*Psychologica Belgica*
1997, 37-1/2, 131-140.

# CATEGORIZATION OF AUTOCORRELATED SEQUENCES

Michaël DUBOIS and Guy LORIES
*Université Catholique de Louvain*

Numerous studies show that human subjects can extract the structure of a set of sequences of discrete stimuli. The dynamic control tasks used by Berry and Broadbent (1984), on the other hand, contain a continuous aspect. Following Lories, Dubois, Gaussin (1997) the present experiment shows that subjects can perceive and reproduce some statistical aspects of the structure of simple continuous time series. Our subjects learn to discriminate positively and negatively autocorrelated data series and transfer this ability to new series. Typicality and variability effects are found that are compatible with a standard categorization capability.

## Introduction

Prediction is an important component of decision making. It plays a role in preparing decisions both in more formal contexts and in everyday life situations. Whether we think of the weather forecast or of predicting variations of the stock exchange, the prediction we are interested in here amounts to estimating at time *t* the value that some numerical value is likely to take at time *t+n*. Although statistical or operations research approaches first come to mind, (Mélard, 1990) and although technical solutions exist, human intervention remains important if only because the figures provided by a computer will have, at some, point to convince a human.

In this perspective, several authors (Bolger & Harvey, 1993; Lawrence & O'Connor, 1992) have started to investigate how human subjects can make predictions using the information contained in a time series. A time series is a sequence of numerical values sampled from, or generated by, a system at regular intervals. Bolger and Harvey (1993) asked their subjects to predict the next 6 values for several simple time series presented graphically. They conclude that their subjects rely on an anchor-and-adjust heuristic. The anchor-and-adjust heuristic attempts to predict a value from a plausible global estimate and to make corrections on that estimate to adapt it to specific features of the situation.

In the same vein, Lories, Dubois and Gaussin (1997) showed that the

Michaël Dubois, Department of Experimental Psychology (ERGO, Unité de psychologie du travail et des organisations); Guy Lories, Department of Experimental Psychology (EXCO, Unité de psychologie cognitive).

Correspondence concerning this article should be addressed to Guy Lories, Department of Psychology, Université catholique de Louvain, Place du Cardinal Mercier 10, 1348 Louvain-la-Neuve. Electronic mail may be sent to lories@exco.ucl.ac.be.

subjects do not do well at minimizing their mean square prediction error when compared with an optimal statistical technique. Actually, when asked to forecast future values of a time-series graphically presented, the subjects do not act as statisticians. A statistical prediction involves minimizing a probable error; the predicted value is chosen in such a way that the difference between predicted and observed values is as small as possible 'on average'. The subjects do not provide the most plausible value in this sense. In the task used by the authors the subjects had to predict the 6 next values that such a series would take given the latest 45 values. In such a case, and considering the series, the proper response would have been to predict a sequence of values slowly returning to the long term mean of the series. What the subjects do instead is to imitate the available data and produce something that *looks* similar to the series presented. To provide a comparison, suppose the subjects were given a sample of 45 individual heights and asked to predict 6 new heights from the same population. The statistically correct strategy is to use the sample mean as prediction for all 6 values; what the subjects did was more similar to providing 6 heights more or less randomly sampled from the same population. In other words, their prediction has more variance than necessary. Now while doing this, the subjects can preserve interesting features of the sample they were exposed to and provide a set of 6 values with a mean and standard deviation close to the mean and standard deviation of the original sample. This is about what was observed.

Although the subjects perform better with some series than others, and although they never make very good predictions (replicating Bolger & Harvey, 1993), they do approximate -and thus, in a sense, perceive- significant aspects of their statistical structure. Lories et al. (1997) showed that some mathematical parameters of the 6-point predictions the subjects produced were a function of the parameters of the to be predicted series. More precisely, Lories et al. (1997) showed that the autocorrelation of the subjects response is a function of the autocorrelation of the original series.

In terms of performance, this result suggests that human subjects can extract regularities not only from a sequence of discrete events or from letter strings (Reber 1989; see also Reber (1993), Cleeremans (1993), Berry & Dienes (1993) for a review) but also from a representation of a continuous serial process. In terms of heuristics it suggests that the subjects are actually applying a representativity heuristics: they try to produce a series that could have been generated by the same underlying process as the original.

To provide an example, suppose we have to predict the next value of a series generated by (1).

$$\mathbf{x}_{(t)} = \phi\, \mathbf{x}_{(t-1)} + \varepsilon_t \qquad \varepsilon_t = \mathrm{NIID}\,(0,\sigma^2) \qquad -1<\phi<1 \qquad (1)$$

Equation 1 was used to generate series similar to the series of Bolger and Harvey (1993) and of Lories et al. (1997)[1]. It defines the value of a random series x at time *t* as a function of the value at time *t-1*. The series is random because $\varepsilon_t$ is a normally distributed random variable, with zero mean. Assuming that the successive values of *x* are independent of each other and of $\varepsilon_t$, we can compute the most appropriate estimator of $\phi$, say $\varphi$. Given this estimator, a statistically reasonable prediction would be:

$$\mathbf{x}_{(t)}=\varphi\, \mathbf{x}_{(t-1)}+E(\phi_t) \tag{2}$$

Because $E(\phi_t)=0$, it reduces to:

$$\mathbf{x}_{(t)}=\varphi\, \mathbf{x}_{(t-1)} \tag{3}$$

It should be noted that $\varepsilon_t$ will usually not be equal, nor even close to zero. As a consequence, the statistical prediction is bound to be erroneous, but it is optimal in the sense that it minimizes an expected error. This is what "prediction" means in this context. Our subjects are not statistically sophisticated and they do not know what model they should use. They cannot make complex computations either. What they can do is draw something that, in a perceptive sense, is similar to what they see. If their notion of similarity preserves the most interesting statistical properties of the series, what they draw will have statistical properties that will be similar to the properties of the original.

This is probably what happened in Lories et al. (1997). Whether the subjects are asked to predict or to imitate the series does not really affect their behavior and, in all cases, they prove more or less capable of producing something that has some of the relevant statistical features. Still, this type of results is only evidence for an adequate perceptive mechanisms that preserves a number of statistical characteristics. It does not show that the subjects have reached a general characterization of the series. One way to demonstrate that they actually can do that, is to prove that they can use their perception of similarities between series to categorize the series, for instance, as positively and negatively autocorrelated. This categorization behavior should also exhibit the usual typicality or similarity effects described in the categorization literature (Medin & Schaffer, 1976; Nosofsky, 1986; Tversky, 1977; etc.).

The following experiment shows that a number of features can apparently be extracted and used effectively to categorize time series generated from equation (1) using different values of $\phi$. The stimuli used are time-series

---

[1]Actually Bolger & Harvey (1993, p. 783) used the equation $X_t = \phi X_{(t-1)} + (1-\phi)\, m + \varepsilon_t$, where m represents the series mean. In our case, we choose to have m = 0. Since $-1 < \phi < 1$ anyway, Bolger and Harvey's equation reduces to equation (1) above.

generated from equation (1). The categories show a typicality gradient (Malt & Smith, 1984; Massaro, 1987; McCloskey & Gluksberg, 1979; Mervis, Catlin, & Rosch, 1975; Nosofsky, 1991; Smith, 1978; Smith, Shoben & Rips, 1974; etc.). The experiment is based on the paradigm of implicit induction described in Fried and Holyoak (1984).

## Experiment

### *Method*

The stimuli were time-series generated from equation (1). Two sets of 48 time-series were generated using $\phi = +0.3$ for category A, and $\phi = -0.3$ for category B. The series were graphed horizontally on a computer screen. The 45 points of the time-series were connected with a black line. The time-series were presented one at a time, each in one 45 point graph.

When the series was displayed on the screen, the subject was asked to identify within 7 seconds the category it came from ('A' or 'B'). The subject pressed one of two keys to answer. When the key was pressed, the computer gave a feedback ('correct' or 'incorrect'). The feedback message was showed for 3 seconds and the next stimulus was presented immediately after that.

The instructions given to the subject were simple. In a first stage (learning), the subject was just asked to categorize time-series in one of two categories: category A and category B. No explanation was given. Nothing was said about how to identify the category. The 96 series were presented, in a random order, repeatedly until a criterion of 10 consecutive successes was met.

In the second stage (transfer), the subjects were asked to categorize a set of new time-series but told that these new series could belong to 3 categories A, B or C, implying that a third category had been added. The subject no longer received any feed-back. The new material was created using more diverse values of the $\phi$ parameter. It went from -0.6 to +0.6 by steps of 0.1 so that 65 series were used (5 series at 0.6; 5 at 0.5, etc.).

The introduction of response 'C' allowed to test the hypothesis that the subject could detect atypical series, series that would not have the expected characteristics. We predicted that the subjects would categorize a series correctly if the series was typical enough for the category. We expected the representation built by the subject to consist in two categories, each defined by a simple prototype around which some dispersion is allowed. When typicality is too low, we expected the series to go into the "other"—category, that is category 'C'. According to these hypotheses, a series with a value of $\phi$ close to zero was expected to be rejected as too a-typical for both categories. On the other hand the subject may also learn to order the series along a continuum and identify a single dimension on which to discriminate between two types of series. In this case, the 'rejections' should not cluster as clearly around the small absolute values of $\phi$.

We also asked the subjects to rate the typicality of the series presented. The subjects were shown a series with the label of the category it came from, and asked to judge how typical it was for that category. The subject provided a rating between 1 (not typical) to 7 (very typical).

The independent variable was the objective typicality of the exemplar series, as estimated by the value of $\phi$ used to build the series. The dependant variables were the classification performance during transfer and the typicality estimated.

The design and the procedure were the same for all subjects. 25 subjects participated to the experiment. At the end of the experiment, we asked the subject to describe the difference(s) between a series coming from category A or B. These comments were noted.

### *Learning Results*

The learning criterion was 10 correctly categorized time-series in a row before the end of the presentation of 192 (2x96) time-series. Six out of 25 subjects never reached the criterion. 19 subjects did. The 19 subjects who managed to master the task took on average 59 trials to reach the criterion. So at the end of this first phase, 19 subjects had learned to discriminate between positively versus negatively autocorrelated series for an absolute autocorrelation value of 0.3.

### *Transfer results*

Performance can be measured as a probability of correct categorization. Categorization is correct when label A is assigned to a series with a positive $\phi$ or when label B is assigned to a series with a negative $\phi$. Label C is never taken as correct.

*Performance and the value of $\phi$.* Figure 1 shows the performance of our subjects (number of correct responses) as a function of $\phi$. The number of successes is the number of correct categorizations across all subjects (itemwise). We computed an average Spearman coefficient (again itemwise, i.e., series by series), between the number of correct responses and the real value of $\phi$. This correlation is significant when $\phi$ is negative ($r= -0,71$, $p<0,01$) but not when $\phi$ is positive ($r= 0,18$, n.s.). Performance is poor and close to random when $\phi$ is close to zero, and the specificity of $\phi$ is reduced.

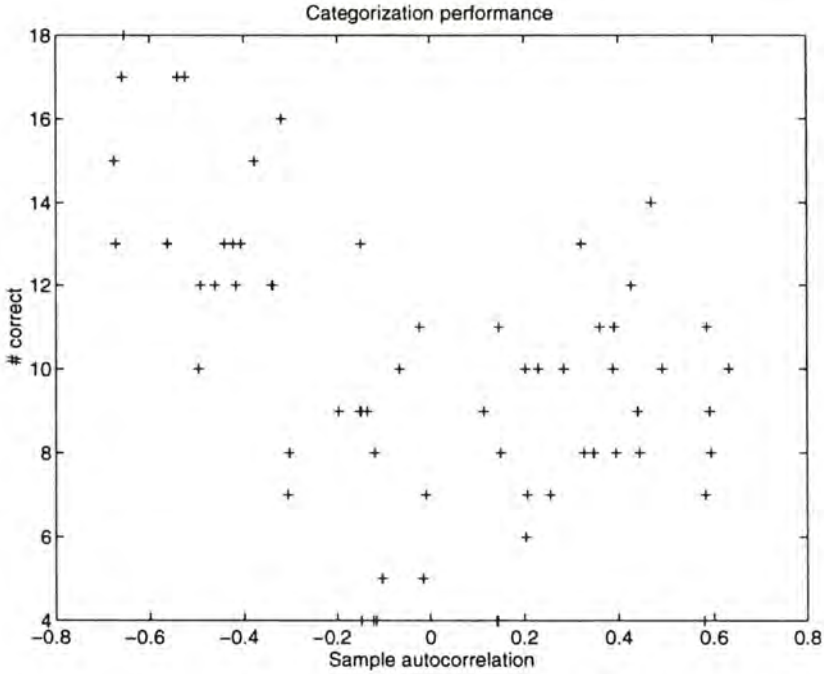


*Figure 1.* Number of successes as a function of the value of $\phi$. Each point represents one time-series.

*Categorisation and the value of $\phi$*: Table 1 shows that negatively autocorrelated series are more often put in the correct category than in the incorrect one (57.95%). On the contrary, positively autocorrelated series are categorized correctly in 43.86% of the cases. They are put about as frequently in category C (40.35%).

Performance is clearly a function of $\phi$. We expect that the higher the value of $\phi$ the more likely it is that the series will receive the A label and vice-versa but we also have to deal with category C. One way to analyze the results globally is to treat C as an intermediate category, and assess the relationship between $\phi$ and the category assignment. We decided to assign a value of 1 to category A, -1 to label B and 0 to label C; next a Goodman-Kruskall gamma coefficient was computed across items for each subject and the coefficients were averaged across subjects. The mean gamma coefficient obtained in this manner is 0.47 (sd=0.28, n=19); this value is significantly larger than 0.

To make sure that the subjects' transfer performance can be attributed unequivocally to what they learned in the learning phase, we had to show that the categorization performance in the transfer phase was not due to prolonged

Table 1
*Distribution of Categorization Decisions*
*as a Function of Sample Autocorrelation Sign (4 missing values)*

| | Series $\phi > 0$ | Series$\phi < 0$ | Total |
|---|---|---|---|
| Cat. A - Series $\phi > 0$ | 275 | 130 | 405 |
| Cat. B - Series $\phi < 0$ | 99 | 350 | 449 |
| Cat. C - Series other | 253 | 124 | 377 |
| Total | 627 | 604 | 1.231 |

learning during the transfer phase. We split the transfer phase into two halves and we analyzed the categorization performance separately for the first and the second half. The gamma was 0.46 (sd=0.29) for the first half and 0.52 (sd=0.33) for the second half. A T-test showed no statistical difference between these values (T=1.121, df=18, p=0.28).

*Typicality rating.* A potentially more sensitive approach rested on the typicality ratings. The mean (subject by subject) correlation between the value of $\phi$ and the typicality judgment is +0,40 (sd=0,18, t= 0,49) when $\phi$ is positive, and -0,47 (sd=0,26, t= -0,42) when $\phi$ is negative. This difference is again significant. The larger the absolute value of $\phi$, the more typical the series is considered, but this is more obvious for negative values of $\phi$ .

## Discussion

The learning phase in this experiment showed that subjects can discriminate the positively from the negatively autocorrelated series. The transfer phase also showed that subjects can generalize what they have learned to series with new values of $\phi$, which is an argument for the psychological reality and generality of the autocorrelation continuum. These results confirm those by Lories et al. (1997), but the ability to make an explicit categorization decision and the transfer results are new.

As for most categorization tasks, typicality effects were to be expected (Rosch, 1978) and they were found. The typicality rating does vary as a function of $\phi$, and the subjects reliably perceive that some exemplars are more representative of their category than others. Likewise, the series are more easily categorized when the values of $\phi$ are more distant from the category border.

While negatively autocorrelated series are usually classified correctly, positively autocorrelated series are often categorized as C. As a consequence, for these series, the association between correct category choice and $\phi$ is not as clear. Because learning involved two categories only, it was enough for the subjects to identify one relevant dimension to make their categorization decision. So, the results are compatible with the idea that subjects use a single dimension to discriminate between the two categories.

The criterion used by the subjects to make this discrimination may translate into a single, possibly perceptive, dimension. If a negative $\phi$ had a few salient perceptive consequences, the subjects may simply learn to recognize the negative autocorrelations from these aspects. By trial and error, we identified at least one reasonably simple perceptive characteristic that may allow for the appropriate categorization. We tabulated the number of jumps made by the series from below to above the mean. It clearly correlates with $\phi$ across our set of stimuli ($r= -0.89$). When the number of jumps is small, the series would be categorized either in the 'positive' category either in the 'other' category almost indifferently.

From the above point of view, the experiment requires that the subject learn to put a verbal label onto something that is, at the beginning, a purely perceptive distinction. This suggests that there is very little that they should be able to tell us about their learning experience. In our post-experimental interviews, the subjects were able to give some information about their behaviour but not much (Dubois, 1996). Some of them did mention the number of jumps across the mean, saying things like 'One category has more oscillations...' but this was about the only explicit comment they made. It is impossible to exactly determine what role this plays in the learning process; in a sense, when the subjects become able to make this kind of comment, learning has already taken place. This is probably related to the nature of the stimuli, which are continuous. When discrete stimuli (letters or digits) are used in a learning task, some verbalization and an explicit hypothesis testing process can be expected. It is more unlikely with graphically presented continuous sequences. In a sense, with such a material, it would even seem necessary for the subject to form some perceptive category before he/she becomes able to put a verbal label on it. Moreover, in this case, the inclusion of a particular sequence in a given category has been shown to be a relative matter, giving raise to a typicality gradient.

For these reasons, the results demonstrate an ability to extract information from the sequence, in a very approximate and possibly implicit manner that may remind us of two paradigms commonly met in the implicit learning literature. One is Reber's paradigm (Cleeremans, 1993; Reber, 1967, 1989, 1990; etc.). It deals with discrete strings of letters generated by a finite state grammar. After having been exposed to grammatical strings, the subjects become able to recognize strings generated by the grammar among distractor strings. In principle, Reber's task allows the subjects to extract verbalizable regularities like 'You should never have a T after two X's' etc. It turns out that this is not what they usually do, which is why this kind of learning has been called implicit. Our task is similar to this paradigm in the sense that a categorisation is required from the subjects; the categorisation as A or B is similar to the categorisation as grammatical or not. Moreover some implicit learning paradigms study the generation of a sequence (Perruchet & Amorim 1992) which is similar to the task of 'continuing' the sequence (Lories et al., 1997). More detailed comparisons could probably be pursued. Unfortunately it is difficult to determine whether learning is implicit in our experiment because there is not much that the subjects could tell about a process that is very perceptive anyway. The important difference, here, is that the stimuli used in our experiment are continuous.

In contrast, the other relevant implicit learning paradigm does use continuous sequences. The process control paradigm of Berry and Broadbent (1984) requires that the subject learn to control a dynamical continuous process by inputting suitable values. In this case also, learning takes place without much verbalization, although it is less surprising because quantitative relationships are difficult to describe verbally anyway. The interesting aspect for our purpose is that the control decisions made by the subjects in Berry and Broadbent's experiments must somehow correspond to the perception of classes of situations. In this sense, they imply the discovery of several categories of situations and of responses appropriate to each. An interesting line of research would be to investigate whether the ability demonstrated by our subjects can play a role in this type of experimental situations by specifying the conditions that must be matched for a given action to be appropriate.

## References

Berry, D.C., & Broadbent, D.E. (1984). On the relationship between task and associated verbalizable knowledge. *Quarterly Journal of Experimental Psychology, 36a,* 209-231.

Berry, D. C., & Dienes, Z. (1993). *Implicit learning: Theoretical and empirical issues.* Hillsdale, NJ: Erlbaum.

Bolger, F., & Harvey, N. (1993). Context-sensitive heuristics in statistical reasoning. *Quarterly Journal of Experimental Psychology, 46A,* 779-811.

Cleeremans, A. (1993). *Mechanisms of implicit learning: Connectionist models of sequence processing.* Cambridge, MA: The MIT Press.

Dubois, M. (1996). *Les fondements de l'anticipation dans le contrôle de processus continu.* Doctoral Dissertation, University of Louvain-la-Neuve, Department of Experimental Psychology.

Fried, L.S., & Holyoak, K.J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory* *and Cognition, 10,* 2, 234-257.

Lawrence, M.J., & O'Connor, M. (1992). Exploring judgmental forecasting. *International Journal of Forecasting, 8,* 15-26.

Lories, G., Dubois, M., & Gaussin, J. (1997). Judgmental forecasting and anticipation in human process control. *Le Travail Humain*, 60, (1), 87-101.

Malt, B.C., & Smith, E.E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior, 23,* 250-269.

Massaro, D. W. (1987). Categorical partition: A fuzzy-logical model of categorization behavior. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 254-283). Cambridge: Cambridge University Press.

McCloskey, M., & Glucksberg, S. (1979). Decision process in verifying category membership statements: Implications for models of semantic memory. *Cognitive Psychology, 11,* 1-37.

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

Mélard, G. (1990). *Méthodes de prévision à court terme.* Bruxelles: Editions de l'Université de Bruxelles.

Mervis, C., Catlin, J., & Rosch, E. (1975). Development of the structure of color names. *Developmental Psychology, 11,* 54-60.

Nosofsky, R.M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39-57.

Nosofsky, R.M. (1991), Typicality in logically-defined categories: exemplar-similarity versus rule instantiation. *Memory and Cognition, 19,* 2, 131-150.

Perruchet, P., & Amorim, M.A. (1992). Conscious knowledge and changes in performance in sequence learning - Evidence against dissociation. *Journal of Experimental Psychology - Learning Memory and Cognition, 18,* 785-800.

Reber, A.S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning & Verbal Behavior, 77,* 317-327.

Reber, A.S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118,* 219-235.

Reber, A. (1990). The primacy of the implicit: A comment on Perruchet and Pacteau. *Journal of Experimental Psychology: General, 119,* 340-342.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive uncounscious.* Oxford: Oxford University Press.

Rosch, E. (1978), Principles of categorization. In E. Rosch & B.B. Lloyd (Eds.), *Cognition and Categorization.* Hillsdale, NJ: Erlbaum.

Smith, E.E. (1978). Theories of semantic memory. In W.K. Estes (Ed.), *Handbook of learning and cognitive processes.* Hillsdale, NJ: Erlbaum.

Smith, E.E., Shoben, E.J., & Rips, L.J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review, 81,* 214-241.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84, 4,* 327-352.

Received April, 1997