Inductive Reasoning: From Carnap to Cognitive Science

Peter Gärdenfors

Lund University Cognitive Science
Kungshuset, S-222 22 Lund Sweden
E-mail: Peter.Gardenfors@fil.lu.se

Abstract

The dominating models of inductive processes have been based on symbolic representations of knowledge. This was the explicit assumption of the Vienna school and most of Carnap's work on induction follows this principle.

However, it is becoming increasingly clear that most cognitive phenomena in humans and animals are based on non-symbolic representations. As an alternative to symbolic representations of information and knowledge, this paper investigates a theory of "conceptual spaces." Such spaces consist of a number of "quality dimensions" which often are derived from perceptual mechanisms. They can be used to describe cognitive processes like concept formation and induction. A 'geometric' model of concept formation is proposed and its relation to prototype theory is discussed. It is shown that Carnap in his later writings was moving towards this approach to induction. It is also be argued that conceptual spaces are suitable for representing the results of information processing in connectionist systems.

1. Program: Three levels of inductive inferences

The goal of this paper is to compare some epistemological models in order to evaluate which is the most appropriate for inductive inferences. It is commonplace that induction is going from single observations to generalizations. But this statement loses its air of triviality if one takes seriously the question of *what* an observation is. It is surprising that this question has received very little attention within the philosophy of science.[1] Following my earlier work (Gärdenfors 1994), I shall argue that there is no unique way of characterizing an observation. I shall distinguish three levels of accounting for observations:

1. The symbolic level:[2] This way of viewing observations consists of describing them in some specified language. The language is assumed to be equipped with a fixed set of primitive predicates and the denotations of these predicates is taken to be known. As will be argued in Section 2, the symbolic approach is a corner-stone of logical positivism.

2. The conceptual level: On this level observations are not defined in relation to some language but characterized in terms of some underlying "conceptual space." The conceptual space, which is more or less connected to perceptual mechanisms, consists of a number of "quality dimensions." Induction is here seen as closely related to *concept formation*. According to the conceptual perspective, inductive inferences show prototype effects, in contrast to the symbolic perspective which operates on Aristotelian concepts (cf. Smith & Medin 1981).

3. The subconceptual level: Observations are here characterized in terms of inputs from *sensory receptors*. The observations are thus described as occurring before

---

1One notable exception is Shapere (1982). See Section 8.

2This was called the "linguistic level" in Gärdenfors (1994).

conceptualization. The inductive process is seen as establishing connections between various types of inputs. One currently popular way of modelling this kind of process is by using neural networks.

Depending on which approach to observations is adopted, thoroughly different considerations about inductive inferences will come into focus.[3] In my opinion there is a multitude of aspects of inductive reasoning and not something that can be identified as *the* problem of induction. What is judged to be the salient features of the inductive process depends to a large extent on *what* an observation is considered to be. In this article, I will focus on inductive processes on the conceptual level, but they will also be contrasted with inductive reasoning on the two other levels.

2. The symbolic level: Logical positivism and the problems of induction

The most ambitious project of analyzing inductive inferences during this century has been that of the logical positivists. As a background to my presentation of conceptual spaces, I shall start by commenting briefly on this research program.

Inductive inferences were important for the logical positivists, since such inferences were necessary for their verificationist aims. The basic objects of study for them were sentences in some more or less regimented language. Ideally, the language was a version of first order logic where the atomic predicates represented observational properties. These observational predicates were taken as *primitive*, unanalysable notions. The main tool used when studying the symbolic expressions was logical analysis. In its most pure form, logical positivism allowed only this tool. A consequence of this methodology was that all observational predicates were treated in the same way since there were no *logical* reasons to differentiate between them. For example, Carnap (1950, Section 18B) requires that the primitive predicates of a language be logically independent of each other.[4]

However, it became apparent that the methodology of the positivists led to serious problems in relation to the problem of induction. The most famous one's are Hempel's (1965) "paradox of confirmation" and Goodman's (1955) "riddle of induction." I will not repeat these well-known problems, but only state my diagnosis of what causes them (for a more detailed account, cf. Gärdenfors (1990)). What I see as the root of the troublesome cases is that if we use *logical* relations alone to determine which inductions are valid, the fact that all predicates are treated on a par induces *symmetries* which are not preserved by our understanding of the inductions: "Raven" is treated on a par with "non-raven," "green" with "grue" etc. What we need is a *non-logical* way of distinguishing those predicates that may be used in inductive inferences from those that may not.

There are several suggestions for such a distinction in the literature. One idea is that some predicates denote "natural kinds" or "natural properties" while others don't, and it is only the former that may be used in inductive reasoning. Natural kinds are normally interpreted realistically, following the Aristotelian tradition, and thus assumed to represent something that exists in reality independently of human cognition. However, when it comes to inductive *inferences* it is not sufficient that the properties exist out there somewhere, but we need to be able to grasp the natural kinds by our minds. In other words, what is needed

3I cannot talk about three ways of *describing* observations, because the very notion of "describing" presumes the linguistic level.

4However, in his later writings, this requirement is abolished (see Carnap 1971, p. 82). The change in his view on induction will be discussed in Section 7.

to understand induction, as performed by humans, is a *conceptualistic* or *cognitive* analysis of natural properties. It is one of the aims of the present paper to outline such an analysis.

Another notion that is relevant here is that of "similarity." Quine (1969) discusses this notion and its relation to that of a natural kind. He notes that similarity "is immediately definable in terms of kind; for things are similar when they are two of a kind" (p. 117). Furthermore, he says about similarity that we

"cannot easily imagine a more familiar or fundamental notion than this, or a notion more ubiquitous in its application. On this score it is like the notions of logic: like identity, negation, alternation, and the rest. And yet, strangely, there is something logically repugnant about it. For we are baffled when we try to relate the general notion of similarity significantly to logical terms." (Quine 1969:117)

The core of the problem, it seems to me, is that the symbolic level is insufficient for handling the problems of induction and similarity. We not only want to know how observational predicates should be combined in the light of inductive evidence, but, much more importantly, *how the basic predicates are inductively established* in the first place. This problem has, more or less, been swept under the rug by the logical positivists. Using logical analysis, the prime tool of positivism, is of no avail for these forms of concept formation. In brief, the symbolic approach to induction sustains no creative inductions, no genuinely new knowledge, and no conceptual discoveries. To do this, we have to go below language.

3. The conceptual level: Conceptual spaces

A conceptual space consists of a number of *quality dimensions.* As examples of quality dimensions let me mention color, pitch, temperature, weight, and the three ordinary spatial dimensions. Some of the dimensions are closely related to what is produced by our sensory receptors, but there are also quality dimensions that are of an abstract non-sensory character. Furthermore, the dimensions are cognitive and *prelinguistic* in the sense that we can think about the qualities of objects, for example when planning an action, without presuming an internal language in which these thoughts can be expressed. The notion of a conceptual space is closely related to Carnap's (1971, 1980) "attribute space" and Quine's (1960) "quality space". These relations will be discussed in Section 7.

The notion of a *dimension* should be understood literally. It is assumed that each of the quality dimensions is endowed with certain topological or metric structures. For example, "time" is a one-dimensional structure which we conceive of as being isomorphic to the line of real numbers. Similarly, "weight" is one-dimensional with a zero point, isomorphic to the half-line of non-negative numbers. Some quality dimensions have a *discrete* structure, i.e., they merely divide objects into classes, e.g., the sex of an individual.[5]

At this point it is important to make a distinction between a *psychological* and a *scientific* interpretation of the quality dimensions. For example, our psychological visual space is not a perfect 3-dimensional Euclidean space since it is not invariant under all linear (Galilean) transformations. Because of gravity, among other things, the vertical dimension is treated differently from the two horizontal dimensions. However, the scientific representation of visual space as a 3-D Euclidean space is an idealization that is mathematically amenable (there is no preferred direction, Galilean transformations preserve the structure, etc.).

---

[5] Discrete dimensions may also have additional structure as, for example, in kinship or biological classifications. The topology of discrete dimensions is further discussed in Gärdenfors (1988).

Similarly, our perception of the weight of objects is not fine enough to justify its representation by the full structure of the positive real numbers, but this scientific representation is motivated by the fact that the mathematics of this structure is well-known, and thus makes it possible to formulate a quantitative theory of weight which is easy to handle computationally. However, when it comes to providing an analysis of human induction, it is, of course, the psychological interpretations of the quality dimensions that are in focus.

A psychologically interesting example of a quality dimension concerns *color perception*.[6] In brief, our cognitive representation of colors can be described by three dimensions. The first dimension is *hue*, which is represented by the familiar *color circle*. The topological structure of this dimension is thus different from the quality dimensions representing time or weight which are isomorphic to the real line. One way of illustrating the differences in topology is by noting that we can talk about psychologically *complementary* colors, i.e. colors that lie *opposite* to each other on the color circle. In contrast it is *not meaningful* to talk about two points of time or two weights being "opposite" to each other. This simple example shows that the topological structure of the cognitive representations of perceptual qualities will have important consequences for the *semantics* of linguistic expressions used to talk about these qualities.

The second psychological dimensions of color is *saturation*, which ranges from gray (zero color intensity) to increasingly greater intensities. This dimension is isomorphic to an interval of the real line. The third dimension is *brightness* which varies from white to black and is thus a linear dimension with end points. Together these three dimensions, one with circular structure and two with linear, make up the color space which is a subspace of our perceptual conceptual space (see Figure 1).[7]

---

6 This example is discussed by Carnap (1971, 1980).

7 For further examples of perceptual quality dimensions, see Gärdenfors (1990).

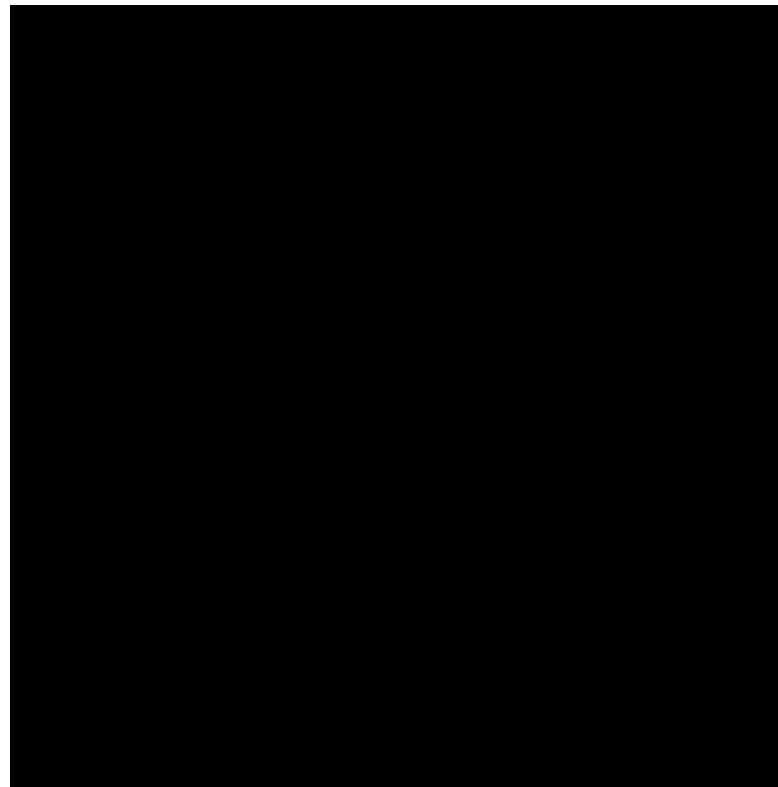

Figure 1. The three-dimensional color space

The neurophysiological mechanisms underlying the mental representation of color space are comparatively well understood. However, the connections between what excites the cones and rods in the retina and what color is *perceived* is far from trivial. According to Land's (1977) results, the perceived color is not a function of radiant energy received by the cones and rods, but rather determined by "lightness" values computed at three wavelengths.

As another example of sensory space representations let me only mention that the human perception of taste appears to be generated from four distinct types of receptors: salt, sour, sweet, and bitter. Thus the quality space representing tastes should be at most 4-dimensional. I do not know much about the actual structure of the gustatory space. Suffice it to say that it quite clearly has some non-trivial metric structure. For instance, we can meaningfully claim that the taste of a walnut is *closer* to the taste of a hazelnut than to the taste of popcorn in the same way as we can say that the color orange is closer to yellow than to blue.

Quine notes that something like a conceptual space is needed to make *learning* possible:

"Without some such prior spacing of qualities, we could never acquire a habit; all stimuli would be equally alike and equally different. These spacings of qualities, on the part of men and other animals, can be explored and mapped in the laboratory by experiments in conditioning and extinction. Needed as they are for all learning, these distinctive spacings cannot themselves all be learned; some must be innate." (Quine 1969:123)

In general, representing information about the world in terms of quality dimensions involves using *geometrical* or *vectorial* notions. In contrast the research program of the logical positivists (and much of current AI) is based on *symbolic* or *logical* modes of representation (cf. Sloman's (1971) distinction between "analogical" and "Fregean"

representation and P. S. Churchland's (1986) criticism of the "sentential paradigm"). The move to a new form of representation is, I believe, necessary; the problems that plagued the logical positivists indicate that we have to go beyond logic and language to find a solution to the problem of projectibility in inductive reasoning.

This concludes my general presentation of conceptual spaces. I believe it can be seen as a generalization of the state space approach, advocated among others by P. M. Churchland (1986) and P. S. Churchland (1986), and of the vector function theories of Foss (1988). Here, my aim is to show its viability as a foundation for an analysis of inductive reasoning.

4. The origin of quality dimensions

How do we know that the inductions generated from the natural properties determined by a conceptual space will be successful? In order to answer this question we must first consider the *origins* of the quality dimensions we use in classifying kinds and judging similarities.

There does not seem to be a unique origin of our quality dimensions. Some of the dimensions are presumably *innate* and to some extent hardwired in our nervous system, as for example color, pitch, and probably also ordinary space. These subspaces are obviously extremely important for basic activities like finding food and getting around in the environment.[8]

Other dimensions are probably *learned*. "Volume" may be an example here. The experiments performed by Piaget and his school indicate that small children have no separate mental dimension of volume: They confuse the volume of a liquid with the *height* of the liquid in its container. Learning new concepts often involves expanding one's conceptual space with new quality dimensions.

This means that many of the quality dimensions of human conceptual spaces are not directly generated from sensory inputs. This is even clearer when we use concepts based on the *functions* of artifacts or the *social roles* of people in a society. Even if we do not know much about the topological structures of these dimensions, it is quite obvious that there is some non-trivial such structure (see e.g. Vaina's (1983) analysis of functional representation).

Still other dimensions may be *culturally* dependent. Take "time," for example: In some cultures time is conceived to be circular – the world keeps returning to the same point in time and the same events occur over and over again; and in other cultures it is hardly meaningful at all to speak of time as a dimension. A sophisticated time dimension with the full metric structure is needed for advanced forms of planning and coordination with other individuals, but is not necessary for the most basic activities of an organism. As a matter of fact, the standard Western conception of time is a comparatively recent phenomenon.

Finally, some quality dimensions are introduced by science. Witness, for example, Newton's distinction between *weight* and *mass*, which is of crucial importance for the development of his celestial mechanics, but which has no correspondence in human perception. To the extent we have mental representations of the masses of objects in distinction to their weights, this is something which clearly has to be learned by adopting Newton's conceptual space in our mental representations.

---

[8] The evolutionary value of conceptual spaces is further discussed in Gärdenfors (1993).

The most drastic changes in science occur when the underlying conceptual space is changed. I believe that most of the "paradigm shifts" discussed by Kuhn (1970) can be understood as shifts of conceptual spaces. I do not see any principal difference between this kind of change and the change involved in the development of a child's conceptual space: Introducing the distinction between "height" and "volume" is the same kind of phenomenon as introducing the distinction between "weight" and "mass." Such changes in the underlying conceptual space will have considerable effects on which concepts are used in inductive processes.

5. Natural properties

Let us now turn to the problem of identifying observations on the conceptual level. Using the notion of conceptual spaces, an observation can be defined as *an assignment to an object of a location in a conceptual space*. For example, the observation that is described on the linguistic level as "x is red" is expressed on the conceptual level by assigning x a point in color space. Since natural languages only divide the color domain into a finite number of categories the information contained in the statements that x is red is much less precise than the information furnished by assigning x a location in color space. In this sense, the conceptual level allows much richer devices for reporting observations.

On the conceptual level one can distinguish between *two types of inductive processes*. One is closely related to *concept formation*: In Gärdenfors (1990), I analysed "natural properties" in terms of conceptual spaces. The key idea is that a natural property is identified with a *convex region* of a given conceptual space. Via the notion of "convexity," the topological properties of the quality dimensions are utilized. A convex region is characterized by the criterion that for every pair o_1 and o_2 of points in the region, all points *between* o_1 and o_2 are also in the region. The definition presumes that the notion of "between" is meaningful for the relevant dimensions. This is, however, a rather weak assumption which demands very little of the underlying topological structure.

As an application of the criterion of a natural property, I conjecture that all color terms in natural languages express natural properties with respect to the psychophysical representation of the quality dimensions of colors. In other words, the conjecture predicts that if some object o_1 is described by the color term C in a given language and another object o_2 is also said to have the color C, then any object o_3 with a color which lies between the color of o_1 and the color of o_2 will also be described by the color term C. It is well-known that different languages carve up the color circle in different ways, but all carvings seem to be done in terms of convex sets. Strong support for this conjecture can be gained from Berlin and Kay (1969), although they do not treat color terms in general but concentrate on basic color terms.

On the basis of this criterion of natural properties, it is now possible to formulate a constraint on induction, which is helpful in solving the conundrums of the linguistic approach:

(P) Only properties corresponding to a convex region of the underlying conceptual space may be used in inductive inferences.

It is only proposed that convexity is a necessary condition, but perhaps not sufficient, for a property to count as natural and thus allowed in inductive inferences. I argue in Gärdenfors (1990) that criterion P solves many of the problems of induction that appear on the symbolic level.

An assumption that is within reach now is that most basic words in natural languages denote convex regions in some conceptual space. From the assumption it follows that the assignment of meanings to the expressions on the linguistic level is far from arbitrary. On the contrary, the semantics (and to some extent even the grammar) of the linguistic constituents is severely constrained by the structure of the underlying conceptual space. This thesis is anathema for the Chomskian tradition within linguistics, but, as a matter of fact, it is one of the central tenets of the recently developed "cognitive" linguistics.[9]

As another sign of the importance of the conceptual level, I submit that most of scientific theorizing takes place at this level. Determining the relevant dimensions involved in the explanation of a phenomenon is a prime scientific activity. And once the conceptual space for a theory has been established, theories, in the form of *equations*, that connect the dimensions can be proposed and tested.[10]

6. Natural properties in cognitive science: Prototype theory

Apart from giving intuitively plausible solutions to the old riddles of induction (cf. Gärdenfors 1990), the definition of natural properties in terms of convex regions of conceptual spaces derives independent support from the *prototype theory* of categorization developed by Rosch and her collaborators (Rosch 1975, 1978, Mervis and Rosch 1981, Lakoff 1987). This theory can also be used to provide a reasonable explication of the notion of similarity that is presupposed in inductive inferences. Quine (1969:119-120) argues that "natural kind" is definable in terms of "similarity" and, as will be shown in the following section, he actually proposes a precursor to prototype theory.

The main idea of prototype theory is that within a category of objects, like those instantiating a property, certain members are judged to be more representative of the category than others. For example robins are judged to be more representative of the category bird than are ravens, penguins and emus; and desk chairs are more typical instances of the category chair than rocking chairs, deck-chairs, and beanbag chairs. The most representative members of a category are called *prototypical* members. It is well-known that some properties, like "red" and "bald" have no sharp boundaries and for these it is perhaps not surprising that one finds prototypical effects. However, these effects have been found for most properties including those with comparatively clear boundaries like "bird" and "chair."

Now, if the traditional definition of a property is adopted, it is very difficult to explain such prototype effects. Either an object is a member of the class assigned to a property or it is not and all members of the class have equal status as category members. Rosch's research has been aimed at showing asymmetries among category members and asymmetric structures within categories. Since the traditional definition of a property does not predict such asymmetries, something else must be going on.

In contrast, if natural properties are defined as convex regions of a conceptual space, prototype effects are indeed to be expected. In a convex region one can describe positions as being more or less central. For example, if color properties are identified with convex subsets of the color space, the central points of these regions would be the most prototypical examples of the color. Rosch has, in a series of experiments, been able to demonstrate the psychological reality of such "focal" colors.

---

[9]Cf. Lakoff (1987), Langacker (1986) and Gärdenfors (1995).

[10]For a discussion of the role of conceptual spaces in science, see Gärdenfors (1990) and (1991).

It should be noted that even if even if different members of a category are judged to be more or less prototypical, it does not follow that some of the existing members must represent "the prototype." If a category is viewed as a convex region of a conceptual space this is easily explained, since the central member of the region specifies a possible individual but need not be among the existing members of the category.

It is possible to argue in the converse direction too and show that if prototype theory is adopted, then the representation of properties as convex regions is to be expected. Assume that some quality dimensions of a conceptual space are given, for example the dimensions of color space, and that we want to partition it into a number of categories, for example color categories. If we start from a set of prototypes p_1, ..., p_n of the categories, for example the focal colors, then these should be the central points in the categories they represent. One way of using this information is to assume that for every point p in the space one can measure the distance from p to each of the p_i's. If we now stipulate that p belongs to the same category as the *closest* prototype p_i, it can be shown that this rule will generate a partitioning of the space that consists of convex areas (convexity is here defined in terms of the assumed distance measure). This is the so called *Voronoi tessellation* (see Figure 2).

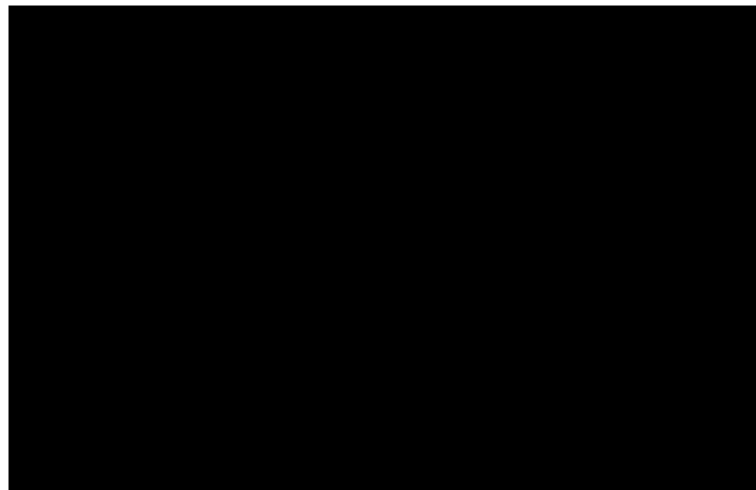

Figure 2: Voronoi tessellation of the plane into convex sets.

Thus, assuming that a *metric* is defined on the subspace that is subject to categorization, a set of prototypes will by this method generate a unique partitioning of the subspace into convex regions. Hence, there is an intimate link between prototype theory and the analysis of this article where properties are defined as convex regions in a conceptual space. Furthermore, such a metric is an obvious candidate for a measure of *similarity* between different objects. In this way, the Voronoi tessellation gives an explicit answer to how a similarity measure determines a set of natural kinds.

7. Two foregoers: Carnap and Quine

In Carnap's later writings (in particular Carnap (1971) and the posthumous Carnap (1980)), one finds some ideas which go against his earlier views on inductive logic and which come quite close to the theory presented here. To explain the difference between his earlier position and the newer one, he introduces a distinction between *pure* and *applied* inductive logic (Carnap 1971, Section 4). The crucial difference is that in pure inductive logic, the language is described in an abstract way without any interpretation of the nonlogical

constants, while in applied inductive logic, the language is interpreted. He says that it is one of the tasks of applied inductive logic to "lay down requirements that primitive attributes and relations must fulfill to be admissible as primitive concepts in an object language … " (1971, p. 70). This means, among other things, the following:

"We specify for each family the modality on which it is based (e.g., color), the attribute space (color space), and the chosen partition of this space. We define a suitable metric for the attribute space, i.e., a distance function based on similarity relations. Furhermore, we give as much information about the regions of the partition as seems relevant for inductive logic." (Carnap, 1971: 69-70)

Here, the distinction between "attribute space" and "partition" corresponds to the distinction between the dimension(s) of a subspace and the regions of that space that have been discussed here. In this context it should be noted that an even earlier forerunner is Johnson's (1921, Ch. XI) distinction between "determinables" (corresponding to dimensions) and "determinates" (corresponding to regions).

Carnap's proposal thus seems close to the theory presented here. However there are some differences. One point is that he never proposes convexity as a criterion for partitioning the space into regions. In Carnap (1980), he discusses which regions of the attribute space are "scientifically useful" or "admissible" for inductive reasoning. In his first "conjecture" (1980, p. 20) he mentions some elementary topological properties that may be relevant. However, he ends up by proposing parallelograms and their n-dimensional generalizations as the "basic regions" and (countable) unions of such basic regions as "admissible regions." In my opinion, when characterizing the regions of a conceptual space that are "scientifically useful," the use of parallelograms as primitives seems too narrow, while on the other hand, allowing arbitrary unions of parallelograms seems too liberal. In this paper, criterion P is proposed as a preferable alternative.

A more fundamental difference is that even if Carnap notes that the similarity comparisons sometimes can be expressed in terms of a metric (1971, p. 79, and 1980, Section 14), he does not take such a metric as a primitive notion. On the contrary, he assumes that the "phenomenological" assumptions ("B-principles") that cannot be formulated in the object language L are expressed in the metalanguage that is used to talk about the models and propositions of L and the inductive functions defined on L. This means that he never leaves the ideal that all knowledge should be expressed symbolically, although he relegates some of the information to a metalanguage.

As typical examples of the "phenomenological" assumptions that Carnap is discussing, one can mention the following (Carnap 1971, p. 78):

"The colors Green and Blue are incompatible, i.e., they cannot occur simultaneously at the same place."
"Green is *similar* to Blue."
"Green is *more similar* to Blue than to Red."

Carnap says about such assumptions:

"There are many controversies among philosophers about the logical nature of statements of these kinds, whether they are analytic or synthetic, and about their epistemological nature, whether they are known a priori, i.e., independently of factual experience, or a posteriori, on the basis of experience. At least the simpler ones […] may be counted as phenomenological. Statements of this kind were regarded by Husserl as synthetic–a priori. In the Vienna Circle we regarded them likewise as a priori, but analytic. I have frequently offered such sentences as examples of A-postulates. For years now, however, my friends and I tend to a more cautious attitude with respect to the epistemological question. We think that so far no satisfactory explication has been given for the concepts of a priori versus empirical (a posteriori) knowledge." (Carnap 1971, p. 79)

It follows from the theory presented here that if we assume that the meanings of the predicates are determined by a mapping into a conceptual space *S*, it follows from criterion P and the topological structure of different quality dimensions that certain statements will become analytically true. For example the fact that comparative relations like "earlier than" are *transitive* follows from the linear structure of the length dimension and is thus an analytic feature of this relation (analytic-in-*S*, that is). Similarly, it is analytic that everything that is green is colored (since "green" refers to a region of the color space) and that nothing is both green and blue. Analytic-in-*S* is thus defined on the basis of the topological and metric structure of the conceptual space *S*. However, different conceptual spaces will yield different notions of analyticity.

Another foregoer is Quine (1969), who formulates a precursor to the psychological prototype theory:

"One may be tempted to picture a kind, suitable to a comparative similarity relation, as any set which is 'qualitatively spherical' in this sense: it takes in exactly the things that differ less than so-and-so much from some central norm. If without serious loss of accuracy we can assume that there are one or more actual things (*paradigm cases*) that nicely exemplify the desired norm, and one or more actual things (*foils*) that deviate just barely too much too be counted into the desired kind at all, then our definition is easy: *the kind with paradigm a and foil b* is the set of all things to which *a* is more similar than *a* is to *b*. More generally, then, a set may be said to be a *kind* if and only if there are *a* and *b*, known or unknown, such that the set is the kind with paradigm *a* and foil *b*." (Quine, 1969:119-120)

Quine notes that, as it stands, this definition of a kind is not satisfactory:

"Thus take red. Let us grant that a central shade of red can be picked as norm. The trouble is that the paradigm cases, objects in just that shade of red, can come in all sorts of shapes, weights, sizes, and smells. Mere degree of overall similarity to any such paradigm case will afford little evidence of degree of redness, since it will depend also on shape, weight, and the rest. If our assumed relation of comparative similarity were just comparative chromatic similarity, then our paradigm-and-foil definition of kind would indeed accommodate redkind. What the definition will not do is distill purely chromatic kinds from mixed similarity." (Quine, 1969:120)

The problem for Quine is that he does not assume anything like a conceptual space to help structuring the relations of similarity. However, if such a structure is given, we need not rely on actual objects as paradigm cases, but can use focal points on a particular quality dimension, like the color dimension, as a basis for comparing chromatic similarity. The shape, weight, and the rest of the qualities of objects will simply not be relevant for such comparisons.

8. The subconceptual level

Quine's problem concerning how similarity is to be determined leads to the question of the origins of perceptual observations. In the most basic sense, an observation is what is received by our sensory organs. An observation can then be identified with what is received by a set of *receptors*. For human beings, these inputs are provided by the sensory receptors, but one can also talk of a machine having observations of this kind via some measuring instruments serving as receptors. The receptors provide "raw" data in the sense that the information is not assumed to be processed in any way, neither in a conceptual space, nor in the form of some linguistic expression.

Within the philosophy of science, it is important to make a distinction between *perception* and *observation*. As Shapere (1982) points out, the term "observation" plays a double role for the traditional philosopher of science. He writes:

"On the one hand, there is the *perceptual* aspect: "observation", as a multitude of philosophical analyses insist, is simply a special kind of perception, usually interpreted as consisting in the addition to the latter of

an extra ingredient of focussed attention. ... On the other hand, there is the *epistemic* aspect of the philosopher’s use of "observation’: the *evidential* role that observation is suppose to play in leading to knowledge or well-grounded belief or in supporting beliefs already attained." (Shapere 1982: 507-508)

Within the empiricist tradition of philosophy of science, the two uses of "observation" have been confounded. However, in modern science it is obvious that it is the epistemic aspect of observation that is of importance. As Shapere (1982: 508) formulates it:

"Science is, after all, concerned with the role of observation as evidence, whereas sense-perception is notoriously untrustworthy ... . Hence, with the recognition that information can be received which is not directly accessible to the senses, *science has come more and more to exclude sense-perception as much as possible from playing a role in the acquisition of observational evidence;* that is, it relies more and more on other appropriate, but dependable, receptors."

Given that we are focussing on the epistemic aspect of observations, let us then consider induction on the subconceptual level. How do we distill sensible information from what is received by a set of receptors? Or, in other words, how do we make the transition from the subconceptual to the conceptual and the symbolic levels? These questions indicate the kinds of inductive problems that occur on the subconceptual level.

The basic problem is that the information received by the receptors is too rich and unstructured. What is needed is some way of transforming and organizing the input into a form that can be handled on the conceptual or linguistic level. There are several methods for treating this kind of problem. Within psychology, various methods of *multidimensional scaling* have been developed (Shepard 1962a, 1962b, 1987).

A currently popular method is the one based on *neural networks.* In a neural network, the receptors and the information they receive can be identified with a set of *input neurons* and their *activity values*. This set of values will be called the *input vector*. In most cases there is a large number of input neurons which means that the dimensionality of the input vector is very high. The purpose of an inductive method at this subconceptual level is to reduce the complexity of the input information in an efficient and systematic way.

For example, we may consider Kohonen’s (1988) *self-organizing feature maps*. The distinguishing property of these maps is that they are able to describe the topological relations of the signals in the input vector using something like a conceptual space with a small number of dimensions. Basically, the mapping can be seen as reducing the dimensionality of the input vector.

A self-organizing feature map is a neural network which consists of an input vector that is connected to an output array of neurons. In most applications, this array is one- or two-dimensional, but in principle it could be of any number of dimensions. The essential property of the network is that the connections between the neurons in the array and the learning function are organized in such a way that *similarities* that occur among different input vectors are *preserved* in the mapping, in the sense that input vectors that have common features are mapped onto *neighbouring* neurons in the map. The degree of similarity between two input vectors is determined by some *distance* measure.

In other words, the mapping from the input vector to the array preserves the topological relations while reducing the dimensionality of the representation space. The low-dimensional "feature map" that results as an output of the process can be viewed as a conceptual space in the sense of the preceding sections. The mapping is *generated* by the network itself via the learning mechanism of the network. In practice, it normally takes quite a large number of learning instances before the network stabilizes enough so that further changes can be ignored.

What are then the drawbacks of using neural networks of the type described here for inductive processes? A fundamental epistemological problem is that even if we know that the network will generate Kohonen surfaces that perform the right kind of job, we may not be able to "describe" *what* the emerging dimensions represent. Even if we, for example, know that a system consisting of three one-dimensional Kohonen surfaces provides a perfect classification of class of objects, this may not help us in *interpreting* the "meaning" of the surfaces, i.e., what overall features of the objects that they *represent.* In other words, we may not be able to make the transition between the subconceptual level and the conceptual level. This kind of level problem is ubiquitous in applications of neural networks. The upshot is that a future theory of neural networks must somehow bridge the gap of going from the subconceptual level to the conceptual level. We may account for the information provided at the subconceptual level in terms of a dimensional space with some topological structure, but there is no general recipe for determining what is the *conceptual* meaning of the dimensions of the space.

9. Conclusion: What is induction?

Where on the three levels that have been described here is real induction to be found? It has been argued here and in Gärdenfors (1994) that there are several kinds of inductive processes. Depending on what perspective one takes on observations, different ways of generalizing the observations become relevant. Traditional philosophy of science has concealed these distinctions by neglecting the conceptual and subconceptual levels. For a complete account of induction, all three levels must be mustered.

What is the relation between the three levels? I hope it has become clear from my presentation that I do not view the three levels as being in conflict with each other. They should rather be regarded as three *perspectives* on observations that complement each other. Different aspects of inductive processes need to be explained on different levels. By disregarding some level one restricts the possibilities for understanding the mechanisms of inductive reasoning.

Even though all three perspectives are necessary to provide a complete picture of different kinds of inductive processes in science and in everyday life, the conceptual level has a special status for several reasons. Firstly, scientific laws and theories are mainly formulated under the assumption of an underlying conceptual space with a specific set of dimensions. Secondly, the conceptual level provides the *semantics* for the expressions on the symbolic level. Finally, the information expressed on this level serves as a comprehensive summary of all what reaches sensors of different kinds at the subconceptual level. Since research on inductive processes from the Vienna Circle and on has focused on the symbolic level, I propose that the conceptual level should be given much more attention in the future.

Acknowledgements

Research for this paper has been supported by the Swedish Council for Research in the Humanities and Social Sciences. The article draws on material in Gärdenfors (1990, 1993, and 1994) and some of the ideas on prototype theory are borrowed from Gärdenfors (1991).

References

Berlin, B. and P. Kay (1969), *Basic Color Terms: Their Universality and Evolution,* Berkeley, CA: University of California Press

Carnap, R. (1950), *Logical Foundations of Probability,* Chicago, IL: Chicago University Press

Carnap, R. (1971), "A basic system of inductive logic", Part 1, in R. Carnap and R. C. Jeffrey (eds.), *Studies in Inductive Logics and Probability, Vol. 1*, University of Berkeley, CA: California Press, pp. 35-165

Carnap, R. (1980), "A basic system of inductive logic", Part 2 in R. C. Jeffrey (ed.), *Studies in Inductive Logics and Probability, Vol. 2*, Berkeley, CA: University of California Press, pp. 7-155

Churchland, P. S. (1986), *Neurophilosophy: Toward a Unified Science of the Mind/Brain*. Bradford Books, Cambridge, MA: MIT Press

Churchland, P. M. (1986), "Some reductive strategies in cognitive neurobiology", in *Mind 95*, no. 379, pp: 279-309

Foss, J. (1988), "The percept and vector function theories of the brain", *Philosophy of Science 55*, pp. 511-537

Gärdenfors, P. (1988), "Semantics, conceptual spaces and the dimensions of music", in *Essays on the Philosophy of Music*, V. Rantala, L. Rowell, and E. Tarasti (eds.) *Acta Philosophica Fennica*, vol. 43, Helsinki, pp. 9-27

Gärdenfors, P. (1990), "Induction, conceptual spaces and AI", in *Philosophy of Science*, *57*, pp. 78-95

Gärdenfors, P. (1991), "Frameworks for properties: Possible worlds vs. conceptual spaces", in *Acta Philosophica Fennica 49*, 383-407

Gärdenfors, P. (1993), "Induction and the evolution of conceptual spaces", in *Charles S. Peirce and the Philosophy of Science*, E. C. Moore (ed.), Tuscaloosa: The University of Alabama Press, pp. 72-88.

Gärdenfors, P. (1994), "Three levels of inductive inference", in *Logic, Methodology and Philosophy of Science IX*, D. Prawitz, B. Skyrms, and D. Westerståhl (eds.), Amsterdam: Elsevier Science, pp. 427-449.

Gärdenfors, P. (1995), "Conceptual spaces as a framework for cognitive semantics", to appear in *Philosophy and Cognitive Science: Categories, Consciousness and Reasoning*, J. Ezquerro (ed.), Dordrecht: Kluwer

Goodman, N. (1955), *Fact, Fiction, and Forecast,* Cambridge, MA: Harvard University Press

Hempel, C. G. (1965), *Aspects of Scientific Explanation, and Other Essays in the Philosophy of* Science, New York: Free Press

Johnson, W. E. (1921), *Logic, Part I*, Cambridge: Cambridge University Press

Kohonen, T. (1988), *Self-Organization and Associative Memory*, Second Edition, Berlin: Springer-Verlag

Kuhn, T. (1970), *The Structure of Scientific Revolutions*, 2nd ed., Chicago, IL: University of Chicago Press

Lakoff, G. (1987), *Women, Fire, and Dangerous Things*, Chicago, IL: University of Chicago Press

Land, E. (1977), "The retinex theory of color vision", in *Scientific American 237*, pp. 108-128

Langacker, R. W. (1986), *Foundations of Cognitive Grammar (Vol. 1)*, Stanford, CA: Stanford University Press

Mervis, C. and E. Rosch (1981), "Categorization of natural objects", in *Annual Review of Psychology 32*, pp. 89-115

Quine, Willard V. O. (1960), *Word and Object*, Cambridge, MA: MIT Press

Quine, Willard V. O. (1969), "Natural kinds", in *Ontological Relativity and Other Essays*, New York: Columbia University Press, pp. 114-138

Rosch, E. (1975), "Cognitive representations of semantic categories", in *Journal of Experimental Psychology: General 104*, pp. 192-233.

Rosch, E. (1978), "Prototype classification and logical classification: The two systems", in *New Trends in Cognitive Representation: Challenges to Piaget's Theory*, E. Scholnik (ed.), Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 73-86.

Shapere, D. (1982), "The concept of observation in science and philosophy", in *Philosophy of Science*, *49*, pp. 485-525.

Shepard, R. N. (1962a), "The analysis of proximities: Multidimensional scaling with an unknown distance function. I", in *Psychometrika*, *27*, pp. 125-140.

Shepard, R. N. (1962b), "The analysis of proximities: Multidimensional scaling with an unknown distance function. II", in *Psychometrika*, *27*, pp. 219-246.

Shepard, R. N. (1987), "Toward a universal law of generalization for psychological science", in *Science, 237*, pp. 1317-1323.

Sloman, A. (1971), "Interactions between philosophy and AI – the role of intuition and non-logical reasoning in intelligence", in *Proceedings 2nd IJCAI*, London, reprinted in *Artificial Intelligence*, vol. 2.

Smith, E. and Medin, D. L. (1981), *Categories and Concepts*, Cambridge, MA: Harvard University Press

Smolensky, P. (1988). On the proper treatment of connectionism, in *Behavioral and Brain Sciences*, *11*, pp. 1-23.

Vaina, L. (1983), "From shapes and movements to objects and actions", in *Synthese 54*, pp. 3-36.