Book Review:

Peter Gärdenfors, *Conceptual Spaces — The Geometry of Thought*

by Willem Labuschagne

The topic addressed by Gärdenfors[1] in *Conceptual Spaces* is representation. How do high-level concepts, represented symbolically (i.e. by language) arise from low-level neural processes? This book details a response based on something new, namely a topological approach in which the notion of spatial nearness plays the central role.

I first heard Gärdenfors describe this approach in a talk at the International Congress on Logic, Methodology and Philosophy of Science in Florence, 1995. Gärdenfors, a Swedish professor of philosophy who achieved fame for his contributions to logic in the area known as belief revision, is a tall, elegant fellow and a charismatic speaker. Still, I left his talk in Florence feeling as if I had dined on candy floss instead of steak — the topic was interesting, the approach novel, but there was a dearth of nitty-gritty details. Five years later, *Conceptual Spaces* appeared and at last we have a feast of details.

In order to place the book in context, two opposing lines of thought in cognitive science should be mentioned. The symbolic tradition views thought as the manipulation of language, and has scant respect for semantics. Daniel Dennett has been one of the most influential advocates of this model, for example writing that[2]
"… brains … can only be ***synta[c]tic engines****, responding only to structural or formal properties. According to the traditional distinction in linguistics, a sentence's form or syntax is one thing and its meaning or semantics is another. Now how does the brain manage to get semantics from syntax? It couldn't."*
In other words, Dennett asserts that the only representations of information in the brain are symbolic representations, and thus that cognition is (only) a matter of processing language with the help of syntactic rules.

The alternative model has been articulated particularly clearly by Stevan Harnad, who espouses a three-level theory of representation[3] in terms of which perception involves first the construction of iconic representations — topographically organised analogs of the sensory input — followed by the construction of higher-level categorical representations by means of some kind of *analog-to-discrete transformation*, culminating finally in symbolic representations. Loosely speaking, the model says that if you look at a landscape you first have a pattern of neural excitation in the brain that resembles the picture on the retina, with bits that are close together in the retinal image corresponding to areas of excitation that are close together in the brain. This iconic representation undergoes a change that exaggerates some differences while discarding others, in much the way that an accurate photo of a landscape (or person) may be caricatured or reduced to a line drawing. This hardening of boundaries changes what was originally continuous variation into discrete parts that may readily be distinguished — here is a tree, there a building. The omission of detail required by the change makes the result more generic, so that it can stand for a whole class of items rather than one specific item. And at this point it becomes possible to attach labels (symbols) to the representation, which brings language into the arena.

The problem with the three-level theory of representation has always been the middle level. How exactly is the analog-to-discrete transformation accomplished? In his book, Gärdenfors approaches the middle level activity from a geometric perspective. However, this is no dry geometry text. Gärdenfors's account draws upon the full panoply of cognitive science — psychology, philosophy, logic, computer science, and linguistics. The lasting impression created by the book is of the extraordinary breadth, both substantive and methodological, that the author has at his command.

So what is a conceptual space?

Crudely speaking, a space is a mathematical structure with dimensions along which measurements may be noted. But conceptual spaces are unlike the spaces used in natural sciences like physics, for Gärdenfors sees conceptual spaces as human, or at least *agent-oriented* (since one would expect our understanding of other animals and even our design of artificial agents to benefit from the theory). Whereas in physics one would measure, for example, the absolute wavelengths and frequencies of light, a conceptual space concerned with vision would be structured according to the way a human perceives the light, which is something rather different from objective measurements of wavelength and frequency. While wavelength still plays a role, the dimensions of the conceptual space are partly determined by the agent architecture, and what is important is not whether the wavelength is so many nanometers but whether the agent perceives it as blue. After all, it is a fact of perception that colour constancy requires a term such as *blue* to be associated with various different wavelengths depending on the context (i.e. depending on the time of day, whether the sky is cloudy, and so on).

(To see how different perception and physics can be, the reader is invited to visit the webpage http://www.scientificpsychic.com/graphics/shadow.jpg, which illustrates how the same shade of grey, i.e. the same absolute wavelength of light, may, depending only on context, be perceived by a human agent either as black or as white.)

One begins the construction of a conceptual space, therefore, with *quality dimensions* — such as hue, saturation, and luminosity, if we continue to take the perception of colour as an example. Such quality dimensions correspond to ways in which an agent can judge stimuli to be the same or different. Gärdenfors gives examples showing that researchers who want to model the conceptual spaces of agents can identify relevant quality dimensions by statistical techniques such as multi-dimensional scaling.

Quality dimensions form clusters called *domains*. For example, a stimulus object cannot have hue without also having a degree of saturation and of luminosity, so those three dimensions together form the colour domain. Dimensions such as weight or pitch would not form part of this domain but would belong to their own clusters.

A conceptual space consists of one or more domains together with a notion of distance (a metric). Nonmathematicians tend to be surprised by the suggestion that there can be more than one kind of distance, but Gärdenfors shows that it is necessary to consider various kinds of metric. Two examples illustrate how notions of distance can differ. The Euclidean metric is (roughly speaking) a way of measuring the distance between two points in the space by taking the length of the straight line connecting those points, whereas the City-Block metric measures the distance by, as it were, following the streets (i.e. going first horizontally the necessary number of steps and then vertically, instead of cutting straight across like the Euclidean metric).

The metric associated with a particular conceptual space has a very important role to play in the construction of categorical representations (the middle level of Harnad's theory of representation). Gärdenfors postulates the following role for the metric.

Given a conceptual space, certain subsets of the space are *convex regions*. Roughly speaking, if you have two measurements in a convex region then any measurements between those two will also fall in the convex region. Or visualise a convex region as a fat blob, like the inside of a circle or a square rather than the inside of a wasp-waisted figure eight. Basically, convex regions are nice things to have, according to Gärdenfors, who suggests that concepts, or properties such as "is blue", correspond to convex regions in the relevant conceptual space.

The mechanism proposed for the construction of categorical representations is a known as a Voronoi tessellation of the space, which is a process that breaks the space up into convex regions. The metric forms the Voronoi tessellation with the aid of *prototypes*.

Here Gärdenfors links up conceptual spaces with a well-established area in cognitive psychology. A prototype may be thought of as a cognitive reference point. Consider the colour categories again. It is known that humans, regardless of ethnicity or culture, agree on certain areas of colour space as the best examples of basic colour terms in their language.[4] That is to say, there are certain small bits of colour space that everyone would agree best exemplify red, green, blue, and so on. (And this is not a theoretical assumption but a thoroughly tested empirical fact, not only for colour categories but for a variety of other `basic' categories[5].)

Let us, following Eleanor Rosch, call these points the prototypes of the colour categories. And let us suppose that we have a conceptual space in which certain points, the prototypes, have been identified. Now the metric associated with the space can be used to annex to each prototype all the points in the space that lie closer to it than to any other prototype. The result is a Voronoi tessellation, a breaking up of the (continuous) space into a discrete collection of convex regions. At last we have a mechanism to accomplish the analog-to-discrete transformation required by Harnad's middle level of representation.

Mindful of his roots as logician, Gärdenfors proceeds to link up conceptual spaces with Jackendoff's views on semantics[6] and with intensional logic. He draws a new philosophical distinction between properties and concepts. He shows that conceptual spaces provide a rich framework unifying various discussions pertaining to artificial intelligence and theory of mind. But for me the key achievement is to cast light on the previously problematic middle level of representation, and thereby to deal a stunning blow to those who, like Dennett, attempt to construe cognition in terms of syntax only, without internally represented symbol-grounding semantics.

---

[1] Peter Gärdenfors, *Conceptual Spaces* (Cambridge, Massachusetts : MIT Press 2000).
[2] Daniel Dennett, *Elbow Room: The Varieties of Free Will Worth Having* (Oxford: Clarendon Press, 1984) p28.

[3] Stevan Harnad (ed.), *Categorical Perception* (Cambridge: Cambridge University Press, 1987) pp535-565.
[4] Eleanor Rosch Heider, "`Focal' Color Areas and the Development of Color Names", *Developmental Psychology*, 1971, 4(3):447-455.
[5] Eleanor Rosch, "Principles of Categorization" in Collins and Smith (editors), *Readings in Cognitive Science* (San Mateo : Morgan Kaufmann 1988) pp312-322.
[6] Ray Jackendoff, *Semantics and Cognition* (Cambridge, Massachusetts : MIT Press 1983).