PATTERN RECOGNITION: HUMAN AND MECHANICAL

SATOSI WATANABE
University of Hawaii

A Wiley-Interscience Publication
JOHN WILEY & SONS
New York • Chichester • Brisbane • Toronto • Singapore

CHAPTER 3 (excerpt)

idea. For instance, he says, "Blue and green are different simple ideas, but are more resembling than blue and scarlet." He notes also when (1) one globe of white marble, (2) one cube of white marble, and (3) one globe of black marble are presented, (1) and (2) resemble each other with regard to the "aspect" of color and (1) and (3) resemble each other with regard to the "aspect" of figure.

3.3 SIMILARITY THEORY

The commonsensical view of classes is that of a collection of similar objects. One might say that after many centuries of scholastic detour, the philosophers with and after Hume have come back to this simple-minded yet robust commonsensical view. The theory of general concepts based on similarity has regained some respectability, but we should note that the most important contribution by Hume, namely, that similarity is a product of association formed by mental habit, is left out of consideration by most of the proponents of the similarity theory.

Those who hold the similarity theory, though diverse in details of their arguments, agree more or less on the following two points:

1. What really exist are particulars, not universals.
2. It is not the case that members of a class covered by a general concept have an identical common element which non-members do not have. (If they had a common element it would justify a realism.) The particulars in a class are bound together only by similarity.

The practitioners of mechanical pattern recognition usually share this view too; however, these two points raise several serious questions. Relegating the most serious objection to this view to the next chapter, we shall limit ourselves here to the criticism already raised by different philosophers, and to some simple counter-examples we can find in our daily life.

Recently, Strawson [S-8] presented cogent arguments against the first point. To quote, "Grant that the existence of universals is no more than the fact, or the possibility, that particular things do, or might, exemplify them. But acknowledge that the fact or possibility of their being exemplified is no less than the fact of their existence." He says that we can apprehend or perceive a particular thing only by referring to its general characters. We can indeed not recognize a particular dog without using a concept of dog, a concept of color, a concept of size, etc. Therefore, the existence of particulars and that of universals are mutually dependent.

Further, Strawson applies Quine's criteria for existence to universals: (a) To be is to be the value of a variable, and (b) To be requires having an identity. In plain language, (a) means that an object that exists has to be a subject of a predicate or predicates. (Remember what we have called the Aristotelian view of the world at the beginning of Chapter 2.) Usually, Quine's criteria are thought to apply only to particular things, but Strawson argues that they apply just as well to universals. According to Strawson, particular things exist only in the space-time, whereas the universals exist only in our thought. This standpoint may be characterized as a special variation of conceptualism with realistic leaning.

There are two ways to interpret the second theme of the theory of similarity. One way is to say that similarity is not an attribute of an object, but a relationship between two objects. If we consider a pair of objects as one particular thing, then similarity (or its absence) can be considered as a property of a pair of things. Those who oppose this view would say that it is a realism in disguise, and that if we want to be consistent, we should apply the similarity view also to the general concept of similarity. But, this similarity view of similarity immediately encounters a serious difficulty, namely, that of defining the similarity between two particular similarities, where we have to resort to similarity of similarity, etc. And we can never end the chain of infinite regression. Of course, there is an argument that we do not need to flee forever, and we can cut the chain somewhere in a finite domain. But this argument is as obscure as the argument of infinite regression itself.

Another question is: Assuming that similarity is not a realistic attribute, do we have, or can we achieve, an objective judgment or at least an intersubjective consensus as to similarity of two things? Many negative answers are at hand.

For instance, suppose we have a whale, an elephant, and a tuna fish. The whale and elephant are similar because they are mammals, but the whale and tuna fish are similar because they live in water. This shows that similarity depends on the aspect we are using in the judgment of similarity.

Persons of a race remote from one's self all look similar, whereas persons of one's own race all look different. Discrimination seems to depend on remoteness. Is this an effect akin to Weber-Fechner's law? In any event similarity depends on perspective and remoteness.

Suppose we have three square wood panels whose sides are (a) 1 cm, (b) 2 cm, and (c) 2.8 cm. Since the difference between (a) and (b) is 1 cm, while the difference between (b) and (c) is 0.8 cm, we may say that (b) and (c) are more similar than (a) and (b). If we take areas, however, instead of the length of the side, we have (a) 1 cm^2, (b) 4 cm^2, and (c) 7.84 cm^2. Therefore, (a) and (b) are more similar than (b) and (c). One may say that this again is a case of dependence of similarity on the aspect. If we want to build a shelf in a narrow space, the aspect of the side may be more useful. If we sell building materials, the aspect of area may be more pertinent.

For Japanese-speaking people green and blue are so similar that they often form a single category, because the Japanese word blue (ao) often covers both green and blue. The distinction between French *saucisse* and *saucisson* does not exist in the minds of English-speaking people. The distinction between the English desk and table is also absent in the German mind. The desk for them is only a subclass of the table, not another class. Rat and mouse have almost opposite emotional connotations in European languages, whereas it is almost impossible to make a difference between them for the Japanese-speaking people. All these aspects show the language dependence of similarity. To admit a language dependence is, of course, different from adherence to nominalism, but there is no such thing as an intersubjective or intercommunal consensus.

In Chapter 2, we learned that there is strong evidence that the origin of general ideas is to be found in the prelingual stage of evolution. In general, the philosophers of linguistic leaning exaggerate the dependence of thought on languages; however, this does not mean that we can ignore the influence of language on the general concepts.

In relation to linguistic influence on human thought, we should not neglect what Wittgenstein said with regard to general concepts (see Sections 66 and 67 of [W-14], Anscomb's version, p. 31). He takes as an example what is called "games." He declares that there is nothing common in various activities called under the same name. He says, like many other similarity-theoreticians, that similarities and relations, etc., bind them together under the same name. What is new is that he points out that some games are bound together by some kind of similarity such as "winning and losing" and some other games are bound together by some other kind of similarity such as "amusement." And these different similarities are overlapping, so that all kinds of activities form a family called games. He writes: "We see a complicated network of similarities overlapping and crisscrossing: sometimes overall similarities, sometimes similarities of detail." He chooses the word *"Familienähnlichkeit"* ("family resemblance") to express his idea.

Therefore Wittgenstein's theory sometimes is taken as a special case of the similarity theory, allowing disjunction of subgroups. Barnbrough wanted to claim that Wittgenstein initiated a new third position between nominalism and realism, but his claim is not well-founded [B-4] p. 207. Although Wittgenstein acknowledges that the family is not bound merely by a common name, he knows that the concept of games, for instance, depends on what people consider as games. And this depends obviously on the linguistic-behavioral community and its habit of classification. For instance, there could very well be a community which does not consider gambling and marathon running as the same kind of activities. And this dependence on the kind of community is probably an undeniable fact of life. Recognition of the societal and linguistic influence on the small details of concepts is different from the recognition of language as the major determinant of human

concepts. See also [H-10] as one of the first books on pattern recognition that referred to Wittgenstein.

Apart from the explication of similarity as an association of ideas, Hume's emphasis on the degree of similarity has a great importance for the problem of general ideas. We can't say that to recognize coincidence of a property between two objects amounts to a concession to realism, because we can always quantize continuous variables and reduce the description to yes-or-no statements of categorical predicates. Conversely, most of the categorical predicates can be considered as approximations to reality. There are two ways of introducing a degree of similarities. When we use categorical predicates, we can use the number of predicates shared (i.e., simultaneously affirmatively satisfied) by two objects is a usual measure of similarity. Two men are said to resemble each other, when both are taller than 5 feet 8 inches, both wear eyeglasses, their eye colors are both non-black, etc.

When we use n continuous variables, we can consider an n-dimensional space and define a formula of distance between two points in the space. The concept of distance which is often more convenient than similarity is such that the larger the distance, the smaller the similarity. Simple examples of the relation between distance D and similarity S are a linearity:

$$S + D = \text{constant} \tag{3.3.1}$$

or inverse proportionality,

$$S \cdot D = \text{constant} \tag{3.3.2}$$

but these are not the only possible ones.

When we define a concept of distance, it is customary to require the so-called triangular relation: If D(A,B) represents the distance between object A and object B, then the distances between A and B, B and C, and C and A should satisfy

$$D(A,B) + D(B,C) \geq D(A,C). \tag{3.3.3}$$

Of course, this rule is still so lenient that all sorts of concepts of distance are allowed. The so-called Euclidean distance is only just one possible case, although it is very commonly used:

$$D(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}, \qquad (3.3.4)$$

where x_i, $i = 1, 2,\ldots, n$ is the i-th coordinate of point x in the n-dimensional space subtended by the n variables; similarly, for point y.

3.4 ABSTRACTION, SIMILARITY, AND CATEGORIZATION

The foregoing discussions are based on the assumption that the world consists of a discrete number of objects subject to a discrete number of attributes, affirmatively or negatively. The logical notation for a proposition P(a) which expresses that object a satisfies property P, reflects this "Aristotelian" view. To give a mathematical expression to this picture of the world, it is convenient to use a matrix representation, which we call Aristotelian table or object-predicate table [W-7].

Let $X = \{x_i\}$, $i = 1, 2,\ldots, m$ be the collection of objects under examination and $Y = \{y_j\}$, $j = 1, 2,\ldots, n$ be the collection of predicates that are used to describe the objects. The Y is assumed to be such that proposition $y_j(x_i)$ is either true or false for each i and each j. To make this assumption be always true, it is sometimes useful to agree that $y_j(x_i)$ is true when the predicate originally meant by y_j is applicable to x_i and found to be true, and $y_j(x_i)$ is false when the predicate originally meant by y_j either is applicable to x_i and found to be false or is not applicable. The object-predicate table A (for Aristotle) is an n x m matrix with m rows and n columns whose element $A(x_i,y_j)$ or A_{ij} is equal to 1 or 0 according to whether the proposition $y_j(x_i)$ is confirmed to be true or false, i.e., according as the predicate y_j is affirmed (satisfied) by x_i or is negated (not satisfied) by x_i. Because of the aforesaid convention we can increase X and Y indefinitely and still define the values of A_{ij}.

We have mentioned before, and we shall discuss again later, the questionable nature of the assumption that we can always identify a

self-identical object. But we shall not take time in this section to analyze this assumption, which is necessary to introduce set X. As far as set Y is concerned, it is mandatory to assume the "compatibility" of the predicates in Y, or, more generally, they should be "adequate" predicates. It is a common occurrence in atomic physics and psychology that two predicates, say, y_j and y_k, are incompatible, i.e., they change the values, when we change the order of observation, i.e., when we observe first y_j and second y_k or first y_k and second y_j. This is a case of what was called "pseudo-predicates" [W-22], [W-15], [W-34], [W-35]. Pseudo-predicates or inadequate predicates are those which cannot be assumed to have definite sets of objects that clearly affirm or negate them. Sallantin claims that non-compatibility exists even among some ordinary predicates (such as determination of medical signs) which are usually assumed to be purely "primary qualities" [S-7]. Y is assumed to be a collection of predicates that are genuine predicates and mutually compatible, i.e., those whose values are definite and do not depend on the order of their observation. See A-5,A-6 for anti-Aristotelian logic.

Second, we have to adhere primarily to an empirical viewpoint. The assignment of value to the object-predicate table must be based on the information obtained by observation. Human sensory organs do complicated manipulation of physical stimuli. Also, logical operations of the observational data are inevitable. For instance, a predicate y_j and its negation $\neg y_j$ are equivalent as information content and are interchangeable--0 becomes 1 and 1 becomes 0. Instead of two variables y_j and y_k we can use $y_j \cap y_k$, and $y_j \cup y_k$ if we are ready to lose part of the information. The loss of information is obvious because we cannot reconstruct the values of y_j and y_k separately from the values of $y_j \cap y_k$ and $y_j \cup y_k$. But, sometimes we want purposely to lose part of the information, to limit our knowledge to only useful information.

It is obvious that if we introduce a new predicate from the old predicate by the use of the logical connectives $(\neg,\cup,\cap)$, the entries for these new predicates will be obtained by the use of the truth-value tables for these connectives. But, if there are some constant (i.e., valid for all members of X) relations between observationally independent predicates, these relations bring some new information. If one observation y_j determines whether an animal has a gill, and another y_k determines whether the animal lives in water most of the time, we shall have $y_j \to y_k$ (if y_j then y_k), although $y_j \neq y_k$ (because a porpoise lives in water). This means that if $A_{ij} = 1$ then $A_{ik} = 1$, but, if $A_{ij} = 0$, A_{ik} can be either 0 or 1. A row of matrix which has $A_{ij} = 1$ and $A_{ik} = 0$ does not appear in the observational object-predicate table.

This constraint is due to an empirical law. More generally, the constraints may be due to an accidental situation created by the finite size of X. On the other extreme, if we increase X as much as possible, the number of constraints will become smaller and smaller. The remaining constraint may be considered purely "analytical." If we include in X all the objects in all possible conceivable worlds, the remaining constraint will be due to logical reasons. From this point of view, we might understand the idea of Quine [Q-2] and White, that the distinction between synthetic truth and analytic truth is not as clear-cut as traditional philosophy usually assumes, because no constraint holds in all possible worlds. If we include all possible worlds, what remains will be considered as (Boolean) logical constraint. If we further include the non-Boolean worlds into our consideration, we shall be left with only very few fundamental constraints. If we go backward, we shall gradually see more and more of logic-dependence, language-dependence, theory-dependence, nature-dependence, and finally pure contingence (see pp. 363-366 of [W-10] for more discussion). At the same place, we can see the difference between our object-predicate table and Carnap's "state description" [C-5].

Now, let us go back to the ordinary cases where X consists of a finite number of accessible objects and Y consists of usually used observations. First, let us consider Locke's idea of abstraction. Suppose G is a subset of X, $G \subset X$, consisting of all available particular objects covered by a complex idea (general concepts) C. Then, the abstraction, according to Locke's idea, must be a collection of those predicates (features) common to all members of G.

$$F = \{y_j | A_{ij} = 1, \quad \text{for all } x_i \in G\}. \tag{3.4.1}$$

The members of G are characterized by the condition that $A_{ij} = 1$ for all $y_j \in F$. We can take the conjunction of all such predicates

$$f = \bigcap_{y_j \in F} y_j \tag{3.4.2}$$

and call f the conjunctive feature of G.

According to the above derivation, it is clear that if $x_i \in G$, then $f(x_i)$, i.e., x_i satisfies f, but it is not clear that if $x_i \in (X - G)$, $\neg f(x_i)$, i.e., x_i does not satisfy f. However, we shall show in the next section that for any arbitrary collection of objects, we can define f such that $f(x_i)$ if, and only if, x_i belongs to the collection. The same general concept C can therefore be seen from either the aspect of collection of objects or from the aspect of collection of predicates. The former, i.e., G, is the extension of C and in the latter, i.e., F or f is the intension of C. The extension expresses the richness in membership, and the intension expresses the richness in characterization. We shall discuss the notions of extension and intension more rigorously in the next section, but here we would like to give a general idea of the quantitative aspects of intension and extension.

We should note that we have to expect that if we increase the number of samples that constitute G, some of the predicates y_j of F will drop out. Therefore, when we have a limited G and guess what F would be in the limiting case of large G, the mental process is a generalization and a case of inductive inference. Some of the predicates will be expected to survive in this process, and these may be considered to be more essential to characterize the general idea C. But, such a characterization is possible only on the assumption that there exists a unique limiting F for a given general idea.

If we call the number of members of G extension of concept C, and the number of members of F intension of concept C, we have obviously

$$\frac{d\ \mathrm{Ext}(C)}{d\ \mathrm{Int}(C)} \leq 0, \tag{3.4.3}$$

where

$$\mathrm{Ext}(C) = \mathrm{Card.}\ G \tag{3.4.4}$$

$$\mathrm{Int}(C) = \mathrm{Card.}\ F \tag{3.4.5}$$

Indeed, if we increase the number of characterization of G, it will tend to drive out some members of G. The equality sign in (3.4.3) can happen because an addition of predicate sometimes does not change the set G. For instance, if F includes in addition to the adjective "--is a biped mammal," the phrase "--is a speaking animal," then this addition would not change the set; disregarding the existence of mums. The set is still the set of human beings.

To count the extension of G, it is more convenient to assume that there are no two rows in the object-predicate table which are exactly the same, because as far as our description is concerned, they are "identical." If we add a new independent member to Y, it may happen that one single row will split in two.

If we denote by F' the feature of another group G' which could be the complementary group X - G or of any other competitive group; i.e., if

$$F' = \{y_j \mid A_{ij} = 1,\ \ x_i \in G'\} \tag{3.4.6}$$

then the discriminating feature F_d of G from G' would be

$$F_d = F - F \wedge F'. \tag{3.4.7}$$

Each term in this expression will become smaller as X increases.

If X consists of several classes and the remainder, i.e., if

$$X = G_1 \vee G_2 \vee \dots \vee G_\mu \vee R$$

$$G_\rho \wedge G_\sigma = \Phi \qquad \rho, \sigma = 1, 2, \dots, \mu$$

$$G_\rho \wedge R = \Phi \qquad \rho = 1, 2, \dots, \mu$$

we can say that the given objects are categorized into μ categories and the remainder.

What has been said so far applies to any kind of collection of objects. It is true that if members of the class do not resemble each other a great deal, the number of common features F will become small, but, we can still formally define F as well as f. In order to give some meaning to a collection of objects, we have to introduce the idea of similarity.

Thus, we have now to interpret the similarity theory in terms of the object-predicate table. Each row A_{ij}, i = fixed, j = 1, 2,..., n, represents the properties of an object. The similarity between two objects x_i and x_k is to be defined in terms of A_{ij} and A_{kj}, with i = fixed, k = fixed, j = 1, 2,..., n. As indicated earlier, it is often convenient to use a distance instead of a similarity, such that similarity S and distance D stand in the relation:

$$dS/dD < 0. \tag{3.4.8}$$

The simplest way is to use the so-called Hamming distance

$$D_{ik} = \sum_{j=1}^{n} |A_{ij} - A_{kj}| \tag{3.4.9}$$

which measures the number of y_j's for which object i and object k have different entries. The so-called Tanimoto distance defined by

$$D_{ik} = \frac{W(i,k)}{W(i) + W(k) - W(i,k)} \tag{3.4.10}$$

with

$$W(i) = \sum_{j=1}^{n} A(x_i, y_j),\quad W(i,k) = \sum_{j=1}^{n} A(x_i, y_j) A(x_k, y_j) \tag{3.4.11}$$

amounts to giving more emphasis to affirmative entries (1) than to negative entries (0). This may, in some cases, be useful, but logically y_j and $\neg y_j$ are perfectly equivalent. Therefore, the ground for adoption of this definition is rather weak. Of course, the choice of definition of distance is not limited to these two. It is usually assumed that $D_{ik} = D_{ki}$.

Now, two objects x_i and x_k may be considered as similar, when the distance between them is less than a threshold θ:

$$D_{ik} \leq \theta. \qquad (3.4.12)$$

If a family is to be defined by similarity, we shall have

$$x_i \varepsilon G \text{ and } x_k \varepsilon G \text{ iff } D_{ik} \leq \theta \qquad (3.4.13)$$

If we want to allow a disjunctive definition of families of similar objects, in accordance with Wittgenstein, we can write

$$G = \vee\, G_\rho, \qquad (3.4.14)$$

where each G_ρ is defined by similarity (3.4.13) and the disjunction is defined by the condition that ($x_i \varepsilon G$ and $x_k \varepsilon G$) means that there exists a series of G's: $G_\alpha, G_\beta, \ldots, G_\omega$ such that

$$x_i \epsilon G_\alpha, \; x_k \epsilon G_\omega$$

and (3.4.15)

$$G_\alpha \wedge G_\beta \neq \phi, \; G_\beta \wedge G_\gamma \neq \phi, \ldots, G_\psi \wedge G_\omega \neq \phi.$$

If we decrease the threshold θ, the similarity increases, as a result of which the condition for membership in G will be more restrictive. This will cause Int(C) to become larger and Ext(C) to become smaller.

3.5 INTENSION AND EXTENSION

In Section 3.4, we have dealt with intension and extension mainly from the point of view of the greater-smaller relation and did not concern ourselves about their actual values. For this limited purpose, it was sufficient to define Y vaguely as the set of predicates used in the description of objects observed, and X as the set of available objects.

But, in defining Y we could include or exclude $\neg y_j$ when y_j is already included in Y without altering the content of the description. This

approach will change the value of Int(C) as defined by (3.4.5). Again, in addition to two predicates y_j and y_k, we may include or exclude $y_j \cap y_k$ and/or $y_j \cup y_k$ in Y, without changing the information content, but this will also alter the value of the Int(C). To have a really fixed value of Int(C), we should probably include all the logical combinations of the starting predicates.

As a matter of fact, the so-called raw data are outputs of sensory organs and observation instruments. They are, as we have seen in Chapter 2, results of complicated functions of physical inputs, including logical operations. What is important is the information content, and not the question as to which are the original predicates. For instance, when there are two original predicates y_j and y_k, we can just as well take y_j and $y_j \uplus y_k$ [$= (y_j \cup y_k) \cap (\neg y_j \cup y_k)$] instead of y_j and y_k as original predicates without changing the information contents. Indeed if we write $z_j = y_j$ and $z_k = y_j \uplus y_k$, then we can write $y_j = z_k$, $y_k = z_j \uplus z_k$. There is no logical ground why y_j and y_k should be considered "atoms", whereas z_j and z_k should be considered "molecules". For if we take z_j and z_k as "atoms", then y_j and y_k become "molecules." (The names "atoms" and "molecules" are used here according to the usage of some philosophers, and they are different from what should be called atoms and molecules according to the lattice theory.) In any event, we can state that there is no logical or empirical reason for the so-called logical atomism as philosophical doctrine.

If we have two object-predicate tables, one with y_j and y_k, the other with z_j and z_k, whereas all the rest is the same, the intention of C calculated from them can be different. Indeed, if the intension of a general concept includes y_k but not y_j, the contribution to Int(C) from these two columns in the table is 1, but if we use z_j and z_k instead of y_k and y_j, the contribution will become 2. To avoid all these anomalies, we should include all the logically non-equivalent combinations of starting predicates.

Another argument for including all logical combinations in Y is that the conjunctive feature f (3.4.2) which is the most important predicate may or may not be found in the original Y. But if we include all possible predicates in Y, the class predicate f will certainly be found among the members of the extended Y.

As the table is originally defined, subset G may contain many objects which have the same description; this will introduce an inconvenient superfluity of objects in definition of Ext(X). In order to fix this ambiguity, we can introduce a convention that we write only one row for all these objects which have the same description by Y. The extension Ext(C) will become the number of different "object-types" included in G in the table. Even if we extend the Y by logical combinations, clearly such an extension will not change Ext(C). Of course, if we introduce a new independent predicate other than logical combinations of the original predicates, one object-type may split into two, changing the extension. But as far as we do not increase the scope (i.e., the content of information), the extension of C will not change.

Let us now call $\hat{Y}$ the full enlargement of Y by including all possible logical combinations. A member of Y, say y, is a member of $\hat{Y}$, but a member of $\hat{Y}$, say $\hat{y}_i$, may or may not be a member of Y. To obtain the entries for the enlarged object-predicate table, we should use the following rule:

$$A(x_i, \neg\hat{y}_j) = 1 - A(x_i \hat{y}_j) \text{ for all } i \tag{3.5.1}$$

$$A(x_i, \hat{y}_j \cap \hat{y}_k) = A(x_i, \hat{y}_j) A(x_i, \hat{y}_k) \text{ for all } i \tag{3.5.2}$$

$$A(x_i, \hat{y}_j \cup \hat{y}_k) = A(x_i, \hat{y}_j) + A(x_i, \hat{y}_k) - A(x_i \hat{y}_j) A(x_i, \hat{y}_k) \tag{3.5.3}$$

If we take $\hat{y}_k = \neg\hat{y}_j$, we get from $A(x_i, \hat{y}_j \cap \neg\hat{y}_j) = A(x_i, \emptyset) = 0$ for all i, and from $A(x_i, \hat{y}_j \cup \neg\hat{y}_j) = A(x_i, \Box) = 1$ for all i. Thus, $\emptyset$ and $\Box$ will be members of $\hat{Y}$.

To express in words what is meant by (3.5.1) through (3.5.3), we can say (1) $\neg\hat{y}_j$ is obtained from $\hat{y}_j$ by exchanging 1 with 0, (2) $\hat{y}_j \cap \hat{y}_k$ is obtained by putting 1 where both $\hat{y}_j$ and $\hat{y}_k$ have 1 and putting 0 otherwise, and (3) $\hat{y}_j \cup \hat{y}_k$ is obtained by putting 1 where either $\hat{y}_j$ or $\hat{y}_k$ or both have 1 and putting 0 otherwise.

From the n original predicates y_j ($\in$ Y), we can formally construct $m = 2^n$ different object types. But, in an actual table some object types may be missing. The reason for the absence of certain types may be classified into four general cases: (1) A sheer accident due to the finiteness of the collection X, (2) A logical reason --for instance, if Y contains a predicate y_j and its negation $\neg y_j$ all the rows with entries 1 in both columns y_j and y_k will be missing, (3) A semantic reason. For instance, if y_j means "is red," and y_k means "is colored" then all the object types with 1 for y_j and 0 for y_k will be missing, (4) An empiric reason--for instance, if y_j means "... has gills" and y_k means "... lives (most of the time) in water," then the type with 1 for y_j and 0 for y_k will be missing for an empiric reason.

Cases that violate (2) above do not really appear, therefore we can ignore them. The absence of certain types due to (1), (3), or (4) above does not show any difference in appearance. They can be written in the form of a logical relation (not logical connective), i.e., either equivalence (=) or implication ($\rightarrow$). Note that one equivalence amounts to two implications. Each time there exists one implication ($\rightarrow$), entries for y's become pairwise equal. For instance, if $y_j \rightarrow y_k$, then pairs: (y_j and $y_j \cap y_k$), ($\neg y_j \cap y_k$, and $y_j \uplus y_k$), ($\neg y_k$ and $\neg y_j \cap \neg y_k$) (y_k and $y_j \cup y_k$), ($y_j \cup y_k$ and $\neg(y_j \uplus y_k)$), ($\neg y_j$ and $\neg y_j \cup \neg y_k$), ($y_j \cap \neg y_k$ and $\emptyset$), and ($\neg y_j \cup y_k$ and $\Box$) become identical (see Table 3.5.1).

As for the possible predicate types in Y, we note that for one y_j, we can have 0 or 1 for each x_i. Therefore, the number of predicate types, i.e., the number of different columns in the table with m objects is 2^m. If all the formally possible object-types for the original Y are available, the number will be 2^{2^n}. Each time there is an "implication" relation among $\hat{y}_j \in \hat{Y}$, this number becomes one half of what would be in the absence of the implication. The collection of all the predicate types constitutes the complete Boolean lattice. The types of predicates in $\hat{Y}$ can be classified into those which have r ones and m-r zeros, with r = 0, 1,..., m. The integer r is called the rank of a predicate; r = 0 corresponds to $\emptyset$ and r = m to $\Box$. There are m $\hat{y}$'s with r = 1; these are atoms in the sense of lattice theory. If we consider a fixed r, there are $\binom{m}{r}$ different ways of placing one in r places out of m possible places. Therefore the total number of possible ways of distributing ones and zeros is $\sum_{r=0}^{m} \binom{m}{r} = (1+1)^m = 2^m$, coming back to the same number of predicate-types. There is a one-to-one correspondence between an atomic predicate and an object type, because for a given atomic predicate $\hat{y}_j$, there is only i such that $A_{ij} = 1$ and for a given i, there is only one $\hat{y}_j$ that has only one 1 along the column and precisely on the i-th row.

It is obvious that the extension of a predicate with rank r, say $\hat{y}_j$, can be considered as a disjunction of r atoms. This is a predicate that applies to these r different object-types of the last section. Therefore r is the extension of the predicate in question. If we apply this description to a collection of r objects G of last section, then the predicate $\hat{y}_j$ of rank r under consideration corresponds to f of last section.

On the other hand, all predicates that are implied by the predicate $\hat{y}_j$ in question correspond to properties shared by all the objects covered by $\hat{y}_j$. Consider a predicate with rank r+1 that is implied by $\hat{y}_j$. Such a predicate has one more one in the column than $\hat{y}_j$. There are (m-r) places to place the one, hence there are (m-r) different predicates of rank r+1 which are implied by $\hat{y}_j$. Similarly, there are $\binom{m-r}{2}$ different predicates of rank r+2 that are implied by $\hat{y}_j$, etc. There are $\binom{m-r}{m-r-1} = \binom{m-r}{1} = (m-r)$ predicates of rank (m-r-1) + r = m-1 that are implied by $\hat{y}_j$, and of rank m we have one predicate □. In total $1 + \binom{m-r}{1} + \binom{m-r}{2} + \ldots \binom{m-r}{1} + 1 = 2^{m-r}$ that are implied by $\hat{y}_j$, including $\hat{y}_j$ itself. The collection of these predicates is the F of the last section. Any of these numbers (except the one with m = r) decrease with increasing r.

We can take any of these numbers that are decreasing the function of r as the intension of $\hat{y}_j$, as far as (3.4.3) is concerned. But for symmetry's sake, we can adopt the number of predicates of rank m-r that are implied by y_j. Thus

$$\mathrm{Int}(C) = m-r$$

$$\mathrm{Ext}(C) = r,$$

where C is the general concept corresponding to the collection of objects G.

We note that this way of defining G and F have nothing to do with the internal relation such as similarity among object-types. Any arbitrary collection of object-types can be dealt with in the same way. The idea of similarity will be discussed in the next chapter.

The reader can test the knowledge he has learned with the help of Table 3.5.1.

	$\hat{Y}$							
	Y							
rank in x	1	2	0	1	1	2	2	3
y	y_1	y_2	ø	$\neg y_2$	$\neg y_1 \cap y_2$	y_1	$y_1 \cup \neg y_2$	□
x	$y_1 \cap y_2$	$y_1 \cup y_2$	$y_1 \cap \neg y_2$	$\neg y_1 \cap \neg y_2$	$y_1 \cup y_2$	$\neg y_1 \cup \neg y_2$	$\neg(y_1 \cup y_2)$	$\neg y_1 \cup y_2$
x_1	0	0	0	1	0	1	1	1
x_2	0	1	0	0	1	1	0	1
x_3	1	1	0	0	0	0	1	1
x_4	1 0	0 1	0 1	1 0	0 1	0 1	1 0	1 0
rank in X	2 1	2 3	0 1	2 1	1 2	2 3	3 2	4 3

Table 3.5.1 Consider first only X which consists of three object-types with $(A_{i1}, A_{i2}) = (0,0)$, $(0,1)$, $(1,1)$, respectively. Since there are two original predicates (n = 2), there could be in principle four $(2^n = 4)$ object-types. But, one of them is missing (m = 3), causing a constraint $y_1 \rightarrow y_2$ As a result, a pair of predicates that can be formed in terms of y_1 and y_2 coincides with another, reducing the total number of elements of the completed Boolean lattice Y from its maximum value sixteen $(2^{2^n} = 16)$ to eight

($2^m = 8$). The atoms corresponding to x_1, x_2, x_3 are respectively $\neg y_2$, $\neg y_1 \cap y_2$ and y_1. Now add x_4 which is characterized by $A_{41} = 1$ and $A_{42} = 0$. The m of this extended $\hat{X}$ is 4. The constraints disappear and the predicate-types become 16. The rank with respect to X and that with respect to $\hat{X}$ are not the same. For instance, y_1 and $y_1 \cap y_2$ have the same rank 1 with respect to X, but they become respectively 2 and 1 with respect to $\hat{X}$. Starting with two predicates y_1 and y_2, which constitute Y_A we can build a total of 16 predicates that form $\hat{Y}$. But as far as the description of X (but not $\hat{X}$) is concerned, two predicates $y_1 \cap y_2$ and $y_1 \cup y_2$ become identical with y_1 and y_2 respectively, thus appearing under the same rubric of Y.