NOTES ON THE DEFINITION

OF A CONCEPT

J. de Leeuw

January 1968

PSYCHOLOGICAL INSTITUTE - UNIVERSITY OF LEIDEN - NETHERLANDS

- 1 -

Let $X_1, X_2, \ldots, X_i, \ldots, X_m$ be random variables, and define the sample space Ω as the cartesian product set of the X_i

$$\Omega = \prod_{i=1}^{m} X_i$$

A typical element of Ω (an ordered m-tuple of values of the X_i) will be denoted by ω_k. If P is a probability measure over Ω then a function f with

$$f(\omega_k) = P(\Omega = \omega_k)$$

is called a probability distribution over . Clearly

$$f(\omega_k) \geqslant 0$$

$$\int_{\Omega} f(\omega)\, dF(\omega) = 1$$

(where integration is meant in the Riemann-Stieltjes sense). Note that f is a function of m variables, because $\omega_k \in \prod_{i=1}^{m} X_i$.

- 2 -

To define a "concept" we interpret the X_i as dimensions (not neccessarily those of the Euclidean space R^m) and Ω as the set of all possible stimuli. We now define a concept C_1 as an ordered pair

$$C_1 = \langle \Omega_1, f_1 \rangle$$

with

$$\Omega_1 \subsetneq \Omega$$

where f_1 is some probability distribution function over Ω_1. Without loss of generality, however, we may take $\Omega_1 = \Omega$ and define C_1 as follows:

$$C_1 = \langle \Omega\ , \ \phi_1 \rangle$$

where ϕ_1 is the product of f_1 and the indicator of Ω_1. This means that $\phi_1(\omega_k) = f_1(\omega_k)$ for each $\omega_k \in \Omega_1$ and $\phi_1 = 0$ elsewhere. Clearly ϕ_1 is a probability distribution with

$$\int_{\Omega} \phi_1(\omega)\, dF(\omega) = \int_{\Omega_1} \phi_1(\omega)\, dF(\omega) = \int_{\Omega} f_1(\omega)\, dF(\omega) = \int_{\Omega_1} f_1(\omega)\, dF(\omega) = 1$$

Because of this it is possible without any ambiguity, given the sample space Ω , to define a concept H_1 over this sample space as a probability distribution over Ω, which assigns to each $\omega_k \in \Omega$ a value $\phi_1(\omega_k)$ on the open interval $(0,1)$. Under the usual interpretation of conditional probability we may write

$$p(\omega_k/H_1) = \phi_1(\omega_k)$$

In words this reduces to the trivial statement: the probability of ω_k, given that the relevant distribution is ϕ_1, equals $\phi_1(\omega_k)$.

- 3 -

Consider as an example

$$X_1 = \{\text{red, green}\}$$
$$X_2 = \{\text{square, circle}\}$$

so

$$\Omega = \{\text{red square, red circle, green square, green circle}\}$$

Five arbitrary concepts are defined in the following table (where the element $(1,k)$ is the $\phi_1(\omega_k)$ value).

	ω_1	ω_2	ω_3	ω_4
ϕ_1	.50	.50	.00	.00
ϕ_2	.00	.00	.50	.50
ϕ_3	.50	.00	.50	.00
ϕ_4	.20	.10	.60	.10
ϕ_5	.40	.40	.10	.10

- 4 -

According to the classical definition a concept is a class of stimuli having some property in common, or

$$C_1 = \left\{ y \;\middle|\; y \in \Omega \;\&\; F_1 y \right\}$$

where F_1 is the relevant predicate that partitions Ω into two mutually exclusive and exhaustive subsets: the positive and negative instances of the concept C_1. In our example we may identify F_1 with the predicate "red", F_2 with "green" and F_3 with "square". In the H_4 case confusion arises. The most reasonable thing to do would be to define F_4 as "red or green" or "square or circle" which implies $C_4 = \Omega$. For the same reasons $C_5 = \Omega$.

- 5 -

In a typical concept learning experiment the subject has to discriminate optimally between the positive and negative instances of a particular concept H_1. Now it is easy to see that in the case of H_4 or H_5 the class of possible positive instances C and the class of possible negative instances $\bar{C}$ are the same:

$$C = \bar{C} = \Omega$$

This means that if we define a concept as a subclass Ω_1 of Ω, then in "probabilistic" cases as for example H_4 and H_5 every stimulus is both a positive and a negative instance of the concept.
If S' task in a particular experiment is to discriminate between the concepts corresponding to H_4 and H_5 then the "class"-definition implies that $C_4 = C_5$. If, on the other hand, we identify concepts with probability distributions, this undesirable consequence does not rise. Moreover, we may apply this definition equally well to the deterministic cases such as H_1, H_2 and H_3.

Compare the following definitions:

A : A concept is a class of stimuli which are specified by dimensions in such a way that the values on each dimension are varied according to probability distributions.

B : A concept is a probability distribution over the total stimulus space.

C : A concept is a point in a m-dimensional space.

We have seen that definition A implies in some cases that $C_1=C_2=\Omega$. It is possible to avoid this undesirable consequence by applying A to the sample from Ω used in a particular experiment, but this is hardly a satisfactory way out (the worse our sampling, the better our definition works, vice versa). Definition B does not postulate, somewhere in the subject's head, a representation of, say, a chair in the form of a probability distribution. It is a description of the way S uses the word "chair", and in artificial laboratory situations a description of the way E is going to use the word "VEC" while giving feedback. Definition C can be interpreted as a special case of A and B:

$$(\exists x)\left\{(x\in\Omega \,\&\, \emptyset_1(x)=1) \;\&\; (y)(y)\neq x \Rightarrow \emptyset_1(y) = 0)\right\}$$

Another, somewhat platonistic, interpretation of C is to define a concept as the expected value (or some other measure of central tendency) of the probability distribution of definition B:

$$C_1 = \int_{\Omega}\omega \quad \emptyset_1(\omega)\,d\,F(\omega)$$

(which presupposes however that all X_1's are measurable).

In summary: our proposal is to define a concept as a probability distribution over the stimulus space Ω. This means that a concept is the description of the way a word (or some other linguistic entity) is used. Another possibility is to define a concept as some kind of central tendency of all the ways in which a particular word is used (all the situations in which it is applied). The choice tetween definitions B and C (both definitions of a concept, i.e. of the denotative meaning of a word) is a matter of philosophical taste, although it is obvious that we throw away useful information if we only consider the central tendency (e.q. the first moment) of the distribution.

+++