Abstractional and Associative Processes in Concept Learning: A Simulation of pigeon data.

Helena Matute
Departamento Psicología Básica
Universidad de Deusto
Apartado 1; 48080 Bilbao
Spain
Email: matute@deusto.es

Eugenio Alberdi
Computing Science Department
University of Aberdeen.-King's College
Old Aberdeen, AB9 2UB
Scotland, U.K.
Email: eugenio@csd.abdn.ac.uk

Abstract

Symbolic and associative theories have been claimed to be able to account for concept learning from examples. Given that there seems to be enough empirical evidence supporting both claims, we have tried to integrate associative and symbolic formulations into a single computational model that abstracts information from empirical data at the same time that it takes into account the strength with which each hypothesis is associated with reward. The model is tested in a simulation of pigeon data in a fuzzy concept learning task, where only a few abstractions are stored in representation of all the training patterns and strengthed or weakened depending on their predictive value.

Introduction

In concept learning from examples, subjects are required to incrementally be able to describe the relevant characteristics of a concept and to correctly classify new instances as either members or nonmembers of the category. For example, after seeing many instances of tall and short people, a child may come up with a useful --though fuzzy-- description of a tall person. Similar experiments have been conducted with pigeons (Pearce, 1988; 1989), and in some sense, the task to which the pigeons were exposed was more complex than that in our example, since the tall and short instances presented to the pigeons were not single objects (like the tall or short people in our example), but rather, each exemplar was a tall or short group of bars.

In general, it can be said that if subjects are exposed to several exemplars of a concept (S+), and several negative exemplars (S-), they will eventually be able to discriminate the relevant characteristics of the concept and to respond differently to members and nonmembers of the category. This result has been widely observed, and reported either as concept or category learning, in both animal and human cognitive research, or as discrimination learning in conditioning experiments conducted mostly with animals (Estes, 1985; Medin & Schaffer, 1978; Pearce, 1989).

At the theoretical level, however, there is no general agreement about the internal processes involved, or even as to whether a single learning process can account for results obtained in concept learning and conditioning experiments. Research in the animal and human traditions has been conducted separately for many years and many would view animal learning as a purely associative mechanism which is much simpler than human cognitive learning (see Catania, 1985). On the other hand, several models have been recently proposed in the attempt to offer a unified view of conditioning and category learning, assuming that the same type of learning takes place when a human is learning concepts from examples, as when an animal is

---

This research was conducted in the context of a joint research project of Deusto University and Labein laboratories (Bilbao) and was carried out while the second author was at Labein supported by a grant from the Engineering School at Bilbao. The research was also partially supported by the Trade and Industry Department of the Basque Government.

learning to respond to positive instances and to avoid responding to negative ones. But while some explore the role of associative learning processess, traditionally studied in conditioning research, when accounting for concept learning tasks (Gluck & Bower, 1988; Pearce, 1989), others prefer to emphasize the symbolic components of conditioning and concept learning, arguing that they can not be reduced to associative learning (Holland et al., 1986; Holyoak, Koh & Nisbett, 1989; Waldmann & Holyoak; 1990).

The dispute is not new. Associative (Spence 1936) and hypothesis testing (Krechevsky 1932) theories have been proposed to account for discrimination learning in animals; studies of human conditioning have been plagued by the controversy between purely associative mechanisms versus awareness, hypothesis testing and symbolic representation of the contingencies (see Boakes, 1989; Davey, 1987). Similarly, two main families of theories of concept learning have been traditionally distinguished. The first theory was the associative account proposed initially by Hull in 1920 and further developed in greater depth by Hull (1943) and Spence (1936). This theory postulated the existence of similar associative mechanisms in animal and human learning, but it was abandoned by most psychologists since the publication of the book by Bruner, Goodnow and Austin (1956), who viewed concept learning in terms of hypothesis testing and emphasized the symbolic aspects of learning and representation. Although Hull had conducted experiments which favored an associative interpretation of concept learning, experiments conducted thereafter by Levine (1975) and others, supported the symbolic account of Bruner, Goodnow and Austin who viewed subjects as hypotheses generators and testers. More recent research has shown that this is not a complete view either, and that most experiments in the hypothesis testing tradition were using well-defined concepts with an all-or-none structure which can be defined by necessary and sufficient conditions, and that this does not correspond to natural concepts which are usually fuzzy, ill-defined and with a graded structure of more typical and less typical exemplars (Rosch, 1978). On the other hand, learning of natural concepts has also been reported in animals (see Herrnstein 1984 for a review) and associative theories have been claimed, once again, to account for the learning of ill-defined concepts (Gluck & Bower, 1988; Pearce, 1989).

Given that enough empirical evidence seems to support each of the above claims, we have tried to integrate associative and symbolic formulations into a single model, called IKASLE[1], which we have implemented in LISP code and tested successfully in simulations of human concept learning during problem solving (Alberdi & Matute, 1991). In order to explore the generality of our model, here we present a simulation of pigeon behavior during concept learning as reported by Pearce (1988), where pigeons were exposed to a fuzzy discrimination task of compound stimuli with overlapping features.

Pearce's Data

Pearce (1988; 1989) reported several experiments conducted with pigeons that were exposed to a series of compound stimuli which were exemplars of the "tall" and "short" categories. Each stimulus was composed of three colored bars against a blue background. In the short category, the mean height of each bar was 3 units (+-2) and the sum of the heights of the three bars was 9 units. For instance, the pattern 3-5-1 is an example of this category (the numbers refer to the heights of each of the three bars). In the tall category, the mean height of each bar was 5 units (+-2) and the sum of their heights was 15 units. An example of this category is the stimulus 7-3-5. There were 36 compound stimuli; 18 exemplars of each category.

In the first experiment (Pearce, 1988), pigeons were randomly allocated to two groups. For group "Category", the short patterns were consistently reinforced in an autoshaping paradigm whereas the tall patterns were never reinforced. For group "Random", half of the tall and half of the short patterns were followed by food. As expected, subjects in group "Category" learned to discriminate between both types of patterns whereas subjects in group "Random" did not show a discriminative behavior.

In a subsequent test phase, Pearce presented new stimuli which were not used during acquisition. The

[1] IKASLE means learner in Basque and stands for "Incremental, Knowledge-independent, Associative and Symbolic Learning from Examples".

stimuli 3-3-3 and 5-5-5 represented the respective means of the short and tall categories. In the test task, however, pigeons in group "Category" showed a greater excitation toward the 1-1-1 than to the 3-3-3 stimulus, and a greater inhibition to the 7-7-7 than to the 5-5-5 stimulus. This "shift of the peak" (Hanson, 1959) was replicated in similar concept learning experiments conducted thereafter (Pearce, 1989) and interpreted in terms of the interaction between the excitatory and inhibitory associative gradients (Spence, 1936) that generalize to similar stimuli from the exemplars stored during the learning phase.

The Elements of the Association

Accepting an associative view of concept learning requires, as Pearce noted, specifying the elements of the association. Several alternative explanations for the above results were discussed by Pearce (1888; 1989), including the association of reinforcement with a single feature (area of blue background) which is constant for all members of a category, and the formation of a prototype (as central tendency of a category), both of which did not seem to be supported by the experimental data. In Pearce's view, two alternative explanations could account for the data: associations appear to be formed, "either between reinforcement and individual *elements* of the patterns, or between reinforcement and separate *configurations* of the elements that represent the different training patterns" (1989, p. 405, italics added). Finally, Pearce argues for an exemplar view of concept learning (Medin & Schaffer, 1978), suggesting that the pigeons remember each pattern and its significance (Pearce, 1988; 1989).

Although it is difficult to empirically determine the elements of the associations, in the simulation described below, we show that, in principle, an associative-symbolic approach, as implemented in IKASLE, is also able to account for the data. By "symbolic" representation we do not necessarily mean that pigeons share the "human" ability to encode relational descriptions (see Pearce, 1988 for data suggesting that pigeons are different from humans in this respect), but rather, that associations can be formed between reinforcement and *abstractions* of the training patterns, instead of between reinforcement and representations of the individual training patterns. In our view, the advantage of this approach is that it allows a more economical treatment of memory when many training patterns are used (Herrnstein 1984), and of generalization gradients, since only a few abstractions are stored and weighted in representation of all the training patterns. Below we present a brief outline of IKASLE (see Alberdi & Matute, 1991 for more details) and its results in the simulation of the above data.

IKASLE

IKASLE is an associative-symbolic computational model of learning which preserves a symbolic representation of events and hypotheses as postulated by cognitive theories while at the same time, in order to strengthen or weaken the alternative hypotheses that the system is forming while learning, it makes use of the associative capabilities demonstrated in animals and humans.

The information provided by positive and negative stimuli --or exemplars-- is summarized in two sets of hypotheses (positive and negative). Hypotheses are abstractions from empirical data and are formed through a generalization process (Michalski, 1983).

Old hypotheses are not abandoned upon the creation of a new one. Instead, the process of hypothesis testing is made more flexible and adaptive by taking into account the predictiveness of each hypothesis, or, in other words, the strength with which each hypothesis becomes associated with reward. IKASLE deals with the assignment and revision of associative strengths implementing the Shanks and Dickinson (1987) adaptation of the Rescorla-Wagner (1972) model of conditioning. Hypotheses are reinforced if they correctly predict the outcome of future trials and lose strength if their prediction is incorrect.

In this way the system is able to cope with inconsistent or noisy data, --or imperfect correlations between events-- as well as with concept drift or changes over time (Schlimmer & Granger, 1986). Whereas acquired concepts are mutable and flexible during the earlier phases of learning, resistance to extinction is increased as the hypotheses generated acquire enough associative strength. Thus, dealing with concept drift will be easier for concepts not yet fully acquired than for concepts already established through enough empirical data supporting them --and hence, with high associative strength. On the other hand, complete lack of correlation between events (such as the treatment of group "Random") can not lead to discriminative behavior since the descriptions which are being formed can never get enough strength.

One of the collateral effects that we have obtained this way has been the simulation of typicality effects: given that the system generates and tests several hypotheses, and stores each of them with a different associative strength, a hierarchical structure of possible descriptions of the concept is formed, and therefore, typical exemplars of the category are predicted with high accuracy by the best descriptions, whereas atypical exemplars are only predicted with low strength by weak hypotheses.

Results of the Simulation

In our attempt to test IKASLE in quite different learning tasks, the program was not modified for this simulation and the way in which IKASLE learned in this experiment was identical to the way it learned concepts in other domains, such as the game of "mus" (a card game similar to poker; Alberdi & Matute, 1991).

Table 1 clarifies the learning process of IKASLE during the simulation of Pearce's experiment. First, IKASLE did not store every single training exemplar. Rather, it abstracted regular characteristics from positive and negative exemplars, thus forming a set of hypotheses to describe both the short and tall --positive and negative-- categories at the same time that it kept a record of empirical validity of each hypotesis through its associative strength. In the absence of more information, the first positive and negative hypotheses are formed by the first positive and negative instances respectively. If a new instance in the next trial is not covered by the current hypotheses, previous feature intervals in the descriptions are generalized so as to cover this new exemplar. This abstraction summarizes information from empirical data. If it correctly predicts the outcome of future trials, its associative strength is augmented. If it incorrectly covers a given exemplar (or predicts incorrectly the outcome of a given trial), its strength is reduced.

The results obtained by group "Category" in the Test phase of our simulation are shown in Figure 1. The overall probability of responding was greater in Pearce's experiment since the pigeons responded in an autoshaping paradigm and we did not attempt a search for the best parameters, but rather, we simply replicated the values used by Shanks & Dickinson (1988) for changes in the strength of the associations. Although with a lower response rate,

S	H+	H-
(1-3-5)+	(1 3 5) .24	
(4-7-4)-		(4 7 4) .24
(5-1-3)+	((1-5)(1-3)(3-5)) .24	
(2-2-5)+	→ .42; .56	
(4-4-7)-		(4 (4-7) (4-7)) .24
(4-3-2)+	((1-5)(1-3)(2-5)) .24	
(7-3-5)-		((4-7)(3-7)(4-7)) .24
(4-2-3)+		
	(1 3 5) .24	(4 7 4) .24
	((1-5)(1-3)(3-5)) .86	(4 (4-7) (4-7)) .24
	((1-5)(1-3)(2-5)) .42	((4-7)(3-7)(4-7)) .88
	((1-5)(1-4)(2-5)) .56	((3-7)(3-7)(4-7)) .56
	((1-5)(1-5)(2-5)) .42	((3-7)(3-7)(3-7)) .56
	((1-5)(1-5)(1-5)) .56	

Table 1: Summary of the behavior of IKASLE when simulating a hypothetical subject in Pearce's experiment. The upper panel shows some of the stimuli to which the subject is exposed and the process of hypothesis generation and weighting. The lower panel shows the descriptions acquired by this subject at the end of the training phase, along with their associative strengths.

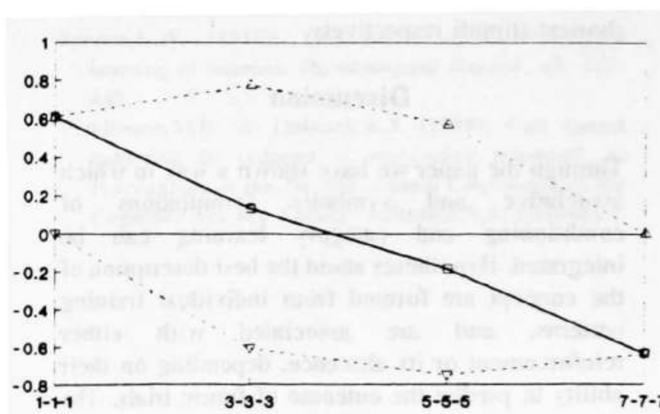

Figure 1: Probability of responding to the test stimuli depending on the difference between the maximum excitatory and inhibitory strength with which the pattern is predicted by both positive and negative descriptions.

the results that we obtained in the simulation showed the same tendency of those of the pigeons reported by Pearce. The shift of the peak is also shown here for group "Category", with a greater probability of responding to the 1-1-1 than to the 3-3-3 stimulus and a greater inhibition to the 7-7-7 than to the 5-5-5 stimulus. Group Random did not show any clear pattern of discrimination.

Peak shift effects ocurred in group "Category" as a consecuence of the interactions between the excitatory and inhibitory strengths of the hypotheses that the system formed abstracting the common feature values for each category. Although not all the possible configurations of the stimuli had the same strength (see Table 1), in general, it can be said that, for the short category, values between 1 and 5 were possible for each of the 3 bars, whereas for the tall category, the possible range of values for each bar was between 3 and 7. Thus, test stimulus 3-3-3 was receiving both excitatory and inhibitory strengths and so, responding to it was less probable than responding to the 1-1-1 pattern which was possible only in the short category. Similarly, for inhibitory or negative stimuli, test stimulus 7-7-7 was clearly a tall --negative-- pattern whereas responding to the 5-5-5 test stimulus was dependent on the interaction of inhibitory and excitatory strengths of both the short and tall categories. The greater inhibitory and excitatory response rates are shown with the tallest and shortest stimuli respectively.

Discussion

Through the paper we have shown a way in which associative and symbolic formulations of conditioning and category learning can be integrated. Hypotheses about the best description of the concept are formed from individual training patterns, and are associated with either reinforcement or its abscence, depending on their ability to predict the outcome of future trials. The advantage of this approach is that it can account for quite different types of data, such as human hypothesis testing in problem solving, pigeon discrimination learning between categories with overlapping features, and typicality effects.

Certainly, we do not claim to account for all of the complexity of concept learning. Selective attention, a greater representational potential, and similarity are just some important factors that should be dealt with.

For instance, for the simulation of the experiment presented above, the introduction of a similarity measurement (Medin & Schaffer, 1978) and a generalization rule for excitatory and inhibitory gradients has not been needed, and IKASLE has been tested in its original version without modification. However, we do not mean to imply that a similarity measurement is not necessary in other tasks. In general, the inclusion of a similarity measurement should permit the system to respond to a new pattern that is not predicted by any of the hypotheses that the system has already generated and tested, as well as to simulate typical conditioning experiments where generalization gradients are reported after training with just one single instance (e.g., Pearce 1989 experiment 2). Note that this approach is not incompatible with the formation and weighting of abstractional descriptions. Given that organisms are exposed to hundreds of training objects and that many of them are certainly very similar to each other, it is probably more adaptive to compute the similarity of a new pattern to a few abstractional representations summarizing information from previous training patterns than to compute the similarity of this new instance to the representations of all the training patterns (Pearce, 1989) or to a random subset of all the individual representations (Medin & Schaffer, 1978).

In this view, and assuming that learning and responding are better understood as separate processes (Miller & Matzel, 1988), subjects can respond to a new pattern not covered by current hypotheses, by means of its similarity to previously stored and weighted descriptions. The information provided by this new instance is then incorporated into previous knowledge, modifying the already existing hypotheses. If, on the other hand, the test patterns are already predicted by the current descriptions, only associative strengths are modified depending on the accuracy of the prediction.

Acknowledgments

Thanks are due to Santi Rementeria and Anselmo DelMoral for their help and suggestions at many stages of the research.

References

Alberdi,E. & Matute,H. (1991): Aprendizaje a partir de ejemplos en un contexto ruidoso de resolución de problemas. [Learning from examples in a noisy context of problem solving]. *Actas del IV Congreso de la Asociación Española para la Inteligencia Artificial.* Madrid: AEPIA.

Boakes,R.A. (1989): How one might find evidence for conditioning in adult humans. In T.Archer & L.G.Nilsson (Eds.): *Aversion, Avoidance and Anxiety*. Hillsdale,N.J.: Erlbaum.

Bruner,J.S., Goodnow,J.J. & Austin,G.A. (1956): *A Study of Thinking*. New York: Wiley.

Catania,A.C. (1985): The two Psychologies of learning: Blind alleys and nonsense syllables. In S.Koch & D.E.Leary (Eds.): *A Century of Psychology as Science*. New York: McGraw-Hill.

Davey,G. (Ed) (1987): *Cognitive Processes and Pavlovian Conditioning in Humans.* Chichester: Wiley.

Estes, W.K. (1985): Some common aspects of models for learning and memory in lower animals and men. In L.G.Nilsson & T.Archer (Eds.): *Perspectives on Learning and Memory.* Hillsdale, N.J.: Erlbaum.

Gluck,M.A. & Bower,G.H. (1988): From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117*. 227-247.

Hanson,H.M. (1959): Effects of discrimination training on stimulus generalization. *Journal of Experimental Psychology, 58*, 321-334.

Herrnstein,R.J. (1984): Objects, categories, and discriminative stimuli. In H.T.Roitblat, T.G.Bever, & H.S.Terrace (Eds.): *Animal Cognition.* Hillsdale,N.J.: Erlbaum.

Holland,J.H., Holyoak,K.J., Nisbett,R.E. & Thagard,P.R. (1986): *Induction: Processes of Inference. Learning and Discovery.* Cambridge, MA: MIT press.

Holyoak,K.J., Koh,K. & Nisbett,R.E. (1989): A theory of conditioning: Inductive learning within rule-based default hierarchies. *Psychological Review, 96*. 315-340.

Hull,C.L. (1920): Quantitative aspects of the evolution of concepts. *Psychological Monographs, 28*. (whole No. 123).

Hull,C.L. (1943): *Principles of Behavior.* New York: Appleton.

Krechevsky,I. (1932): "Hypotheses" in rats. *Psychological Review, 39*. 516-532.

Levine,M (1975): *A Cognitive Theory of Learning.* Hillsdale,N.J.: Erlbaum.

Medin,D.L. & Schaffer,M.M. (1978): Context theory of classification learning. *Psychological Review, 85*. 207-238.

Michalski,R.S. (1983): A theory and methodology of inductive learning. In *Machine Learning: An Artificial Intelligence Approach, vol.1.* Los Altos,CA: Morgan Kaufmann.

Miller,R.R. & Matzel,L.D. (1988): The comparator hypothesis: A response rule for the expression of associations. In G.H. Bower (Ed): *The Psychology of Learning and Motivation, vol.22.* New York: Academic.

Pearce,J.M. (1988): Stimulus generalization and the acquisition of categories by pigeons. In Weiskrantz (Ed.): *Thought without language.* Oxford: Clarendon.

Pearce,J.M. (1989): The acquisition of an artificial category by pigeons. *Quarterly Journal of Experimental Psychology, 41B*. 381-406.

Rescorla,R.A. & Wagner,A.R. (1972): A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A.H. Black & W.F. Prokasy (Eds.): *Classical Conditioning II: Current Research and Theory.* New York: Appleton.

Rosch,E. (1978): The principles of categorization. In E.Rosch & B.B. Lloyd (Eds.): *Cognition and Categorization.* Hillsdale,N.J.: Erlbaum.

Schlimmer,J.C & Granger,R.H. (1986): Beyond incremental processing: Tracking concept drift. In *Proceedings of the National Conference on Artificial Intelligence.* Los Altos,CA: Morgan Kaufmann.

Shanks,D.R. & Dickinson,A. (1987): Associative accounts of causality judgment. In G.H. Bower (Ed.): *The Psychology of Learning and Motivation, vol.21.* New York: Academic.

Shanks,D.R. & Dickinson,A. (1988): The role of selective attribution in causality judgment. In D.J.Hilton (Ed.): *Contemporary Science and Natural Explanation.* Sussex: Harvester.

Spence,K.W. (1936): The nature of discrimination learning in animals. *Psychological Review, 43*. 427-449.

Waldmann,M.R. & Holyoak,K.J. (1990): Can causal induction be reduced to associative learning?. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society.* Hillsdale,N.J.: Erlbaum.