CATEGORIZATION AND ADAPTIVE BEHAVIOR: THE ROLE OF ASSOCIATIVE PROCESSES IN SYMBOLIC CONCEPT LEARNING[1]

E. Alberdi[1] and H. Matute[2]

1 Computing Science Department, King's College, University of Aberdeen
Old Aberdeen, AB9 2UE, Scotland (UK)
Email: eugenio@csd.abdn.ac.uk
2 Dpto. Psicología, Universidad de Deusto
Apartado 1, 48080 Bilbao, Spain
Email: matute_helena@euskom.spritel.es

1. Introduction

1.1. Associationism and concept learning

A new tendency is emerging in cognitive science that pays attention to the possible commonalities between elementary processes of learning (such as classical or instrumental conditioning) and more complex inductive processes (like categorization or causality judgements) (see, e.g., Estes, 1985; Gluck & Bower, 1988; Holland et al., 1986; Pearce, 1988; Shanks & Dickinson, 1987; Schlimmer & Granger, 1986a,b; Waldmann and Holyoak, 1990).

For example, it has been suggested that complex kinds of Pavlovian learning, like compound conditioning (Schlimmer & Granger, 1986b, Gluck et al., 1989) where subjects are conditioned to Boolean combinations of stimuli, might be explained by the same learning principles as concept acquisition (Schlimmer & Granger, 1986a). Both types of learning could be viewed as a process of covariation detection where an organism (animal or human) detects the conditions (either compound stimuli or features expressed in a concept description) under which a determinate response (whether Pavlovian or classificatory) must be given. A similar interpretation, in which learning is viewed as a process of adaptation to the changing configurations of the environment, could be applied to causal induction (Shanks & Dickinson, 1987).

As a consequence of this tendency, associative theories of learning, that traditionally had been postulated to account for conditioning processes, are now proposed to explain human categorization (Gluck & Bower, 1988). This approach can be understood as a consequence of the renewed appeal of connectionist (or neural) networks (Rumelhart &

---

1. The research reported in this paper was carried out while the first author worked at the AI Department of Labein Laboratories in Bilbao (Spain), supported by a grant from the Engineering School from Alava, Guipuzcoa and Vizcaya. This research has also been funded, in part, by Deusto University (Bilbao), Labein and the Department of Trade and Industry of the Basque Government. We would like to thank Santi Rementeria for providing discussions and helpful advice on the topics discussed in this paper, and, likewise, Derek Sleeman for his useful comments and suggestions to earlier drafts of the paper, and Norman Wetherick for his helping insights.

In Proceedings of The European Conference on Cognitive Science. Le Chesnay, France: INRIA, 1995 (pp. 39-46)

McClelland, 1986) as plausible models for cognitive simulation and contrasts with the symbolic tradition that explained concept formation as a process of hypothesis generation and testing (Bruner et al., 1956). Both connectionist models and associative theories have proved successful when accounting for many of the adaptive features of learning (see, e.g., Grossberg, 1988; Rescorla, 1985), a process that these models interpret primarily in terms of the revision of multiple associative or connective strengths between elements of diverse complexity. Significantly, much of the computer modeling of conditioning phenomena and their relationship to higher-level cognitive processes is made using neural or connectionist networks (e.g., Gluck & Bower, 1988; Gluck et al., 1989; Henry, 1986; Sutton & Barto, 1981;).

In contrast with this tendency, Waldmann and Holyoak (1990) have recently reported empirical evidence that suggests the inadequacy of associative accounts to explain integrally the learning processes that take place in human induction. They view classical conditioning as a learning process characteristic of lower animals which differs in important ways from human categorization tasks. In particular, they propose that to account properly for causal induction, some mentalistic constructs need to be postulated, such as abstract causal models, that describe information about relations between the input cues during learning.

Although Waldmann and Holyoak's claim that induction cannot be reduced to associative learning is plausible, there are reasons to believe that associative processes do play an important role in higher-order types of learning. We suggest that, even if purely associative accounts do not explain all the complexities of categorization, elementary processes of learning should be considered, together with other explanatory constructs. On the one hand, conditioning is not contrived to lower animal species: it has also been detected in people (see, e.g., Davey, 1987) and constitutes an important aspect of human adaptive behavior. Let us note furthermore that there is early empirical evidence that supports an associative interpretation of concept learning (Hull, 1920). On the other hand, the symbolic simulation of inductive concept acquisition, which deals with the creation and revision of abstract descriptions from positive and negative examples (see Dietterich and Michalski, 1983), lacks important adaptive features of learning emphasized by associative accounts. For example, one of the traditional problems for symbolic machine learning methods, has been their inability of learning in an incremental fashion, specially when dealing with 'noise' (i.e., inconsistent or imperfect data, or, in other terms, lack of perfect correlation between events caused by lack of information, problems in the communication or other uncertainty factors).

Our particular claim is that a plausible alternative way to understand adaptive and efficient classificatory behavior implies the combination of associative mechanisms with abstract representations of events. Let us remark as a tendency in accordance with this approach the growing emphasis given recently to hybrid symbolic-connectionist models of cognition (e.g., Hall & Romaniuk, 1990; Lange et al., 1989; Lee et al., 1989; Rose & Bellew, 1989; Touretzky, 1986).

1.2. Associationism and Weighting Mechanisms in Machine Learning

Weights, or numeric values attached to hypotheses generated symbolically, are used in inductive machine learning in order to reflect information about the number of examples correctly or incorrectly classified by a learning system (Langley, 1987). The use of weighting methods in symbolic learning from examples can allow the learner to revise its hypotheses incrementally and cope efficiently with noise.

Traditional approaches to weighting in Machine Learning have included simple counting of instances (as in Michalski et al., 1986) and probabilistic or Bayesian treatment of data as involved in, for instance, inductive decision trees (Quinlan, 1986) and different concept formation methods (Anderson, 1990; Gennari et al., 1989).

Another interesting example of the use of a Bayesian approach is the system STAGGER (Schlimmer and Granger 1986a,b), in which contributions from the study of conditioning phenomena are taken into account. STAGGER's weighting mechanisms are based on contingency theory (Rescorla, 1968). This theory explains conditioning in terms of a Bayesian computing of the events perceived by the learner and the relationship among them. Although Rescorla's contingency model was very influential, an alternative account was presented some years later by Rescorla and Wagner (1972) with the purpose of overcoming some of the limitations of contingency theory. As Papini and Bitterman (1990) have suggested, sophisticated probabilistic or Bayesian computations are unlikely to characterize plausibly the way organisms organize their experience in order to respond adaptively to their environment. Rescorla and Wagner's (1972) model represents an associative account of the incremental aspects of learning and of the ability of organisms to deal with variable environments.

In the following sections we describe the performance of a concept learning algorithm, IKASLE[2], which incorporates some weighting mechanisms based on a recent adaptation of Rescorla-Wagner's associative model. The combination of symbolic and associative procedures allows the system to simulate two different learning tasks that involve respectively human and animal categorization behavior.

2. IKASLE' s Learning Procedures

IKASLE is a data-driven bottom-up learning algorithm capable of incrementally creating conjunctive descriptions, consisting of sets of attribute-value pairs, from positive and negative examples. The system has been designed to progressively revise its hypotheses and recover the most useful and predictive ones whenever inconsistencies in the data cause an impasse in its performance or when the concept changes over time. This will be achieved through the three procedures we describe below: conservation of earlier hypotheses, credit assignment and the creation of negative concepts.

2.1 Conservation of Earlier Versions of the Concept

In contrast to hill-climbing approaches in concept learning (see Gennari et al., 1989), where only one hypothesis is conserved at each learning trial, IKASLE characterizes a concept as a set of hypotheses which is called *H*. The set *H* consists of two subsets of hypotheses: H_0 and H_1. H_1 contains the hypotheses which represent the current representation of the concept and which are used by the system to classify a new instance. H_1 is updated whenever one of its hypotheses does not match a positive instance. Such a hypothesis (considered too specific) is replaced in H_1 by its most specific generalization(s) and is transferred to H_0. So H_0 stores earlier reliable versions of the

2. IKASLE stands for Incremental Knowledge-independent Associative and Symbolic Learning from Examples and means "learner" or "student" in Basque.(Alberdi & Matute, 1991)

concept. Whenever the presence of inconsistencies in the data leads the system to a dead-end, and a backtracking is imposed, the system searches in H_0 for earlier hypotheses which match the data. IKASLE is able to detect if a hypothesis is useful or reliable, through the numeric values which are attached to every description in H and are assigned through the weighting procedures explained next.

2.2 Weighting Procedures

IKASLE's weighting procedures simulate Dickinson and Shank's (1985) associative theory of human causality judgment. This theory accounts for the learning processes by which people detect the influence of a given event, the putative cause, in predicting the appearance of an outcome, the effect.

In IKASLE, weights are viewed as associative strengths. IKASLE deals with the assignment and revision of weights, implementing the acquisition function with which Dickinson and Shanks explain the increment or decrement of the associative strengths of the events as a function of the presence or absence of the outcome in a training trial[3].

When a description predicts correctly a positive instance (i.e., the description matches the example) its weight is augmented: it will be said that the hypothesis has been reinforced. If a counterexample is incorrectly matched by a hypothesis, the hypothesis loses associative strength: the hypothesis is said to be inhibited.

A description in H with a high strength is a hypothesis which has been confirmed by many examples and has received little negative evidence. The higher the weight of a hypothesis, the higher its predictive power and usefulness. Whenever the weight of a description is lower than a given threshold, the description is removed from H (either from H_0 or H_1). This hypothesis is supposed to be useless or with little predictive value.

2.3 The Negative Concept

IKASLE keeps track of the negative evidence (counterexamples, nonreinforced stimuli) basically through two mechanisms: the inhibitory processes outlined above and, secondly, the creation of what we call the negative concept. The main purpose of the negative concept is to prevent the overgeneralization from positive instances. IKASLE does not specialize its hypotheses.

Created simultaneously with H and through the same learning procedures (conservation of earlier abstractions and weighting), the negative concept (hereafter H^-) contains two sets of hypotheses, H_1^- and H_0^-, obtained from the generalization of negative examples. A positive example of the concept (which was considered above as a positive instance of H) is considered at the same time by IKASLE as a negative instance of H^-. A counterexample (a negative instance of H) is considered as a positive instance of H^-.

If a positive example of the concept is matched by some of the hypotheses in H^-, the

3. Dickinson and Shanks explain changes in associative strength by the standard linear operator equation: $\Delta V = \alpha\beta\ (\lambda - V)$, where ΔV represents the changes in the associative strength of an event, α is a learning rate parameter associated with the putative cause, β is an equivalent parameter for the outcome, and λ is the asymptote of the associative strength. The asymptote is assumed to be 1 if the outcome occurs, and 0 in its absence. V is the current strength of an event.

system recognizes that a similar instance has been considered negative in the past. Therefore the hypotheses in H_1 that do not match the instance will not be generalized. Likewise a positive example for H^- (i.e., a counterexample of the concept) does not cause generalizations in H_1^- if is matched by some hypothesis in H. In any case, the weights of any of the hypotheses in H or H^- that do match the new example (positive or negative) are always altered (increased or decreased) following the logic of the weighting mechanisms described above.

3. Testing of IKASLE

Two different domains were used in order to test, on the one hand, the validity of the combination of associative mechanisms and symbolic processing in IKASLE and, on the other, the ability of the system to simulate different types of learning phenomena. One of the testing domains is a human problem solving task: the card game *mus*, a game similar to poker. The other is a categorization task executed by pigeons in experiments where J.M. Pearce (1988) studied concept learning as a result of the generalization and discrimination processes detected in conditioning. In both domains, the concepts to be learned are not defined by explicit boundaries, and incrementality, uncertainty and noise play an important role.

3.1 Card Game *Mus*

IKASLE simulated the acquisition of one of the most important kinds of information required for playing *mus* successfully: the knowledge or concepts that define the configurations of cards which are good enough to risk a bet (see Alberdi & Matute, 1991). The instances accepted by the system were combinations of four cards; once each hand was over, the system compared "its cards" with the opponent's cards. If a card combination gave successful results in the game, it was considered a positive instance, otherwise it was considered a counterexample.

The system based its betting decisions on the acquired concepts. Whenever IKASLE found that its current cards were matched by the descriptions held in H_1, it decided to bid or to accept a bet. If, however, the cards were matched by H_1^-, it decided to pass or to reject a bet. When the cards were not matched by either of them, IKASLE made its decisions at random.

In order to verify the usefulness of the weighting mechanisms implemented in the system, IKASLE's performance was compared with the performance of another program that possessed the same learning procedures as IKASLE except the weighting mechanisms. Both programs played against an opponent which made all of its decisions at random, and both possessed at the beginning of each experiment the same initial knowledge. It can be observed in Figure 1 that although the profiles of IKASLE and of the no-weighting program are similar, the behavior of IKASLE is remarkably better, almost reaching 90% of correct decisions. Furthermore the results of these two versions in which some kind of learning is performed contrast with the reference curve at the bottom of the figure which reflects the purely random results of a system that does not acquire new knowledge (no-learning).

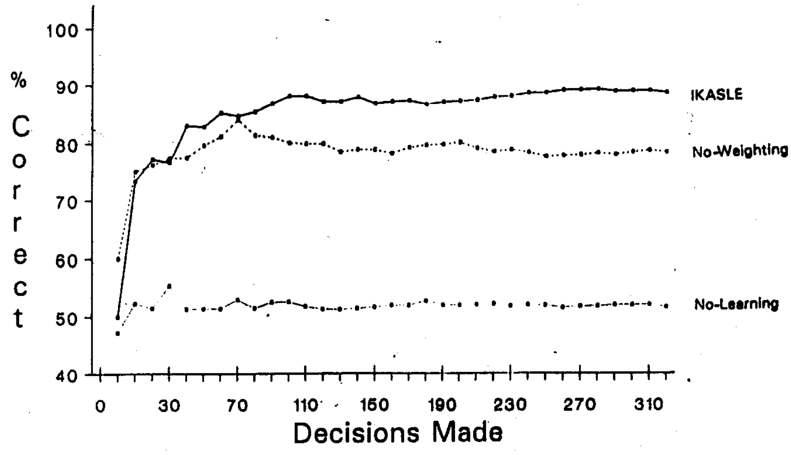

Figure 1: Evolution of the correct classification percentages as a function of the total of decisions made in *mus* game by: i) IKASLE, ii) a version of the algorithm without weighting mechanisms (no-weighting) and iii) a program that does not learn (no-learning) and makes all of its decisions at random. (From Alberdi & Matute, 1991).

3.2 Simulation of Pearce's experiments.

In the experiments described by Pearce (1988), the pigeons were exposed to a number of compound stimuli that were called "tall" and "short". Each stimulus consisted of three colored bars. The categories were defined as follows: in the "short" category, the mean height of each bar was 3 units (+/- 2) and the sum of the heights was 9 units. In the "tall" category, the average height was 5 units (+/- 2) and the sum of the heights was 15. Pecking responses to the stimuli belonging to the "short" category were reinforced and responses belonging to the "tall" category were not reinforced.

In the test phase, Pearce presented new stimuli not used during the acquisition phase and whose patterns consisted of three bars of the same size: e.g., 1-1-1, 3-3-3, 5-5-5, 7-7-7. The patterns 3-3-3 and 5-5-5 represent the averages of the "short" and the "tall" categories respectively. However during the test, the excitation (i.e., number of key pecks) was higher for 1-1-1 than for 3-3-3; and likewise the inhibition was higher for 7-7-7 than for 5-5-5. In summary, a peak shift took place: pigeons did not respond to the average stimuli of each category, but to the most extreme ones (the "shortest" one).

In our simulation of these experiments, IKASLE learned as its target concept the description defining the "short" category, and the description defining the "tall" category was considered the negative concept. During the test phase, the response probability for each stimulus was obtained by subtracting the associative strength of the negative hypothesis that predicted the stimulus from the strength of the positive hypothesis that matched it. The results obtained in this simulation show the same tendency as those obtained by Pearce, and a shift of the peak is also observed (see more details of this simulation in Matute & Alberdi, 1992).

4. Concluding Remarks

The combination of associative mechanisms with higher-order symbolic processing in IKASLE, has permitted the simulation of adaptive aspects of learning, such as the incremental acquisition of reliable and useful concepts in domains where data were imprecise or inconsistent.

We do not claim, certainly, that the associative processes take part in concept learning exactly in the same way as we have implemented them in IKASLE, but the results of our simulation, initially satisfactory, encourage us to continue investigating an approach to symbolic concept acquisition where associative learning processes are considered. Further empirical research should be carried out to support, on the one hand, the plausibility of integrating associative and symbolic processes in concept learning and, on the other hand, to determine: (a) the precise role that associative processes play in categorization or causal induction, and (b) the advantages of the associative model implemented over alternative theories of learning.

5. References

Alberdi, E. & Matute, H. (1991): Aprendizaje incremental a partir de ejemplos en un contexto ruidoso de resolución de problemas. In *Actas del IV Congreso de la Asociación Española para la Inteligencia Artificial (AEPIA-91)*. Madrid: AEPIA.

Anderson, J. R. (1990): *The adaptive character of thought.* Hillsdale, NJ: Lawrence Erlbaum.

Bruner, J. S., Goodnow, J. J. & Austin, G. A. (1956): *A study of thinking*. New York, NY: Wiley.

Davey, G. (Ed.) (1987): *Cognitive process and Pavlovian conditioning in humans.* Chichester: Wiley.

Dickinson, A. & Shanks, D. (1985): Animal conditioning and human causality judgment. In L. G. Nilsson & T. Archer (Eds.) *Perspectives on learning and memory.* Hillsdale, NJ: Erlbaum.

Dietterich, T. G. & Michalski, R. S. (1983): A comparative review of selected methods for learning from examples. In R. S. Michalski, J. G. Carbonell & T. M. Mitchell (Eds.): *Machine learning: an artificial intelligence approach.* Los Altos, CA: Morgan-Kaufmann.

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, NJ: Erlbaum.

Gennari, J. H., Langley, P. & Fisher, D. (1989): Models of incremental concept formation. *Artificial Intelligence, 40,* 11-61.

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

Gluck, M. A., Hee, M. R. & Bower, G. H. (1989): A configural-cue network model of animal and human associative learning. In *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Grossberg, S. (1988): *Neural networks and natural intelligence.* Cambridge, Mass: MIT Press.

Hall, L. O. & Romaniuk, S. G. (1990): A hybrid symbolic-connectionist learning system. In *Proceedings of the Eighth National Conference on Artificial Intelligence.* Cambridge, Mass: MIT Press.

Henry, H. N. (1986): A neural simulation of classical conditioning in aplysia. In *Proceedings of the Eighth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Holland, J. H., Holyoak, K. J., Nissbet, R. E. & Thagard, P. R. (1986): *Induction: processes of Inference, Learning, and Discovery.* Cambridge, Mass: MIT Press.

Hull, C. L. (1920): Quantitative aspects of the evolution of concepts. *Psychological Monographs, Vol. 28 (whole No. 123).* 86 pp.

Lange, T. E., Hodges, J. B., Fuenmayor, M. E. & Belyaev, L. V. (1989): DESCARTES: Development environment for simulating hybrid connectionist architectures. In *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Lee, G. Flowers, M. & Dyer, M. G. (1989): A symbolic/connectionist script applier mechanism. In *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Langley, P. (1987): A general theory of discrimination learning. In D. Klahr, P. Langley & R. Neches (Eds.): *Production system models of learning and development.* Cambridge, Mass: MIT Press.

Matute, H. & Alberdi, E. (1992): Abstractional and associative processes in concept learning. A simulation of pigeons data. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Michalski, R. S., Mozetic, I., Hong, J. & Lavrac, N. (1986): The multipurpose incremental learning system AQ15 and its testing application to three medical domains. In *Proceedings of the fifth National Conference on Artificial Intelligence.* Los altos, CA: Morgan Kaufmann.

Papini, M. R. & Bitterman, M. E. (1990): The role of contingency in classical conditioning. *Psychological Review, 97,* 396-403.

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

Quinlan, J. R. (1986): The effect of noise in concept learning. In R. S. Michalski, J. G. Carbonell & T. M. Mitchell (Eds.): *Machine learning: An artificial intelligence approach.* Los Altos, CA: Morgan Kaufmann.

Rescorla, R. A. (1968): Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology, 66,* 1-5.

Rescorla, R. A.(1985): Associationism in animal learning. In L.G. Nilsson & T. Archer (Eds.): *Perspectives on learning and memory.* Hillsdale, NJ: Erlbaum.

Rescorla, R. A. & Wagner, A. R. (1972): A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.): *Classical conditioning II: Current research and theory.* NY: Appleton.

Rose, D. E. & Belew, R. K. (1989): A case for symbolic-subsymbolic hybrid. In *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Rumelhart, D. E. & McClelland, J. L. (Eds.): (1986): *Parallel Distributed Processing.* Cambridge, Mass: MIT Press.

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

Schlimmer, J. C. & Granger, R. H. (1986b): Simultaneous configural classical conditioning. In *Proceedings of the Eighth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

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

Sutton, R. S. & Barto, A. G. (1981): Toward a modern theory of adaptive networks: expectation and prediction. *Psychological Review, 88,* 135-170.

Touretzky, D. (1986): BoltzCONS: Reconciling connectionism with the recursive nature of stacks and trees. In *Proceedings of the Eighth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

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, NJ: Erlbaum.