TR/03/85 January 1985

Knowledge Representation for Basic Visual Categories

by

R. I. Phelps and P. B. Musgrove

KNOWLEDGE REPRESENTATION FOR BASIC VISUAL CATEGORIES

R I Phelps - Brunei University
P B Mangrove - Poly - South Bank

Keywords: Learning, categorization, prototypes, knowledge representation.

This paper reports work on a model of machine learning which is based on the psychological theory of prototypical concepts. This theory is that concepts learnt naturally from interaction with the environment (basic categories) are not structured or defined in logical terms but are clustered in accordance with their similarity to a central prototype, representing the "most typical" member.

In attempting to operationalize this theory in a machine learning programme it has been necessary to study object descriptions of visual data. This has led to the development of a knowledge representation structure based on the natural properties inherent in visual images, shape and parts. Methods developed to achieve this are discussed.

The organisation of knowledge derived from this approach differs fundamentally from existing K:R formalisms in leading to a view of concepts as perceptual structures at the basic level,with relations between concepts being described via a superordinate level of abstract concepts. This has implications both for the differential treatment of concrete and abstract concepts and in suggesting a feature based retrieval mechanism.

This paper reports a proposed design for a scheme of knowledge representation which has evolved during work on a machine learning system. This scheme differs in many respects from current rule-based, semantic net and frame-based designs. The motivation for this scheme springs from work in progress to construct a program with a learning capability in the field of object recognition, and in particular at the fundamental level, that of natural object categorization. The structure of this paper is that firstly the background in this work is described, then the description method for encoding visual shapes, then the knowledge representation scheme to which we are led by these considerations. Finally, differences from existing schemes and implications for psychological investigations are discussed.

As this work is still in progress, many details of the scheme are at present undecided but the general outline and operational approach are clear and will not be significantly affected by whatever form the detailed mechanisms finally take. It is the intention of this paper to present and provoke debate upon the underlying principles of this scheme.

BACKGROUND

Stimulus situations are unique but living creatures do not treat them uniquely - they respond on the basis of past learning and categories. These categories are given labels (by humans) which represent a single mental concept for all individuals in that category, the assumption being made that the same structural and processing principles hold in both perceptual and conceptual realms. Consequently parallel theoretical developments have taken place in psychological research on perceptual and conceptual categorization.

In contrast to most work in AI to date, one school of psychologists believes that most natural concepts are ill-defined, that is, there is no rule that can determine membership for all members of a category. Furthermore, not all members of a concept have equal status- Members judged to be typical of a concept (e.g. apples for the concept 'fruit') can be categorized faster and more accurately than members judged less typical (e.g. tomato). This line of thought led Rosch et al [1] to develop 'prototype' theories of concepts in which membership of a category is determined by the typicality of a particular object to an ideal member of the concept which has the average attributes of all class members. This is not the whole story, however, as any object may be categorized at each of several different levels, higher levels being abstract and lower levels more detailed and specific, e.g. a chair may be classified as an inorganic object, a piece of furniture, a chair or a kitchen chair. These psychologists have argued that the most cognitively efficient and therefore most basic level of categorization is that level at which the categories produced provide the most distinct clusters of objects, i.e. the level which maximizes the similarity of objects within a category and maximises the differences between objects in different categories. Thus of the classifications suggested for a chair, the basic level category is 'chair' because chairs are quite similar amongst themselves and distinct from tables, tomatoes, etc, whereas items of furniture are not very similar amongst themselves and kitchen chairs do not differ sharply from other chairs. Tversky and Hemingway [2] provide evidence that this basic level of categorization has the following properties:

1) it is the most abstract level at which instances have similar shapes.

2) it is the most abstract level at which instances have similar parts.

3) it is the most abstract level at which a mental image can reflect the appearance of the entire category.

4) objects are recognized more quickly as members of basic level categories than as members of categories at other levels.

5) it is the level at which humans spontaneously name an object.

The overall intention of the work of which this paper forms a part is to construct a model of the categorization process which can learn basic level 'natural' categories from simple visual data without ins truction. The approach adopted is to produce a model that operationalizes the above psychological findings.

VISUAL INPUT - SHAPE DESCRIPTION

The domain for learning in this work was chosen to be that of 2-D silhouettes of 3-D objec ts, such that the difference between convex and concave parts of a silhouette represents properties of the 3-D surface and where the surface looks continuous in 2-D it really is continuous in 3-D. Marr [3] shown that if these conditions are met the 3-D surface can often be successfully inferred from the silhouette. In limit ing ourselves to this domain we are ignoring various sources of visual informat ion, e.g. colour, texture, depth, motion, so we might expect some deficiencies in what is considered a basic category (e.g. silhouettes of oranges and apples are very similar although they form different basic level categories). However, these properties could easily be incorporated within the knowledge representation scheme to be proposed, and the key properties of basic categories, those of shape and parts, will remain the same if we are judicious in our choice of projections. For these reasons it was thought reasonable to explore learning and representation in the 2-D domain which is much easier to handle as regards object description and segmentation.

Shapes and parts seem a good starting point for visual-categorization, both on the psychological grounds advanced previously and intuitively. They are therefore taken as the fundamental descriptors of visual object perceptions. This raises the question of precisely what is meant by 'parts'. Marr's research in machine vision has taken the view that objects are most naturally segmented into convex parts [3] and we have followed this line of thought but refined it so that parts need only be 'psuedo-convex' in the sense that further dividing than into more convex subparts does not significantly increase the measure of convexity. This approach is reported elsewhere in more detail [4]. The effect is illustrated by the resulting 'parts' of a horse in FIG 1.

The description of the visual image of a horse is achieved in stages. Firstly, the horse is described holistically by a set of descriptors including such measures as principal axis, axis extension ratio, compactness (perimeter/area), size, etc., applied to the whole image [5]. The precise set of descriptors used is unimportant as long as it contains a rough description of the shape. Only a rough description is necessary as more accurate descriptions are provided by successive stages. At the second stage the horse is decomposed as in FIG 1 into its primary subparts. Each subpart is now described by the same set of descriptors, and the relative position of each subpart is also stored. This process is now repeated to any desired number of stages, the subparts being successively divided and described in increasing detail. The result is a hierarchial description as in FIG 2.

The level of detail in which a given object is analysed would be determined sequentially. If it can be recognized at the holistic level there is no need for any decomposition into parts. Decomposition continues until the object can be recognized or until the attempt is given up.

KNOWLEDGE REPRESENTATION

Evolutionarily, the earliest forms of nervous systems are reflex systems giving a 'hard-wired' response to a recognized stimulus. Evolutionary development has resulted in the change from a fixed response to a more flexible knowledge-based response, but it seems reasonable to asume that knowledge representation structures developed earlier in connection with recognition provide the basic forms for the later representation of knowledge of other types. We therefore consider the representation needed for the efficient categorization of visual data and extend this form of storage to other levels of thought.

Given that the incoming data is primarily encoded in terms of part/shape hierarchies as outlined in the previous section, recognition must be achieved by matching this against stored representations of recognizable categories. The most direct way to achieve this is for the category representations in memory to be structured in the same way. However, different horses have slightly different shapes and it Is clearly inefficient to store a representation of every horse ever seen. A more efficient approach would be to store a 'typical' or 'average' horse representation, and look for a reasonably close match to this. Such an 'averaged' description only makes sense if the instances being averaged are sufficiently perceptually similar and if the descriptions do not descend to a level of detail where large differences appear. e.g. it is possible to visualize an 'average' banana but not an 'average' fruit. Further, the search tree of stored representations can be more efficiently searched the earlier it begins branching, i.e. if these averaged representations are representative of a large a class of instances as possible. Thus for recognition efficiency we would expect stored hierarchical shape/part descriptions with 'typical' or 'averaged' measure values to be formed at the maximum level of generality consistent with such averages being representative of their instances, Being representative of their instances means that within a category the instances cluster closely around the average compared to their distances from other categories. Thus the categories will be as distinct as possible subject to their averages making perceptual sense. It follows that we expect the structure of this knowledge representation scheme at the immediate recognition level to display the properties of Rosen's basic level.

These arguments suggest that there is a basic level of knowledge representation at which initial recognition is accomplished. It consists of prototypical representations which are 'averages' of category members and structured as in FIG 2. In order to be maximally general, these prototypes are not very specific, i.e. contain a relatively small degree of decomposition and thus of detail. Proposed mechanisms by which these prototypical categories might be formed (learnt) are reported elsewhere [6].

Within the knowledge representation scheme it is also necessary to accomplish more specific recognition by use of more detailed information than is required at the initial prototype matching stage. In keeping with our earlier remarks on recognition level organizations providing the form for representing other knowledge, and in keeping with Rosch's results [1], we assume that structure for recognizing more specific categories, e.g.. kitchen chair, replicates that at basic level. Thus within the class of objects assigned to a given basic category, e.g. 'horse', more detailed prototypical images are formed, each representing *a* smaller class of instances than 'horse' but again the prototypes being constructed to form classes of maximum generality and distinctness subject to the 'averages' being representative at this more detailed level. 'Horse' at basic level might be stored simply in terms of possessing a body, neck, head, 4 legs, tail and their rough shapes. At a subordinate level 'Racehorse' might have different shape descriptors for these parts and be further decomposed, e.g. leg into upper leg, lower leg and hoof. Since the shape measures of 'racehorse' are not the same as those of 'horse' in the parts they both include in their hierarchies, e.g. body, the subordinate levels are not simply lower branches on the basic level tree. (If they were, we would have to identify 'racehorse' only from its details e.g. hoof shape, ignoring differences in larger parts such as its body shape). Instead the structure consists of successive levels of prototypes, each prototype being stored as a hierarchical decomposition into parts and shapes, with the number of stages of decomposition and the specificity of the measures becoming greater at each subordinate level of prototype, FIG 3.

With this structure recognition is achieved in stages. Firstly a categorization is made at the basic level, by matching the object representation against the stored prototypes. Secondly categorization is made at the first subordinate level by matching the object representation against prototypes of categories directly subordinate to the recognized basic level category. This process may be repeated until a sufficiently specific categorization has been reached.

In this scheme there is no inheritance of features from a higher level e.g. horse to a lower level e.g. racehorse. Instead there is a separate complete prototypical representation on each level. However, this does not duplicate storage requirements, since as previously mentioned, the descriptor measures of 'horse', applying to a wider set of instances than 'racehorse', are different. This separate prototypes structure has the additional advantage that once an object has been categorized as a horse, a quick check on which subordinate category it fits may be made by checking only the initial levels of the subordinate prototypes against the object. If recognizable differences exist between subordinate prototypes at these initial levels of decomposition, it is not necessary to check the complete decoraposit ion of the prototype at this level.

The questions of how matching is evaluated between object and prototype representations and of how to order searches through the prototype space are present topics of investigation. Because of their nature as 'averages' only a 'close' match can be expected between representations. Thus a measure of similarity between representations is required. Such a measure induces both a 'degree of membership' of an object in a category and a set of similarity relations between the prototypes. These latter could be used in a technique such as multidimensional scaling [7] to induce a low dimensional ordering on the prototype space which might be used, together with context-driven constraints on lilcely prototypes, to design an ordered search procedure.

So far we have dealt with knowledge representation from the basic level down. Above the basic level there are more abstract superordinate levels. At these levels, a category, e.g. fruit, cannot be represented by a prototype and therefore has no representation as a decomposition into parts and shapes. The established pattern of clusters of instances according to within cluster similarity and between cluster differences must therefore be based on relationships other than overall perceptual similarity. In line with this these abstract categories are taken to consist of sets of concrete prototypes. Thus 'fruit' consists of 'apple', 'orange', 'pear', etc., and the similarity relations might include 'eaten by animals', 'grows on plants', 'juicy'. In general the similarity relations might be perceptual as in 'furry animals', based on actions or uses as in 'eaten by animals' or based on emotional or cognitive effects as in 'beautiful flowers', 'intended for storage'.

For such similarity relations to be utilized it is necessary to revise our picture of basic and subordinate leve1 categories to include with their prototypical perceptual descriptions also prototypical information on their uses, actions, interactions, emotional inplications, cognitive implications and any other important non-perceptual effects.

Categorization on the basis of non-perceptual features is aided by the existence of episodic memory, storing time sequence descriptions of recognized objects, their physical interactions and the observer's feelings and inferences about them [8]. At these abstract levels of categorization there is no natural division into distinct categories corresponding to the natural division into prototypical categories at basic level, and so the abstract classes formed are essentially arbitrary. They will be formed and survive as part of the knowledge structure depending on their usefulness, which is in associating input data with inferences and actions, since rule-type knowledge is most efficiently stored and used when represented in terms of the most general categories to which it applies. Thus the abstract categories formed will depend upon the environment, the decision processes being used and on feedback received

The complete picture of knowledge representation for categories at all levels is now given in FIG 4.

Note chat in a sense the abstract categories do not have any content. They are merely a collection of pointers to concrete prototypes. Also, just as at the basic level some instances are closer to the prototype and thus have a greater degree of membership of the category than others, some concrete prototypes will have a greater degree of membership in an abstract category than others.

DISCUSSION

The knowledge representation scheme for categories outlined in this paper contrasts in many respects with existing schemes [9, 10]. Firstly, there is no logical definition of categories. All categories in this scheme consist of clusters of instances or clusters of prototypes. An instance may have a greater or lesser degree of membership of a category, but there is no sharp division between members and non-members. This implies that the use of logical deduction as a reasoning mechanism with such a scheme must be very limited if used at all. Secondly, the representation structure is not hierarchical. The use of terms like superordinate and subordinate levels is merely illustrative; there is no inheritance of properties from more abstract to less abstract levels. Each category comes complete with its own properties represented in its prototype or, for abstract categories, its set of prototypes. Thus in recognition an object is not recognized as firstly animal, then fish and then goldfish but instead firstly at the basic level, say fish, and then depending on need, as goldfish or animal. Thirdly, the nature of abstract categories or collections of prototypes means that thinking about such categories will automatically be in terms of concrete 'examples' of the abstract category, the prototypes. This accords with the exemplar view of categories [11]. Fourthly, there is not necessarily a one to one correspondence between the categories (concepts) in the representation scheme and verbal labels for concepts. For example, a penguin and a sparrow will not be in the same prototype class at basic level, since they do not look alike. Thus there are at least two basic level categories 'penguins' and 'ordinary birds', say. However, most people would say that both a penguin and a sparrow are 'birds'. Thus 'birds' is a superordinate (abstract) category to which the verbal label corresponds, but there appears to be no verbal label ordinarily used for the 'ordinary birds' category. Since superordinate categories are different in nature from basic and subordinate ones, psychological experiments treating 'bird' on the same level as 'penguins' may be treating separate measures of similarity as identical, e.g. Tversky & Smith [12] compare similarities between different sorts of fruits such as cherry, apple, etc., with the similarities between cherry and fruit,apple and fruit, etc.

This knowledge representation scheme has been based on considerations of the fundamental cognitive processes, categorization and recognition. It does not attempt to deal with advanced reasoning processes which have formed the model for most existing schemes.

For some specific reasoning tasks this structure will seem inefficient, and the prototypical organization in some ways carries redundant information. However, redundancy implies flexibility which is certainly needed in a general knowledge representation scheme. If, as is usual in evolutionary biology, cognitive structures developed earlier for fundamental purposes are found to form the basis of more advanced processes, it is necessary that any model of human reasoning should be compatible with an underlying categorization structure. It is hoped that this paper will serve the purpose of initiating. debate on this underlying structure and its implications for more advanced reasoning.

References

[1] Basic objects in natural categories. E. Rosch, C.B. Mevis, W.Gray, D. Johnson. Cognit.Psychology 8 pp. 382-439. 1976.

[2] Objects, parts and categories. B. Tversky & K. Hemingway. J. Exp- Psychology 113/2 pp. 169-190. 1984.

[3] Vision. D. Marr. 1982.

[4] Brunei Technical Report, in preparation.

[5] "Relational Descriptions in Picture Processing" H.G. Barrow & R.J. Popplestone, Machine Intelligence 6. 1971 Edinburg University Press Ed 8 Moltzer & D Michie.

[6] Brunei Technical Report, in preparat ion.

[7] Mult idimensional Sea 1ing by optimizing goodness of fit to a non metric hypothesis. J.B. Kruskal. Psychometrika 29, pp 1-28. 1964.

[8] Episodic & Semantic memory. E . Tulving in Organization of memory, ed.
E. Tulving & W. Donaldson. Academic Press. 1972.

[9] Knowledge representation in The Handbook of AI, ed. A. Barr & E.A. Feigenbaum. Pitman. 1981.

[10] Conceptual Structures. J.F. Sowa. Addison-Wesley. 1984.

[11] Categories & concepts. E.E. Smith & D.L. Medin. Harvard University Press. 1981.

[12] Tversky and Smith reported in [11] .

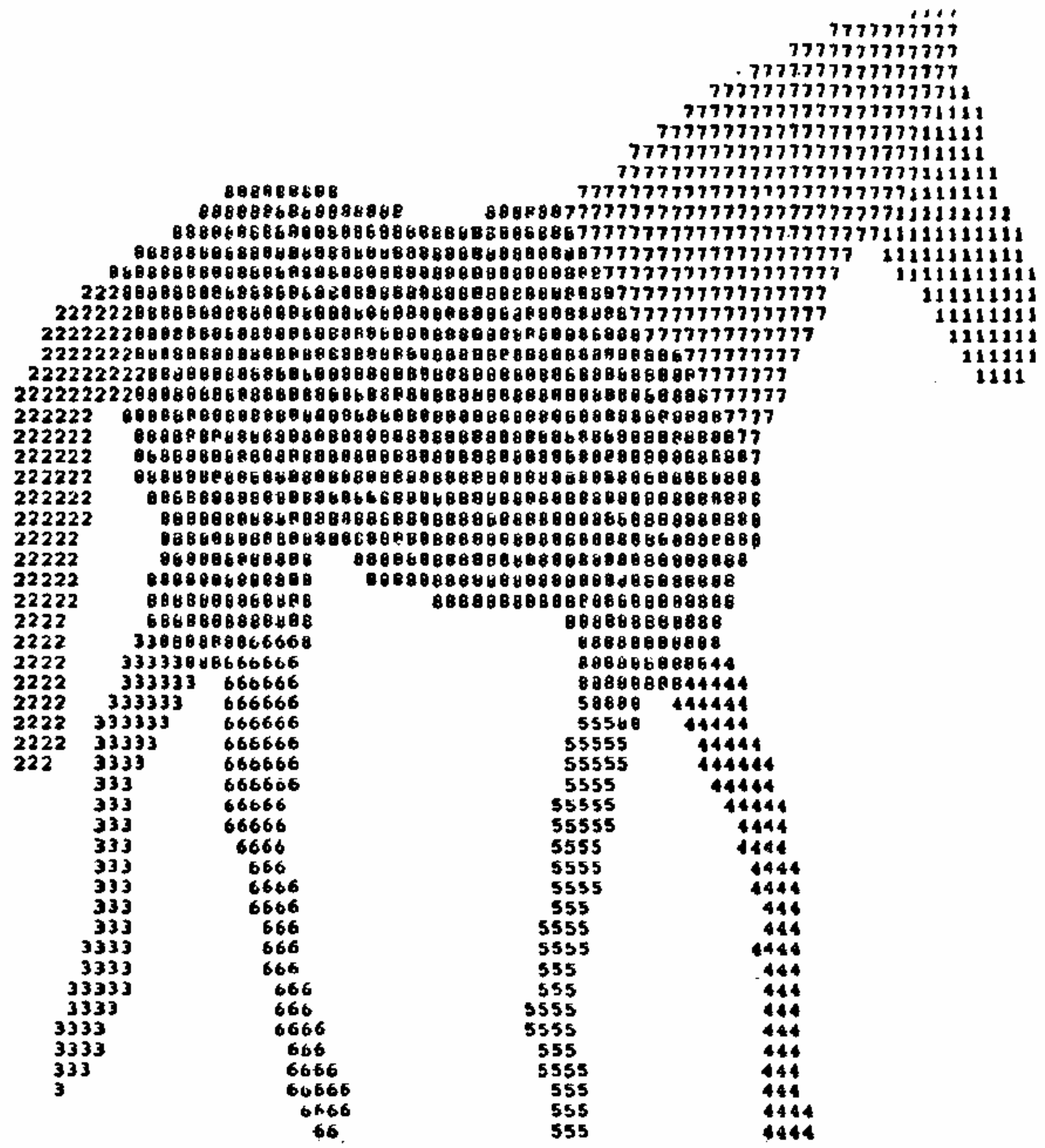


FIG. 1

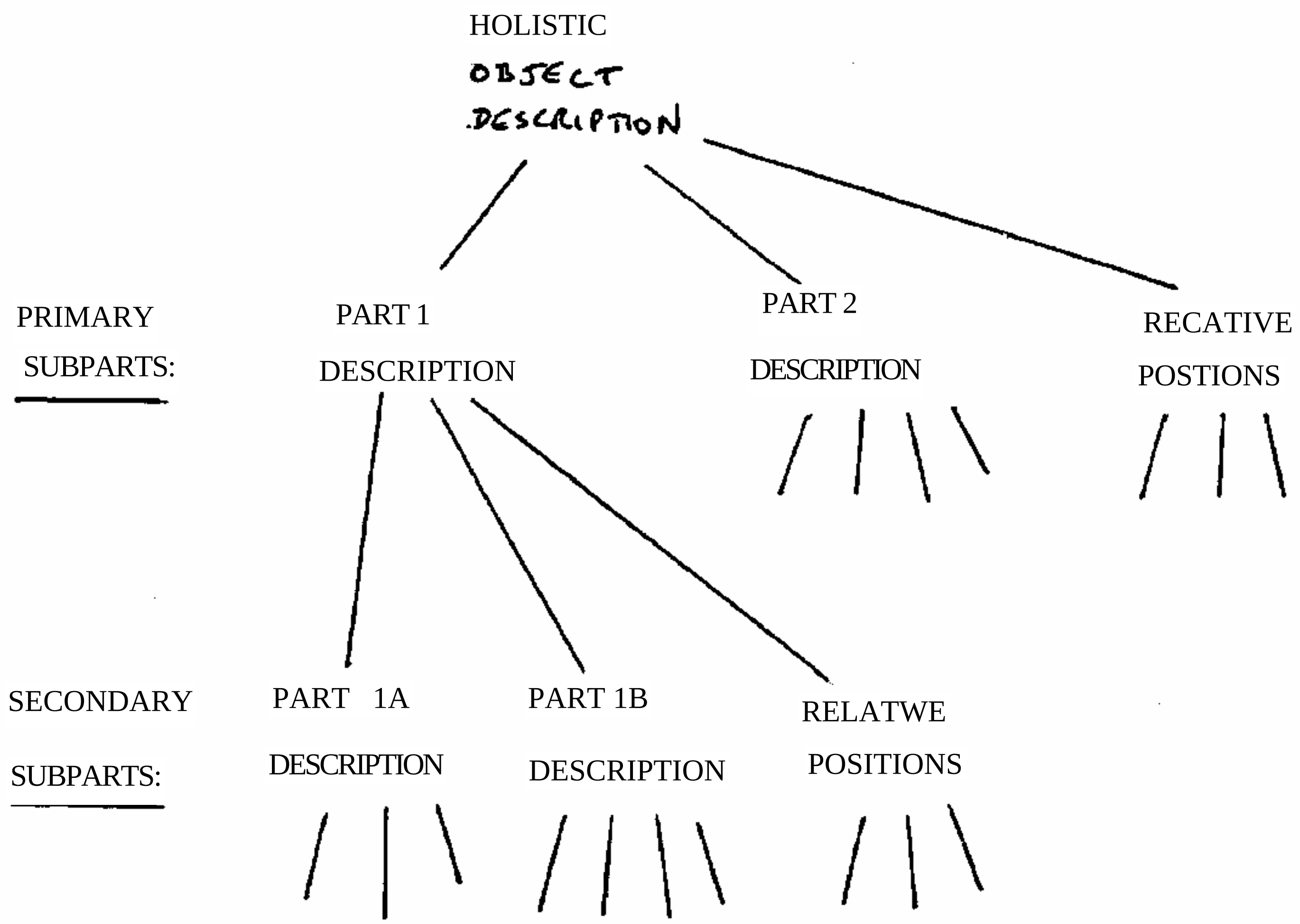
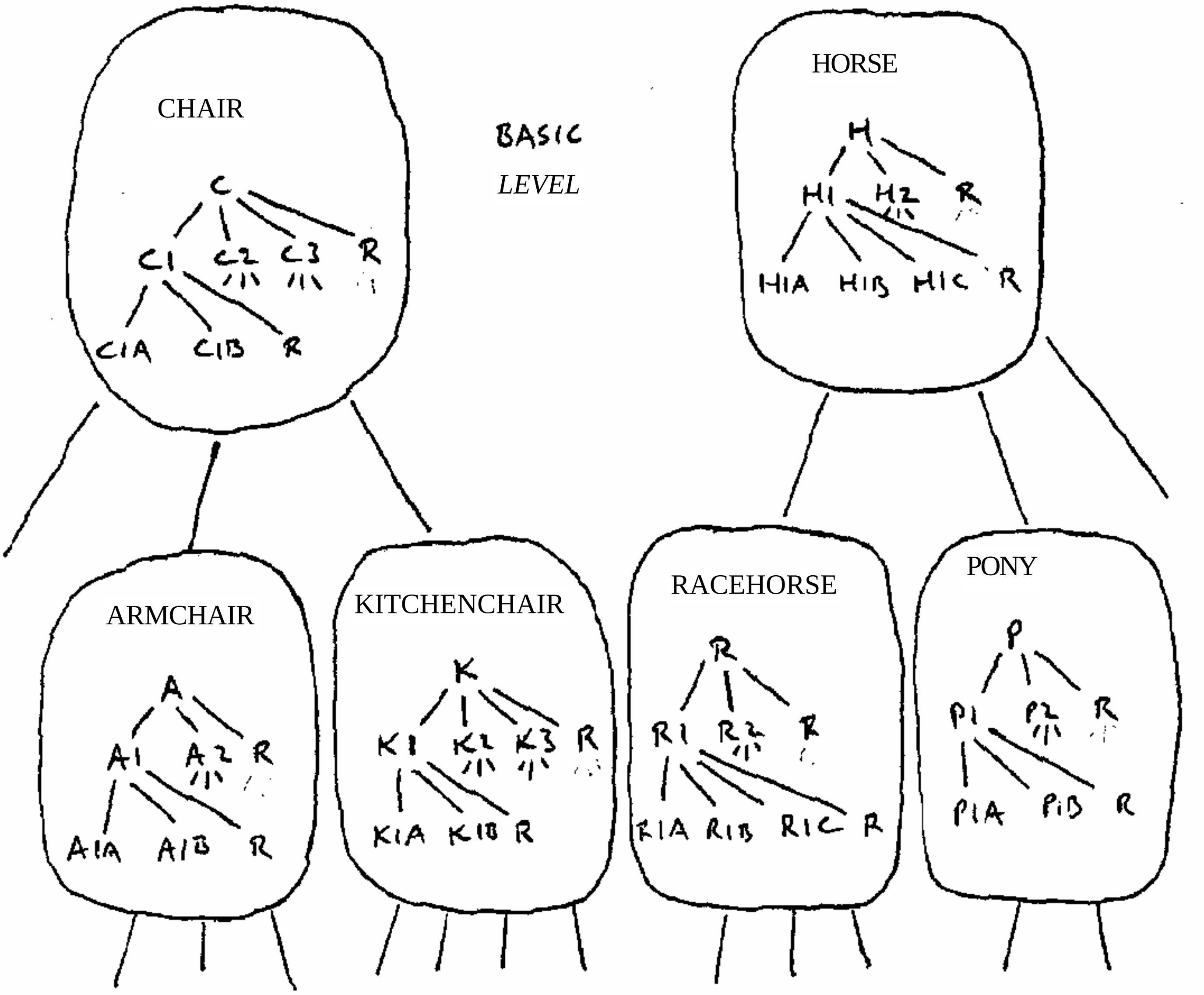
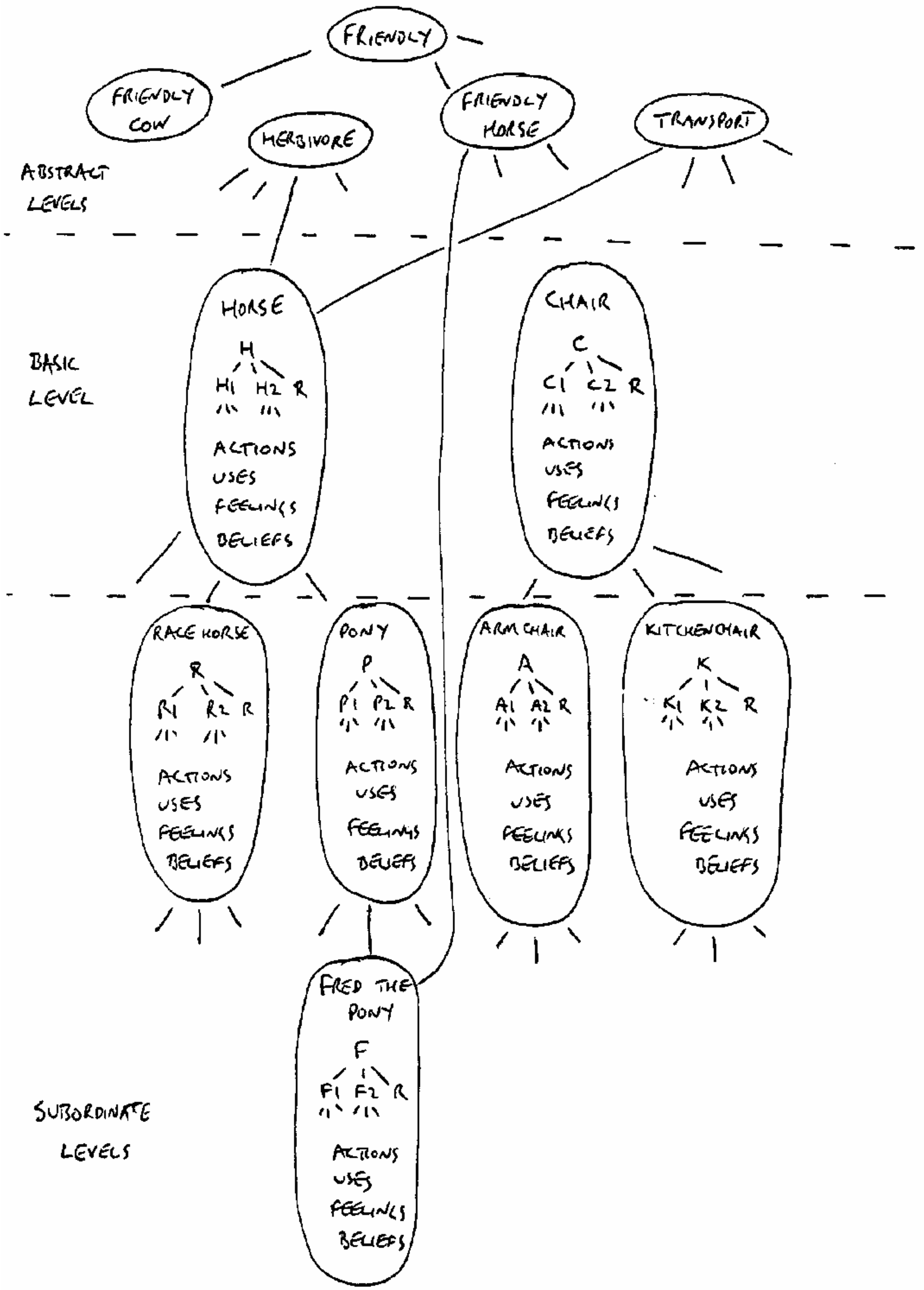

HOLISTIC
OBJECT
DESCRIPTION
PRIMARY
SUBPARTS:
PART 1
DESCRIPTION
PART 2
DESCRIPTION
RECATIVE
POSTIONS
SECONDARY
SUBPARTS:
PART 1A
DESCRIPTION
PART 1B
DESCRIPTION
RELATWE
POSITIONS

FIG 2.



SUBORDINATE LEVELS

FIG 3



FRIENDLY
FRIENDLY COW
HERBIVORE
FRIENDLY HORSE
TRANSPORT
ABSTRACT LEVELS
HORSE
H
H1 H2 R
ACTIONS
USES
FEELINGS
BELIEFS
CHAIR
C
C1 C2 R
ACTIONS
USES
FEELINGS
BELIEFS
BASIC LEVEL
RACE HORSE
R
R1 R2 R
ACTIONS
USES
FEELINGS
BELIEFS
PONY
P
P1 P2 R
ACTIONS
USES
FEELINGS
BELIEFS
ARM CHAIR
A
A1 A2 R
ACTIONS
USES
FEELINGS
BELIEFS
KITCHEN CHAIR
K
K1 K2 R
ACTIONS
USES
FEELINGS
BELIEFS
FRED THE PONY
F
F1 F2 R
ACTIONS
USES
FEELINGS
BELIEFS
SUBORDINATE LEVELS

FIG 4