transformations is that space is *constant* over time. When we move or turn ourselves around we actually perform a Galilean transformation of the perceptual input, so it is very natural that an efficient neural system picks up the invariants and uses the represented space as a basis for the actions of the individual. Gibson (1966, 264) made this point a long time ago: "An individual who explores a strange place by locomotion produces transformations of the optic array for the very purpose of isolating what remains invariant during these transformations" (see also Agrawal et al. 2015). Our movements occur mainly in the two horizontal dimensions, less so in the vertical. As a consequence, our perception of the vertical dimension is 'flattened' in relation to a Euclidean space (Kaufman and Kaufman 2000).

3.3 Objects

The question of how infants represent and reason about objects is central for an analysis of primary forms of perception. Several constraints have been offered in the literature. For example, Spelke et al. (Spelke et al. 1992, 606) propose the following: (i) *continuity* (objects move in continuous paths), (ii) *solidity* (objects move only on unobstructed paths and, consequently, no two objects occupy the same place), (iii) *gravity* (if not supported, objects fall downwards), and (iv) *inertia* (objects do not change their motion abruptly). In my opinion, at least the last two constraints are not constitutive of objects *per se*, but rather concern the behaviour of objects (the inertia constraint is, to some extent, violated by objects that are *agents*). A special case of continuity is *object permanence*, which means that objects do not disappear from a place even if they are not perceived at the moment. Another central constraint, not mentioned by Spelke, is that objects have a *shape* (see section 4.3).

Although I cannot fill in the details, I submit that the relevant constraints can be derived from invariants of perceptual properties along the lines outlined above. First of all, the relative locations of different parts of an object exhibit different types of invariants. For a solid object, the invariants are total. For an object with movable parts, the invariants of the locations within each part is total and so are the locations of the points where the different parts are connected. Johansson (1964) formulates this as a 'rigidity principle' – a constraint of the visual process that generates a perception of rigidity whenever equal motions in a series of simultaneous proximal elements are detected (cf. Marr's (1982) representation of shapes). For deformable objects – such as cushions, towels and dough – the invariants of relative locations are less stable, but the

changes of relative locations are continuous. (A dough is on the verge of being a mass rather than an object.) Another aspect of continuity is that objects 'hang together' in the sense that if you pull at one end of an object, the other parts will follow. Clouds are therefore marginal as objects.[8]

Solidity or relative solidity is but one type of invariants that apply to objects. There are many other types. For example, the *size* of an object is typically invariant, something which helps our visual system to efficiently judge the distance to an object. Murray et al. (2005) show that size invariance is evident already in the dorsal retinotopic visual area V3. Another salient domain is colour. The colour pattern of an object is not invariant since it varies with the illumination. In most cases, however, the perceptual *relations* between the colours of an object are invariant (Land 1977). For many kinds of objects, for example, different species of birds, the patterns of colours are characteristic features.

It is still unknown how the brain picks up the invariants that are relevant for generating a space that represents objects. Again, perceiving objects involves a considerable reduction of dimensions in the sensory input. There exist a number of computational procedures for dimension reduction, for example Principal Component Analysis (Abdi and Williams 2010) and Multidimensional Scaling (Kruskal and Wish 1978; Borg and Groenen 2005) but it is not known to what extent brain processes match these procedures. However, Wiskott and Sejnowski (2002) have constructed an artificial neural network based on 'slow feature analysis' that, to a large extent, can learn translation, size, rotation, contrast and illumination invariances of objects. A particularly interesting feature of their model is that the 'what' and the 'where' components get represented in separate components of the system. This supports my hypothesis that the space and object invariants are of different kinds (see section 3.5).

3.4 Actions

The human brain is extremely efficient at identifying different kinds of actions. For example, you see immediately whether somebody is walking or jogging, even if the leg movements look quite similar. Furthermore, the amount of information you need to perform such a categorization is very limited. This point was established by Johansson in a series of ground breaking psychophysical experiments in the 1950's (Johansson 1973). He developed a patch-light technique for analysing biological motion where no direct shape information is available. He attached light bulbs to the joints of actors who were dressed in black and moved in a black room. The actors were filmed performing actions such as walking, running, and dancing. Subjects who watched the movements of the lights (but saw nothing else) categorized the actions within a fraction of a second.

These experiments show that that the surfaces of the agents performing the action are not required for identifying and categorising the actions. A movie containing only stick figures performing the same movements is sufficient. (In passing, it should be mentioned that this observation confirms Johansson's rigidity principle.) So what kind of information is used in such a categorisation?

Runesson (Runesson 1994, pp. 386–387; see also Wolff 2008) claims that people can directly perceive the forces that control different kinds of motion:

---

[8] Doughs and clouds suggest that there may be grades of objecthood.

"The fact is that we can *see* the weight of an object handled by a person. The fundamental reason we are able to do so is exactly the same as for seeing the size and shape of the person's nose or the colour of his shirt in normal illumination, namely that *information* about all these properties is available in the optic array."

He summarizes this as that the kinematics of a movement contains sufficient information to identify the underlying dynamic force patterns. This thesis is formulated with respect to biological motion. I speculate that it extends to other forms of motion as well. I have hypothesized that the brain automatically extracts the *forces* that lie behind different kinds of movements and other actions (Gärdenfors and Warglien 2012; Gärdenfors 2014). Furthermore, the process is automatic: one cannot help but perceive the forces. For example, the pattern of forces involved in the movements of a person running is different from the pattern of forces of a person walking; likewise, the pattern of forces for saluting is different from the pattern of forces for throwing.[9] Just as for shapes, the space within which force patterns are located can be treated as a separate perceptual domain, with its unique structure of similarities. Of course, the perception of forces is not perfect; people are prone to illusions, just as in all types of perception (Johansson 1964, 1973).

An important consequence of this hypothesis is that the individuals or objects involved in an action are *not* part of the representation of the action, but only the forces are involved. I speak of patterns of forces since, for bodily motions, several body parts are involved; and thus, several force vectors are interacting (by analogy with Marr and Vaina's (1982) differential equations). Again, these patterns form the invariants that I submit generate the structure of actions. However, the invariants that pertain to actions are different from both those for objects and those for space. In particular, the patterns are neither dependent on the location of the acting object, nor of its surface properties. However, the more precise structure of action space remains to be investigated. As for space and objects, the structure generated by the invariants involves a considerable reduction in dimensions.

It should be noted that similar arguments can be applied to *speech*. Gibson (1966, 93) identifies some of the invariants of speech: "[P]honemes are *transposable* over the dimensions of pitch, loudness and duration, and […] the stimulus information for detecting them is invariant under the transformations of frequency, intensity and time." Browman and Goldstein (1990) describe the act of uttering a word as a 'score of gestures' where the gestures are performed, not by the hands, but by the five vocal organs of velum, tongue tip, tongue body, lips, and glottis. They then describe the utterance of a word as a temporal sequence – a score – of activation of these organs. Such a score can be re-described as a temporal pattern of force vectors. Browman and Goldstein's description of the patterns as 'vocal gestures' underlines this analogy.

---

[9] An example of data that can be used to study force patterns comes from Wang et al. (2004). They collected data from the walking patterns of humans under different conditions. Using the methods of Giese et al. (2008), these patterns can be used to calculate the similarity of the different gaits in terms of the underlying forces. Gharaee et al. (2017) have applied the force dynamic model in a robotic system that has been constructed for categorizing actions.

3.5 The Brain is Prepared to Find Invariances

The main conclusion to be drawn from the preceding subsections is that the primary domains for space, objects and actions can be generated from the invariants that apply to each of the three domains. Thus the same method has been used to identify the domains. It should be noted, however, that the sets of invariants are distinct for the three structures: For space, the main invariants are relative distances that are also invariant of time. Object locations may change rapidly, but object identity changes rarely, or slowly. Thus object categories are invariant of location in space. Furthermore, the relative positions of the parts of objects show more or less strict invariants. Other properties of objects, such as relative colours, may, also be invariant. For actions, finally, the invariants pertain to force patterns. In brief, the set of invariants for the three primary knowledge domains are more or less disjoint, which is an argument for why the domains are represented separately.[10] This analysis must be developed in more detail, but if valid, it would provide a strong argument for why these domains are indeed primary and universal among humans.

Although I cannot provide any conclusive arguments at this stage, I submit that the invariants that determine the domains for space, objects and actions are the ones that are most easily picked up by the sensory system of an infant. If this can be substantiated, it would provide a strong argument for why places, objects and actions are fundamental cognitive domains. My position is basically empiricist since the invariances must be learned.

An important question is now whether there are other primary domains that can be identified via the proposed method of searching for invariants. I will return to this question in the concluding section.

A follow-up question would be: *Why* are the invariants that determine places, objects and actions the ones that are the easiest to learn? At the bottom, this question would need an argument in terms of evolutionary epistemology. The process turning sensations into perceptions by identifying invariances takes the different kinds of energy hitting our sensory receptors and turns them into something that represents structures in the environment. In brief, some regularities in the world have been evolutionarily more important than the amounts of energy at sensory surfaces.

A part of the argument would build on that human infants are not born as blank slates (Pinker 2002). Evolution has made the brain *prepared* for picking up the most relevant invariants. To this extent there is a nativist element in my analysis. In particular, the space representation is generated in the dorsal stream of the cortex (the where pathway), object representation is generated in the dorsal stream (what pathway) and action representation in the dorsal stream (how pathway). However, even if the pathways in the brain are to some extent prepared, the infant must still *learn* which invariants generate the most useful perceptual structures. Even after the invariants have been learned, the brain exhibits an amazing plasticity that supports relearning: For example, if a person is given goggles that turn the visual field upside-down, it is

---

[10] This argument can be used to provide and alternative definition of separability: Two domains are separable if the sets of invariants determining the domains are disjoint. Such a definition presumes, however, that the determining invariants have been identified.

possible to relearn the mapping so that, after a few weeks, the world is perceived in the 'normal' way (Kohler 1951).

Gibson (1979) favoured a bottom-up approach to how the invariants are acquired, claiming that the information is picked up directly, so that no intervening mental processes are necessary for visual perception, but this position has been criticised. For example, Gregory (1970) argued that top-down processes must mediate perception. Goldstein (1981, 193) writes:

> "The problem comes with Gibson's statement that what an object affords is specified in the light, and his failure to deal adequately with the fact that affordances must be learned. A wooden chair may afford sitting for a human, but something to gnaw on for a beaver, even though the information provided by the light is the same for both."

While useful information may exist directly in the ambient light, Gibson presents no account of the mechanisms of how this information is picked up. In contrast to his view, the sensory information received is often incomplete and, consequently, the brain must 'construct' a perception.

4 Concept Formation

An old philosophical question is whether supposedly natural concepts, such as 'red', 'gold', and 'cat', reflect real divisions in nature that exist independently of our thinking and theorizing, or whether their meanings are dependent on our minds. The first position is called realism, the second conceptualism. Without further ado, I here adopt the conceptualist position about concepts. For some arguments, see (Gärdenfors 2000, 2014).

A crucial factor is what concepts *are for*. There are three main uses of concepts: (i) for categorization; (ii) for communication; and (iii) for reasoning. Here I focus on our need to categorize entities. For example, we must be able to distinguish edible things from non-edible ones. The most important cognitive function of a system of concepts is to provide a mapping from perceptions to actions. In the case of simple reflex mechanisms, the mapping is more or less fixed and automatic. In most cases, however, the mapping has to be learned and it is a function not only of the current perception, but also of memory and context. It is central that such a mapping can be learnable in an efficient way. In earlier works (Gärdenfors 2000), I have argued that *similarity* should be a fundamental notion when modelling the concepts that mediate perceptions and actions.[11] In this section, I show how similarities in the primary knowledge domains can be used when learning the content of concepts.

4.1 Clusters of Sensory Information

I now turn to the second general learning process – the one generating concepts. Given a perceptual domain of the kind discussed in the previous sections, concepts can be

---

[11] Not everybody agrees, for example Carey (2009). I will return to this topic in section 5.

built up from perceptual mechanisms (to some extent combined with memory), based on the information contained in the instances of a concept. Here follows a proposal for how this learning process works.

The key idea is that perceptual information is not random but information comes in clusters. Work by Billman (1983) and Billman and Knutson (1996) indicates that humans are quite good at detecting *covariations* that cluster several dimensions, in spite of our limitations in detecting isolated correlations between variables (see also Kornblith 1993, 96–105). For example, singing covariates with having feathers, flying, laying eggs and building nests. In other words, we have a sensitivity to features that tend to be found together.

A plausible explanation of this phenomenon is that our perceptions of 'natural' objects show covariations along multiple dimensions, and, as a result of natural selection, we have developed a competence to detect such clustered covariations. Kornblith (1993), pp. 105–6) provides a similar argument:

> "It is thus safe to say that we have a sensitivity to the features of objects which reside in homeostatic clusters. Indeed, the way in which we detect covariations is precisely tailored to the structure of natural kinds. […] we conceptualize kinds in such a way in order to separate the properties of the members of a kind which are projectable from those which are not. We are aided in this task by our ability to detect clustered covariation."

Billman and Knutson (1996, 459) identify two structural principles in such covariations that help category learning:

> *Value systematicity*: If one property value (e. g. that the form of locomotion is flying) predicts the value of a second property (that the limb is a wing), then that same value should predict values of other (for instance that the covering of the limb is feathers).
> *Value contrast*: If one value of a property (that the form of locomotion is flying) predicts the value of a second property (that the limb is a wing), then other values of the same property (that the form of locomotion is walking, swimming or crawling) should also be predictive.

When investigating covariation learning, Billman used a technique called *focused sampling* both in her computer models and in her and Heit's study of human subjects (Billman and Heit 1988). In this process, the material consists of a large class of objects, each of which is characterized by a large number of properties. Because of the large number, a complete survey of the objects and the corresponding properties is impossible both for a computer and a human. Correlations must therefore be detected from samples of the objects. Rather than performing a random search, focused sampling preferentially selects those objects that have properties that have already proven to be connected. So if properties C and D have been found to correlate, objects with these properties are more likely to be studied. If C and D correlate with a further property E, this technique will reinforce itself and rapidly detect clusters of properties that correlate. The upshot is that the more properties objects have in common, the more similar they will be, and, consequently, the smaller will be the size of the cluster they form.

A central part of the theory of conceptual spaces is that concepts can be modelled as *convex regions* in a domain or a set of domains (Gärdenfors 2000, 2014). For example, even though different languages carve up the colour domain in different ways, it seems to be a universal principle that colour concepts form convex regions (Jäger 2010).

A set of clusters in a conceptual space can be used to *partition* the space into regions, where the elements of a cluster are central in a region. The clusters form the *extensions* while the regions are the *intensions* of the concepts. Assuming that the space has a metric, there are several computational methods for determining such a partitioning, for example, *K*-means, self-organizing maps and neural gas (see e.g. Filippone et al. 2008). For another example, (Gärdenfors 2000) proposes to take the mean of each cluster as a prototype of a concept and then use the prototype to generate a so-called Voronoi tessellation.[12]

A problem is that clusters can be identified at several levels of coarseness. For example the set of scotch terriers forms a cluster that is a subset of the cluster of dogs, which in turn is a subset of the cluster of mammals. Depending on the size of the cluster chosen, different superordinate or subordinate concepts can therefore be generated. I will return to this in connection with my discussion of prototype theory.

I next turn to a description of the concept learning process for each of the spatial structures connected with the primary domains of space, objects and actions.

4.2 Space Concepts

General spatial concepts are not common. The most obvious examples are *places*, which literally are regions of physical space. Common examples are forests, mountains, lakes, beaches, and villages.

Concepts for spatial *relations* form a richer system. In language prepositions are used to express such relations, for example locative prepositions – such as *inside*, *near*, *far*, *above*, *in front of*, and *beside* – and directional prepositions – such as *to*, *from*, and *through*. Zwarts and Gärdenfors (2016) show that locative prepositions can be represented by (convex) regions in ordinary space and that directional prepositions can be represented by (convex) regions in the space of paths.[13]

A special type of spatial concepts is *landmarks* that are objects the locations of which are invariant. It must be possible to sense the landmark (by visual, olfactory or auditory means) from a distance that is large relative to the movements of an individual. Animals are surprisingly skilled at maintaining a precise representation of their location in relation to landmarks in the environment (Gallistel 1990).

4.3 Object Concepts

The space of objects is rich and it contains a number of subdomains (properties) that have their own structure, each with their own invariants. However, this richness helps the child to detect similarities between objects – similarities that determine the

---

[12] In passing, I note that by using this method to generate concepts, the learner can learn a concept from a few examples and she need not be informed about examples that do *not* fall under the concept.
[13] An interesting detail is that Zwarts and Gärdenfors (2016) use *polar* coordinates in their model rather than the standard Euclidean one.

clustering of objects, and thereby the formation of object concepts. In particular the invariants of mereonomic structure and rigidity that apply to a single object – solid or partially solid – are central for how infants judge object similarities. These similarities will group objects into clusters of things with similar *shape* (Zhu and Yuille 1996). In support of this argument, it has been established that children show a strong shape bias when learning object categories (e. g. Billman and Heit 1988; Smith 1995). My explanation for this bias is thus that the shape invariants are among the most important features when objects are clustered.

There are, however, often other types of similarities that are combined with shapes when an object is categorized. For example, even though many songbirds have similar shapes, it is sometimes possible to categorize them based on their colouring patterns that are similar for a species. Or if a colouring pattern is also indistinct, the song of the bird – that for many species forms a highly specific pattern – further helps to categorize the bird. Given that these properties also show strong covariations, clear clusters of objects can be identified, which then can generate the regions that represent the corresponding concepts.

As part of prototype theory, Rosch (Rosch 1975, 1978; Mervis and Rosch 1981) introduces the *basic level* of a hierarchy of object categories as a particularly salient level of concept formation. She presents a number of criteria for what distinguishes the basic level from superordinate or subordinate levels. One criterion says that superordinate categories contain much fewer common properties than the basic level and the subordinate levels contain hardly any additional common properties. For example, *cat* has many more characteristic properties than *mammal*, but not many more than *abyssinian*. In support of this analysis, Hunn (1976) has argued that the basic level is the only level at which category membership can be determined by an overall configurational Gestalt perception.

A strong argument for the importance of meronomic relations in concept formation comes from Tversky and Hemenway (1984). They show that part terms occur frequently when subjects describe categories at the basic level, but are rare on superordinate levels. Basic level objects are often distinguished from each other by the configuration of their parts. Furthermore, subordinate categories typically share the part structure with the basic level, but differ from one another on other domains.

I have now given some arguments for why object concepts can be generated from different types of covariances of properties along the lines of Billman's criteria. However, the outline I have provided needs to be connected to research concerning how infants form object concepts (see e.g. Carey 1985, 2009; Landau et al. 1998; Mandler 2004; Smith 2005; Spelke 2000, 2004;).

4.4 Action Concepts

In section 3.4, I argued that the structure of the action domain is determined by invariants of force patterns. In order to identify the relevant clusters and regions of the action space, *similarities* between force patterns should be determined. The dynamic properties of actions can be judged with respect to similarities: for example, walking is more similar to running than to waving. This can be accomplished by basically the same psychological methods used for investigating similarities between objects. I submit that the similarities between actions are determined via the

covariances of the movement patterns of different body parts. In earlier works I have proposed the thesis that an action concept can be described as a (convex) region of such patterns (Gärdenfors and Warglien 2012; Gärdenfors 2014).

In analogy with shapes, force patterns also have *meronomic* structure. For example, a dog with short legs moves in a different way than a dog with long legs. Furthermore, there are strong reasons to believe that actions exhibit many of the prototype effects that Rosch (1975) presented for object categories. For example, Hemeren (Hemeren 1997, 2008) showed that action categories show a similar hierarchical structure and have similar typicality effects as object concepts.

One example of analytic work along these lines is Giese and Lappe (2002). Using Johansson's (1973) patch-light technique, they started from video recordings of natural actions such as walking, running, limping, and marching. By creating linear combinations of the dot positions in the videos, they then made films that were morphs of the recorded actions. Subjects watched the morphed videos and were asked to categorize them as instances of walking, running, limping, or marching, as well as to judge the naturalness of the actions. In accordance with the proposal made in (Gärdenfors and Warglien 2012; Gärdenfors 2014), prototypes could be found and the categorization identified convex regions of the underlying space.

4.5 Concepts in Primary Knowledge Domains and the Semantics of Word Classes

In this section I have outlined how the primary domains can be seen as the fundaments on which concepts can be erected. The main ideas have been that concept formation is based on discovering covariations in the knowledge domains and that the clusters of covariations are used to partition conceptual spaces into regions that represent concepts. I next want to argue that this process is central also for language learning.

When infants begin to extract patterns in the sounds emitted by people in their environment (some of which will later be identified as words), they have no idea that these patterns stand for different types of entities. The patterns will, however, form part of the sensory input that is used to identify covariances. For example, the sound pattern "kitty" covaries with the presence of cats, toy cats, or pictures of cats (although the word may be uttered also in other contexts). In particular, when a parent is establishing joint attention with the infants to such objects, the covariation is strong. The sound pattern thus become *part of* the perceptual clusters that generate the concepts. Only later does the infant learn that the sound patterns can be used to trigger the corresponding concepts in the minds of others even when no entity falling under the concept is present. They then learn that *words refer to regions of conceptual spaces* (that in turn are determined by clusters). This principle can be seen as a linguistic 'meta-invariant' that is picked up from their communicative interactions with others.[14]

Our words express our concepts. Hence a theory of semantics should be founded on a theory of concepts. Croft (2001, 364) makes the connection as follows:

> The categories defined by constructions in human languages may vary from one language to the next, but they are mapped onto a common conceptual space,

---

[14] This principle only applies to 'content' words and not to syntactic markers. A wild speculation is that this may be the reason why children learn syntactic markers later than a considerable number of content words.

which represents a common cognitive heritage, indeed the geography of the human mind […] which can be read in the facts of the world's languages in a way that the most advanced brain scanning techniques cannot ever offer us.

In this article the focus is, however, not on the geography of the mind, but on its *geometry*. However, as I have already mentioned in relation to colour concepts, different languages carve up the domains in different ways. A similar point is made by Mandler (1991, 414):[15]

> "Language is unlikely to be mapped *directly* onto sensorimotor schemas. There is a missing link: A conceptual system that has already done some of the work required for a mapping to take place."

The work that she mentions has been performed by the first learning process that generates the primary domains.

Even if the concepts defined on a domain (and their corresponding words) are not universal, my analysis in section 3 suggests that at least the primary domains are universal in human cognition. If this is correct, they should somehow be reflected in the structure of language (a related argument is presented by Strickland 2017).

Indeed, the three primary domains I have identified in section 3 correspond to three of the main *word classes* in languages: Concepts based on the object knowledge domain are typically expressed by *nouns*; concepts based on the action domain are expressed by *verbs*; and relational concepts based on the space domain are expressed by *prepositions* (although many languages use other means to express spatial relations).

These connections between knowledge domains and word classes help children learn language more efficiently (Bloom 2000; Gärdenfors 2014). Most languages use different kinds of syntactic markers for the main word classes. These markers help identify the relevant primary domain for the word. Lupyan and Dale (2010, p. 8) make "the paradoxical prediction that morphological overspecification, while clearly difficult for adults facilitates infant language acquisition". Mandler (2004, p. 281) argues along the same lines:

> "Many of the grammatical aspects of language seem impossibly abstract for the very young child to master. But when the concepts that underlie them are analyzed in terms of notions that children have already conceptualized, not only does the linguistic problem facing the child seem more tractable but also the types of errors that are made become more predictable. The invention of grammatical forms to express conceptual notions that are salient in a young child's conceptualization of events seems especially informative."

The upshot is that the underlying structures in form of word classes that are common to languages in the world have strong connections to the primary knowledge domains. This parallel deserves further investigations.

---

[15] Mandler (1991) proposes image schemas from cognitive semantics as the underlying conceptual system. I believe that her proposal is consistent with the one made in this article since image schemas can be seen as an alternative way (albeit less systematic) of representing invariants.

5 Properties Emerge Via Dimensionalization

5.1 Context Dependence of Similarity

I argued in section 4.3 that objects are grouped by their overall similarity.[16] There I assumed that similarity is determined from the structures of the primary domains. However, similarity judgments are not constant over time, but as children learn more about the structure of the world (and more of their mother tongue), their perception of similarity develops into a complex system that, among other things, becomes dependent on the categorization context.

Smith (1989, p. 159) points out that similarity judgments are holistic at the beginning, but are then *separated* into dimensions:[17]

> "[T]here is a *dimensionalization* of the knowledge system. […] Children's early word acquisitions suggest such a trend. Among the first words acquired by children are the names for basic categories–categories such as *dog* and *chair*, which seem well organized by overall similarities. Words that refer to superordinate categories (e.g., *animal*) are not well organized by overall similarity, and the words that refer to dimensional relations themselves (e.g., *red* or *tall*) appear to be understood relatively late […] School-age children consistently assign objects to groups by single dimensions, categorizing reds versus blues, bigs versus littles. Children under 5 do not […]; instead they classify objects by their similarity overall."

In section 4, I argued that the primary domains can be represented as conceptual spaces. The object domain consists of several subdomains, for example, shape, size, colour and weight. A domain of such a space is a set of dimensions that are integral. What happens in children's development is that one dimension after the other is separated out in perception and can be attended to. For example, two-year-olds can represent object categories, but they cannot reason about the dimensions of those objects. One way to express the development is to say that children go from judgments of similarities to judgments of *kinds* of similarities.

In line with this, Goldstone and Barsalou (1998, 252) note:

> "Evidence suggests that dimensions that are easily separated by adults, such as the brightness and size of a square, are treated as fused together for children […] . For example, children have difficulty identifying whether two objects differ on their brightness or size even though they can easily see that they differ in some way. Both differentiation and dimensionalization occur throughout one's lifetime."

---

[16] The similarity need not be exclusively perceptual. For example for functional categories, such as chairs and watches, children also use the actions performed by an object as a cue to its categorization, in addition to shape and other static domains (see, for example, Smith 2005; Gärdenfors (2007)). Carey (2009, 275) argues that infatns sometimes categorize on the basis of global kind rather than by perceptual similarity. However, her examples concerns animals that are similar with respect to a number of properties, even if they are not directly perceptual.

[17] See also Smith and Sera (1992, 132).

An example of dimensionalization is seen in Piaget's (1972) conservation task. Children under the age of five cannot separate the volume of a liquid from its height. When choosing between two glasses of lemonade, they pick the glass with the highest level of lemonade even though that glass is very narrow and the other is wide. Only later do they learn that the volume of a liquid is conserved between containers and not always correlated with height. In other words, volume is an invariant of liquids (which height is not). When this invariant is discovered, children learn to separate the domain of volume from that of height. A related phenomenon from child language is that adjectives that denote contrasts within one adult domain are often used for other domains as well. Thus, three- and four-year-olds confuse *high* with *tall*, *big* with *bright*, *small* with *dim* etc. (Carey 1985). This is an indication that the domains are not yet sufficiently separated in the minds of the children.

The separation into dimensions (domains) means that children learn to *focus* on certain properties of objects. Only when they, for example, can attend to the colour of objects (instead of, say, shape or size) is it possible for them to learn the full meaning of the colour words (see section 6).

5.2 Properties Expressed by Adjectives

In Gärdenfors (1990, 2000), *properties* are identified with convex regions of single domains. For example, the property *red* is a convex region of the colour domain and the property *hot* is a convex region of the temperature domain. Properties are thus special cases of concepts.

One of the first domains that is separated out in perception is that of shape (Smith 1989). Shapes are multimodal since they can be perceived by both vision and touch and they remain invariant through a large class of transformations. Interestingly, Fölster and Hansson (2017) show that the capacity for shape perception in children at the age of 24 months correlates with their linguistic competence at the age of 6 or 7 years.

In language, properties are typically expressed by *adjectives*. Thus, the semantics of yet another central word class is given a cognitive grounding via the proposed account of properties as concepts that depend only on a single domain (in contrast to the meaning of concrete nouns that depend on covariations between several domains).

If property concepts are learned later than object concepts, then it should be expected that adjectives should be learned later than nouns. There is strong evidence from language development supporting this conclusion (e. g. Dromi 1987; Jackson-Maldonado et al. 1993; Sandhofer and Smith 2007). For example, Mintz and Gleitman (Mintz and Gleitman 2002, 269) note:

> "Glaring asymmetries in noun vs. adjective (and verb) frequencies in novice vocabularies … persist until about their third birthday […]. [O]ne potential explanation for why acquiring adjectives is hard has to do with the possibility that they fall into a variety of conceptual classes whose conflation under a lexical categorization […] is more arbitrary than natural."

Their phrase 'conceptual class' corresponds to my 'domain'. I will return to this phenomenon in the following section in relation to the complex first paradox.

Mintz and Gleitman (2002) show, however, that if the adjective comes together with a noun that already is understood, then even 2-year-olds can learn the meanings of new adjectives quickly (see also Waxman and Markow 1998). Mintz and Gleitman (2002, 285) conclude that "24- and 36-month-olds do not seem to map novel adjectives to object properties without the support of a full noun".

6 The Complex-First Paradox

In the previous section, I have outlined a mechanism for concept formation that constitutes the basis for word learning. Such a proposal is not uncontroversial. One potential counterargument that recently has been suggested is the 'complex-first paradox' that was formulated by Werning (2010). The paradox derives basically from the clash of two facts: (i) Children learn noun concepts such as *cat*, *cup*, and *chair* earlier than adjectives like *red*, *hot* and *short* (Bloom 2000; Mintz and Gleitman 2002). (ii) The meanings of nouns are 'semantically thick' since they comprise multidimensional information while the meanings of adjectives are 'thin' since they cannot be decomposed. Nouns should therefore be more difficult to learn than adjectives. The second statement is supported by findings from neuroscience showing that the cortical correlates of nouns are more complex than those of adjectives (Werning 2010, 1097).

An elegant solution to the complex-first paradox, based on conceptual spaces, has been presented by Poth (2016). Her key idea is that entities denoted by concrete nouns show a greater *overall similarity* than those denoted by adjectives. The reason for this is that entities falling under a concrete noun show greater covariances than entities falling under an adjective. This idea thus depends on the *size* of the regions that are associated with a word, for example a noun or an adjective. She notes that children's language learning seems to follow a general 'size principle' saying that the meaning of a word should be determined from the cluster with the *smallest* size that the observed entities belong to.[18]

To spell out this idea, let me make a proposal concerning the learning mechanism involved. A problem that I noted earlier is that clusters of objects can be identified on different levels of coarseness. For example, assume that the child has heard the word 'dog' a few times referring to, say, a cocker spaniel, a Scotch terrier and a German shepherd. The child then identifies the smallest cluster to which theses objects belong, that is, the cluster of dogs and the meaning of 'dog' with the region covered by this cluster. Even though all the objects also belong to the cluster of objects corresponding to 'mammal', this cluster will not be selected since the cluster of dogs has a smaller size.[19] However, if all the observed objects in the cluster happen to be cocker spaniels, then the size principle would predict that the child instead associates 'dog' with the region determined from the cluster of cocker spaniels.

In contrast to nouns, words denoting adjectives, such as 'brown' apply to objects that do not show much overall similarity. For example, a brown shoe is not particularly

---

[18] See Poth (2016) for a discussion of a probabilistic version of this principle and its relation to a proposal by Xu and Tenenbaum (2007).

[19] Poth (2016) also assumes that the instances of the objects associated with 'dog' is a *random sampling* from the corresponding cluster. However, in my opinion, this assumption is not required for the learning mechanism.

similar to a brown cow or a brown log. Thus the size of the region of object space that is associated with a colour term is considerably larger and more weakly clustered than those for nouns. Consequently, more instances of objects with a particular colour are required for a child to learn the appropriate extension of the corresponding colour word.

It is only when children have gone through a dimensionalization that separates out a particular class of properties, say colours, that the child can learn to see similarities *with respect to* colours and thereby learn the meanings of colour terms. When the colour domain is focused on, brown things form a cluster in this domain and this cluster determines a region of the domain. Thus the learning strategy used to generate children's early conceptual space offer, via the size principle, an explanation of why the meanings of nouns are easier to learn than the meanings of adjectives. A seemingly counterintuitive fact is that the semantic 'thickness' of nouns actually contributes to making the size of the corresponding concepts smaller. However, this fact contains the solution to the complex-first paradox.

This argument also explains the finding from Mintz and Gleitman (2002), that if the adjective comes together with a noun, then even young children can learn the meanings of new adjectives quickly. In this case the colour domain must be identified as a substructure *within* the region of object space associated with the noun. For example, brown shoes forms a sub-cluster among shoes that can be distinguished from clusters of black, blue and red shoes. This task is cognitively considerably easier than learning to identify the colour contrasts between all objects, which would amount to identifying the colour domain in the full object space.

Poth (2016) formulates her arguments in a Bayesian framework. In this section I have tried to show that the central idea of her solution can be formulated without relying on probabilities – using sizes of regions is sufficient. Instead of probabilistic representations, it therefore seems possible to rely directly on the structure of the underlying conceptual space (see Gärdenfors 2000, 2014).

7 Conclusion

The main question I have addressed in this article is how the infant mind develops from the initial 'blooming buzzing confusion' to a mind full of sensory concepts and categories. I have outlined a process that has three main steps:

(1)   The brain reduces sensory information into more manageable structures. The most efficient way to do this is to extract different kinds of *invariants*. I have argued that by identifying such invariants, primary perceptual domains are constructed, at least those related to space, objects and actions. The knowledge domains can be modelled as conceptual spaces that reflect similarity judgments.

(2)   Once the primary domains are in place, the brain is efficient in finding covariances of different features. Such covariances generate clusters of entities. These clusters then determine regions of the underlying conceptual space and the regions can be taken as the intensions of the concepts. This analysis also explains that when certain instances are more central in the regions, they are perceived as being more prototypical.

(3)   A part of the sensory input is the language spoken around the infant. These sound patterns form part of the data for detecting covariances so the infants learns to bring in sound patterns (or other communicative signs) as part of the cluster formation. Thereby the infant eventually learns to associate sound patterns with concepts. I am aware that this form of word learning is not the full story of language acquisition, but it forms a seed for coupling words to meanings that can later be expanded by other methods (see Bloom 2000).

In this article, I have focussed on the primary domains concerning space, objects and actions. There are, however, other domains that should be considered when studying how sensory concepts are learned. I conclude by briefly presenting some of the main candidates that I leave for further analysis in the future.

A first example is the domain of *numbers* that has been proposed by several researchers (Dehaene 1996; Spelke 2000, 2004; Carey 2009). Number cognition can be divided into two subsystems: approximate magnitudes and discrete numbers (Dehaene 1996). It should be noted that numbers relate to *collections* of objects and thus to a different ontological category. Furthermore, it is clear that both approximate and discrete numbers are governed by invariances (Harbour 2014). For example, the number of objects in a collection is invariant under the spatial location of the objects and under replacement of one object by another.

I would also like to suggest *events* as a fundamental domain for structuring sensory information (see also Strickland 2017). Already Gibson (1979, 100) describes events as primary realities. More recently, Radvansky and Zacks (2014, Ch. 10) present a review of experiments concerning children's development of event cognition. I have argued that the semantic reference of a basic sentence is an event (Gärdenfors 2014). This explains why sentences are natural units in language. Knowledge about event structure brings in the core 'thematic roles' – agent, patient, recipient, instrument, cause and effect – that help the child understand the construction of sentences. For example, Papafragou (2015, 338) compares how speakers of Greek and English describe events and she concludes: "Basic patterns in event perception are independent from one's native languages". It is also clear that our understanding of causality is related to event structure (Gärdenfors and Warglien 2012; Warglien et al. 2012; Gärdenfors 2014). Given all this, it would be an interesting task to find out what are the central invariants in our perception of events.

It is often proposed that cognitive representations of events presupposes representing time. Consequently, *time* would be an even more primary domain. However, the abstract conceptual domain of time is not culturally universal, but the product of systems for measuring time intervals, and hence a socio-historical construction (Sinha and Gärdenfors 2014). In addition to this argument, children understand events earlier than they understand time as a separate entity, which supports my claim that knowledge about event structures is more primitive.

References

Abdi, H., and L.J. Williams. 2010. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics* 2 (4): 433–459.

Agrawal, P., Carreira, J., and Malik, J. 2015. Learning to see by moving, *The IEEE International Conference on Computer Vision, 2015,* 37–45.

Billman, D. O. 1983. *Procedures for learning syntactic structure: A model and test with artificial grammars.* Dissertation, University of Michigan.

Billman, D.O., and E. Heit. 1988. Observational learning from internal feedback: A simulation of an adaptive learning method. *Cognitive Science* 12: 587–825.

Billman, D.O., and J. Knutson. 1996. Unsupervised concept learning and value systematicity: A complex whole aids learning the parts. *Journal of Experimental Psychology: Learning, Memory and Cognition* 22: 458–475.

Bloom, P. 2000. *How children learn the meaning of words.* Cambridge: MIT Press.

Borg, I., and P.J. Groenen. 2005. *Modern multidimensional scaling: Theory and applications.* Berlin: Springer Science & Business Media.

Browman, C.P., and L.M. Goldstein. 1990. Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics* 18: 299–320.

Carey, S. 1985. *Conceptual change in childhood.* Cambridge: MIT Press.

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

Croft, W. 2001. *Radical construction grammar: Syntactic theory in typological perspective.* Oxford: Oxford University Press.

Cutting, J.E. 1986. *Perception with an eye for motion.* Cambridge: MIT Press.

Dehaene, S. 1996. *The number sense. How the mind creates mathematics.* Oxford: Oxford University Press.

Dromi, E. 1987. *Early lexical development.* New York: Cambridge University Press.

Filippone, M., F. Camastra, and F. Masulli. 2008. A survey of kernel and spectral methods for clustering. *Pattern Recognition* 41 (1): 176–190.

Fölster, A., and Hansson, J. 2017. Tidigt ordförråd och formigenkänningsförmåga kan förutsäga språklig förmåga i 6–7 årsåldern. Master thesis, Department of Logopedy, phoniatrics and audiology, Lund University.

Gallistel, C.R. 1990. *The organization of learning.* Cambridge: MIT Press.

Gärdenfors, P. 1990. Induction, conceptual spaces and AI. *Philosophy of Science*, *57*(1), 78–95.

Gärdenfors, P. 2000. *Conceptual spaces: The geometry of thought.* Cambridge: MIT Press.

Gärdenfors, P. 2003. *How Homo became Sapiens: On the evolution of thnking.* Oxford: Oxford University Press.

Gärdenfors, P. 2007. Representing actions and functional properties in conceptual spaces. In Body, Language and Mind, Volume 1: Embodiment, ed. by T. Ziemke, J. Zlatev and R.M. Frank, 167–195. Mouton de Gruyter: Berlin.

Gärdenfors, P. 2014. *Geometry of meaning: Semantics based on conceptual spaces.* Cambridge: MIT Press.

Gärdenfors, P., and M. Warglien. 2012. Using conceptual spaces to model actions and events. *Journal of Semantics*, *29*, 487–519.

Gärdenfors, P., and S. Löhndorf. 2013. "What is a domain? – Dimensional structure versus meronomic relations". *Cognitive Linguistics*, *24*(3), 437–456.

Gharaee, Z., P. Gärdenfors, and M. Johnsson. 2017 First and second order dynamics in a hierarchical SOM system for action recognition. *Applied Soft Computing*, *59*, 574–585.

Gibson, J.J. 1966. *The senses considered as perceptual systems.* Oxford: Houghton Mifflin.

Gibson, J.J. 1979. *The ecological approach to visual perception.* Hillsdale: Lawrence Erlbaum.

Giese, M.A., and M. Lappe. 2002. Measurement of generalization fields for the recognition of biological motion. *Vision Research* 42: 1847–1858.

Giese, M., I. Thornton, and S. Edelman. 2008. Metrics of the perception of body movement. *Journal of Vision* 8: 1–18.

Goldstein, E.B. 1981. The ecology of J. J. *Gibson's perception, Leonardo* 14: 191–195.

Goldstone, R.L., and L. Barsalou. 1998. Reuniting perception and conception. *Cognition* 65: 231–262.

Gregory, R. 1970. *The intelligent eye*. London: Weidenfeld and Nicolson.

Harbour, D. 2014. Paucity, abundance, and the theory of number. *Language* 90 (1): 185–229.

Harnad, S. 1990. The symbol grounding problem. *Physica D* 42: 335–346.

Held, R., and A. Hein. 1963. Movement-produced stimulation in the development of visually guided behavior. *Journal of Comparative and Physiological Psychology* 56 (5): 872–876.

Hemeren, P.E. 1997. Typicality and context effects in action categories. In *Proceedings of the 19th Annual Conference of the Cognitive Science Society*, 949. Stanford: Lawrence Erlbaum Associates.

Hemeren, P. E. 2008. *Mind in action*. Lund: Lund University Cognitive Studies 140.

Humphrey, N.K. 1993. *A history of the mind*. London: Vintage Books.

Hunn, E. 1976. A measure of the degree of correspondence of folk to scientific biological classification. *American Ethnologist* 2: 309–327.

Jackson-Maldonado, D., D. Thal, V. Marchman, E. Bates, and V. Gutierrez-Clellen. 1993. Early lexical development in Spanish-speaking infants and toddlers. *Journal of Child Language* 20: 523–549.

Jäger, G. 2010. Natural color categories are convex sets, *Amsterdam Colloquium 2009, LNAI 6042*, 11–20.

James, W. 1890. *The principles of psychology*. New York: Holt.

Johansson, G. 1964. Perception of motion and changing form: A study of visual perception from continuous transformations of a solid angle of light at the eye. *Scandinavian Journal of Psychology* 5: 181–208.

Johansson, G. 1973. Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics* 14: 201–211.

Johansson, I. 2015. Collection as one-and-many: On the nature of numbers. *Grazer Philosophische Studien* 91: 17–58.

Kaila. E. 1939/2014. *Inhimillinen tieto,* Helsinki: Otava. English translation by A. Korhonen: *Human knowledge*. Chicago: Open Court.

Kaufman, L., and J. Kaufman. 2000. Explaining the moon illusion. *Proceedings of the National Academy of Sciences* 97: 500–504.

Kohler, I. 1951. Formation and transformation of the perceptual world. *Psychological Issues* 3 (4): 1–173.

Kornblith, H. 1993. *Inductive inference and its natural ground: An essay in naturalistic epistemology.* Cambridge: MIT Press.

Kruskal, J.B., and M. Wish. 1978. *Multidimensional scaling*. Thousand Oaks: Sage Publising.

Kumaran, D., D. Hassabis, and J.L. McClelland. 2016. What learning systems do intelligent agents need? Complementary learning systems theory updated. *Trends in Cognitive Science* 20: 512–534.

Land, E.H. 1977. The retinex theory of color vision. *Scientific American* 237 (6): 108–128.

Landau, B., L. Smith, and S. Jones. 1998. Object perception and object naming in early development. *Trends in Cognitive Science* 2: 19–24.

Lupyan, G., and R. Dale. 2010. Language structure is partly determined by social structure. *PLoS One* 5 (1): e8559. https://doi.org/10.1371/journal.pone.0008559.

Maddox, W.T. 1992. Perceptual and decisional separability. In *Multidimensional Models of Perception and Cognition*, ed. G.F. Ashby, 147–180. Hillsdale: Lawrence Erlbaum.

Mandler, J. M. 1991. Prelinguistic primitives. *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on The Grammar of Event Structure* (pp. 414–425). Berkeley, C: Berkeley Linguistics Society.

Mandler, J.M. 2004. *The foundations of mind: Origins of conceptual thought*. New York: Oxford University Press.

Marr, D. 1982. *Vision: A computational approach*. San Fransisco: Freeman.

Marr, D., and Vaina, L. 1982. Representation and recognition of the movements of shapes. *Proceedings of the Royal Society in London*, B214, 501–524.

Melara, R.D. 1992. The concept of perceptual similarity: From psychophysics to cognitive psychology. In *Psychophysical Approaches to Cognition*, ed. D. Algom, 303–388. Elsevier: Amsterdam.

Mervis, C., and E. Rosch. 1981. Categorization of natural objects. *Annual Review of Psychology* 32: 89–115.

Mintz, T.B., and L.R. Gleitman. 2002. Adjectives really do modify nouns: The incremental and restricted nature of early adjective acquisition. *Cognition* 84: 267–293.

Murray, S.O., H. Boyaci, and D.J. Kersten. 2005. The emergence of object size invariance in the human visual cortex. *Journal of Vision* 5: 744–744. https://doi.org/10.1167/5.8.744.

Papafragou, A. 2015. The representation of events in language and cognition. In E. Margolis, & S. Laurence (Eds.) The Conceptual Mind: New Directions in the Study of Concepts. Cambridge: MIT Press.

Piaget, J. 1954. *The construction of reality in the child*. New York: Basic Books.

Piaget, J. 1972. *The psychology of the child*. New York: Basic Books.

Pinker, S. 2002. *The blank slate: The modern denial of human nature*. New York: Viking.

Poth, N. 2016. A Bayesian approach towards concept learning, Master thesis, Department of Psychology, Bohr University Bochum.

Radvansky, G.A., and J.M. Zacks. 2014. *Event cognition*. Oxford: Oxford University Press.

Rosch, E. 1975. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General* 104: 192–233.

Rosch, E. 1978. Prototype classification and logical classification: the two systems. In *New trends in cognitive representation: Challenges to Piaget's theory*, ed. E. Scholnik, 73–86. Hillsdale: Lawrence Erlbaum Associates.

Runesson, S. 1994. Perception of biological motion: The KSD-principle and the implications of a distal versus proximal approach. In *Perceiving evens and objects*, ed. G. Jansson, S.-S. Bergström, and W. Epstein, 383–405. Hillsdale: Lawrence Erlbaum.

Sandhofer, C., and L.B. Smith. 2007. Learning adjectives in the real world: How learning nouns impedes learning adjectives. *Language Learning and Development* 3 (3): 233–267.

Sinha, C., and P. Gärdenfors. 2014. Time, space, and events in language and cognition: a comparative view. *Annals of the New York Academy of Sciences*, *1326*(1), 72–81.

Smith, L.B. 1989. From global similarities to kinds of similarities – the construction of dimensions in development. In S. Vosniadou, S., & Ortony, A. (Eds.), *Similarity and analogical reasoning* (pp. 146–178). Cambridge: Cambridge University Press.

Smith, L. 1995. Self-organizing processes on learning to learn words: Development is not induction. In *Basic and Applied Perspectives on Learning, Cognition, and Development*, ed. C.A. Nelson, vol. 28, 1–32. Mahwah: Lawrence Erlbaum.

Smith, L.B. 2005. Action alters shape categories. *Cognitive Science* 29: 665–679.

Smith, L.B., and M.D. Sera. 1992. A developmental analysis of the polar structure of dimensions. *Cognitive Psychology* 24: 99–142.

Spelke, E. S. 2000. Core knowledge. *American Psychologist*, *November 2000*, 1233–1243.

Spelke, E.S. 2004. Core knowledge. In *Attention and performance, vol. 20: Functional neuroimaging of visual cognition*, ed. N. Kanwisher and J. Duncan. Oxford: Oxford University Press.

Spelke, E.S., K. Breinlinger, J. Macomber, and K. Jacobson. 1992. Origins of knowledge. *Psychological Review* 99: 605–632.

Spelke, E.S., and K.D. Kinzler. 2007. Core knowledge. *Developmental Science* 10 (1): 89–96. https://doi.org/10.1111/j.1467-7687.2007.00569.x.

Strickland, B. 2017. Language reflects "core" cognition: A new theory about the origin of cross-linguistic regularities. *Cognitive Science* 41: 70–101.

Thelen, E., and L.B. Smith. 1994. *A dynamic systems approach to the development of cognition and action*. Cambridge: MIT Press.

Tversky, B., and K. Hemenway. 1984. Objects, parts, and categories. *Journal of Experimental Psychology: General* 113: 169–191.

Wang, W., R.H. Crompton, T.S. Carey, M.M. Günther, Y. Li, R. Savage, and W.I. Sellers. 2004. Comparison of inverse-dynamics musculo-skeletal models of AL 288-1 Australopithecus afarensis and KNM-WT 15000 Homo ergaster to modern humans, with implications for the evolution of bipedalism. *Journal of Human Evolution* 47: 453–478.

Warglien, M., P. Gärdenfors, and M. Westera. 2012. Event structure, conceptual spaces and the semantics of verbs. *Theoretical Linguistics*, *38*, 159–193.

Waxman, S.R., and D.B. Markow. 1998. Object properties and object kind: Twenty-one-month-old infants' extension of novel adjectives. *Child Development* 69: 1313–1329.

Werning, M. 2010. Complex first? On the evolutionary and developmental priority of semantically thick words. *Philosophy of Science* 77 (5): 1096–1108.

Wiskott, L., and T.J. Sejnowski. 2002. Slow feature analysis: Unsupervised learning of invariances. *Neural Computation* 14: 715–770.

Wolff, P. 2008. Dynamics and the perception of causal events. In *Understanding events: How humans see, represent, and act on events*, ed. T. Shipley and J. Zacks, 555–587. Oxford: Oxford University Press.

Xu, F., and J.B. Tenenbaum. 2007. Sensitivity to sampling in Bayesian word learning. *Developmental Science* 10 (3): 288–297.

Zhu, S.C., and A.L. Yuille. 1996. FORMS: A flexible object recognition and modelling system. *International Journal of Computer Vision* 20: 187–212.

Zwarts, J., and P. Gärdenfors. 2016. Locative and directional prepositions in conceptual spaces: The role of polar convexity. *Journal of Logic, Language and Information*, *25*, 109–138.