Frameworks for Properties
Possible Worlds *vs.* Conceptual Spaces

Peter Gärdenfors

1. Program

I would like to argue that the traditional kind of model theory used in intensional semantics is not the right kind of framework for its purpose. Instead, I shall put forward some ideas from what is becoming known as *cognitive semantics* and try to show that this kind of framework has a better chance of doing the job.

My discussion will be focussed on the notion of *property*. I first present the standard intensional definition of a property, which is formulated in terms of possible worlds, and outline some of the philosophical problems this definition leads to. Then I introduce the notion of *conceptual space* and show how such spaces can be used as a basis for a new criterion of what a property is. This criterion will be shown to elude the problems of the traditional approach. Furthermore, it will be argued that the criterion is useful for understanding *prototype effects* of properties and the role of properties in *non-monotonic reasoning*.

Even though I concentrate on the notion of property in this paper, I believe that conceptual spaces have a more general bearing for the development of a cognitive semantics capable of handling intensional notions. Here I will only give some general remarks in that direction.

2. The traditional definition of a property and its problems

What *is* a property ? Intuitively, it is something that objects can *have* in common. Often we can *perceive* that an object has a specific property.

These intuitions are not mirrored in the traditional definitions of the notion of property. In the classical *extensional* type of definition, like in Tarski's model theory for first order logic, a property is defined as *a set of objects*, to wit, the set of objects which has the property. Formally, this is done with the aid of a mapping from a language L to a model structure,

representing a world, where each one-place predicate in L is mapped onto a subset of the individuals in the model structure.

However, it was apparent that many so-called intensional properties did not fit this definition. A typical example is 'small' : an emu is a bird, but a small emu is not a small bird ; hence, the property of being small cannot be identified by a set of 'small' objects.

In order to handle the problems concerning intensional properties and other intensional concepts, the classical semantic theories were extended to so-called *intensional semantics*. Pioneers in this development were [Hintikka, 1961], [Kanger, 1957], and [Kripke, 1959], and as an analysis of natural language it reaches its peak with [Montague, 1974]. Here the language L is mapped onto a *set of possible worlds* instead of only a single world. Possible worlds and their associated sets of individuals are the only primitive semantical elements of the model theory. Other semantical notions are defined as *functions* on individuals and possible worlds. For example, a *proposition* is defined as a function from possible worlds to truth values. Such a function thus determines the *set of worlds* where the proposition is true. According to traditional intensional semantics, this is all there is to say about the meaning of a proposition.

In this kind of semantics, a property is something that relates individuals to possible worlds. In general terms, a property can be seen as a many-many relation P between individuals and possible worlds such that iPw holds just when individual i has the property in world w.

In intensional semantics, functions are preferred to many-many relations. There are two ways of turning the relation P into a function. Firstly, it may be described as a *propositional function*, i.e. a function *from individuals to propositions*. Since a proposition is identified with a set of possible worlds, this means that a property is a rule which for each individual determines a corresponding set of possible worlds. But we can also turn the table around to get an equivalent function out of P : for each possible world w, a property will determine a set of individuals which has w as an element of the sets of possible worlds the individuals are assigned (cf. **Figure 1**, on the opposite page). This means that an equivalent definition of property is that it is *a function from possible worlds to sets of individuals*. This alternative definition shows the correspondence between the extensional and the intensional definition of a property.

I now want to show that the standard definition of a property within intensional semantics leads to a number of serious problems. First of all, this definition is highly counterintuitive since properties become very abstract things. Bealer has the following remarks :

> "How implausible that familiar sensible properties are functions — the color of this ink, the aroma of coffee, the shape of your hand, the special painfulness of a burn or itchiness of a mosquito bite. No function is a color, a smell, a shape, or a feeling" [Bealer, 1989, p. 1].

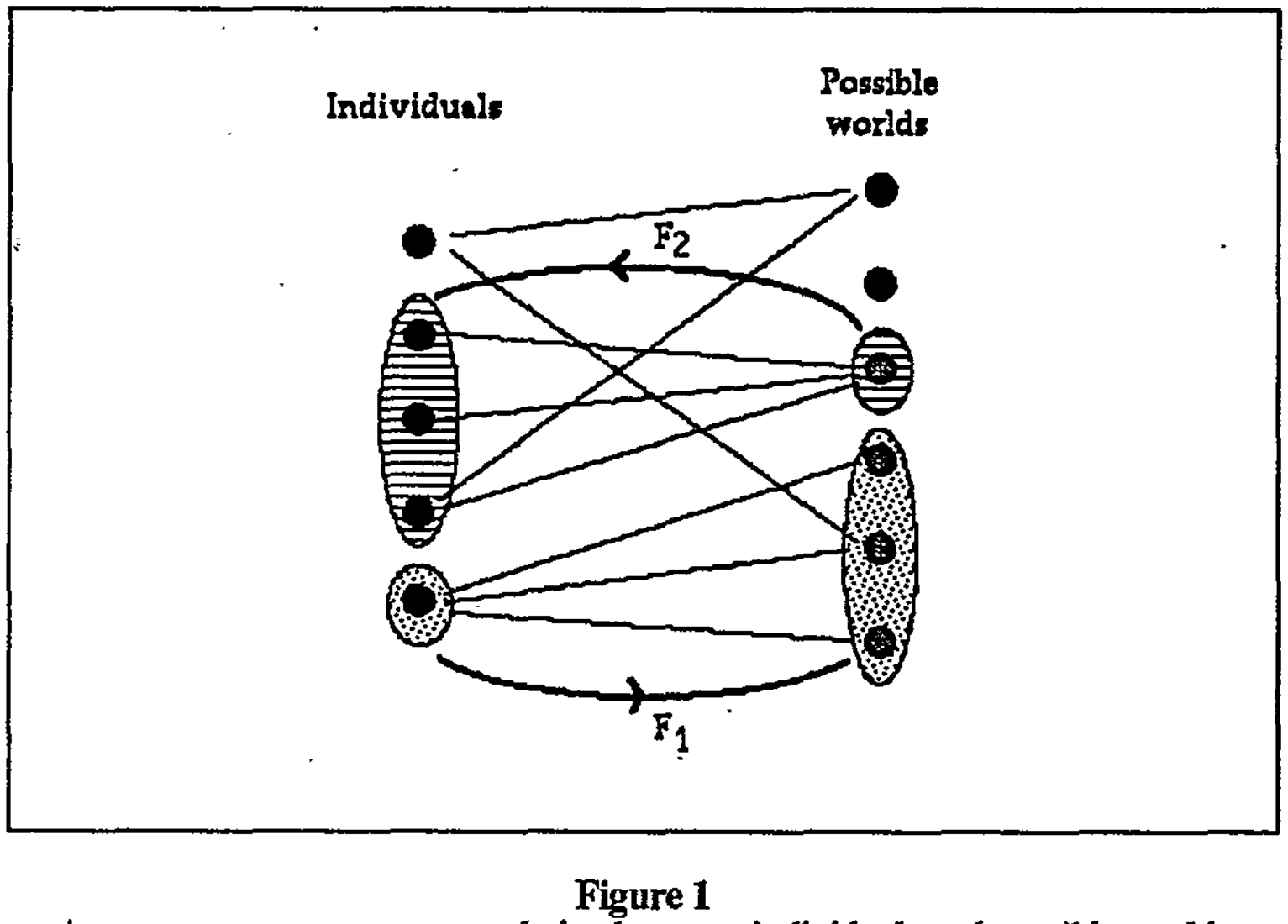

Figure 1
A property as a many-many relation between individuals and possible worlds

F_1 : 'Propositional' function mapping individuals on propositions
F_2 : 'Extensional' function mapping possible worlds on classes

The definition is certainly not helpful for cognitive psychologists who try to explain what happens when a person *perceives* that two objects have the same property in common or, for example, why certain colors look *similar*. However, the fact that the definition is counterintuitive is not a decisive argument ; it may be argued that the abstract character of properties is merely a cosmetic feature of the intensional semantics — as long as semantics produces the right results, the technical form of semantic concepts is not so important.

A related, but more serious, problem for the traditional definition of a property is that it can hardly account for *inductive reasoning*. An inductive inference generally consists in *connecting* two properties to each other, as when one concludes that all living things have chromosomes. This connection is obtained from a number of instances of individuals exhibiting the relevant properties. If a property is defined as a function from possible worlds to sets of individuals (or equivalently as a function from individuals to sets of possible worlds), then in order to determine which properties are instantiated by a particular individual (or a set of individuals), one has to determine *which functions* have the individual (or the set of individuals) as value in the actual world. Apart from problems concerning how we determine which is the actual world, this recipe will in general give us *too many* properties. For example, if we are examining a particular emerald it will instantiate a large class of Goodman-type properties like 'grue' apart from standard properties like 'green'.

If the only thing we know about properties is that they are some kind of abstract functions, then we have no way of distinguishing natural and inductively projectible properties like 'green' from inductively useless properties like 'grue'. What is needed is a criterion for separating the projectible sheep from the non-projectible goats. However, classical intensional semantics does not provide us with such a criterion[1].

[1]*For further criticism of the classical approach to properties in relation to inductive reasoning, cf. [Gärdenfors, 1990a].*

A third problem that arises from defining properties as abstract functions concerns the difficulty of expressing an *anti-essentialistic* doctrine as has been pointed out by [Stalnaker, 1981]. Anti-essentialism is the doctrine that things have none of their properties necessarily. In that paper, Stalnaker's aim is not to defend the doctrine, but to show that on the traditional account of properties it cannot be coherently expressed.

Some kinds of essential properties are unobjectionable to an anti-essentialist : any property that is necessarily an essential property of everything, like 'being self-identical' or 'being either human or non-human' ; certain relational properties that are defined in terms of specific individuals may be essential to that individual, like 'being the same age as Ingmar Bergman' which is essential to Ingmar Bergman ; and "world-indexed" properties, like 'being an emu in the actual world' which an individual has essentially if and only if it actually is an emu. The problem for an anti-essentialist is how to find a criterion which can distinguish between such innocent essential properties and the ontologically dangerous kinds of essential properties. Stalnaker formulates the problem as follows :

> "In terms of this extensional account of properties (extensional in the sense that properties are defined by their extensions in different possible worlds), what corresponds to the intuitive distinctions between referential and purely qualitative properties, and between world-indexed and world-independent properties? Nothing. All properties are referential in the sense that they are defined in terms of the specific individuals that have them. All properties are world-indexed in the sense that they are defined in terms of the specific possible worlds in which things have them. While one can, of course, make a distinction between essential and accidental attributes in terms of the standard semantical framework, one cannot find any independent distinctions corresponding to the intuitive ones needed to state a coherent version of the anti-essentialist thesis. Thus there is no satisfactory way, without adding to the primitive basis of the semantical theory, to state the thesis as a further semantical constraint on legitimate interpretations of the language of modal logic" [Stalkaner, 1981, p. 346].

Stalnaker concludes that "[w]hat the standard semantics lacks is an account of properties that defines them independently of possible worlds and individuals. (...) [A] property must be not just a rule for grouping individuals, but a feature of individuals in virtue of which they may be grouped (...)" [Stalnaker, 1981, p. 347].

The final problem that I shall point out for the functional definition of properties is perhaps the most serious one. [Putnam, 1981] has shown that the standard model-theoretic definition of "property" which has been

given here does not work as a theory of the *meaning* of properties. In proving this result, Putnam makes two assumptions about "the received view" of meaning : (1) The meaning of a sentence is a function which assigns a truth value to the sentence in each possible world ; and (2) the meaning of the parts of a sentence cannot be changed without changing the meaning of the whole sentence.

Putnam's general proof is quite technical, but the thrust of the construction can be illustrated by his example [Putnam, 1981, p. 33-35]. He begins with the sentence :

(1) A cat is on a mat.

where "cat" refers to *cats* and "mat" to *mats* as usual. He then shows how to give (1) a new interpretation.

(2) A cat* is on a mat*.

The definitions of the properties cat* and mat* make use of three cases :

(a) Some cat is on some mat and some cherry is on some tree.

(b) Some cat is on some mat and no cherry is on any tree.

(c) Neither (a) nor (b) holds.

Here are Putnam's definitions :

• DEFINITION OF 'CAT*'

x is a cat* if and only if case (a) holds and x is a cherry or case (b) holds and x is a cat ; or case (c) holds and x is a cherry.

• DEFINITION OF 'MAT*'

x is a mat* if and only if case (a) holds and x is a tree or case (b) holds and x is a mat ; or case (c) holds and x is a quark.

Given these definitions it turns out that the sentence (1) is true in exactly those possible worlds where (2) is true. Thus, according to the received view of meaning, these sentences will have the same meaning. In the appendix to his book, Putnam shows that a more complicated reinterpretation of this kind can be constructed for all the sentences of a language. He concludes that "there are always inifinitely many different interpretations of the predicates of a language which assign the «correct» truth-values to the sentences in all possible worlds, *no matter how these «correct» truth-values are singled out*". Thus "(...) *truth-conditions for whole sentences* underdetermine reference" [Putnam, 1981, p. 35]. Again, the underlying reason is that there are *too many* potential properties if they are defined as functions from individuals to propositions, i.e. in terms of

possible worlds and truth values. Cat* and mat* are just two examples from this large class.

Putnam's own diagnosis of the problem is that it occurs because of viewing a language as separate from its interpretation[2]:

> "The predicament only *is* a predicament because we did two things : first, we gave an account of understanding the language in terms of programs and procedures for *using* the language (what else ?) ; then, secondly, we asked what the possible «models» for the language were, thinking of the models as existing «out there» *independent of any description*. At this point, something really weird had already happened, had we stopped to notice. On any view, the understanding of the language must determine the reference of the terms, or, rather, must determine the reference given the context of use. If the use, even in a fixed context does not determine reference, then use is not understanding. The language, on the perspective we talked ourselves into, has a full program of use ; but it still lacks an *interpretation*.
>
> This is the fatal step. To adopt a theory of meaning according to which a language whose whole use is specified still lacks something — viz. its «interpretation» — is to accept a problem which *can* only have crazy solutions. To speak as if *this* were my problem, «I know how to use my language, but, now, how shall I single out an interpretation ?» is to speak nonsense. Either the use *already* fixes the «interpretation» or nothing can.
>
> Nor do «causal theories of reference», etc., help. Basically, trying to get out of this predicament by *these* means is hoping that the *world* will pick one definite extension for each of our terms even if *we* cannot. But the world does not pick models or interpret languages. *We* interpret our languages or nothing does" [Putnam 1980, p. 481-482].

[2] *For further discussion of Putnam's theorem and its relevance to semantics, cf. [Lakoff, 1987, ch. 15].*

I have here presented four different arguments against the traditional definition of the notion of a property. The upshot is that there is something rotten in the kingdom of semantics. What is needed is a completely different way of defining properties. I shall argue that a cognitively oriented approach will do the work[3].

[3] *Prime examples of works in the tradition of cognitive semantics are [Lakoff, 1987] and [Langacker, 1986]. Related versions can be found in the writings of [Jackendoff, 1983, 1990], [Johnson-Laird, 1983], [Fauconnier, 1985], [Talmy, 1988], [Sweetser, 1990] and many others. There is also a French linguistic and semiotic tradition, exemplified by [Desclés, 1985] and [Petitot, 1985, 1992], which shares many features with the American (mainly Californian) group.*

3. Conceptual spaces as a basis for a new criterion of properties

On my view, the semantics for a language is primarily *a relation between the language and a cognitive structure*. The *meaning* of an expression is determined by what it corresponds to in such a cognitive structure. The external world enters the picture only when the relation between it and the conceptual structure is considered. This means that the *truth* of sentences is, at best, a secondary feature of a semantic theory. Questions of meaning must be answered *before* we can raise any questions about truth.

As a framework for a cognitive structure used in describing a semantics I want to put forward the notion of a *conceptual space*. 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[4]. The dimensions are taken to be

cognitive and infra-linguistic in the sense that we (and other animals) can represent the qualities of objects, for example when planning an action, without presuming an internal language in which these qualities are expressed. 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[5].

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[6]. 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[7].

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.). 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 a semantics for a natural language, 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.* 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.

[4]The notion of a quality dimension is closely related to the 'domains' in Langacker's semantic theory [Langacker, 1986].

[5]Some further examples will be given in the following section.

[6]To some extent the representation of time is culturally dependent, so that other cultures have a different time dimension as a part of their cognitive structure. Cf. [Gärdenfors, 1990b] for a discussion of how this influences the structure of language.

[7]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, 1990b].

The second psychological dimension 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 2**)[8].

[8]For further examples of perceptual quality dimensions, cf [Gärdenfors, 1990b].

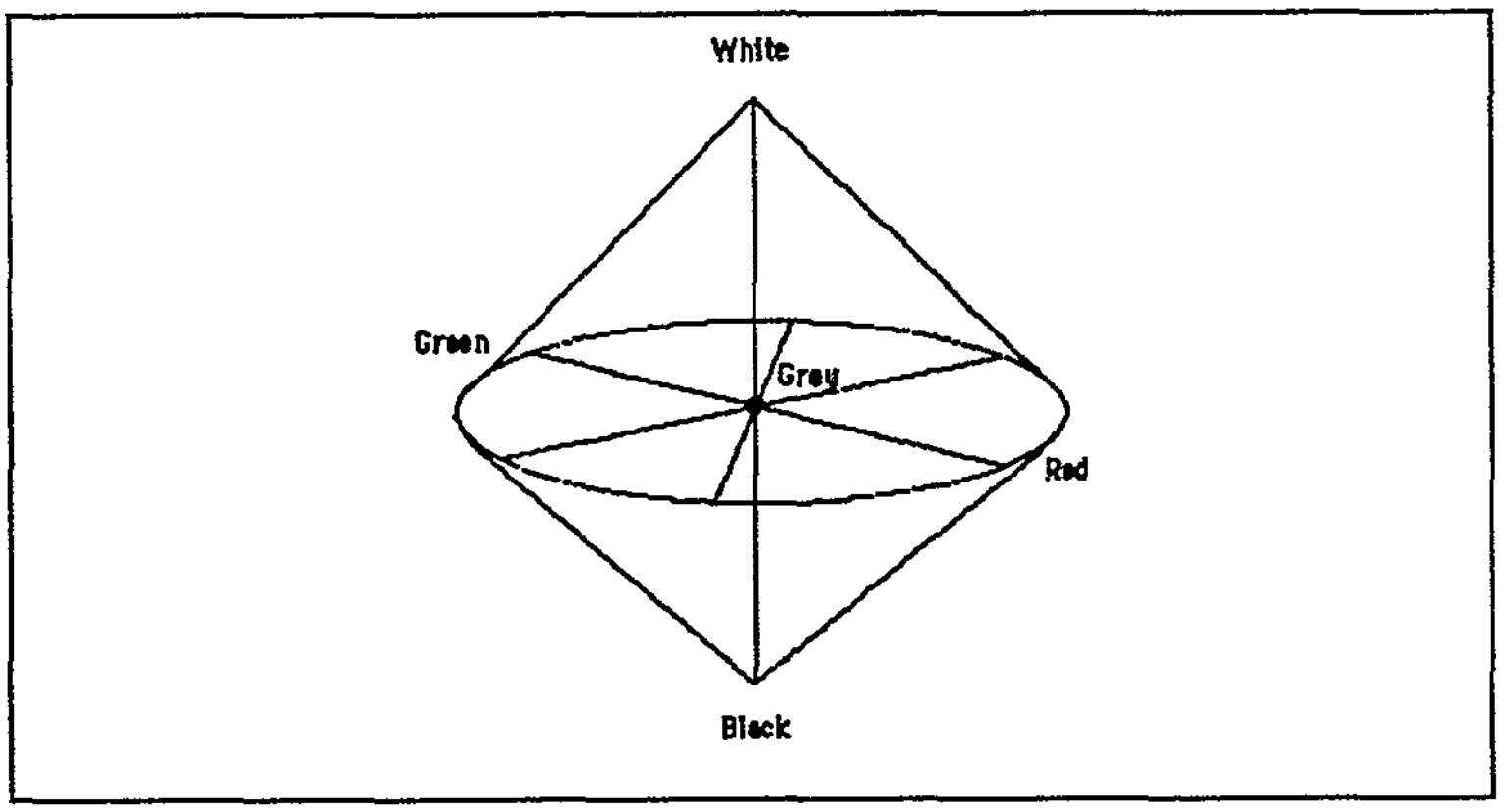

Figure 2
The full color space

I cannot provide a complete list of the quality dimensions involved in our conceptual spaces. Some of the dimensions seem to be *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. Other dimensions are presumably *learned*. Learning new concepts often involves expanding one's conceptual space with new quality dimensions. *Functional* properties used for describing artifacts may be an example here. Such properties are characterized by a set of associated *actions*. Even if we do not know very much about the conceptual dimensions underlying actions, it is quite obvious that there is such a non-trivial structure[9]. Still other dimensions may be *culturally* dependent. 'Time' is a good example — in contrast to our linear conception of time, some cultures conceive of time as circular so that the world keeps returning to the same point in time, and in other cultures it is hardly meaningful at all to speak of time as a dimension[10]. Finally, some quality dimensions are introduced by *science*. As an example, let me mention the distinction between temperature and heat, which is central for thermodynamics, but which has no correspondence in human perception.

[9]See e. g. Vaina's analysis of functional representation [Vaina, 1983].

[10]Another example of the cultural relativity of time dimension is discussed in [Gärdenfors, 1990b, section 6].

(Human perception of heat is basically determined by the amount of heat transferred from an object to the skin rather than by the temperature of the object.)

This concludes my general presentation of conceptual spaces. It can be seen as a generalization of the *state space approach*, advocated among others by [Churchland (P. M.), 1986] and [Churchland (P. S.), 1986], and of the *vector function theories* of [Foss, 1988]. To some extent conceptual spaces also function like the *domains* in Langacker's semantic theory [Langacker, 1987]. The theory of conceptual spaces is a *theory for representing information*, not a psychological or neurological theory, which I believe can be applied to a number of philosophical problems in epistemology and semantics. Here, my primary aim is to show its viability as a foundation for intensional semantics.

4. Properties described with the aid of conceptual spaces

In more abstract terms, a conceptual space S consists of a class $D_1, ..., D_n$ of quality dimensions. A point in S is represented by a vector $v = <d_1,..., d_n>$ with one index for each dimension. Each of the dimensions is endowed with a certain topological or metrical structure.

A first rough idea is to describe a property as a *region* of a conceptual space S, where "region" should be understood as a spatial notion determined by the topology and metric of S. For example, the point in the time dimension representing 'now' divides this dimension, and thus the space of vectors, into two regions corresponding to the properties 'past' and 'future'. In contrast to the traditional definition in intensional semantics that was presented in Section 2, this definition presumes *neither* the concept of an individual *nor* the concept of a possible world. But the proposal suffers from a lack of precision as regards the notion of a "region". A more precise and powerful idea is the following criterion where the topological characteristics of the quality dimensions are utilized to introduce a spatial structure on properties :

Criterion P : A *natural property* is a convex region of a conceptual space.

A *convex* region is characterized by the criterion that for a very pair of points v_1 and v_2 in the region all points in between v_1 and v_2 are also in the region. The motivation for the criterion is that if some objects which are located at v_1 and v_2 in relation to some quality dimension (or several dimensions) both are examples of the property P, then any object that is located between v_1 and v_2 on the quality dimension(s) will also be an example of P. I shall argue later that this criterion is psychologically realistic. Criterion P presumes that the notion of *betweenness* is meaningful for the relevant quality dimensions. This is, however, a rather

weak assumption which demands very little of the underlying topological structure[11].

[11] *See [Birkhoff, 1967] for an axiomatic analysis of "between".*

Most properties expressed by simple words in natural languages are natural properties in the sense specified here. For instance, I conjecture that all *color terms* in natural languages express natural properties with respect to the psychological representation of the three color dimensions. 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 color C, then any object o_3 with a color that lies between the color of o_1 and that 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 seems to be done in terms of convex sets. Strong support for this conjecture can be found in [Berlin & Kay, 1969], although they do not treat color terms in general but concentrate on basic color terms. On the other hand, the reference of an artificial color term like 'grue' will not be a convex region in the ordinary conceptual space and thus it is not a natural property according to Criterion *P*[12].

[12] *For an extended analysis of this example, see [Gärdenfors, 1990a].*

Another illustration of how the convexity of regions determines properties and categorizations is the phonetic identification of *vowels* in various languages. According to phonetic theory, what determines a vowel are the relations between the basic frequency of the sound and its formants (higher frequencies that are present at the same time). In general, the first two formants F_1 and F_2 are sufficient to identify a vowel. This means that the coordinates of two-dimensional space spanned by F_1 and F_2 (in relation to a fixed basic pitch F_0) can be used as a fairly accurate description of a vowel. [Fairbanks & Grubb, 1961] investigated how people produce and recognize vowels in "General American" speech. **Figure 3** (see on the opposite page) summarizes some of their findings. The scales of the abscissa and ordinate are the logarithm of the frequencies of F_1 and F_2 (the basic frequency of the vowels was 130 cps). As can be seen from the diagram, the preferred, identified and self-approved examples of different vowels form convex subregions of the space determined by F_1 and F_2 with the given scales[13]. As in the case of color terms, different languages carve up the phonetic space in different ways (the number of vowels identified in different languages varies considerably), but I conjecture again that each vowel in a language will correspond to a convex region of the formant space.

[13] *A self-approved vowel is one that was produced by the speaker and later approved of as an example of the intended kind. An identified sample of a vowel is one that was correctly identified by 75% of the observers. The preferred samples of a vowel are those which are "the most representative samples from among the most readily identified samples" [Fairbanks & Grubb, 1961, p. 210].*

Criterion *P* provides an account of properties that is independent of both possible worlds and individuals and it satisfies Stalnaker's *desideratum* that a property "(...) must be not just a rule for grouping individuals, but a feature of individuals in virtue of which they may be grouped" [Stalnaker, 1981, p. 347]. However it should be emphasized that I only view the criterion as a *necessary* but perhaps not sufficient

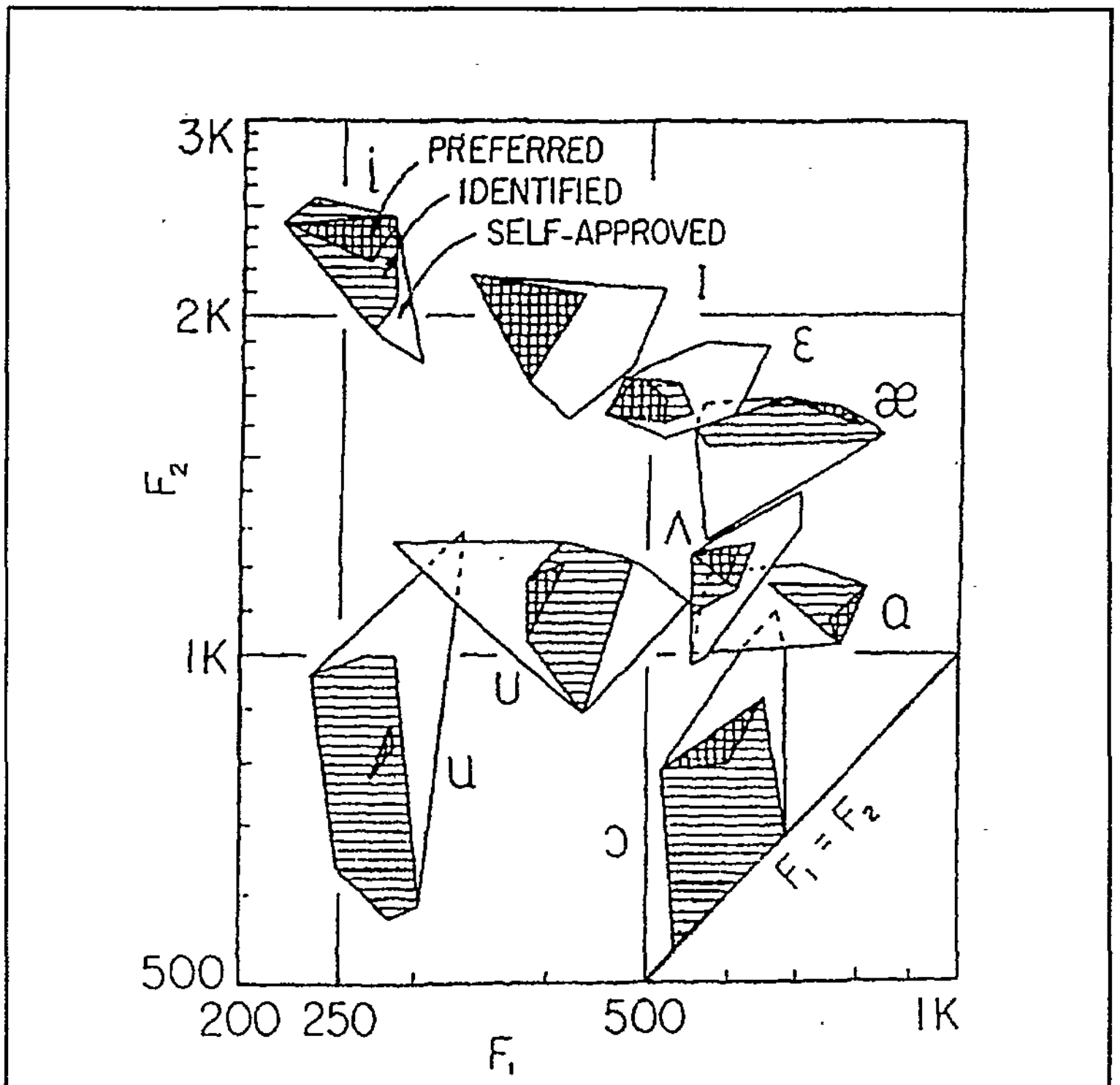

Figure 3
Frequency areas of different vowels in the two-dimensional space generated by the first two formants *(values in cps)*
(From [Fairbanks & Grubb, 1961])

condition on a natural property. The criterion delimits the class of properties that are useful for cognitive purposes, but it may not be sufficiently restrictive.

5. Basic elements of an intensional semantics

Introducing the notion of a natural property is only the first step in developing an intensional semantics based on conceptual spaces. According to my view, semantics is a relation between language and a conceptual space. An *interpretation* for a language *L* consists of a mapping of the components of *L* onto a conceptual space. As a first element of such a mapping, *individual constants* are assigned vectors (i.e. points in the conceptual space) or partial vectors (i.e. points with some

arguments undetermined). In this way each name (referring to an individual) is allocated a specific color, spatial position, weight, temperature, etc[14]. If a constant is assigned to a partial vector, this means that not all the properties are known or have been determined. Following [Stalnaker, 1981, p. 347], a function which maps the individuals into a conceptual space will be called a *location function*.

[14]Abstract entities may be assigned values on a different set of quality dimensions.

As a second element of the interpretation mapping, the *predicates* of the language that denote primary properties are assigned convex regions in the conceptual space in accordance with Criterion *P*. Such a predicate is *satisfied* by an individual just in case the location function locates the individual at one of the points included in the region assigned to a predicate. Some of the so called intensional predicates, like 'tall', 'former' or 'alleged', do not denote primary properties in the sense that their regions can be described independently of other properties. Such secondary predicates, which are "parasitical" on other properties, can be described in terms of the regions assigned to the primary properties[15]. *Relations* (primary and secondary) can be treated in a similar way[16].

[15]See [Gärdenfors, 1988, section 4] for an analysis of the semantics of 'secondary' properties.

[16]Again see [Gärdenfors, 1988, Section 4] and [Holmqvist, 1988] for some of the details.

If we assume that an individual is completely determined by its set of properties, then all points in the conceptual space can be taken to represent *possible individuals*. On this account, a possible individual is a *cognitive* notion that need not have any form of reference in the external world. This construction will avoid many of the problems that have plagued other philosophical accounts of possible individuals. A point in a conceptual space will always have an internally consistent set of properties — since e.g. 'blue' and 'yellow' are disjoint properties in the color space, it is not possible that any individual will be both blue and yellow (all over). There is *no need for meaning postulates* or their ilk in order to exclude such contradictory properties.

What has been accomplished here are only the first steps in the construction of an intensional semantics based on conceptual spaces. One important contrast to the traditional intensional semantics is that the new one does not presume the concept of a *possible world*. However, different location functions describe alternative ways that individuals may be located in a conceptual space. Thus these location functions have the same role as possible world in the traditional semantics. This means that we can *define* the notion of a possible world as a possible location function and this can be done without introducing any new semantical primitives to the theory (cf. [Stalnaker, 1981, p. 348]). In this way most of the constructions from traditional intensional semantics will be available, should we want them.

If we assume that the meanings of the predicates, among other things in a language *L*, 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 (independent of empirical considerations). 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 red and green. Analytic-in-*S* is thus defined on the basis of the topological and metric structure of the conceptual space *S*. A consequence of this definition is that an analytic statement will be satisfied for all location functions. However, different conceptual spaces will yield different notions of analyticity.

6. How the problems for the traditional definition of "property" are avoided

I now want to argue that the criterion of "natural property" given above eludes all the problems of Section 2 that tainted the traditional definition. First of all, Criterion *P* makes many properties *perceptually grounded*. Since the basic quality dimensions are more or less determined by our perceptual mechanisms, there is a direct link between properties described as regions of such dimensions and perceptions. In other words, many of the elements of the location functions will be determined by perceptions. This means that the present criterion of properties will be much more useful for cognitive psychologists than the traditional definition. In the following section it will be shown that there are close connections between the idea of describing properties in terms of conceptual spaces and the *prototype theory* of categorization.

Secondly, describing properties in terms of conceptual spaces makes it much easier to understand how inductive inferences are made. The fundamental criterion is that *in inductive inferences we only allow predicates denoting natural properties* (as described here). This criterion gives us a tool for separating projectible predicates from non-projectible ones. In this way we cut down, in a non-arbitrary way, the immense class of properties (among them the Goodman-type properties) that will be available if properties are defined as functions from possible worlds to individuals.

In [Gärdenfors, 1990a] it is argued that the traditional problems in the analysis of induction have arisen because philosophers (and AI researchers) have confined themselves to linguistic representations of knowledge[17]. However, if we want to understand human inductive reasoning, we must go deeper than language, i.e. down to conceptual spaces. The same applies if we strive to construct computer programs that, even in a limited way, mirror the human inductive capacity. Imposing a *spatial structure* on properties, as is done here, will open up for new kinds of programming techniques. Traditional approaches to mechanized induction basically use formula manipulations. On the other hand, an

[17]The paper contains among other things an analysis of Hempel's and Goodman's paradoxes of induction.

implementation of the theory of conceptual spaces would use more directly "mathematical" operations, like vector calculations.

Third, the problems for the anti-essentialistic doctrine raised by Stalnaker now disappear. Much of the inspiration for the theory of conceptual spaces comes from what [Stalnaker, 1981] calls *logical space*. The main difference between his notion and mine, is that he does not emphasize the role of the topological and metric structure of the quality dimensions and he can thus not talk about "convexity" and similar notions. Here is Stalnaker's solution to the problem of expressing the anti-essentialistic doctrine :

> "It should be clear that every *property* (every region of logical space) determines a propositional function in the following way : given any property, the value of the corresponding propositional function, relative to a given possible world, will be the class of individuals that have the property in that possible world — the individuals that are located in that region of logical space by the location function that represents the possible world. Thus every property determines a unique propositional function and the correspondence is one-one : distinct properties never determine the same propositional function. But it is not the case that every propositional function corresponds to an intrinsic property, for the classes of individuals selected by a propositional function in the different possible worlds need not all come from the same region of logical space. Among the propositional functions, or properties in the broad sense, that do not correspond to regions of logical space are, of course, just those that the anti-essentialist wants to distinguish from full-fledged intrinsic properties. For example, referential properties such as *being the same weight as Babe Ruth* will clearly not correspond to regions of logical space" [Stalnaker, 1981, p. 348-349].

Thus the possible worlds semantics *generated* from conceptual space semantics is rich enough to represent the distinctions needed to make sense of the appropriate kind of anti-essentialism.

Fourth and finally, the problems that Putnam's theorem causes the traditional definition of "property" dissolve into thin air under the new criterion. 'Cat' denotes a region of a conceptual space (relating to the class of possible animals, be they determined in terms of shape, biological functions, or whatever. In the next section a classification of animals in terms of shapes will be outlined). This region would, at least partly, be determined by the perceptual features of cats. We cannot create a new natural property 'cat*' by relating it to what facts are true in various possible worlds : 'cat*' as introduced by Putnam is indeed a propositional function, but definitely not a natural property and thus not an eligible candidate for an interpretation function that maps a language into a conceptual space.

7. Two further applications : Prototype theory and non-monotonic reasoning

Having argued that describing properties in terms of conceptual spaces solves all the traditional problems, I want to show in this section that such

a criterion can throw new light on other areas of research where the notion of a property is central.

As a first application I want to show that describing properties as convex regions of conceptual spaces fits very well with the so called *prototype theory* of categorization developed by Rosch and her collaborators ([Rosch, 1975, 1978] , [Mervis & Rosch, 1981], [Lakoff, 1987]). 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 either the extensional or the intensional version of 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 (relative to a given possible world) 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 neither predicts nor explains such asymmetries, something else must be going on.

In contrast, if properties are described 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. In a series of experiments, Rosch has been able to demonstrate the psychological reality of such "focal" colors. For another illustration we can return to the categorization of vowels presented in Section 4. Here the structure of the subjects' different kinds of responses show clear prototype effects.

For more complex categories like 'bird' it is perhaps more difficult to describe the underlying conceptual space. However, if something like Marr and Nishihara's analysis of shapes is adopted [Marr & Nishihara, 1978], we can begin to see how such a space would appear[18]. Their scheme for describing biological forms uses hierarchies of cylinder-like modelling primitives. Each cylinder is described by two coordinates (length and width). Cylinders are combined by determining the angle between the dominating cylinder and the added one (two polar

[18]This analysis is expanded in [Marr, 1982, ch. 5]. A related model, together with some psychological grounding, is presented by [Biederman, 1987].

coordinates) and the position of the added cylinder in relation to the dominating one (two coordinates). The details of the representation are not important in the present context, but it is worth noting that on each level of the hierarchy an object is described by a comparatively small number of coordinates based on lengths and angles. Thus the object can be identified as a hierarchically structured vector in a (higher order) conceptual space. **Figure 4** provides an illustration of the hierarchical structure of their representations.

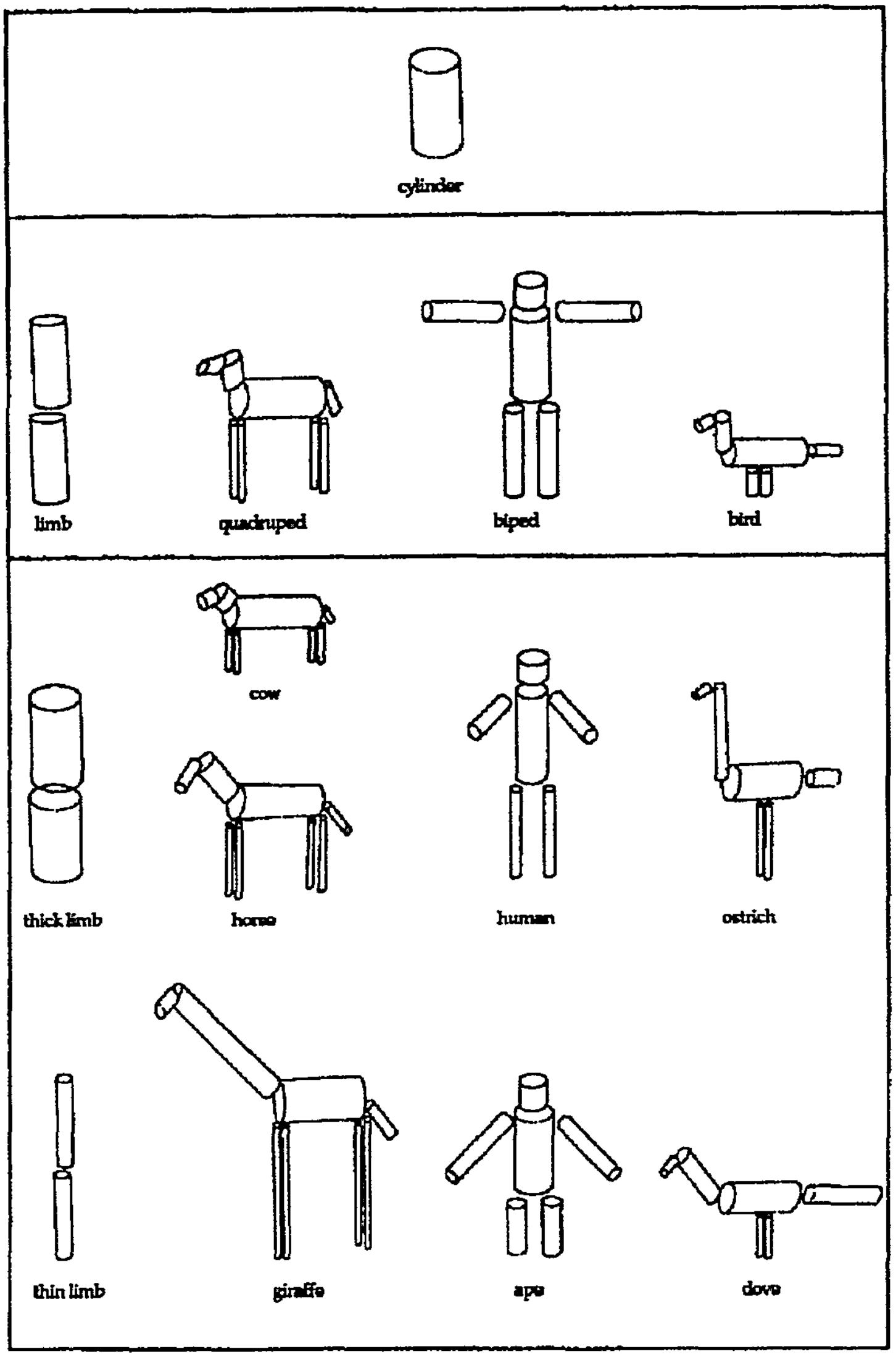

Figure 4
Hierarchical representation of animal shapes
using cylinders as modelling primitives
(from [Marr, 1982])

It should be noted that even if different members of a category are judged to be more or less prototypical, it does not follow that some of the existing objects must represent "the prototype"[19]. If a category is viewed as a convex region of a conceptual space this is easily explained, since the central member of the region (if unique) is a possible individual in the sense discussed above (if all its properties are specified) but need not be among the existing members of the category. Such a prototype point in the region need not be completely described as an individual, but is normally represented as a partial vector, where only the values of the dimensions that are relevant to the category have been determined. For example, the general shape of the prototypical bird would be included in the vector, but its color or age would presumably not.

[19]Rosch has been misunderstood on this point. Cf. [Lakoff, 1987, ch. 2].

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*, a two-dimensional example of which is illustrated in **Figure 5.**

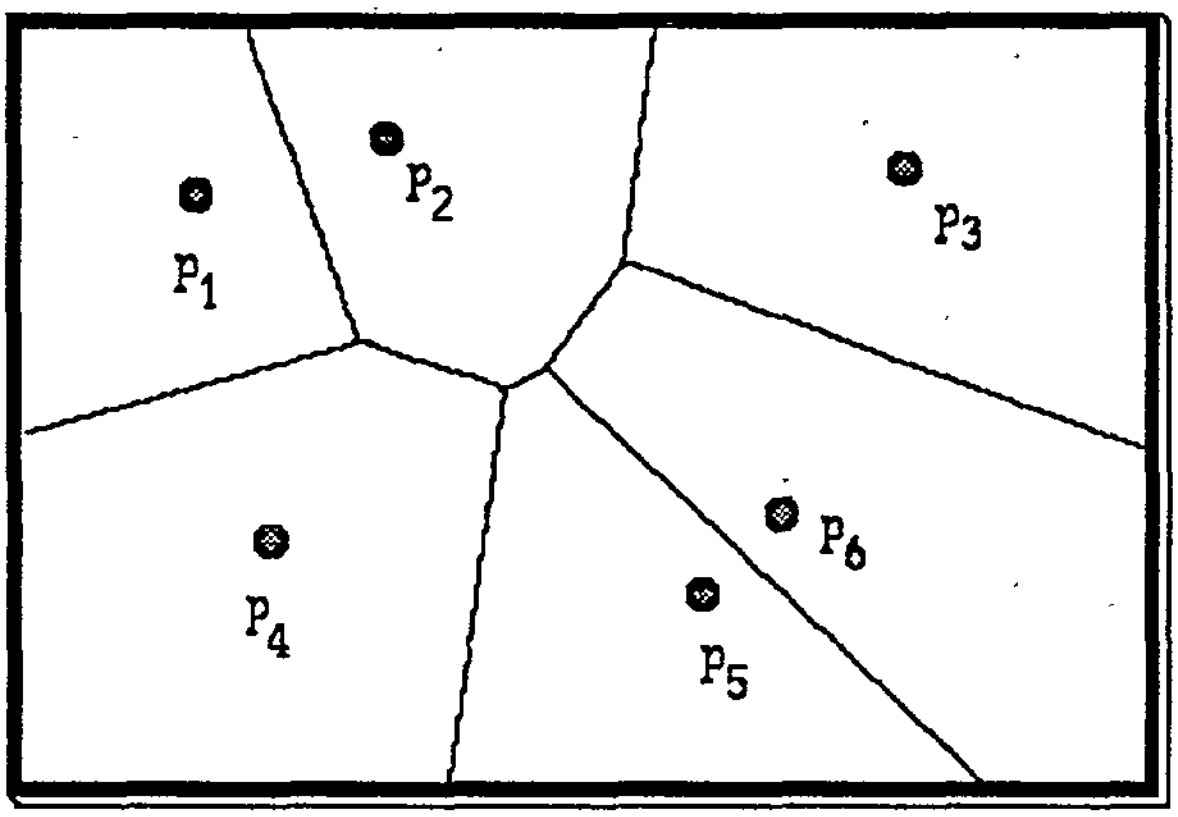

Figure 5
Voronoi tessellation of the plane into convex sets

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 described as convex regions in a conceptual space.

A second application of the proposed criterion of properties concerns *non-monotonic reasoning*. The deductive reasoning of traditional logic is *monotonic* in the sense that when a proposition A can be inferred from a set S of sentences, then A can be inferred from any set that contains S. However, everyday reasoning, which is in general based on assumptions about what is "normally" the case, is often non-monotonic. For example, if I learn that Gonzo is a bird, then with the aid of the presumption that birds normally fly, I may conclude that Gonzo flies. But if my knowledge is expanded by the new information that Gonzo is an emu, this conclusion is no longer drawn.

Non-monotonic reasoning is one of the hottest topics within AI[20]. However, most of the research efforts have been concentrated on finding the appropriate *logical rules* that govern non-monotonic reasoning. This means that a *propositional representation* of the relevant knowledge is already presumed. I believe (for essentially the same reasons as for the case of inductive reasoning) that in order to understand non-monotonic reasoning one must go beyond linguistic formulations of information.

Here the representation of properties as convex regions in a conceptual space may be useful. If the first thing I ever hear about the individual Gonzo is that it is a bird, I will naturally locate it in the conceptual space as a more or less prototypical bird, i.e. at the center of the region representing birds[21]. And in that area of the conceptual space, birds do fly, i.e. almost all individuals located there also have the ability to fly. However, if I then learn that Gonzo is an emu, I must *revise* my earlier location function and put Gonzo in the emu region, which is a subset of the bird region but presumably lies at the outskirts of that region [see note 6]. And in the emu region of the conceptual space almost all individuals do not fly.

This simple example only hints at how the *correlation* between different parts of a region representing a property and regions representing other properties can be used in understanding non-monotonic reasoning. For this analysis the spatial structure of properties is essential. Such correlations will only be formulated in an *ad hoc* manner if a propositional representation of information is used, where the spatial structure cannot be utilized. But I am convinced that developing the program sketched here will show how the idea of representing properties and other categories in terms of conceptual spaces can explain various phenomena of non-monotonic reasoning[22].

[20] See [Reinfrank et al., 1988] for a survey of current research areas.

[21] The relevant conceptual space may be something like a 17-dimensional hierarchical space of coordinates in Marr and Nishihara 's style [Marr & Nishihara, 1978].

[22] See [Holmqvist, 1989] for some further examples and ideas in this direction.

8. Conclusion

This article contains one critical and one more constructive part. In the critical part, I examined the definition of "property" that is standard within traditional intensional semantics and argued that this definition leads to serious philosophical problems, let alone problems in other areas of cognitive science.

As a remedy, I proposed, in the constructive part of the paper, that conceptual spaces be used as a framework for representing information. I outlined the first steps of an alternative intensional semantics based on conceptual spaces and, in particular, I suggested that the notion of "natural property" be described as a convex region in a conceptual space. It was then shown that this criterion avoids the problems that tainted the traditional definition. Furthermore, I argued that the new criterion is useful for understanding some other areas involving categorization, namely prototype theory and non-monotonic reasoning.

Department of Philosophy
Lund University

Acknowledgements
Research for this paper has been supported by the Swedish Council for Research in the Humanities and Social Sciences, by Erik Philip-Sörensens Foundation, and by Magnus Bergvall's Foundation. I wish to thank Johan van Benthem, Jaakko Hintikka, Ernest Sosa, Lucia Vaina, and the participants at the *Cognitive Science seminar* in Lund for helpful comments on the manuscript.

References

BEALER (G.)
1989, "On the Identification of Properties and Propositional Functions", *Linguistics and Philosophy*, 12, p. 1-14.

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

BIEDERMAN (I.)
1987, "Recognition-by-components : A Theory of Human Image Understanding", *Psychological Review*, 94, p. 115-147.

BIRKHOFF (G.)
1967, *Lattice Theory*, Providence, American Mathematical Society.

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

CHURCHLAND (P. M.)
1986, "Some Reductive Strategies in Cognitive Neurobiology", *Mind*, 95, n°379, p. 279-309.

DESCLÉS (J. -P.)
1985, "Représentation des connaissances", *Actes Sémiotiques-Documents*, VII, 69-70, Paris, Institut National de la Langue Française-CNRS.

FAIRBANKS (G.) & GRUBB (P.)
1961, "A Psychophysical Investigation of Vowel Formants", *Journal of Speech and Hearing Research*, 4, p. 203-219.

FAUCONNIER (G.)
1985, *Mental Spaces*, Cambridge (MA), MIT Press.

FOSS (J.)
1988, "The Percept and Vector Function Theories of the Brain", *Philosophy of Science*, 55, p. 511-537.

GÄRDENFORS (P.)
1988, "Semantics, Conceptual Spaces and the Dimensions of Music", p. 9-27, *in Essays on the Philosophy of Music*, V. Rantala, L. Rowell, and E. Tarasti eds., (*Acta Philosophica Fennica*, vol. 43), Helsinki.

1990a, "Induction, Conceptual Spaces and AI", *Philosophy of Science*, n°57, p. 78-95.

1990b, "Mental Representation, Conceptual Spaces and Metaphors", forthcoming in *Synthese*.

HINTIKKA (J.)
1961, "Modality and Quantification", *Theoria*, 27, p. 110-128.

HOLMQVIST (K.)
1988, *Aspects of Parameterizing Concept Representations, manuscript*, Department of Philosophy, Lund University.

1989, *«Non-monotonic» Reasoning from an Image Schema Semantics Point of View, manuscript,* Department of Cognitive Science, Lund University.

JACKENDOFF (R.)
1983, *Semantics and Cognition,* Cambridge (MA), MIT Press.

1990, *Semantic Structures,* Cambridge (MA), MIT Press.

JOHNSON-LAIRD (P.)
1983, *Mental Models,* Cambridge University Press.

KANGER (S.)
1957, *Provability in Logic,* Stockholm, Almqvist & Wiksell.

KRIPKE (S.)
1959, "A Completeness Theorem in Modal Logic", *Journal of Symbolic Logic,* 24, p. 1-24.

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

LANGACKER (R. W.)
1987, *Foundations of Cognitive Grammar,* vol. 1, Stanford University Press.

MARR (D.)
1982, *Vision,* San Francisco, Freeman.

MARR (D.) & NISHIHARA (H. K.)
1978, "Representation and Recognition of the Spatial Organization of Three-dimensional Shapes", *Proceedings of the Royal Society in London, B 200,* p. 269-294.

MERVIS (C.) & ROSCH (E.)
1981, "Categorization of Natural Objects", *Annual Review of Psychology,* 32, p. 89-115.

MONTAGUE (R.)
1974, *Formal Philosophy,* ed. by R. H. Thomason, New Haven, Yale University Press.

PETITOT (J.)
1985, *Morphogenèse du Sens,* Paris, PUF.

1992, *Physique du Sens,* Paris, Éditions du CNRS.

PUTNAM (H.)
1980, "Models and Reality", *Journal of Symbolic Logic,* 45, p. 464-482.

1981, *Reason, Truth, and History,* Cambridge University Press.

[Reinfrank *et al.*]
REINFRANK (M.) & de KLEER (J.) & GINSBERG (M.) & SANDEWALL (E.), eds.
1988, *Non-Monotonic Reasoning,* Berlin, Springer-Verlag.

ROSCH (E.)
1975, "Cognitive Representations of Semantic Categories", *Journal of Experimental Psychology : General,* 104, p. 192-233.

1978, "Prototype Classification and Logical Classification : The Two Systems", p. 73-86, *in New Trends in Cognitive Representation : Challenges to Piaget's Theory*, E. Scholnik ed., Hillsdale (NJ), Lawrence Erlbaum Associates.

STALNAKER (R.)
1981, "Antiessentialism", *Midwest Studies of Philosophy*, 4, p. 343-355.

TALMY (L.)
1988, "Force dynamics in language and cognition", *Cognitive Science*, 12, p. 49-100.

SWEETSER (E.)
1990, *From Etymology to Pragmatics*, Cambridge University Press.

VAINA (L.)
1983, "From Shapes and Movements to Objects and Actions", *Synthese*, 54, p. 3-36.