



Primary Cognitive Categories Are Determined by Their Invariances

Peter Gärdenfors^{1,2*}

¹ Cognitive Science, Department of Philosophy, Lund University, Lund, Sweden, ² Faculty of Humanities, Palaeo-Research Institute, University of Johannesburg, Johannesburg, South Africa

The world as we perceive it is structured into objects, actions and places that form parts of events. In this article, my aim is to explain why these categories are cognitively primary. From an empiricist and evolutionary standpoint, it is argued that the reduction of the complexity of sensory signals is based on the brain's capacity to identify various types of invariances that are evolutionarily relevant for the activities of the organism. The first aim of the article is to explain why places, object and actions are primary cognitive categories in our constructions of the external world. It is shown that the invariances that determine these categories have their separate characteristics and that they are, by and large, independent of each other. This separation is supported by what is known about the neural mechanisms. The second aim is to show that the category of events can be analyzed as being constituted of the primary categories. The category of numbers is briefly discussed. Some implications for computational models of the categories are also presented.

Keywords: category, invariance, space, object, place, event, number

OPEN ACCESS

Edited by:

Guy Dove,
University of Louisville, United States

Reviewed by:

Ute Schmid,
University of Bamberg, Germany
Daniel Weiskopf,
Georgia State University,
United States

*Correspondence:

Peter Gärdenfors
peter.gardenfors@lucs.lu.se

Specialty section:

This article was submitted to
Cognitive Science,
a section of the journal
Frontiers in Psychology

Received: 16 July 2020

Accepted: 13 November 2020

Published: 08 December 2020

Citation:

Gärdenfors P (2020) Primary
Cognitive Categories Are Determined
by Their Invariances.
Front. Psychol. 11:584017.
doi: 10.3389/fpsyg.2020.584017

WHAT DETERMINES THE CATEGORICAL STRUCTURE OF OUR PERCEPTIONS?

The world as we perceive it is structured into *objects*, *places*, and *actions* that form parts of *events*. We have a strong tendency to be realists, that is, to believe that these categories exist out there in the world. Kant taught us, however, to distinguish between “das Ding an sich” and “das Ding für uns.” According to him, and much of modern cognitive science (e.g., Marr, 1982; Humphrey, 1993; Anderson et al., 1998; Von Glasersfeld, 2005; Hoffman, 2019), we cannot know external reality but only how our minds construct the world. For such a constructivist position, a fundamental question is why our mental constructs end up with categories of objects, places and actions. The answer, as always, should be grounded in the evolutionary mechanisms that have molded our perceptual systems and in how the brain handles the information presented by these systems.

The senses generate an extremely rich and unstructured mass of signals. When trying to understand what happens to the sensory information in our brains, it is standard to distinguish between *sensations* and *perceptions*. Our subjective world is full of colors and patterns that we see, things that we taste and smell, itches, pains, and sensations of cold that we feel. In philosophy such sensations are called *qualia*. The evolutionary value of sensations is that they inform us about what is happening right now to our bodies (Humphrey, 1993).

An individual that also receives signals about what is going on in the world and not only what is happening to its body will be better prepared to foresee the future and thus to survive in a challenging environment. This is the purpose of perceptions. In order to make sense of

the sensations, the perceptions result from processes in the brain that reduce their complexity by structuring them into kinds of entities. In this article, I argue that this complexity reduction is based on the brain's capacity to identify various types of *invariances* in the sensory signals—invariances that are evolutionarily relevant for the activities of the organism. My aims are, firstly, to explain why places, objects and actions are *primary cognitive categories* in our constructions of the external world, and, secondly, how these components generate cognitive representations of *events*.

Traditionally, there are two approaches to the functioning of the mechanisms of our brain that generate the primary categories: (1) nativism: the categories are *innate*; and (2) empiricism: the categories are *learned*. Spelke and Carey (Spelke, 2000, 2004; Spelke and Kinzler, 2007; Carey, 2009) propose objects, actions, space and numbers as “core knowledge domains,” which form the framework of perceptual categories. They defend a nativist position in relation to child development. In contrast, my solution will be empiricist (Gärdenfors, 2018), although I will suggest that the structure of the brain imposes constraints on how the categories are learned. There is thus a nativist element in my analysis, albeit of a different kind than that advocated by Spelke and Carey. Following them, I will also briefly discuss to what extent numbers form another primary cognitive category.

A part of an evolutionarily grounded argument builds on the fact that human (and other mammal's) infants are not born as blank slates (Pinker, 2002). By evolutionary processes, the brain is prepared to pick the most relevant invariances (see e.g., Leibo et al., 2015). As examples of how the brain organizes invariances, the dorsal stream of the cortex handles space representation (the where pathway), the ventral stream generates object representation (what pathway) and the dorsal stream accounts for action representation in (how pathway). Even though these pathways are to some extent neurologically given, the infant must, however, *learn* to identify the invariances that create the most relevant cognitive categories. After the invariances have been learned, the plasticity of the cortex still supports considerable relearning: An amazing example is that a person who is given goggles turning the visual field upside-down, will, after a few weeks, be able to relearn the projection from the visual cortex so that the perceived world is “normal” again (Kohler, 1951).

The strong capacity to detect invariances that the brain has, leads to the crucial question concerning which cognitive categories that are the most fundamental. A central question for the analysis becomes: *Why* are the invariances that determine places, objects and actions cognitively primary?¹ This is, in a sense, a neo-Kantian epistemological question, seeking the “forms of perception” (“Anschauungsformen”) that generate the framework for more specific categorizations.

By using an analysis in terms of invariances, I will show that each of the categories of places, objects and actions has its separate characteristics and that they are, by and large,

independent of each other. A preliminary attempt to identify primary cognitive categories in terms of invariances for space, objects and actions was made in Gärdenfors (2018). That paper dealt with two learning processes: how the primary categories are learned and how concepts that are grounded in the categories are learned. This paper presents a more detailed analysis of the role of invariances and also analyses the categories of numbers and events.

EXTRACTING STRUCTURE: INVARIANCES IN PERCEPTION

The primary categories build up our perceptual structures. My thesis is that the sensations, at an early stage of the process in the brain, become perceptions that are organized along primary ontological categories, in particular space, objects and actions. By saying that the categories are primary, I mean that they form the fundament from which specific concepts are constructed, for example, places as regions of the space, object categories as determined by specific properties or part-whole relations, etc. Since they are founded in the mechanisms of the human brain, they are also seen as common to all humans.

My approach to perception is in some respects similar to Gibson's (1966, 1979) “ecological approach.” 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). A useful metaphor is that the brain *resonates* with the sensory information. (Gibson, 1966, 201) defines an invariant as a “non-change” that persists during change. This definition is not very useful for identifying invariances so I will instead rely on well-known types of invariances, some taken from physics and some from analyses of children's cognitive processes. Following Breidbach and Jost (2006), I outline in this section how a theory of perceptual invariances can explain our primary categories. A central type of perceptual information is what remains invariant when an agent moves through the environment and interacts with objects in it (see also Cutting, 1986).

Unlike Gibson, I take a constructivist position and do not claim that invariances are “out there,” ready to be “picked up” by the brain. In contrast, I view invariances as something that is constructed by various processes in the brain. Not all possible invariances are constructed—only those that are relevant for survival. Over the millennia, evolution has selected the invariances that are most salient for the activities of the organism.

One central notion for the analysis of invariances is *fungibility*², that is, replacements of equivalents. For example, a place remains the same independently of which objects are located at the place. In other words, objects are fungible with respect to places. Similarly, an object remains the same independently of which place it is allocated at, so places are fungible with respect to objects. These two types of fungibility form the main reason why the place and the object categories are independent³.

¹I avoid Spelke's use of “core” knowledge structures [and Carey's (2009) “core” cognition] since it is connected with an nativist position, and instead speak of primary categories (see Gärdenfors, 2018).

²This term is borrowed from economics.

³Fungibility will also be central in my analysis of the number category.

SPACE

According to Gibson's approach, the visual field is determined from invariances such as texture gradients, occlusions and visual flow. To a large extent the visual flow is determined by the movements of our bodies. Turning our heads and letting our eyes follow along, for example, leads to vary rapid changes in the image that reaches the retina. However, our brain simultaneously produces a representation of the surrounding space that remains still relative to the direction of our body.

During the first months of life, an infant learns how to coordinate sensory input—vision, hearing, and touch—with motor activities (Thelen and Smith, 1994). The infant engages in “motor babbling” that generates an egocentric representation of space, coordinating it with its actions. As Gibson (1979: 2) writes, “the environment to be perceived [...] is not the world of physics but the world at the level of ecology.” The space we perceive can be divided into *peripersonal* space—the region immediately surrounding our bodies (di Pellegrino and Làdavas, 2015)—and *extrapersonal* space, which is the space beyond our reach.

The peripersonal space makes it possible for an individual to see its *field of action*. Moving only the head and not the rest of the body, an individual's potential to act does not change. Since the hand actions of the individual occur in front of the body, it's more efficient if the brain creates a space that is constant in relation to the body direction. The peripersonal representation of space is therefore invariant of the direction of the eyes and the head. The space that is constructed is a three-dimensional space where the body determines its origo and principal direction⁴.

The representation of visual space then expands during the child's development. Firstly, when the auditory input is coordinated with the visual, the represented space extends beyond the child's current visual field to cover the entire surrounding space. The child is then able to direct its attention outside its peripersonal field and it becomes extrapersonal. Importantly, the egocentric representation of space that results from this extension is no longer just visual, but an *amodal* representation based on visual, auditory, tactile, and perhaps even olfactory sensations.

The adult visuo-spatial category should thus be seen as a combination of a peripersonal and an extrapersonal space. The two representations have different basic functions: The peripersonal is used for reaching and interacting with objects, and the extrapersonal for surveillance and navigation (Gallistel, 1990).

There are several experiments supporting that the space category is not an innate structure. It must be learned through *interaction* with the world, where a first step is eye-hand coordination (e.g., Held and Hein, 1963; Agrawal et al., 2015). This process must learn how visual (and auditory) information can be used to create meaningful fields of action. For example, getting a new pair of glasses with stronger lenses changes the

conditions for this process. Further experience is required before the brain has construed an adjusted space and can provide the perceptions needed for carrying out precise actions, such as walking down stairs without stumbling.

A second extension of the space representation involves the ability to represent an *allocentric* space. This is an *imagined* space where the location of the individual is no longer a fixed point. The allocentric representation makes it possible for the an individual to abandon the egocentric perspectives and instead imagine how the world looks like from another point of view⁵. The allocentric space representation is not just invariant of eye and head orientation but also of the *orientation and location* of the body. The primary role of the allocentric space is to allow planning for movements through space. Piaget and Inhelder's (1967) three mountain test was developed to determine when children master problem solving using representations of allocentric space. For a survey of how humans represent space, see Tversky (2003).

In the brain, a self-centered representation of location is transformed into an allocentric representation by a network involving the posterior parietal cortex, the medial retrosplenial complex and the hippocampal formation (hippocampus and entorhinal cortex) (Nau et al., 2018). The allocentric representation in the hippocampal formation then projects allocentric coordinates back to guide navigation.

Importantly, by extracting the various forms of invariances, the egocentric and allocentric spaces that are generated considerably *reduce* the complexity of the information that hits the retinas. If the constructed allocentric space were perfectly invariant under rotations and translations (so-called Galilean transformations, Levy-Leblond, 1971), it would follow that the resulting visual space is three-dimensional Euclidean. However, since our movements mainly take place in the two horizontal dimensions, the vertical dimension is less important for our perception. Consequently, our perception of the vertical dimension is “flattened” (Kaufman and Kaufman, 2000).

An important aspect of the representation of space is that it is invariant of *time*. When we move or turn around, we perform rotational and translational transformations of the perceptual input. If these transformations were not invariant over time, it would not be possible to use the represented space as a basis for actions. This point was made already by Gibson (1966, 264): “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).

The domain of space can be divided into regions or *places*. The identity of a place is determined by its relation to a set of *landmarks* and not by its location in relation to some fixed coordinate system. For example, from my perspective your location may be in the passenger seat of my car that is moving through the landscape. The landmark is the car that determines the relative places inside it. For an extreme case, consider that the earth is rotating around the sun at a very high speed.

⁴The space is generally thought to have a Euclidean geometry and be based on a Cartesian coordinate system, but there is some linguistic evidence that *polar* coordinates might give a better description of its perceptual geometry (e.g., Gärdenfors, 2014; Zwartz and Gärdenfors, 2016).

⁵The distinction between egocentric and allocentric corresponds to Gibson's (1966) distinction between “perspective structure” and “invariant structure.”

Nevertheless, we take the earth to be the landmark and say that Sweden is located in northern Europe.

A place is also, to a large extent, invariant of the objects located there⁶. If somebody else sits in the passenger seat of my car, it will still be the same place. If Sweden, due to severe climate changes, turns into a desert, its identity as a place does not change. As mentioned earlier, we can say that objects are *fungible* with respect to places. Similarly, actions are fungible with respect to places—the identity of a place does not depend on what is done there.

Sometimes, other properties than a set of landmarks are used to identify a place, for example its *function*. For example, in 1988 the Australian parliament moved from its old house to a new one in Canberra. Still one can refer to “the parliament” as a location. A more exotic example is that the entire town of Kiruna in northern Sweden will be moved two miles to the east because there is a risk that the extensive iron mining under the town will lead to a collapse of the ground. New streets will be laid out and many of the historic houses and official buildings will be moved to the new location and, but the spatial relations between the buildings will not be preserved. Still the identity of the town will be preserved for most practical purposes.

OBJECTS

There are many kinds of objects, but I will focus on physical objects, since they have been the most important in the evolution of our cognitive systems. A central property of physical objects is that they have a *shape* (although it may vary over time). This means that the relative locations of different parts of an object can be described in terms of different types of invariances. For a rigid object, the invariances are total. The directions of the parts may change as the object moves, but all the spatial relations between the parts are invariant. For an object with movable parts such as animals, the relations between the locations within each part is more or less invariant and so are the relative locations of the points where the different parts are connected⁷. For example, the parts of your upper leg don't change their relative distances and the connection point between your leg and your body remains invariant. Johansson (1964) calls this type of invariance the “rigidity principle” that functions as a constraint of the visual process: Whenever equal motions in a series of simultaneous proximal elements are detected, the result is a perception of rigidity. Marr (1982) uses this principle extensively in his representation of shapes (for a computationally implemented model see Zhu and Yuille, 1996).

In addition to rigidity or relative rigidity, there are many other types of invariants that apply to objects. The *size* of an object is, for example, typically invariant—at least over short periods of time. This invariance makes it possible to accurately judge the distance to an object. Murray et al. (2005) show that size invariance has been picked up already in the dorsal retinotopic

visual area V3. Another property exhibiting invariance is *color*. For many kinds of objects, for example, different species of birds, the patterns of colors are characteristic features. The absolute colors of objects are not invariant, however, since they vary with the illumination. However, the perceptual *relations* between the colors of an object are, in most cases, invariant (Land, 1977).

Some objects are deformable, for example cushions, towels and doughs. Even though invariances of relative locations are less stable for such objects, the changes of relative locations are still continuous. This is what distinguishes objects from masses. Another general type of invariance is that objects are *cohesive*: if you pull at one end of an object, the other parts will follow. Clouds, flames and shadows are therefore marginal as objects. Leslie (1996) argues that infants just a few months old perceive the world as consisting of cohesive objects that keep much of the same form even when moving.

Clouds, flames and shadows indicate that there are grades of objecthood: they have properties that make them come close to being masses rather than objects. The characteristic distinctions between masses and objects is that masses, such as water and sand, (i) do not have a constant shape, (ii) are variable in size and (iii) are homogenous in material. Linguistically, the distinction shows up in that mass nouns are not countable—one does not say “two sands”—but nouns for objects are⁸.

Neuroscientific support for the thesis about invariances determining the object category is becoming stronger. In particular, Leibo et al. (2015) and Anselmi et al. (2016) present a neural model of object and face recognition based on invariances that builds on the idea that the main task of the ventral stream of visual processing is to compute a “signature for recognition” that is invariant of translations and rotations. They also show that when the relevant transformations have been learned for some objects it generalizes to other objects. For example, if we see a new face in a frontal position, we can accurately predict how it will look like if turned to the side. The grouping of objects is done by their transformation compatibility, that is, the class of transformations that preserve their identity. Another type of support comes from Kriegeskorte et al. (2008) who show that the inferior temporal cortex of monkeys and human share a common code for representing objects, in particular concerning major distinctions such as animate–inanimate and face–body. The response patterns in the cortex form category clusters that match between monkeys and humans.

The perception of objects also involves an extensive reduction of the dimensions of the sensory input. Several computational procedures for dimension reduction have been proposed, for example Principal Component Analysis (Abdi and Williams, 2010) and Multidimensional Scaling (Kruskal and Wish, 1978; Borg and Groenen, 2005). It is not known, however, how similar these procedures are to real brain processes. Wiskott and Sejnowski (2002) have developed an artificial neural network based on “slow feature analysis” that is able to pick up translation, size, rotation, illumination and contrast invariances of objects.

⁶If a place is determined by landmarks, however, the place will change if the landmarks change (Gallistel, 1990).

⁷When categorizing objects with parts, the relations of the parts can be modeled in a “structure space” (Fiorini et al., 2014).

⁸What counts as a mass noun is to some extent language dependent. For example, “furniture” (mass noun in English) is a count noun in French (meubles) and German (Möbel).

From a neuro-cognitive point of view, an interesting feature of the neural network is that the “what” and the “where” components become represented in separate components of the network. This provides indirect support for my hypothesis that the space and object invariances can be separated. The invariances that lead to the dimension reduction, both in Wiskott and Sejnowski’s model and in that of Anselmi et al. (2016), show that the dimensional structure that is represented is closely related to a 3D Euclidean space. This is congenial with proposals that the hippocampal formation is not solely used to represent spatial information, but is also exploited to represent other types of conceptual spaces (Eichenbaum and Cohen, 2014; Bellmund et al., 2018).

When describing how infants represent objects, Spelke et al. (1992, 606) suggest the following criteria: (i) *continuity* (objects move in continuous paths), (ii) *solidity* (objects move only on unobstructed paths and therefore different objects do not occupy the same place), (iii) *gravity* (objects fall downwards, if not supported), and (iv) *inertia* (objects do not change their motion abruptly).

Except for solidity, which I have discussed above, these constraints do not concern invariances of objects. The last two are not about objects *per se*, but rather describe the behavior of objects. Furthermore, objects that are *agents* violate the inertia constraint. Surprisingly, the list of criteria proposed by Spelke et al. (1992) does not contain shape, despite the fact that children’s categorizations of objects have a clear shape bias (e.g., Landau et al., 1998; Smith and Samuelson, 2006).

A consequence of the representation of the continuity of objects is *object permanence* (Piaget, 1952), which means that objects are represented as being located at the place where they were last perceived, even if they currently do not produce any sensations. This means that the object is represented (imagined) in the inner world as located at a particular place, even if it is not perceived. The ability to keep an object in mind is not innate; human infants acquire it around 5 months of age (which is later than among other animal species) (Baillargeon and DeVos, 1991).

ACTIONS

The third primary category of our perceptions involves actions. Humans are exceptionally efficient at categorizing actions. For example, it is easy to instantly judge whether somebody is walking or jogging, even if the movements of the body parts are rather similar. Furthermore, only a very limited amount of information is needed to make such a categorization. The efficiency of action perception was shown by Johansson in a series of classical perception studies in the 1950’s (Johansson, 1973). The patch-light technique that he invented for analysing biological motion contains no direct shape information. Light bulbs were attached to the joints of actors who were dressed in black and moved in a black room. The actors were performing different actions such as walking, running, and dancing while being filmed. Subjects who then watched the films saw the movements of the light bulbs (but nothing else).

They were able to correctly categorize the actions within a few hundred milliseconds.

Experiments of this kind indicate that that seeing the surfaces of the agents performing actions is not necessary for categorizing the actions (Hemerik, 2008). A movie that contains stick figures or only dots moving in the same way is sufficient. These observations give additional support to Johansson’s rigidity principle. The question now is what kind of invariances are involved in action categorizations.

Working in the tradition of Gibson, Runesson (1994, pp. 386–387; see also Wolff, 2008) argues that people can directly perceive the forces that generate different types of motion:

“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 color of his shirt in normal illumination, namely that *information* about all these properties is available in the optic array.”

Runesson formulates this as that the kinematics of an action is sufficient to identify the underlying force patterns. For example, the pattern of forces involved in saluting is different from the pattern of forces involved throwing even if the actions are perceptually rather similar. Johansson and Runesson mainly apply their principles to biological motion. I hypothesize, however, that they can be applied to other forms of action as well. I have argued that the brain extracts the invariances that represent the *forces* that generate different kinds of actions (Gärdenfors, 2007, 2014). The process extracting the invariances is automatic: an individual cannot help perceiving the forces (Wolff, 2008; Wolff and Shepard, 2013; Wolff and Thorstad, 2017). Just as for objects, the space of force patterns can therefore be seen as a perceptual category with a unique structure of similarities and defined by its own class of invariances. 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 example of an empirical study of force patterns is that of Wang et al. (2004). Based on data from the walking patterns of humans collected under different conditions and using the methods of Giese et al. (2008), the force patterns that were extracted were used to calculate the similarity of the different types of walking⁹.

A particular action is, of course, performed by a particular agent (a special kind of object) at a particular place. For the categorization of an action, however, a central invariance is that only the forces, but *not* the individuals or objects performing the action, are involved in the representation of the action. More generally, patterns of forces should be considered since several body parts are typically involved; and several force vectors are consequently interacting. This is analogous to Marr and Vaina’s (1982) differential equations for actions. Such force patterns form the invariances that I submit generate the structure of action categories. However, the invariances that apply to actions are neither the same as those for objects, nor for those for space.

⁹Gharaee et al. (2017a) have applied the force dynamic model in a robotic system that has been constructed for categorizing actions.

To wit, the patterns for actions are neither dependent on the location of the acting object, nor on its object properties such as color or weight. This means that the objects and places are fungible with respect to actions and thus that the action category is independent of the object and space categories. In line with the situation for space and objects, the force patterns determined by the invariances involve a considerable reduction in dimensions. However, the empirical data concerning how actions are perceived is still limited so the precise structure of action space should be further investigated.

Human understanding of actions, however, does not only involve physical movements and their underlying forces, but often also the *intention* behind the action. For example, “blink” and “wink” cover the same kinds of physical eye movements, but the second action is intentional. Accounting for the intentionality of actions also involves representations of a *goal space* in the agent that is attributing the intention (Gärdenfors, 2014, pp. 194–197). It might be argued that such a goal space should also be included in among the primitive cognitive categories. The main reason for not counting the goal space to the primary categories is that representing intentional actions presumes the capacity to represent actions. This position is supported by recent experiments by Ganglmayer et al. (2019). In contrast to what has been claimed previously (Woodward, 2009, 2013), their results indicate that 11-12-month-old infants anticipate the movement path rather than the goal of an action.

To sum up: The three basic categories place, object and action are mutually fungible relative to each other. As a consequence these three categories are, to a large extent, cognitively independent: Space can be characterized independently of the objects and actions present; objects can be characterized independently of where they are located and which actions are performed on them; and actions can be characterized independently of where they are performed and who (what) performs them. These mutual invariances support my thesis that they form independent primary categories for our cognitive processes.

Following the strategy in Breidbach and Jost (2006), *sub-categories* can then be identified by adding the relevant invariances that characterize them. I have already mentioned the distinction between rigid and non-rigid objects, where the rigid objects are characterized by all distances between points on an object being invariant over time. Another example is the distinction between agents and non-agents, where agents are characterized as being objects that are capable of exerting forces. This distinction will be relevant for the model of events that will be presented below.

The primary categories show up in the structure of language, in particular in how it divides words into classes. Gärdenfors (2014, 2018) has argued in some detail that semantic representations of nouns build on the category of objects, and that verbs build on actions. Furthermore, many prepositions express spatial relations. Different languages have different word classes, but all of them have means to denote objects, actions, and spatial relations. This universality of linguistic structure is a further indication that these categories are indeed cognitively primary.

EVENTS

Even though space, object and actions form categorical structures that are determined by separate sets of invariances, it is obvious that there are interactions between these categories. They are all parts of *events*. Therefore, I suggest events as an overarching category for combining different perceptual categories (see also Strickland, 2017). Already Gibson (1979, 100) describes events as primary realities. There is an extensive amount of research on how children’s event cognition develops (e.g., Radvansky and Zacks, 2014, Ch. 10; Papafragou, 2015).

The cognitive structure of events is relational, gluing together objects, actions and locations. In earlier work (Gärdenfors and Warglien, 2012; Warglien et al., 2012; Gärdenfors, 2014), I have suggested an approach to event categorization based on some geometric notions. The key idea is to represent event structures in terms of conceptual spaces—one for actions and one for results—and *mappings* between these spaces (see **Figure 1**).

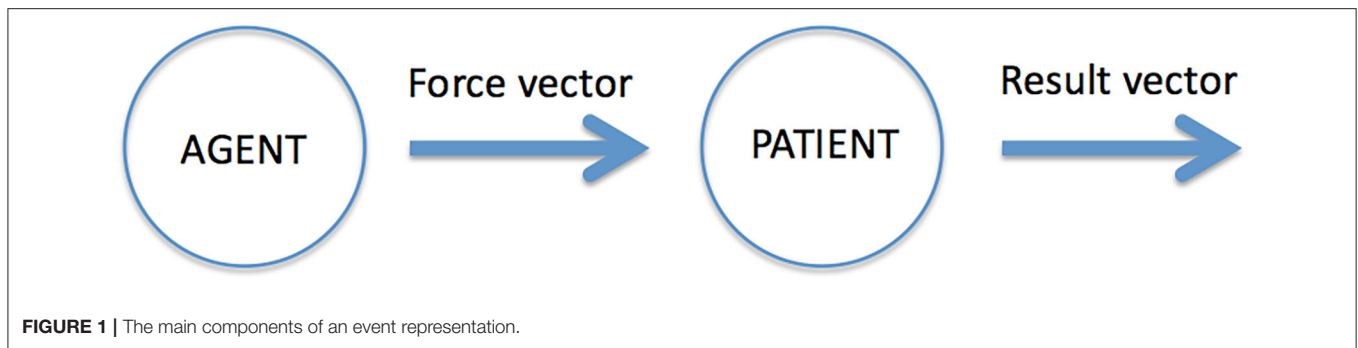
Following the previous section, the action space is represented as a space of forces (or force patterns) acting upon some object. As mentioned above, I view non-intentional actions as primary. Modeling intentional actions would require adding a goal space to represent the aim of the action. The result space of the event represents changes in the properties of the target. This space can therefore be modeled as a vector space where the two ends of a result vector represent the properties of the object acted upon before and after the action¹⁰. The results of actions are typically changes of location (that is, the space category) or changes of object properties. For example, when Donald pushes the table, the agent Donald exerts a force vector (action) on the table that leads to a change of the position of the table (result). Or in the event of heavy rain undermining a road, the force of the rain (action) leads to a change of the shape property of the road (result). More complicated to represent mathematically are events of breaking or dividing when the object acted upon changes into two or more, and events of construction where different objects are combined into a new one¹¹.

A consequence of characterizing an event as a combination of an action space and a result space is that the time domain is not defining for events, but it emerges from the relations between the components of an event. This position contrasts with, for example, Zacks and Tversky (2001) who focus on the temporal structure of events, in particular on how events are segmented. It is often suggested that cognitive representations of events presuppose representing time (Radvansky and Zacks, 2014; Hoerl and McCormack, 2019). For example, Zacks and Tversky (2001, p. 3) write that an archetypical event is “a segment of time at a given location that is conceived by an observer to have a beginning and an end.” If this were correct, time would also be primary category. Of course, the circadian system of our bodies in a sense represents the day and night cycle¹². However,

¹⁰Slightly more mathematically, an event can be represented as a product space of these two spaces.

¹¹Thom’s (1972) work on catastrophe theory presents a general way of characterizing such disruptive changes.

¹²Viera (2020) argues that the circadian system allows us to sense time, but he does not consider the role of the time dimension in cognitive event representation.



this system is inflexible and not involved in our representations of events. In fact, there is linguistic evidence that indicates that time is not a primary cognitive category: The abstract time dimension is not used by all human societies but it is the product of cultural systems for measuring time intervals, and hence time is a socio-historical construction (Sinha and Gärdenfors, 2014). For example, the South American language Amondawa does not have an explicit representation of time. This language employs a time interval system that represents seasonal and diurnal events, but it has no calendric terms, including terms such as month and year. Furthermore, children understand events earlier in their development than they understand time as a separate dimension. The model of events presented here does not explicitly represent the time dimension. However, temporality is implicit in the model since actions and events are dynamic entities—they unfold over time.

The action space and the result space represent different categories: forces have a different nature than changes in object properties. In the limiting case when the result vector is the null vector, that is, when nothing changes, the event is a *state*. As can be seen from this two-vector model of events, it combines the three primary categories of objects, actions and physical space into a relational structure: An event can be characterized as a mapping from an action on an object to a result.

In linguistics, the target entity of the event is called the *patient*. The object that creates the action vector is called the *agent*. The concept of an agent thus combines the object category with the action category. There exist, however, events without agents, for example events of falling, drowning, dying, growing and raining. An event may also include other “thematic roles” (Dowty, 1991), such as recipient and instrument, but they are not components of all events.

As for actions, a particular event is, of course caused by a particular agent (a special kind of object) at a particular place. An event category is, in general, invariant of the location where it is performed and on which object (patient) the action is performed. Gärdenfors and Warglien (2012) define an event category as a structure (product space) that represents the mapping from the action space to the result space. An example is the event category of pushing a table, which is constituted by the force (exerted by some agent, human or non-human) on the table resulting in a movement (change in space) of the table.

Causal relations can also be represented using the event structure (Wolff, 2007, 2008; Gärdenfors and Warglien, 2012; Gärdenfors, 2020a; Gärdenfors and Lombard, 2020): The action causes the result. Most accounts of causation analyse the relation between the action and the effect as a relation between two events (see e.g., Zacks and Tversky, 2001; Radvansky and Zacks, 2014). In contrast, the model presented here views causation as a relation *within* an event by introducing a distinction between forces and changes of states (cf. Wolff, 2007, 2008, 2012; Wolff and Thorstad, 2017). In contrast to many other theories, causes and effects are not treated as symmetrical entities: they belong to different categories—causes to the forces that are applied on objects and results to change in location (in the case of movements) or in some property of objects (color, size, weight, temperature, etc.).

The characteristic part of an event is the mapping between the force space and the result space. For example, pushing a table sometimes results in the table moving, sometimes not; aiming a dart at the bull’s eye sometimes hit it, sometimes not. In such cases the mapping between the force vector and the result vector represents two different events. Gärdenfors et al. (2018) analyse three general constraints on event mappings:

- (i) Larger forces lead to larger results (monotonicity constraint).
- (ii) Small changes in the force lead to small changes of the result (continuity constraint).
- (iii) Intermediate results are caused by intermediate forces (convexity preserving constraint).

Even though it is not the aim of the article to propose computational models of how various forms of invariances can be used in cognitive systems, the event model lends itself to some recommendations for how such models can be constructed [for more details, see Gärdenfors (2019, 2020b) and Gärdenfors et al. (2019)]. There exist several efficient methods for constructing a computational model of space from video, laser range and other forms of input (see e.g., Wyeth and Milford, 2009). Recent advances in deep learning have also led to good methods for object categorization (see e.g., Zhao et al., 2017). It should be noted that these methods depend on the *appearance* of the objects. For robotic interaction with objects, however, these aspects are not the most important. Gibson (1979, Ch. 8) writes that “what we perceive when we look at objects are their

affordances, not their qualities.” In other words, it is what we can *do* with objects that matter, not how they look. Shanahan et al. (2020) discuss this problem. As an example, they take the concept of a “container” that is central to much human interaction with the world. The appearance of containers can vary widely, but it is their affordances that are crucial for how we interact with them. There seems to be no good model of how to capture the affordances of objects from, say, a video stream (Shanahan et al., 2020). As regards actions, they are understudied in robotics. The attempts have focused on the results of actions. For example, the algorithms for learning verb meanings developed by Kalkan et al. (2014) are based on “affordance relations” between entities, behaviors, and effects. Most attempts to computationally categorize actions in terms of manner have been based on stored data, but Gharaee et al. (2017b) present an online, real time algorithm.

There thus exists partly successful work in computer science and robotics that generate models of each of the basic cognitive categories space, objects and action. However, there are very few models of how to combine these models to generate representations of events. The one that comes closest to the approach presented here is Hinaut and Dominey’s (2013) model of “reservoir computing.” In Gärdenfors (2020b), I make a programmatic attempt to describe a computational approach to events and illustrate it with a partial implementation, based on reservoir computing, in an iCub robot (Mealier et al., 2016).

Finally, a comment on the relation between event representations and language. Gärdenfors (2014) has argued that a declarative sentence typically describes the main components of an event. This thesis provides an explanation of why sentences are natural units in language. The event structure connects naturally to the core “thematic roles” —agent, patient, recipient, instrument, cause and effect, that help children understand how sentences are constructed and what their meanings are. For example, Papafragou (2015, 338) compares how speakers of Greek and English describe events and she concludes that basic patterns in event perception are independent of the language one speaks. Another example is Fernandes et al. (2006) who show that toddlers already in their third year have an understanding of the abstract categories “agent” and “patient.”

NUMBER

Another cognitive category is *number*. Even though I do not view it as primary, I will discuss it briefly since it belongs to the core knowledge domains proposed by Spelke (2000, 2004) and Carey (2009). Theories of number cognition distinguish between magnitude (“a large bag of beans”), numerosity (“many sheep”) and number (“five cows”) (Gemel and Quinon, 2019). The underlying cognitive processes are divided into two subsystems that handle approximate magnitudes and discrete numbers respectively (Dehaene, 1996). Non-human animals have an approximate number system that allow them to estimate the relative magnitude of two collections, sometimes with surprising precision (Gallistel, 1990). The discrete number system is used only by humans and it must be learned. There exist human cultures, for example the Amazonian tribe of Pirahã, who do not have a discrete number system (Everett, 2017). Thus, like

time, number is a cultural construct and not as fundamental cognitively as the space, object and action categories are. This goes against Spelke’s and Carey’s position that number is a core knowledge domain.

Nevertheless, approximate as well as discrete numbers are governed by invariances (Harbour, 2014). When judging the invariances that determine the categories of numbers, it should first be noted that number is a property of a *collection*. Collections form an abstract type of objects that can have different properties. Some such properties are shared by physical objects, for example weight and location: “These beans weigh 500 grams.” “The radishes are in the plastic bowl in the fridge.” Many properties are, however, unique to collections: For example, collections can be ordered or unordered, uniform (consisting of the same type of objects) or mixed, dense or spread out. In particular, collections have *cardinality*, that is, they contain a certain number of elements. The cardinality of a finite collection is expressed by a natural number.

Numerical invariances of collections have been studied extensively in developmental psychology (e.g., Gelman and Gallistel, 1986; Fuson, 1988; Sarnecka and Carey, 2008). In a series early experiments concerning “conservation tasks,” Piaget (1952) tested children in order to understand which properties of collections they perceive as being invariant. In one experiment, two equinumerous collections of objects, for example marbles, are placed into two parallel lines that are equally long. Then the objects in one line are spread out. A child that has not understood cardinality will say that there are more objects in the longer line. Failing the Piaget conservation tasks means that a child has not understood that a number is a property of a collection that is invariant of its spatial layout (see Gelman and Gallistel, 1986). In other words, number is fungible with respect to the location of the objects in a collection.

The characteristic invariance of the number category is, however, the *fungibility of objects*: If an object in a collection is exchanged for another object, the collection will still contain the same number of objects. Other properties of collections do not fulfill this criterion: If an object (an apple, say) replaces one of the objects in a uniform collection (of oranges, say), the resulting collection is not uniform any more.

The number of elements of a collection is also, to a large extent, invariant under actions, at least in the sense that independently of what kinds of actions the elements perform (for example, the movements of a football team), their number will still be the same. Similarly, number is typically invariant under actions performed on the objects (as long as the actions do not destroy the objects).

CONCLUSION

In this article I have used the notion of invariances to explain why the categories of space, object and action are fundamental cognitive structures. In philosophical terms this is a version of the neo-Kantian program of describing the “Anschauungsformen” of our perception. The analysis of the primary categories in terms of invariances can be seen as an explanation of such forms of perception. As a part of the explanation, an evolutionary perspective connects the categories to the success of the activities

of an organism. I have also argued that all three categories are central elements in the more abstract category of events. The analysis of the category of numbers that I have presented indicates that, also for non-primary categories, different forms of invariances can be used to characterize a category.

Although the evidence for the invariances that I have presented in this article comes mainly from experiments with human subjects, the perceptual systems of, at least, mammals are sufficiently similar to warrant the conclusion that space and objects are also primary categories for them. As regards actions (and, consequently, events), the situation is less clear¹³.

The enterprise of identifying cognitively primary categories is not only of philosophical and psychological interest. It leads to new questions to cognitive neuroscience. The most pressing one concerns how the invariances are picked up by the brain (e.g., Nau et al., 2018). Understanding these processes may help

¹³One reason for this caveat is that animals have difficulties reasoning about causality that depends on external forces (Tommasello and Call, 1997; Povinelli, 2000; Gärdenfors and Lombard, 2020).

REFERENCES

- Abdi, H., and Williams, L. J. (2010). Principal component analysis. *Wiley Interdiscipl. Rev. Comput. Stat.* 2, 433–459. doi: 10.1002/wics.101
- Agrawal, P., Carreira, J., and Malik, J. (2015). Learning to see by moving. *IEEE Int. Conf. Comput. Vis.* 2015, 37–45. doi: 10.1109/ICCV.2015.13
- Anderson, J. R., Reder, L. M., Simon, H. A., Ericsson, K. A., and Glaser, R. (1998). Radical constructivism and cognitive psychology. *Brookings Papers Educ. Policy* 1, 227–278.
- Anselmi, F., Leibo, J. Z., Rosasco, L., Mutch, J., Tacchetti, A., and Poggio, T. (2016). Unsupervised learning of invariant representations. *Theor. Comput. Sci.* 633, 112–121. doi: 10.1016/j.tcs.2015.06.048
- Baillargeon, R., and DeVos, J. (1991). Object permanence in young infants: further evidence. *Child Dev.* 62, 1227–1246. doi: 10.2307/1130803
- Bellmund, J., Gärdenfors, P., Moser, E., and Doeller, C. (2018). Navigating cognition: spatial codes for human thinking. *Science* 362:eaat6766. doi: 10.1126/science.aat6766
- Borg, I., and Groenen, P. J. (2005). *Modern Multidimensional Scaling: Theory and Applications*. Berlin: Springer Science and Business Media.
- Breidbach, O., and Jost, J. (2006). On the gestalt concept. *Theory Biosci.* 125, 19–36. doi: 10.1016/j.thbio.2006.02.001
- Carey, S. (2009). *The Origin of Concepts*. Oxford: Oxford University Press.
- Cutting, J. E. (1986). *Perception With an Eye for Motion*. Cambridge, MA: MIT Press.
- Dehaene, S. (1996). *The Number Sense. How the Mind Creates Mathematics*. Oxford: Oxford University Press.
- di Pellegrino, G., and Làdavas, E. (2015). Peripersonal space in the brain. *Neuropsychologia* 66, 126–133. doi: 10.1016/j.neuropsychologia.2014.11.011
- Dowty, D. (1991). Thematic proto-roles and argument selection. *Language* 67, 547–619. doi: 10.1353/lan.1991.0021
- Eichenbaum, H., and Cohen, N. H. (2014). Can we reconcile the declarative memory and spatial navigation views on hippocampal function? *Neuron* 83, 764–770. doi: 10.1016/j.neuron.2014.07.032
- Everett, C. (2017). *Numbers and the Making of Us: Counting and the Course of Human Cultures*. Cambridge, MA: Harvard University Press. doi: 10.4159/9780674979185
- Fernandes, K. J., Marcus, G. F., Di Nubila, J. A., and Vouloumanos, A. (2006). From semantics to syntax and back again: argument structure in the third year of life. *Cognition* 100, B10–B20. doi: 10.1016/j.cognition.2005.08.003
- Fiorini, S. R., Gärdenfors, P., and Abel, M. (2014). Representing part-whole relations in conceptual spaces. *Cogn. Process.* 15, 127–142. doi: 10.1007/s10339-013-0585-x
- Fuson, K. C. (1988). *Children's Counting and Concepts of Number*. New York, NY: Springer. doi: 10.1007/978-1-4612-3754-9
- Gallistel, C. R. (1990). *The Organization of Learning*. Cambridge, MA: MIT Press.
- Ganglmayer, K., Attig, M., Daum, M. M., and Paulus, M. (2019). Infants' perception of goal-directed actions: A multi-lab replication reveals that infants anticipate paths and not goals. *Infant Behav. Dev.* 57:101340.
- Gärdenfors, P. (2007). "Representing actions and functional properties in conceptual spaces," in *Body, Language and Mind, Volume 1: Embodiment*, eds T. Ziemke, J. Zlatev, and R. M. Frank (Berlin: Mouton de Gruyter), 167–195.
- Gärdenfors, P. (2014). *Geometry of Meaning: Semantics Based on Conceptual Spaces*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9629.001.0001
- Gärdenfors, P. (2018). From sensations to concepts: a proposal for two learning processes. *Rev. Philos. Psychol.* 10, 441–464. doi: 10.1007/s13164-017-0379-7
- Gärdenfors, P. (2019). Using event representations to generate robot semantics. *ACM Trans. Hum. Robot Interact.* 8, 1–21. doi: 10.1145/3341167
- Gärdenfors, P. (2020a). Events and causal mappings modeled in conceptual spaces. *Front. Psychol.* 11:630. doi: 10.3389/fpsyg.2020.00630
- Gärdenfors, P. (2020b). An epigenetic approach to semantic domains. *IEEE Trans. Cogn. Dev. Syst.* 12, 139–147. doi: 10.1109/TCDS.2018.2833387
- Gärdenfors, P., Jost, J., and Warglien, M. (2018). From actions to events: three constraints on event mappings. *Front. Psychol.* 14:1391. doi: 10.3389/fpsyg.2018.01391
- Gärdenfors, P., and Lombard, M. (2020). Technology led to more abstract causal reasoning. *Biol. Philos.* 35, 1–23. doi: 10.1007/s10539-020-09757-z
- Gärdenfors, P., and Warglien, M. (2012). Using conceptual spaces to model actions and events. *J. Semant.* 29, 487–519. doi: 10.1093/jos/ffs007
- Gärdenfors, P. M., Williams, M.-A., Johnston, B., Billingsley, R., Vitale, J., Peppas, P., et al. (2019). "Event boards as tools for holistic AI," in *Proceedings of the 6th International Workshop on Artificial Intelligence and Cognition, CEUR Workshop Proceedings, Vol. 2418*, eds A. Chella, I. Infantino, and A. Lieto (Aachen), 1–10.
- Gelman, R., and Gallistel, C. R. (1986). *The Child's Understanding of Number*. Cambridge, MA: Harvard University Press.
- Gemel, A., and Quinon, P. (2019). "Magnitude and number sensitivity of the approximate number system in conceptual spaces," in *Conceptual Spaces: Elaborations and Applications*, eds M. Kaipainen, F. Zenker,

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

The author acknowledges support from Lund University.

- A. Hautamäki, and P. Gärdenfors (Cham: Springer Nature), 183–203. doi: 10.1007/978-3-030-12800-5_10
- Gharraee, Z., Gärdenfors, P., and Johnsson, M. (2017a). First and second order dynamics in a hierarchical SOM system for action recognition. *Appl. Soft Comput.* 59, 574–585. doi: 10.1016/j.asoc.2017.06.007
- Gharraee, Z., Gärdenfors, P., and Johnsson, M. (2017b). Online recognition of actions involving objects. *Biol. Inspired Cogn. Architectures* 22, 10–19. doi: 10.1016/j.bica.2017.09.007
- Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Oxford: Houghton Mifflin.
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Hillsdale, NJ: Lawrence Erlbaum.
- Giese, M., Thornton, I., and Edelman, S. (2008). Metrics of the perception of body movement. *J. Vis.* 8, 1–18. doi: 10.1167/8.9.13
- Harbour, D. (2014). Paucity, abundance, and the theory of number. *Language* 90, 185–229. doi: 10.1353/lan.2014.0003
- Held, R., and Hein, A. (1963). Movement-produced stimulation in the development of visually guided behavior. *J. Comp. Physiol. Psychol.* 56, 872–876. doi: 10.1037/h0040546
- Hemerik, P. E. (2008). *Mind in Action*. Lund: Lund University Cognitive Studies. p. 140.
- Hinaut, X., and Dominey, P. F. (2013). Real-time parallel processing of grammatical structure in the fronto-striatal system: a recurrent network simulation study using reservoir computing. *PLoS ONE* 8:e52946. doi: 10.1371/journal.pone.0052946
- Hoerl, C., and McCormack, T. (2019). Thinking in and about time: a dual systems perspective on temporal cognition. *Behav. Brain Sci.* 42, 1–16. doi: 10.1017/S0140525X18002157
- Hoffman, D. D. (2019). *The Case Against Reality*. New York, NY: W. W. Norton.
- Humphrey, N. K. (1993). *A History of the Mind*. London: Vintage Books.
- 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. *Scand. J. Psychol.* 5, 181–208. doi: 10.1111/j.1467-9450.1964.tb01425.x
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Percept. Psychophys.* 14, 201–211. doi: 10.3758/BF03212378
- Kalkan, S., Dag, N., Türüten, O. A., Borghi, M., and Sahin, E. (2014). Verb concepts from affordances. *Interact. Stud.* 15, 1–37. doi: 10.1075/is.15.1.01kal
- Kaufman, L., and Kaufman, J. (2000). Explaining the moon illusion. *Proc. Natl. Acad. Sci. U.S.A.* 97, 500–504. doi: 10.1073/pnas.97.1.500
- Kohler, I. (1951). Formation and transformation of the perceptual world. *Psychol. Issues* 3, 1–173.
- Kriegeskorte, N., Mur, M., Ruff, D. R., Kiani, R., Bodurka, J., Esteky, H., et al. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60, 1126–1141. doi: 10.1016/j.neuron.2008.10.043
- Kruskal, J. B., and Wish, M. (1978). *Multidimensional Scaling*. Thousand Oaks, CA: Sage Publishing. doi: 10.4135/9781412985130
- Land, E. H. (1977). The retinex theory of color vision. *Sci. Am.* 237, 108–128. doi: 10.1038/scientificamerican1277-108
- Landau, B., Smith, L., and Jones, S. (1998). Object perception and object naming in early development. *Trends Cogn. Sci.* 2, 19–24. doi: 10.1016/S1364-6613(97)01111-X
- Leibo, J. Z., Liao, Q., Anselmi, F., and Poggio, T. (2015). The invariance hypothesis implies domain-specific regions in visual cortex. *PLoS Comput. Biol.* 11:e1004390. doi: 10.1371/journal.pcbi.1004390
- Leslie, A. (1996). *A Theory of Agency*. New Brunswick, NJ: Rutgers University Center for Cognitive Science.
- Levy-Leblond, J. M. (1971). “Galilei group and galilean invariance,” in *Group Theory and its Applications* (New York, NY: Academic Press), 221–299. doi: 10.1016/B978-0-12-455152-7.50011-2
- Marr, D. (1982). *Vision: A Computational Approach*. San Francisco, CA: Freeman.
- Marr, D., and Vaina, L. (1982). Representation and recognition of the movements of shapes. *Proc. Royal Soc. London B* 214, 501–524. doi: 10.1098/rspb.1982.0024
- Mealier, A. L., Pointeau, G., Gärdenfors, P., and Dominey, P. F. (2016). Construals of meaning: the role of attention in robotic language production. *Interact. Stud.* 17, 41–69. doi: 10.1075/is.17.1.03mea
- Murray, S. O., Boyaci, H., and Kersten, D. J. (2005). The emergence of object size invariance in the human visual cortex. *J. Vis.* 5, 744–744. doi: 10.1167/5.8.744
- Nau, M., Julian, J. B., and Doeller, C. F. (2018). How the brain’s navigation system shapes our visual experience. *Trends Cogn. Sci.* 22, 810–825. doi: 10.1016/j.tics.2018.06.008
- Papfragou, A. (2015). “The representation of events in language and cognition,” in *The Conceptual Mind: New Directions in the Study of Concepts*, eds E. Margolis, and S. Laurence (Cambridge, MA: MIT Press), 327–345.
- Piaget, J. (1952). *The Origins of Intelligence in Children*. New York, NY: International Universities Press. doi: 10.1037/11494-000
- Piaget, J., and Inhelder, B. (1967). *The Child’s Conception of Space*. New York, NY: Norton.
- Pinker, S. (2002). *The Blank Slate: The Modern Denial of Human Nature*. New York, NY: Viking.
- Povinelli, D. J. (2000). *Folk Physics for Apes: The Chimpanzee’s Theory of How the World Works (Vol. 7)*. Oxford: Oxford University Press.
- Radvansky, G. A., and Zacks, J. M. (2014). *Event Cognition*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199898138.001.0001
- Runeson, S. (1994). “Perception of biological motion: the KSD-principle and the implications of a distal versus proximal approach,” in *Perceiving Events and Objects*, eds G. Jansson, S.-S. Bergström, and W. Epstein (Hillsdale, NJ: Lawrence Erlbaum), 383–405.
- Sarnecka, B. W., and Carey, S. (2008). How counting represents number: what children must learn and when they learn it. *Cognition* 108, 662–674. doi: 10.1016/j.cognition.2008.05.007
- Shanahan, M., Crosby, M., Beyret, B., and Cheke, L. (2020). Artificial intelligence and the common sense of animals. *Trends Cogn. Sci.* 24, 862–872. doi: 10.1016/j.tics.2020.09.002
- Sinha, C., and Gärdenfors, P. (2014). Time, space, and events in language and cognition: a comparative view. *Ann. NY Acad. Sci.* 1326, 72–81. doi: 10.1111/nyas.12491
- Smith, L. B., and Samuelson, L. (2006). An attentional learning account of the shape bias: reply to Cimpian and Markman (2005) and Booth, Waxman, and Huang (2005). *Dev. Psychol.* 42, 1339–1343. doi: 10.1037/0012-1649.42.6.1339
- Spelke, E. S. (2000). Core knowledge. *Am. Psychol.* 2000, 1233–1243. doi: 10.1037/0003-066X.55.11.1233
- Spelke, E. S. (2004). Core knowledge. In *Attention and Performance, Vol. 20: Functional Neuroimaging of Visual Cognition*, eds N. Kanwisher, and J. Duncan (Oxford: Oxford University Press), 29–56.
- Spelke, E. S., Breinlinger, K., Macomber, J., and Jacobson, K. (1992). Origins of knowledge. *Psychol. Rev.* 99, 605–632. doi: 10.1037/0033-295X.99.4.605
- Spelke, E. S., and Kinzler, K. D. (2007). Core knowledge. *Dev. Sci.* 10, 89–96. doi: 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. *Cogn. Sci.* 41, 70–101. doi: 10.1111/cogs.12332
- Thelen, E., and Smith, L. B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press.
- Thom, R. (1972). *Structural Stability and Morphogenesis*. New York, NY: W. A. Benjamin Inc.
- Tomasello, M., and Call, J. (1997). *Primate Cognition*. New York, NY: Oxford University Press.
- Tversky, B. (2003). Structures of mental spaces: how people think about space. *Environ. Behav.* 35, 66–80. doi: 10.1177/0013916502238865
- Viera, G. (2020). The sense of time. *Br. J. Philos. Sci.* 71, 443–469. doi: 10.1093/bjps/axy019
- Von Glaserfeld, E. (2005). Thirty years constructivism. *Constr. Found.* 1, 9–12.
- Wang, W., Crompton, R. H., Carey, T. S., Günther, M. M., Li, Y., Savage, R., et al. (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. *J. Hum. Evol.* 47, 453–478. doi: 10.1016/j.jhevol.2004.08.007
- Warglien, M., Gärdenfors, P., and Westera, M. (2012). Event structure, conceptual spaces and the semantics of verbs. *Theor. Linguist.* 38, 159–193.
- Wiskott, L., and Sejnowski, T. J. (2002). Slow feature analysis: unsupervised learning of invariances. *Neural Comput.* 14, 715–770. doi: 10.1162/089976602317318938
- Wolff, P. (2007). Representing causation. *J. Exp. Psychol.* 136, 82–111. doi: 10.1037/0096-3445.136.1.82

- Wolff, P. (2008). "Dynamics and the perception of causal events," In *Understanding Events: How Humans See, Represent, and Act on Events*, eds T. Shipley, and J. Zacks (Oxford: Oxford University Press), 555–587. doi: 10.1093/acprof:oso/9780195188370.003.0023
- Wolff, P. (2012). Representing verbs with force vectors. *Theor. Linguist.* 38, 237–248. doi: 10.1515/tl-2012-0015
- Wolff, P., and Shepard, J. (2013). "Causation, touch, and the perception of force," in *The Psychology of Learning and Motivation, Vol. 58*, ed B. H. Ross (New York, NY: Academic Press), 167–202. doi: 10.1016/B978-0-12-407237-4.00005-0
- Wolff, P., and Thorstad, R. (2017). "Force dynamics," in *The Oxford Handbook of Causal Reasoning*, ed M. R. Waldmann (New York, NY: Oxford University Press), 147–167.
- Woodward, A. L. (2009). Infants' grasp of others' intentions. *Curr. Dir. Psychol. Sci.* 18, 53–57. doi: 10.1111/j.1467-8721.2009.01605.x
- Woodward, A. L. (2013). "Infant foundations of intentional understanding," in *Navigating the Social World: What Infants, Children, and Other Species Can Teach Us*, eds M. R. Banaji, and S. A. Gelman (Oxford: Oxford University Press), 75–80. doi: 10.1093/acprof:oso/9780199890712.003.0015
- Wyeth, G., and Milford, M. (2009). Spatial cognition for robots. *IEEE Robot. Autom. Mag.* 16, 24–32. doi: 10.1109/MRA.2009.933620
- Zacks, J. M., and Tversky, B. (2001). Event structures in perception and conception. *Psychol. Bull.* 127, 3–21. doi: 10.1037/0033-2909.127.1.3
- Zhao, B., Feng, J., Wu, X., and Yan, S. (2017). A survey on deep learning-based fine-grained object classification and semantic segmentation. *Int. J. Autom. Comput.* 14, 119–135. doi: 10.1007/s11633-017-1053-3
- Zhu, S. C., and Yuille, A. L. (1996). FORMS: a flexible object recognition and modeling system. *Int. J. Comput. Vis.* 20, 187–212. doi: 10.1007/BF00208719
- Zwarts, J., and Gärdenfors, P. (2016). Locative and directional prepositions in conceptual spaces: the role of polar convexity. *J. Logic Lang. Inform.* 25, 109–138. doi: 10.1007/s10849-015-9224-5

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Gärdenfors. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.