Chapter 1
Dimensions

1.1 *The Problem of Modeling Representations*

1.1.1 *Three Levels of Representation*

Cognitive science has two overarching goals. One is *explanatory*: by studying the cognitive activities of humans and other animals, the scientist formulates *theories* of different aspects of cognition. The theories are tested by experiments or by computer simulations. The other goal is *constructive*: by building *artifacts* like robots, animats, chess-playing programs, and so forth, cognitive scientists aspire to construct systems that can accomplish various cognitive tasks. A key problem for both kinds of goals is how the *representations* used by the cognitive system are to be modeled in an appropriate way.

Within cognitive science, there are currently two dominating approaches to the problem of modeling representations. The *symbolic* approach starts from the assumption that cognitive systems can be described as Turing machines. From this view, cognition is seen as essentially being *computation*, involving symbol manipulation. The second approach is *associationism*, where associations among different kinds of information elements carry the main burden of representation.[1] *Connectionism* is a special case of associationism that models associations using artificial neuron networks. Both the symbolic and the associationistic approaches have their advantages and disadvantages. They are often presented as competing paradigms, but since they attack cognitive problems on different levels, I argue later that they should rather be seen as complementary methodologies.

There are aspects of cognitive phenomena, however, for which neither symbolic representation nor associationism appear to offer appropriate modeling tools. In particular it appears that mechanisms of *concept acquisition*, which are paramount for the understanding of many cognitive phenomena, cannot be given a satisfactory treatment in any of these representational forms. Concept learning is closely tied to the notion of *similarity*, which has turned out to be problematic for the symbolic and associationistic approaches.

Here, I advocate a third form of representing information that is based on using *geometrical* structures rather than symbols or connections among neurons. On the basis of these structures, similarity relations can be modeled in a natural way. I call my way of representing information the *conceptual* form because I believe that the essential aspects of concept formation are best described using this kind of representation.

The geometrical form of representation has already been used in several areas of the cognitive sciences. In particular, dimensional representations are frequently employed within cognitive psychology. As will be seen later in the book, many models of concept formation and learning are based on spatial structures. Suppes et al. (1989) present the general mathematics that are applied in such models. But geometrical and topological notions also have been exploited in linguistics. There is a French tradition exemplified by Thom (1970), who very early applied catastrophe theory to linguistics, and Petitot (1985, 1989, 1995). And there is a more recent development within cognitive linguistics where researchers like Langacker (1987), Lakoff (1987), and Talmy (1988) initiated a study of the spatial and dynamic structure of "image schemas," which clearly are of a conceptual form.[2] As will be seen in the following chapter, several spatial models have also been proposed within the neurosciences.

The conceptual form of representions, however, has to a large extent been neglected in the foundational discussions of representations. It has been a common prejudice in cognitive science that the brain is either a Turing machine working with symbols or a connectionist system using neural networks. One of my objectives here is to show that a conceptual mode based on geometrical and topological representations deserves at least as much attention in cognitive science as the symbolic and the associationistic approaches.

Again, the conceptual representations should not be seen as competing with symbolic or connectionist (associationist) representations. There is no unique correct way of describing cognition. Rather, the three kinds mentioned here can be seen as three *levels* of representations of cognition with different scales of resolution.[3] Which level provides the best explanation or ground for technical constructions depends on the cognitive problem area that is being modeled.

1.1.2 Synopsis

This is a book about the geometry of thought. A theory of *conceptual spaces* will be developed as a particular framework for representing information on the conceptual level. A conceptual space is built upon geometrical structures based on a number of *quality dimensions*. The main applications of the theory will be on the constructive side of cognitive science. I believe, however, that the theory can also explain several aspects of what is known about representations in various biological systems. Hence, I also attempt to connect the theory of conceptual spaces to empirical findings in psychology and neuroscience.

Chapter 1 presents the basic theory of conceptual spaces and, in a rather informal manner, some of the underlying mathematical notions. In chapter 2, representations in conceptual spaces are contrasted to those in symbolic and connectionistic models. It argues that symbolic and connectionistic representations are not sufficient for the aims of cognitive science; many representational problems are best handled by using geometrical structures on the conceptual level.

In the remainder of the book, the theory of conceptual spaces is used as a basis for a constructive analysis of several fundamental notions in philosophy and cognitive science. In chapter 3 is argued that the traditional analysis of *properties* in terms of possible worlds semantics is misguided and that a much more natural account can be given with the aid of conceptual spaces. In chapter 4, this analysis is extended to *concepts* in general. Some experimental results about concept formation will be presented in this chapter. In both chapters 3 and 4, the notion of similarity will be central.

In chapter 5, a general theory for cognitive *semantics* based on conceptual spaces is outlined. In contrast to traditional philosophical theories, this kind of semantics is connected to perception, imagination, memory, communication, and other cognitive mechanisms.

The problem of *induction* is an enigma for the philosophy of science, and it has turned out to be a problem also for systems within artificial intelligence. This is the topic of chapter 6 where it is argued that the classical riddles of induction can be circumvented, if inductive reasoning is studied on the conceptual level of representation instead of on the symbolic level.

The three levels of representation will motivate different types of computations. Chapter 7 is devoted to some computational aspects with the conceptual mode of representation as the focus. Finally, in chapter 8 the research program associated with representations in conceptual spaces is summarized and a general methodological program is proposed.

As can be seen from this overview, I throw my net widely around several problem areas within the cognitive science. The book has two main aims. One is to argue that the conceptual level is the best mode of representation for many problem areas within cognitive science. The other aim is more specific; I want to establish that conceptual spaces can serve as a framework for a number of empirical theories, in particular concerning concept formation, induction, and semantics. I also claim that conceptual spaces are useful representational tools for the constructive side of cognitive science. As an independent issue, I argue that conceptual representations serve as a bridge between symbolic and connectionist ones. In support of this position, Jackendoff (1983, 17) writes: "There is a single level of mental representation, *conceptual structure*, at which linguistic, sensory, and motor information are compatible." The upshot is that the conceptual level of representation ought to be given much more emphasis in future research on cognition.

It should be obvious by now that it is well nigh impossible to give a thorough treatment of all the areas mentioned above within the covers of a single book. Much of my presentation will, unavoidably, be programmatic and some arguments will, no doubt, be seen as rhetorical. I hope, however, that the examples of applications of conceptual spaces presented in this book inspire new investigations into the conceptual forms of representation and further discussions of representations within the cognitive sciences.

1.2 Conceptual Spaces as a Framework for Representations

We frequently compare the experiences we are currently having to memories of earlier episodes. Sometimes, we experience something entirely new, but most of the time what we see or hear is, more or less, the same as what we have already encountered. This cognitive capacity shows that we can judge, consciously or not, various relations among our experiences. In particular, we can tell how *similar* a new phenomenon is to an old one.

With the capacity for such judgments of similarity as a background, philosophers have proposed different kinds of theories about how humans concepts are structured. For example, Armstrong (1978, 116) presents the following desiderata for *an analysis* of what unites concepts:[4]

> If we consider the class of shapes and the class of colours, then both classes exhibit the following interesting but puzzling characteristics which it should be able to understand:
> (a) the members of the two classes all have something in common (they are all shapes, they are all colours)
> (b) but while they have something in common, they differ in that very respect (they all differ as shapes, they all differ as colours)
> (c) they exhibit a resemblance order based upon their intrinsic nature (*triangularity* is like *circularity*, *redness* is more like *orangeness* than *redness* is like *blueness*), where closeness of resemblance has a limit in identity

(d) they form a set of incompatibles (the same particular cannot be simultaneously triangular and circular, or red and blue all over).

The epistemological role of the theory of conceptual spaces to be presented here is to serve as a tool in modeling various *relations* among our experiences, that is, what we perceive, remember, or imagine. In particular, the theory will satisfy Armstrong's desiderata as shown in chapter 3. In contrast, it appears that in symbolic representations the notion of similarity has been severely downplayed. Judgments of similarity, however, are central for a large number of cognitive processes. As will be seen later in this chapter, such judgments reveal the *dimensions* of our perceptions and their structures (compare Austen Clark 1993).

When attacking the problem of representing concepts, an important aspect is that the concepts are not independent of each other but can be structured into *domains*; spatial concepts belong to one domain, concepts for colors to a different domain, kinship relations to a third, concepts for sounds to a fourth, and so on. For many modeling applications within cognitive science it will turn out to be necessary to separate the information to be represented into different domains.

The key notion in the conceptual framework to be presented is that of a *quality dimension*. The fundamental role of the quality dimensions is to build up the domains needed for representing concepts. Quality dimensions will be introduced in the following section via some basal examples.

The structure of many quality dimensions of a conceptual space will make it possible to talk about *distances* along the dimensions. There is a tight connection between distances in a conceptual space and similarity judgments: the smaller the distances is between the representations of two objects, the more similar they are. In this way, the similarity of two objects can be defined via the distance between their representing points in the space. Consequently, conceptual spaces provide us with a natural way of representing similarities.

Depending on whether the explanatory or the constructive goal of cognitive science is in focus, two different interpretations of the quality dimensions will be relevant. One is *phenomenal*, aimed at describing the psychological structure of the perceptions and memories of humans and animals. Under this interpretation the theory of conceptual space will be seen as a theory with testable consequences in human and animal behavior.

The other interpretation is *scientific* where the structure of the dimensions used is often taken from some scientific theory. Under this interpretation the dimensions are not assumed to have any psychological validity but are seen as instruments for predictions. This interpretation is oriented more toward the constructive goals of cognitive science. The two interpretations of the quality dimensions are discussed in section 1.4.

1.3 Quality Dimensions

As first examples of quality dimensions, one can mention *temperature*, *weight*, *brightness*, *pitch* and the three ordinary spatial dimensions *height*, *width*, and *depth*. I have chosen these examples because they are closely connected to what is produced by our sensory receptors (Schiffman 1982). The spatial dimensions height, width, and depth as well as brightness are perceived by the visual sensory system,[5] pitch by the auditory system, temperature by thermal sensors and weight, finally, by the kinaesthetic sensors. As explained later in this chapter, however, there is also a wealth of quality dimensions that are of an abstract non-sensory character.

The primary function of the quality dimensions is to represent various "qualities" of objects.[6] The dimensions correspond to the different ways stimuli are judged to be similar or different.[7] In most cases, judgments of similarity and difference generate an *ordering relation* of stimuli. For example, one can judge tones by their pitch, which will generate an ordering from "low" to "high" of the perceptions.

The dimensions form the framework used to assign *properties* to objects and to specify *relations* among them. The coordinates of a point within a conceptual space represent particular instances of each dimension, for example, a particular temperature, a particular weight, and so forth. Chapter 3 will be devoted to how properties can be described with the aid of quality dimensions in conceptual spaces. The main idea is that a property corresponds to a *region* of a domain of a space.

The notion of a dimension should be understood literally. It is assumed that each of the quality dimensions is endowed with certain *geometrical* structures (in some cases they are *topological* or *ordering* structures). I take the dimension of "time" as a first example to illustrate such a structure (see figure 1.1). In science, time is modeled as a one-dimensional structure that is isomorphic to the line of real

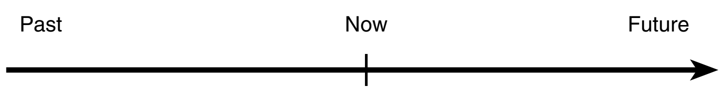

Figure 1.1
The time dimension.

numbers. If "now" is seen as the zero point on the line, the future corresponds to the infinite positive real line and the past to the infinite negative line.

This representation of time is not phenomenally given but is to some extent culturally dependent. People in other cultures have a different time dimension as a part of their cognitive structures. For example, in some cultural contexts, time is viewed as a *circular* structure. There is, in general, no unique way of choosing a dimension to represent a particular quality but a wide array of possibilities.

Another example is the dimension of "weight" which is one-dimensional with a zero point and thus isomorphic to the half-line of nonnegative numbers (see figure 1.2). A basic constraint on this dimension that is commonly made in science is that there are no negative weights.[8]

It should be noted that some quality "dimensions" have only a *discrete* structure, that is, they merely divide objects into disjoint classes. Two examples are classifications of biological species and kinship relations in a human society. One example of a phylogenetic tree of the kind found in biology is shown in figure 1.3. Here the nodes represent different species in the evolution of, for example, a family of organisms, where nodes higher up in the tree represent evolutionarily older (extinct) species.

The distance between two nodes can be measured by the length of the path that connects them. This means that even for discrete

Figure 1.2
The weight dimension.

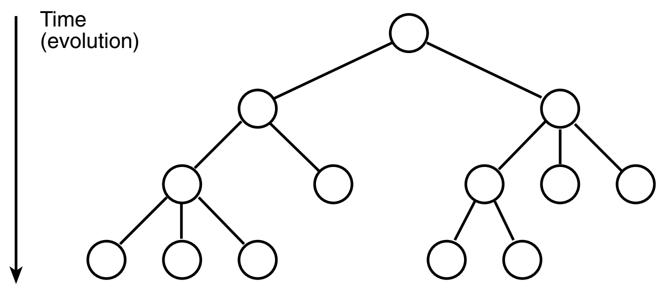

Figure 1.3
A phylogenetic tree.

dimensions one can distinguish a rudimentary geometrical structure. For example, in the phylogenetic classification of animals, it is meaningful to say that birds and reptiles are *more closely related* than reptiles and crocodiles. Some of the properties of discrete dimensions, in particular in graphs, are further discussed in section 1.6 where a general mathematical framework for describing the structures of different quality dimensions will be provided.

1.4 Phenomenal and Scientific Interpretations of Dimensions

To separate different uses of quality dimensions it is important to introduce a distinction between a *phenomenal* (or *psychological*) and a *scientific* (or *theoretical*) interpretation (compare Jackendoff 1983, 31–34). The phenomenal interpretation concerns the cognitive structures (perceptions, memories, etc.) of humans or other organisms. The scientific interpretation, on the other hand, treats dimensions as a part of a scientific theory.[9]

As an example of the distinction, our phenomenal visual space is not a perfect 3-D Euclidean space, since it is not invariant under all linear transformations. Partly because of the effects of gravity on our perception, the vertical dimension (height) is, in general, overestimated in relation to the two horizontal dimensions. That is why the moon looks bigger when it is closer to the horizon, while it in fact has the same "objective" size all the time. The scientific representation of visual space as a 3-D Euclidean space, however, is an idealization that is mathematically amenable. Under this description, all spatial directions have the same status while "verticality" is treated differently under the phenomenal interpretation. As a consequence, all linear coordinate changes of the scientific space preserve the structure of the space.

Another example of the distinction is color which is supported here by Gallistel (1990, 518–519) who writes:

> The facts about color vision suggest how deeply the nervous system may be committed to representing stimuli as points in descriptive spaces of modest dimensionality. It does this even for spectral compositions, which does not lend itself to such a representation. The resulting lack of correspondence between the psychological representation of spectral composition and spectral composition itself is a source of confusion and misunderstanding in scientific discussions of color. Scientists persist in refering to the physical characteristics of the stimulus and to the tuning characteristics of the transducers (the cones) as if psychological color terms like *red*, *green*, and *blue* had some straightforward translation into physical reality, when in fact they do not.

Gallistel's warning against confusion and misunderstanding of the two types of representation should be taken seriously.[10] It is very easy to confound what science says about the characteristics of reality and our perceptions of it.

The distinction between the phenomenal and the scientific interpretation is relevant in relation to the two goals of cognitive science presented above. When the dimensions are seen as cognitive entities—that is, when the goal is to explain naturally occuring cognitive processes—their geometrical structure should not be derived from scientific theories that attempt to give a "realistic" description of the world, but from *psychophysical* measurements that determine how our phenomenal spaces are structured. Furthermore, when it comes to providing a semantics for a natural language, it is the phenomenal interpretations of the quality dimensions that are in focus, as argued in chapter 5.

On the other hand, when we are *constructing* an artificial system, the function of sensors, effectors, and various control devices are in general described in scientifically modeled dimensions. For example, the input variables of a robot may be a small number of physically measured magnitudes, like the brightness of a patch from a video image, the delay of a radar echo, or the pressure from a mechanical grip. Driven by the programmed goals of the robot, these variables can then be transformed into a number of physical output magnitudes, for example, as the voltages of the motors controlling the left and the right wheels.

1.5 Three Sensory Examples: Color, Sound, and Taste

A phenomenally interesting example of a set of quality dimensions concerns *color perception*. According to the most common perceptual models, our cognitive representation of colors can be described by three dimensions: hue, chromaticness, and brightness. These dimensions are given slightly different mathematical mappings in different models. Here, I focus on the Swedish natural color system (NCS) (Hård and Sivik 1981) which is extensively discussed by Hardin (1988, chapter 3). NCS is a descriptive model—it represents the phenomenal structure of colors, not their scientific properties.

The first dimension of NCS is *hue*, which is represented by the familiar *color circle*. The value of this dimension is given by a *polar* coordinate describing the angle of the color around the circle (see figure 1.4). The geometrical 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 geometry is to note that we can talk about phenomenologically *complementary*

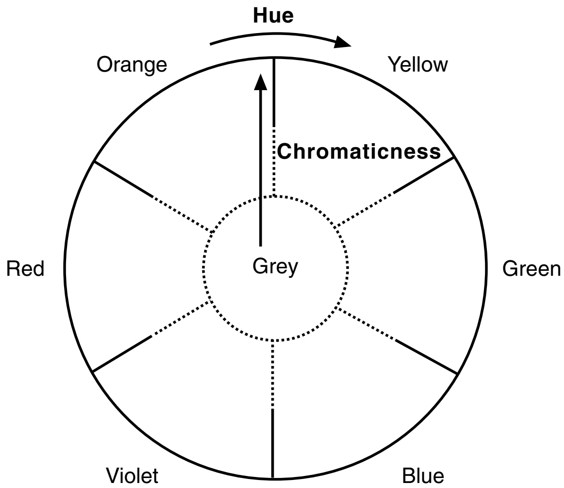

Figure 1.4
The color circle.

colors—colors that lie *opposite* each other on the color circle. In contrast it is *not meaningful* to talk about two points of time or two weights being "opposite" each other.

The second phenomenal dimension of color is *chromaticness* (saturation), which ranges from grey (zero color intensity) to increasingly greater intensities. This dimension is isomorphic to an interval of the real line.[11] The third dimension is *brightness* which varies from white to black and is thus a linear dimension with two end points. The two latter dimensions are not totally independent, since the possible variation of the chromaticness dimension decreases as the values of the brightness dimension approaches the extreme points of black and white, respectively. In other words, for an almost white or almost black color, there can be very little variation in its chromaticness. This is modeled by letting that chromaticness and brightness dimension together generate a triangular representation (see figure 1.5). Together these three dimensions, one with circular structure and two with linear, make up the color space. This space is often illustrated by the so called *color spindle* (see figure 1.6).

The color circle of figure 1.4 can be obtained by making a *horizontal* cut in the spindle. Different triangles like the one in figure 1.5 can be generated by making a *vertical* cut along the central axis of the color spindle.

As mentioned above, the NCS representation is not the only mathematical model of color space (see Hardin 1988 and Rott 1997 for some

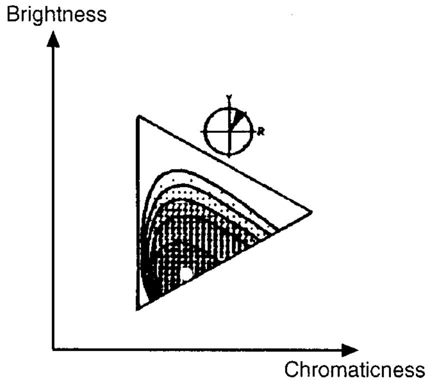

Figure 1.5
The chromaticness-brightness triangle of the NCS (from Sivik and Taft 1994, 150). The small circle marks which sector of the color spindle has been cut out.

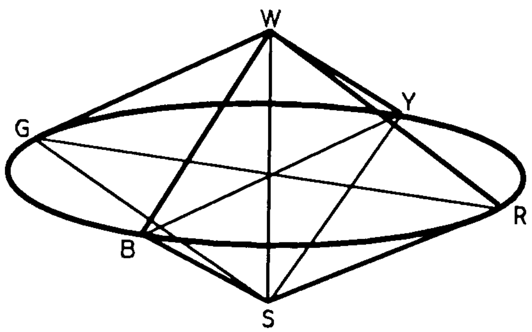

Figure 1.6
The NCS color spindle (from Sivik and Taft 1994, 148).

alternatives). All the alternative models use dimensions, however, and all of them are three-dimensional. Some alternatives replace the circular hue by a structure with corners. A controversy exists over which geometry of the color space best represents human perception. There is no unique answer, since the evaluation partly depends on the aims of the model. By focusing on the NCS color spindle in my applications, I do not claim that this is the optimal representation, but only that it is suitable for illustrating some aspects of color perception and of conceptual spaces in general.

The color spindle represents the phenomenal color space. Austen Clark (1993, 181) argues that physical properties of light are not relevant when describing color space. His distinction between intrinsic and extrinsic features in the following quotation corresponds to the distinction between phenomenal features and those described by scientific theories:

> [A]n analysis of sensory qualities should mention only intrinsic features of the quality space: extrinsic features can be no part of the analysis.
>
> This suggestion implies that the meaning of a colour predicate can be given only in terms of its relations to other colour predicates. The place of the colour in the psychological colour solid is defined by those relations, and it is only its place in the solid that is relevant to its identity. . . .

More general support for the second part of the quotation have been given by Shepard and Chipman (1970, 2) who point out that what is important about a representation is not how it relates to *what* is represented, but how it relates to *other* representations:[12]

> [T]he isomorphism should be sought—not in the first-order relation between (a) an individual object, and (b) its corresponding internal representation—but in the second-order relation between (a) the relations among alternative external objects, and (b) the relations among their corresponding internal representations. Thus, although the internal representation need not itself be square, it should (whatever it is) at least have a closer functional relation to the internal representation for a rectangle than to that, say, for a green flash or the taste of persimmon.

The "functional relation" they refer to concerns the tendency of different responses to be activated together. Such tendencies typically show up in similarity judgments. Thus, because of the structure of the color space, we judge that red is more similar to purple than to yellow, for example, even though we cannot say what it is in the subjective experience of the colors that causes this judgment.[13]

Nevertheless, there are interesting connections between phenomenal and physical dimensions, even if they are not perfectly matched. The hue of a color is related to the wavelengths of light, which thus is the main dimension used in the scientific description of color. Visible light occurs in the range of 420–700 nm. The geometrical structure of the (scientific) wavelength dimension is thus linear, in contrast to the circular structure of the (phenomenal) hue dimension.

The neurophysiological mechanisms underlying the mental representation of color space are comparatively well understood. In particular, it has been established that human color vision is mediated by the cones in the retina which contain three kinds of pigments. These pigments are maximally sensitive at 445 nm (blue-violet), 535 nm (green)

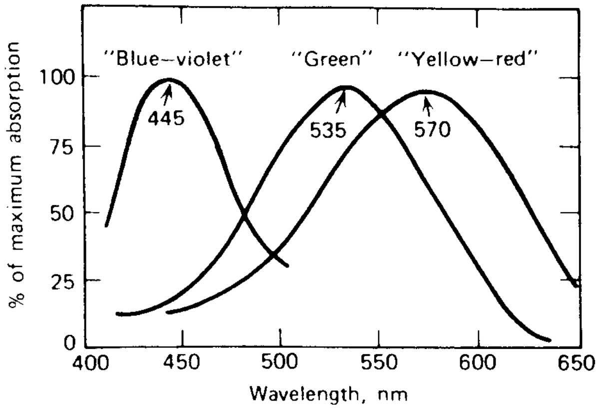

Figure 1.7
Absorption spectra for three types of cone pigments (from Buss 1973, 203).

and 570 nm (yellow-red) (see figure 1.7). The perceived color emerges as a mixture of input from different kinds of cones. For instance, "pure" red is generated by a mixture of signals from the blue-violet and the yellow-red sensitive cones.

The connections between what excites the cones and rods in the retina, however, and what color is *perceived* is far from trivial. According to Land's (1977) results, the perceived color is not directly a function of radiant energy received by the cones and rods, but rather it is determined by "lightness" values computed at three wavelengths.[14]

Human color vision is thus trichromatic. In the animal kingdom we find a large variation of color systems (see for example Thompson 1995); many mammals are dichromats, while others (like goldfish and turtles) appear to be tetrachromats; and some may even be pentachromats (pigeons and ducks). The precise geometric structures of the color spaces of the different species remain to be established (research which will involve very laborious empirical work). Here, it suffices to say that the human color space is but one of many evolutionary solutions to color perception.

We can also find related spatial structures for other sensory qualities. For example, consider the quality dimension of *pitch*, which is basically a continuous one-dimensional structure going from low tones to high. This representation is directly connected to the neurophysiology of pitch perception (see section 2.5).

Apart from the basic frequency dimension of tones, we can find some interesting further structure in the cognitive representation of tones. Natural tones are not simple sinusoidal tones of one frequency only but constituted of a number of higher harmonics. The timbre of a tone, which is a phenomenal dimension, is determined by the relative strength of the higher harmonics of the fundamental frequency of the tone. An interesting perceptual phenomenon is "the case of the missing fundamental." This means that if the fundamental frequency is removed by artificial methods from a complex physical tone, the phenomenal pitch of the tone is still perceived as that corresponding to the removed fundamental.[15] Apparently, the fundamental frequency is not indispensable for pitch perception, but the perceived pitch is determined by a combination of the lower harmonics (compare the "vowel space" presented in section 3.8).

Thus, the harmonics of a tone are essential for how it is perceived: tones that share a number of harmonics will be perceived to be similar. The tone that shares the most harmonics with a given tone is its octave, the second most similar is the fifth, the third most similar is the fourth, and so on. This additional "geometrical" structure on the pitch dimension, which can be derived from the wave structure of tones, provides the foundational explanation for the perception of musical *intervals*.[16] This is an example of higher level structures of conceptual spaces to be discussed in section 3.10.

As a third example of sensory space representations, 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 taste could be described as a four-dimensional space. One such model was put forward by Henning (1916), who suggested that phenomenal gustatory space could be described as a tetrahedron (see figure 1.8). Henning speculated that any taste could be described as a mixture of only three primaries. This means that any taste can be rep-

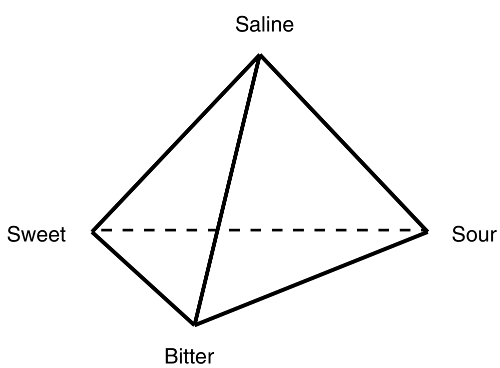

Figure 1.8
Henning's taste tetrahedron.

resented as a point on one of the *planes* of the tetrahedron, so that no taste is mapped onto the interior of the tetrahedron.

There are other models, however, that propose more than four fundamental tastes.[17] Which is the best model of the phenomenal gustatory space remains to be established. This will involve sophisticated psychophysical measurement techniques. Suffice it to say that the gustatory space quite clearly has some nontrivial geometrical 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.

1.6 Some Mathematical Notions

The dimensions of conceptual spaces, as illustrated in these examples, are supposed to satisfy certain structural constraints. In this section, some of the mathematical concepts that will be used in the following chapters are presented in greater detail. Since most of the examples of quality dimensions will have geometrical structures, I focus here on some fundamental notions of geometry.[18]

An axiomatic system for geometry can, in principle, be constructed from two primitive relations, namely *betweenness* and *equidistance* defined over a space of *points*. In most treatments, however, *lines* and *planes* are also taken to be primitive concepts (see, for example, Borsuk and Szmielew 1960), but these notions will only play a marginal role here.

1.6.1 Betweenness

One of the fundamental geometrical relations is betweenness, a concept frequently applied in this book. Let S denote the set of all points in a space. The betweenness relation is a ternary relation $B(a, b, c)$ over points in S, which is read as "point b lies between points a and c." The relation is supposed to satisfy some fundamental axioms. The simplest ones are the following:[19]

> *B0*: If $B(a, b, c)$, then a, b and c are distinct points.
> *B1*: If $B(a, b, c)$, then $B(c, b, a)$.

In words: "If b is between a and c, then b is between c and a."

> *B2*: If $B(a, b, c)$, then not $B(b, a, c)$.

"If b is between a and c, then a is not between c and b."

> *B3*: If $B(a, b, c)$ and $B(b, c, d)$, then $B(a, b, d)$.

"If b is between a and c and c is between b and d, then b is between a and d."

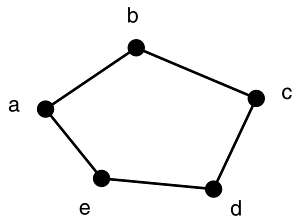

Figure 1.9
Graph violating axiom B3 when *B*(*a*, *b*, *c*) is defined as "*b* is on the *shortest* path from *a* to *c*."

B4: If *B*(*a*, *b*, *d*) and *B*(*b*, *c*, *d*), then *B*(*a*, *b*, *c*).

"If *b* is between *a* and *d* and *c* is between *b* and *d*, then *b* is between *a* and *c*."

These axioms are satisfied for a large number of ordered structures.[20] It is easy to see that they are true of ordinary Euclidean geometry; however, they may also be valid in some "weaker" structures like graphs. If we define *B*(*a*, *b*, *c*) as "there is some path from *a* to *c* that passes through *b*," then axioms B1, B3, and B4 are all valid. B2 is also valid if the graph is a tree (that is, if it does not contain any loops).

In contrast, if *B*(*a*, *b*, *c*) is defined as "*b* is on the *shortest* path from *a* to *c*," then axiom B3 need not be valid in all graphs as figure 1.9 shows. In this figure, *b* is on the shortest path from *a* to *c*, and *c* is on the shortest path from *b* to *d*, but *b* is not on the shortest path from *a* to *d* (nor is *c*).

This example shows that for a given ordered structure there may be *more than one way* of defining a betweenness relation. I will come back to this point in the following chapters, as it is important for an analysis of concept formation.

From B1–B4 it immediately follows:

LEMMA 1.1 (i) If *B*(*a*, *b*, *c*) and *B*(*b*, *c*, *d*), then *B*(*a*, *c*, *d*); (ii) If *B*(*a*, *b*, *d*) and *B*(*b*, *c*, *d*), then *B*(*a*, *c*, *d*).

In principle, the notion of a *line* can be defined with the aid of the betweenness relation (Borsuk and Szmielew 1960, 57), by saying that the line through points *a* and *c*, in symbols L_{ac}, consists of the set of points *b* such that *B*(*a*, *b*, *c*) or *B*(*b*, *a*, *c*) or *B*(*a*, *c*, *b*) (together with the points *a* and *c* themselves). Unless further assumptions are made, however, concerning the structure of the set *S* of points, the lines defined in this way may not look like the lines we know from ordinary geometry. For example, the line between *a* and *d* in figure 1.9 will consist of all the points *a*, *b*, *c*, *d* and *e*.[21] Still, one can prove the following property of all lines:

LEMMA 1.2 If a, b, c and d are four points on a line and $B(a, b, d)$ and $B(a, c, d)$, then either $B(a, b, c)$ or $B(a, c, b)$ or $b = c$.

Furthermore, a *plane* can be defined with the aid of lines and the betweenness relation.[22] Once we have the notions of lines and planes, most of traditional geometry can be constructed.

The basic axioms for betweenness can be supplemented with an axiom for *density*:

> *B5*: For any two points a and c in S, there is some point b such that $B(a, b, c)$.

Of course, there are quality dimensions; for example, all discrete dimensions, for which axiom B5 is not valid.

As is well known from the theory of rational numbers, density does not imply *continuity*.[23]

1.6.2 Equidistance

The second primitive notion of geometry is that of *equidistance*. It is a four-place relation $E(a, b, c, d)$ which is read as "point a is just as far from point b as point c is from point d." The basic axioms for the relation E are the following (Borsuk and Szmielew 1960, 60):

> *E1*: If $E(a, a, p, q)$, then $p = q$.
> *E2*: $E(a, b, b, a)$.
> *E3*: If $E(a, b, c, d)$ and $E(a, b, e, f)$, then $E(c, d, e, f)$.

LEMMA 1.3 (i) If $E(a, b, c, d)$ and $E(c, d, e, f)$, then $E(a, b, e, f)$. (ii) If $E(a, b, c, d)$, then $E(a, b, d, c)$.[24]

The following axiom connects the betweenness relation B with the equidistance relation E:

> *E4*: If $B(a, b, c)$, $B(d, e, f)$, $E(a, b, d, e)$ and $E(b, c, e, f)$, then $E(a, c, d, f)$.

E4 says essentially that if b is between a and c, then the distance between a and c is the sum of the distance between a and b and the distances between b and c. Because sums of distances cannot be defined explicitly using only the relations B and E, however, the condition is expressed in a purely relational way.

1.6.3 Metric Spaces

The equidistance relation is a qualitative notion of distance. A stronger notion is that of a *metric* space. A real-valued function $d(a,b)$ is said to be a *distance function* for the space S if it satisfies the following conditions for all points a, b, and c in S:

D1: $d(a, b) \geq 0$ and $d(a, b) = 0$ only if $a = b$. (minimality)
D2: $d(a, b) = d(b, a)$. (symmetry)
D3: $d(a, b) + d(b, c) \geq d(a, c)$. (triangle inequality)

A space that has a distance function is called a metric space. For example, in the two-dimensional space R^2, the *Euclidean* distance $d_E(x, y) = \sqrt{((x_1 - y_1)^2 + (x_2 - y_2)^2)}$ satisfies D1–D3. Also a finite graph where the distance between points *a* and *b* is defined as the number of steps on the shortest path between *a* and *b* is a metric space.

In a metric space, one can *define* a betweenness relation *B* and an equidistance relation *E* in the following way:

Def B: $B(a, b, c)$ if and only if $d(a, b) + d(b, c) = d(a, c)$.
Def E: $E(a, b, c, d)$ if and only if $d(a, b) = d(c, d)$.

It is easy to show that if *d* satisfies D1–D3, then *B* and *E* defined in this way satisfies B1, B2, B4, and E1–E4. B3 is not valid in general as is shown by the graph in figure 1.9, where the distance between points *a* and *b* is defined as the number of steps on the shortest path between *a* and *b*. B3, however, is valid in tree graphs.

1.6.4 Euclidean and City-Block Metrics

For the Euclidean distance function, the betweenness relation defined by Def B, results in the standard meaning so that all points between *a* and *b* are the ones on the *straight line* between *a* and *b*. As illustrated in figure 1.10, equidistance can be represented by *circles* in the sense that the set of points at distance *d* from *a* point *c* form *a* circle with *c* as center and *d* as the radius.

There is more then one way, however, of defining a metric on R^2. Another common metric is the so called *city-block* metric, defined as follows, where $|x_1 - y_1|$ denotes the absolute distance between x_1 and y_1:

$$d_C(x, y) = |x_1 - y_1| + |x_2 - y_2|. \tag{1.1}$$

For the city-block measure, the set of points at distance *d* from *a* point *c* form a *diamond* with *c* as center (see figure 1.11).

It should be noted that the city-block metric depends on the *direction* of the *x* and *y* axes in R^2, in contrast to the Euclidean metric, which is invariant under all rotations of the axes. The set of points between points *a* and *b*, as given by Def B, is not a straight line for the city-block metric, but the rectangle generated by *a* and *b* and the directions of the axes (see figure 1.12).

It follows that, for a given space, there is not a unique meaning of "between"; different metrics generate different betweenness relations. Further examples of this are given in chapter 3.

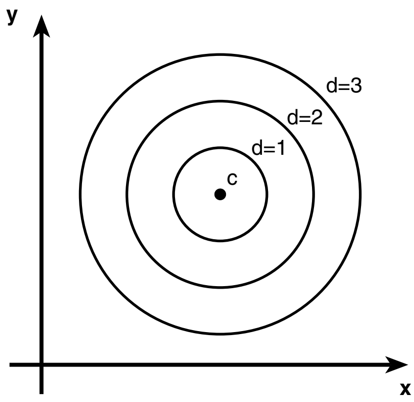

Figure 1.10
Equidistances under the Euclidean metric.

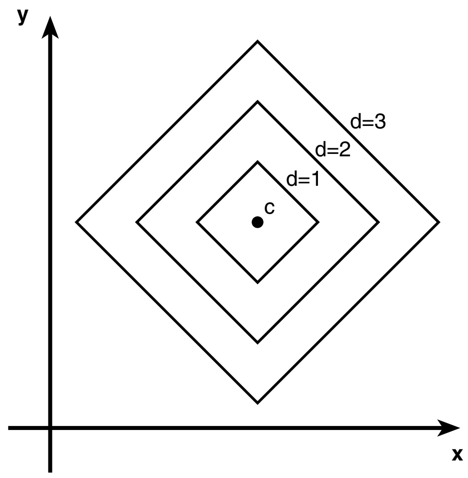

Figure 1.11
Equidistances under the city-block metric.

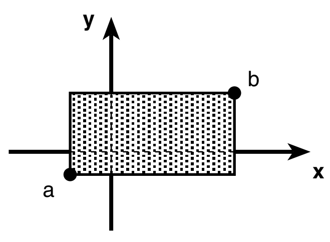

Figure 1.12
The set of points *between* *a* and *b* defined by the city-block metric.

The Euclidean and city-block metrics can be generalized in a straightforward way to the n-dimensional Cartesian space R^n by the following equations:

$$d_E(x, y) = \sqrt{\Sigma_i (x_i - y_i)^2}. \tag{1.2}$$

$$d_C(x, y) = \Sigma_i |x_i - y_i|. \tag{1.3}$$

They are special cases of the class of *Minkowski metrics* defined by

$$d_k(x, y) = \sqrt[k]{\Sigma_i |x_i - y_i|^k}, \tag{1.4}$$

where we thus have as special cases $d_E(x, y) = d_2(x, y)$ and $d_C(x, y) = d_1(x, y)$.[25]

Equations (1.2)–(1.4) presume that the *scales* of the different dimensions are identical so that the distance measured along one of the axes is the same as that measured along another. In psychological contexts, however, this assumption is often violated. A more general form of distance is obtained by putting a weight w_i on the distance measured along dimension i (see, for example, Nosofsky 1986):

$$d_E(x, y) = \sqrt{\Sigma_i w_i \cdot (x_i - y_i)^2}. \tag{1.5}$$

$$d_C(x, y) = \Sigma_i w_i \cdot |x_i - y_i|. \tag{1.6}$$

In these equations, w_i is the "attention-weight" given to dimension i (the role of attention-weights in determining the "salience" of dimensions discussed in section 4.2). Large values of w_i "stretch" the conceptual space along dimension i, while small values of w_i will "shrink" the space along that dimension. In the following, I refer to the more general definitions given by equations (1.5) and (1.6) when Euclidean or city-block distances are mentioned.

1.6.5 Similarity as a Function of Distance

In studies of categorization and concept formation, it is often assumed that the *similarity* of two stimuli can be determined from the distances between the representations of the stimuli in the underlying psychological space. But then what is this functional relation between similarity and distance? A common assumption in the psychological literature (Shepard 1987, Nosofsky 1988a, 1988b, 1992, Hahn and Chater 1997) is that similarity is an *exponentially decaying function* of distance. If s_{ij} expresses the similarity between two objects i and j and d_{ij} their distance, then the following formula, where c is a general "sensitivity" parameter, expresses the relation between the two measures:

$$s_{ij} = e^{-c \cdot d_{ij}}. \tag{1.7}$$

Shepard (1987) calls this the *universal law of generalization* and he argues that it captures the similarity-based generalization performances of subjects in a variety of settings. An underlying motivation for the equation is that matching and mismatching properties are combined multiplicatively rather than additively (Nosofsky 1992, Medin, Goldstone, and Gentner 1993, 258). Given some additional mathematical assumptions, this corresponds to an exponential decay of similarity.

Nosofsky (1986) argues that the exponential function in (1.7) should be replaced by a Gaussian function of the following form:

$$s_{ij} = e^{-c \cdot d_{ij}^2}. \tag{1.8}$$

I will not enter the debate on which of these two functional forms is the more generally valid. Here it suffices to notice that for both equations the similarity between two objects drops quickly when the distance between the objects is relatively small, while it drops much more slowly when the distance is relatively large.

The mathematical notions that have been introduced in this section will prove their usefulness when the theories of properties and concepts are presented in chapters 3 and 4.

1.7 How Dimensions Are Identified

In a conceptual space that is used as a framework for a scientific theory or for construction of an artificial cognitive system, the geometrical or topological structures of the dimensions are *chosen* by the scientist proposing the theory or the constructor building the system. The structures of the dimensions are tightly connected to the *measurement methods* employed to determine the values on the dimensions in experimental situations (compare Sneed 1971 and Suppes et al. 1989). Thus, the choice of dimensions in a given constructive situation will partly depend on what sensors are assumed to be used and their function.

In contrast, the dimensions of a *phenomenal* conceptual space are not obtainable immediately from the perceptions or actions of the subjects, but have to be *infered* from their behavior. There are a number of statistical techniques for identifying the dimensions of a phenomenal space. Here, I only introduce one of the most well-known methods, namely *multidimensional scaling* (MDS).[26] In section 4.10, a different technique will be presented in connection with an analysis of "shell space," and in section 6.5 a method based on artificial neuron networks is described.

If the coordinates of two points are known for all dimensions of a metric conceptual space, it is easy to calculate the *distance* between the points using the metric that goes along with the space. MDS concerns the reverse problem; starting from subjects' judgments about the similarities of two stimuli, MDS is used to determine the number of dimensions in the underlying phenomenal space and the scaling of the space. The goal is to obtain as high a correlation as possible between the similarity judgments of the subjects and the corresponding distances in the estimated dimensional space.

A MDS analysis starts out from a set of data concerning judgments of similarities of a class of stimuli. The similarity judgments can be numerical, but they are often given in an ordinal form obtained from a scale ranging from "very similar" to "very dissimilar." The judgments given by the individual subjects are normally averaged before they are fed into a MDS algorithm (such as Kruskal's 1964 KYST).

The investigator chooses the number *n* of dimensions in the space to be estimated and the metric (normally Euclidean or city-block) to be used in defining distances in the space. Starting from an initial assignment of coordinates to the stimuli in an *n*-dimensional space, the MDS algorithm then systematically adjusts the coordinates to achieve a progressively better fit to the data from the similarity judgments. The degree of misfit between the data and the estimated space is normally measured by a "stress function."[27] The algorithm stops when the stress of the estimated space no longer decreases.

As an example, Shepard (1962a,b) applied "proximity analysis," which is a variant of MDS to a selection of fourteen hues. Subjects were asked to rate the similarity of each pair of hues on an ordinal scale from 1 to 5. The result of the analysis was the structure shown in figure 1.13. In this figure, the MDS program placed the stimuli points in a two-dimensional space. The curved line connecting them was drawn by Shepard. As can be seen, it forms a circle based on two opponent axes, red-green and blue-yellow, in almost perfect fit with the hypothesized color circle.

The higher the dimension *n* of the estimated space, the smaller is the resulting minimal stress. Thus, arbitrarily good fit can be achieved by increasing *n*. This means that it is a methodological problem to decide what value of *n* to use in a model. Unless there are strong a priori reasons for assuming that the underlying phenomenal space has a certain number of dimensions, one often looks for a (rather small) number *n* where the computed stress of $n + 1$ dimensions is not significantly smaller than the stress of *n* dimensions.[28]

One problem with MDS is that it can be difficult to give a psychological *interpretation* of the dimensions generated by the algorithm (given a choice of *n*). Austen Clark (1993, 124) formulates the problem as follows:

> The number of dimensions of the MDS space corresponds to the number of independent ways in which stimuli in that modality

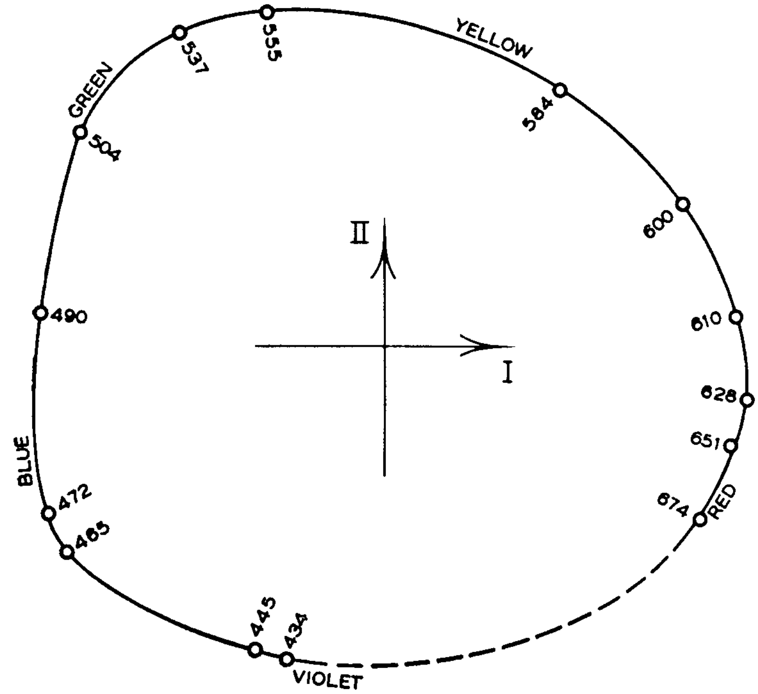

Figure 1.13
A MDS analysis of hues. Stimuli are marked by wavelength in nanometers (from Shepard 1962b, 236).

> can be sensed to resemble or differ, but the dimensions *per se* have no meaning. Indeed, it will be seen that a key step in explaining a quality space is to find interpretable axes. Sometimes one can provide them with a neurophysiological interpretation. Only then can one claim to have determined *what* the differentiative attributes of encodings are, as opposed to knowing simply how many of them there are.

Apart from the neurophysiological interpretations that Clark mentions, investigators often have hypotheses about quality dimensions that could generate the subjects' similarity judgments. So-called "property vector fitting" can then be used to verify the presence of such hypothesized dimensions.[29] This means that, by regression analysis, a stimulus attribute can be correlated with a vector in the space generated by an MDS method. If a high correlation can be found, this indicates that a dimension corresponding to the stimulus attribute is represented in the space. This technique does not guarantee, however, that all the dimensions in a space generated by MDS can be given a meaningful psychological interpretation.

1.8 Integral and Separable Dimensions

A conceptual space is defined here as a set of quality dimensions with a geometrical structure. Now it is time to consider the relations among the dimensions in a conceptual space.

Certain quality dimensions are *integral* in the sense that one cannot assign an object a value on one dimension without giving it a value on the other (Garner 1974, Maddox 1992, Melara 1992). For example, an object cannot be given a hue without also giving it a brightness value. Or a pitch of a sound always goes along with a certain loudness. Dimensions that are not integral are said to be *separable*, as for example the size and hue dimensions.[30] The distinction between integral and separable dimensions will play an important role in the analysis of properties and concepts in chapters 3 and 4. Melara (1992, 274) presents the distinction as follows:

> What is the difference psychologically, then, between interacting [integral] and separable dimensions? In my view, these dimensions differ in their similarity relations. Specifically, interacting and separable dimensions differ in their degree of *cross-dimensional similarity*, a construct defined as the phenomenal similarity of one dimension of experience with another. I propose that interacting dimensions are higher in cross-dimensional similarity than separable dimensions.

Several empirical tests have been proposed to decide whether two perceptual dimensions are separable or integral (see Maddox 1992 for an excellent survey and analysis of these tests). One test is called "speeded classification." The stimuli in this test consists of four combinations of two levels of two dimensions x and y (say size and hue). If the x-levels are x_1 and x_2 (for example, large and small) and the y-values y_1 and y_2 (for example, green and yellow), we can denote the four stimuli (x_1, y_1), (x_1, y_2), (x_2, y_1), and (x_2, y_2) respectively. In the *control* condition, subjects are asked to categorize, as quickly as possible, the level of one dimension, say x, while the other is held constant, by being presented with either (x_1, y_1) or (x_2, y_1) as stimulus (alternatively, (x_1, y_2) or (x_2, y_2)). In the *filtering* condition, the subjects are asked to categorize the level of the same dimension while the other is varied independently. In this condition, the stimulus set thus consists of all four stimuli. Now, if the mean reaction time in the filtering condition is longer than in the control condition, the irrelevant dimension, in this case y, is said to *interfere* with the test dimension x. According to the speeded classification test, x and y are then classified as integral. The underlying assumption is that two separable dimensions can be attended *selectively*, while this is difficult for two integral dimensions; in separable dimensions, the subjects can "filter out" information from the irrelevant dimension.

Another test is the "redundancy task" (Garner 1974). The stimuli and the control condition are the same as in the previous test. In the *redundancy* condition, only two of the four stimuli are utilized, either (x_1, y_1) and (x_2, y_2) or (x_1, y_2) and (x_2, y_1). The values of the two dimension are thus correlated so that the value of one allows the subject to predict the value of the other. The subject is presented with one of the two stimuli and is also here asked to categorize, as quickly as possible, the value of one dimension, say x. If the mean reaction time is shorter than in the control condition, the subjects are said to exhibit a *redundancy gain*. According to the redundancy task, the dimensions are then classified as integral.

A third test, the so-called "direct dissimilarity scaling," concerns the *metric* of the conceptual space that best explains how subjects judge the similarity of different stimuli that vary along the two dimensions (Attneave 1950, Shepard 1964). In this test, the stimuli consist of all combinations of several levels of the two dimensions x and y. The subjects are then presented with all possible pairs of stimuli, one at a time, and are asked to judge the *dissimilarity* of the stimuli on a scale from 1 to 10. This test is an operational way of deciding the distance function in a metric perceptual space.

Using MDS or some other method, the data are fitted into a two-dimensional space. If the Euclidean metric fits the data best, the dimensions are classified as integral; while if the city-block metric gives the best result, they are classified as separable. If two dimensions are separable, the dissimilarity of two stimuli is obtained by adding the dissimilarity along each of the two dimensions, as is done in the city-block metric. In contrast, when the dimensions are integral, the dissimilarity is determined by both dimensions taken together, which motivates a Euclidean metric (compare the above quotation from Melara on cross-dimensional similarity).

Conversely, suppose we are in the constructive mode and attempt to design a conceptual space for solving some cognitive task. If we decide that two dimensions are integral, we should use the Euclidean metric to determine distances and degrees of similarity; while if we decide that the dimensions are separable, the city-block metric should be used instead. Even in scientific theories the relations among dimensions can vary. For example, time and space are treated as separable in

Newtonian mechanics, while the four-dimensional space-time forms an integral set of dimensions in relativity theory (with its own special Minkowski metric).

The notion of a *domain* is central in this book, and it is used in connection with concept formation in chapter 4, with cognitive semantics in chapter 5, and with induction in chapter 6. Using the concepts of this section, I can now define a domain as *a set of integral dimensions that are separable from all other dimensions*. The three-color dimensions are a prime example of a domain in this sense since hue, chromaticness, and brightness are integral dimensions that presumably are separable from other quality dimensions.[31] Another example could be the *tone* domain with the basic dimensions of *pitch* and *loudness*.[32] The most fundamental reason for decomposing a cognitive structure into domains is the assumption that an object can be assigned certain properties *independently* of other properties. An object can be assigned the weight of "one kilo" independently of its temperature or color.

A *conceptual space* can then be defined as a collection of one or more domains. It should be emphasized that not all domains in conceptual spaces are assumed to be metric. Sometimes a domain is just an ordering or a graph with no distance defined. And even if distances are defined for the different domains of a conceptual space, the domains may be "incommensurable" in the sense that there is no common scale to express distances on the entire space.

The domains of a conceptual space should not be seen as totally independent entities, but they are *correlated* in various ways since the properties of the objects modeled in the space covary. For example, ripeness and color domains covary in the space of fruits. These correlations are discussed in connection with the model of concepts in section 4.3 and in connection with induction in section 6.6.[33]

Conceptual spaces will be the focus of my study of representations on the conceptual level. A *point* in a space represents a possible object (see section 4.8). The properties of the objects are determined by its location in the space. As will be argued in chapter 3, properties are represented by *regions* of a domain. As was seen in section 1.5, however, what is important is not the exact form of the representation but rather the *relations* between different areas of a conceptual space.

1.9 On the Origins of Quality Dimensions

In the previous sections I have given several examples of quality dimensions from different kinds of domains. There appears to be different types of dimensions, so a warranted question is: Where do the dimensions come from? I do not believe there is a unique answer to this question. In this section, I will try to trace the origins of different kinds of quality dimensions.

First, some of the quality dimensions appear to be *innate* or developed very early in life. They are to some extent hard-wired in our nervous system, as for example the sensory dimensions presented in section 1.5. This probably also applies to our representations of ordinary space. Since domains of this kind are obviously extremely important for basic activities like getting around in the environment, finding food, and avoiding danger, there is evolutionary justification for the innateness assumption. Humans and other animals that did not have a sufficiently adequate representation of the spatial structure of the external world were disadvantaged by natural selection.

The brains of humans and animals contain topographic areas that map different kinds of sense modalities onto spatial areas (see section 2.5 for more connections to neuroscience). The structuring principles of these mappings are basically innate, even if the fine tuning is established during the development of the human or animal.[34] The same principles appear to govern most of the animal kingdom. Gallistel (1990, 105) argues:

> [T]he intuitive belief that the cognitive maps of "lower" animals are weaker than our own is not well founded. They may be impoverished relative to our own (have less on them) but they are not weaker in their formal characteristics. There is experimental evidence that even insect maps are metric maps.

Quine (1969, 123) notes that something like innate quality dimensions are 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.

The point is that without an initial structure, the world would be just a "blooming, buzzing confusion" (James 1890). We need some dimensions to get learning going.

Once the process has started, however, new dimensions can be added by the learning process.[35] One kind of example comes from studies of children's cognitive development. Smith (1989, 146–47) argues that working out a system of perceptual dimension, a system of *kinds* of similarities, may be one of the major intellectual achievements of early childhood. . . . The basic developmental notion is one of differentiation, from global syncretic classes of perceptual resemblance and magnitude to dimensionally specific kinds of sameness and magnitude.

Two-year-olds can represent whole objects, but they cannot reason about the dimensions of these objects. Goldstone and Barsalou (1998, 252) note:[36]

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

Consequently, learning new concepts is often connected with *expanding* one's conceptual space with new quality dimensions. For example, consider the (phenomenal) dimension of *volume*. The experiments on "conservation" performed by Piaget and his followers indicate that small children have no separate representation of volume; they confuse the volume of a liquid with the *height* of the liquid in its container. It is only at about the age of five years that they learn to represent the two dimensions separately. Similarly, three- and four-year-olds confuse *high* with *tall*, *big* with *bright*, and so forth (Carey 1978).[37]

Along the same lines, Shepp (1983) argues that the developmental shift is from integral dimensions to separable:

> [Y]ounger children have been described as perceiving objects as unitary wholes and failing to attend selectively. This characterization is strikingly similar to the perception and attention of an adult when performing with integral dimensions. In contrast, older children are characterized as perceiving objects according to values on specific dimensions and as succeeding in selective attention. Such a description accurately describes an adult's perception and attention when confronted with separable dimensions. On the basis of these parallels, we have suggested the hypothesis that dimensional combinations that are perceived as separable by the older child and adult are perceived as integral by the young child.

Still other dimensions may be *culturally* dependent.[38] 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 a full metric structure, is needed for advanced forms of planning and coordination with other individuals, but it 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 (Toulmin and Goodfield 1965).

The examples given here indicate that many of the quality dimensions of human conceptual spaces are not directly generated from sensory inputs.[39] 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 geometrical structures of these dimensions, it is quite obvious that there is some such nontrivial structure. This has been argued by Marr and Vaina (1982) and Vaina (1983), who give an analysis of functional representation where functions of an object are determined by the actions it allows. I return to the analysis of actions and functional properties in section 3.10.3.

Culture, in the form of interactions among people, may in itself generate constraints on conceptual spaces. For example, Freyd (1983, 193–194) puts forward the intriguing proposal that conceptual spaces may evolve as a representational form in a community just because people have to *share* knowledge:

> There have been a number of different approaches towards analyzing the structures in semantic domains, but what these approaches have in common is the goal of discovering constraints on knowledge representation. I argue that the structures the different semantic analyses uncover may stem from shareability constraints on knowledge representation. . . . So, if a set of terms can be shown to behave as if they are represented in a three-dimensional space, one inference that is often made is that there is both some psychological reality to the spatial reality (or some formally equivalent formulation) and some innate necessity to it. But it might be that the structural properties of the knowledge domain came about because such structural properties provide for the most efficient sharing of concepts. That is, we cannot be sure that the regularities tell us anything about how the brain can represent things, or even "prefer" to, if it didn't have to share concepts with other brains.

Here Freyd hints at an *economic* explanation of why we have conceptual spaces; they facilitate the sharing of knowledge.[40] Section 5.8 shows that since efficient sharing of knowledge is one of the fundamental requirements of *communication*, Freyd's argument will provide an independent justification for the representational role of conceptual spaces.

Finally, some quality dimensions are introduced by *science*. Witness, for example, Newton's distinction between *weight* and *mass*, which is of pivotal importance for the development of his celestial mechanics but which has hardly any correspondence in human perception. To the extent we have mental representations of the masses of objects in distinction to their weights, these are not given by the senses but have to be learned by adopting the conceptual space of Newtonian mechanics in our representations. The role of new dimensions in science will be further discussed in section 6.4.

1.10 Conclusion

The main purpose of this chapter has been to present the notions of dimensions and domains that constitute the fundamentals of the theory of conceptual spaces. Throughout the book, I apply constructions using conceptual spaces to material from several research areas like semantics, cognitive psychology, philosophy of science, neuroscience, neural networks, and machine learning. I hope that these constructions will establish the viability of the conceptual level of representation.

So what *kind* of theory is the theory of conceptual spaces? Is it an empirical, normative, computational, psychological, neuroscientific, or linguistic theory? The answer is that the theory of conceptual spaces is used in two ways in this book. On a general level, it is a *framework for cognitive representations*. It should be seen as a complement to the symbolic and the connectionist approaches that forms a bridge between these two forms of representation. On a more specific level, the framework of conceptual spaces can then be turned into *empirically testable theories* or *constructive models* by filling in specific dimensions with certain geometrical structures, specific measurement methods, specific connections to other empirical phenomena, and so forth.

Cognitive science has two predominant goals: to *explain* cognitive phenomena and to *construct* artificial systems that can solve various cognitive tasks. My primary aim here is to use conceptual spaces in constructive tasks. In the following chapters, I outline how they can be used in computational models of *concept formation* and *induction* and I also show that they are useful for representing the *meanings* of different kinds of linguistic expressions in a computational approach to semantics. The confidence in the aptness of the theory of conceptual spaces should increase, however, if the theory can also be used to explain various empirical phenomena. Consequently, I also connect the theory to empirical material from psychology, neuroscience, and linguistics, even though I do not attempt to give a complete evaluation of the empirical potency of the theory.

Conceptual spaces are static in the sense that they only describe the *structure* of representations. A full model of cognitive mechanisms not only includes the representational form, but also a description of the *processes* operating on the representations. A particular conceptual space is, in general, compatible with several types of processes, and it must therefore be complemented with a description of the *dynamics* of the representations to generate testable predictions (see, for example, Port and van Gelder 1995, Scott Kelso 1995, van Gelder 1998). This topic is treated in chapter 7.

Finally, a philosophical question: What is the ontological status of conceptual spaces? I view conceptual spaces as *theoretical entities* that can be used to explain and predict various empirical phenomena concerning concept formation (the role of theoretical entities is discussed further in section 6.4). In particular, the distances associated with metric space should be seen as theoretical terms. And if similarity is defined by distances via equation (1.7) or (1.8), similarity will be a theoretical term, too (compare Medin, Goldstone, and Gentner 1993, 255). Since my basic methodological position is *instrumentalistic*, I avoid questions about how real the dimensions of conceptual spaces are but view them as instruments for predictive and constructive purposes (compare, for example, the question of what the equator is).

Some of the neurophysiological correlates of the phenomenal dimensions are presented in section 2.5. By being correlated to empirical phenomena in different ways, the assumptions about the dimensions will have testable empirical consequences. These consequences can then be seen as defining the *content* of the theory (compare Sneed 1971 and Stegmüller 1976). Furthermore, when constructing artificial systems, the dimensions of the conceptual spaces will function as the framework for the architecture of the systems that, for example, will determine the role of the sensors in the system.

In brief, my instrumentalist standing means that I eschew philosophical discussions of how "real" conceptual spaces are. The important thing is that we can *do* things with them. To manifest this is the objective of the rest of the book.