British Journal of Psychology (1997), **88**, 441–458 *Printed in Great Britain*

10

Abstraction of covariations in incidental learning and covariation bias

Pierre Perruchet*
Université de Bourgogne, LEAD, Faculté des sciences, 6 Bd Gabriel, 21000 Dijon, France

Chantal Pacteau
Université René-Descartes, Paris

Jorge Gallego
Université Pierre et Marie Curie, Paris

Experiment 1 was devised to distinguish, in a given set of features composing drawn robots, those whose variations were related *a priori* for participants from those whose variations were *a priori* independent. In Expt 2, correlations were experimentally induced between *a priori*-related features for one group of participants (*pre-primed* group), and between *a priori*-independent features for another group (*arbitrary* group), in incidental learning conditions. A subsequent transfer phase revealed that participants' performances were sensitive to experimentally induced correlations in both groups. However, only the performances of the pre-primed group accurately matched the predictions of a statistical model devised by K. Richardson (e.g. Richardson & Carthy, 1990), postulating the acquisition of genuine knowledge of the correlational structure. Participants' sensitivity to arbitrary correlations appeared to be a by-product of the memory of specific study exemplars. These results demand the reinterpretation of some prior experimental evidence for covariation abstraction, and more generally, are consonant with a current view of implicit learning which emphasizes the role of specific prior episodes in complex learning situations.

There is a general consensus that learning about contingency relationships between events is crucial for adaptive behaviour, even if its role in the attribution of causality may be less important that was once believed (e.g. Ahn, Kalish, Medin & Gelman, 1995; White, 1995). This form of learning has been investigated in several sets of laboratory studies, which differ on two main aspects. The first aspect pertains to the nature, intentional or incidental, of the instructions given to participants during the encoding of the material. In some studies, they are explicitly instructed to search for the strength of the association between two (or more) events on which information is provided, either by direct exposure or through condensed information about feature occurrence frequencies. In other studies, people are asked to process

* Requests for reprints (e-mail: perruche@u-bourgogne.Fr).

individual instances without regard for their relationships or, more generally, without attempting to build any form of summary representation of the displayed material. The second aspect is related to the test devised to assess the resulting knowledge. Participants' resulting knowledge about covariations is assessed either by direct rating of the strength of the relation, or through performance in a task specially designed to be sensitive to this form of knowledge, but which does not require any form of explicit retrieval. Most studies involve both intentional learning instructions and direct assessment of covariation knowledge (for a review, see Alloy & Tabachnik, 1984). The present paper is rooted in the opposite approach, which is intended to match natural conditions more closely. In this approach, participants receive incidental instructions during the encoding session, then perform a task which measures covariation knowledge indirectly.

Prior studies involving these conditions paint a contrasting picture. On the one hand, some authors report that individuals were able to abstract covariations in incidental learning conditions, even though these covariations were made difficult to detect due to the inherent complexity of the situation (e.g. Berry & Broadbent, 1988; Kushner, Cleeremans & Reber, 1991; Lewicki, 1986; Richardson & Carthy, 1990; Younger & Cohen, 1985). On the other hand, several other studies have failed to observe incidental covariation abstraction (e.g. Wattenmaker, 1991, 1993). It is worth noting that some initial evidence favouring incidental covariation detection has been reinterpreted in other terms. For instance, Perruchet (1994) proposed and validated an alternative account of the Kushner *et al.* (1991) data, which does not involve the powerful abstractive mechanisms involved in the original interpretation. The present paper is aimed at a similar reappraisal, but concerns the work carried out in K. Richardson's laboratory (Richardson, 1986, 1987; Richardson & Carthy, 1989, 1990), work which represents one of the most cogent demonstrations of incidental covariation learning provided to date.

In Richardson's studies, participants are first asked to study and remember a set of pictures representing robots, houses or schematic faces. These stimuli differ with respect to three (exceptionally, four) variables (for instance, in the case of robots: the length of the arms, the size of the body and the size of the head). Each variable has three possible values, which are typically small, medium and large. Participants are exposed to a set of exemplars which instantiate a moderately high level of covariation between these variables. In a subsequent test phase, participants are shown the full set of instances which may be generated by the exhaustive combination of variable values, and are asked to make recognition (e.g. Is this robot a member of the previously viewed set?) or typicality (Is this robot representative of the previously viewed set?) judgments for each of them.

Participants' scores are compared with the values predicted by a covariation coding model, based on the application of log-linear statistics. The method of computation is described at length in all the Richardson and associates studies, and it will not be reproduced here in order to save space. For the purpose of this paper, it suffices to point out that a log-linear model allows for the simulation of the performance of an individual who would have abstracted the covariations embedded in the study material, and would make recognition or typicality judgments on the sole basis of this relational information. As regards recognition judgments, for instance, the covariation coding model predicts a tendency to make false recognitions of new items respecting the correlational structure, and likewise, to fail to detect old items which violate this structure. The values predicted by the covariation coding model were found to fit participants' observed performance fairly well, with Pearson's r between predicted and observed values falling within the range .50–.70, depending on experiments and conditions. According to Richardson, these findings demonstrate that people abstract knowledge about covariation among variables while studying exemplars and use this abstract knowledge in subsequent recognition or typicality judgments.

In a previous paper, we (Perruchet, Pacteau & Gallego, 1993) pointed out that the validity of the inference Richardson draws from the good predictive power of the covariation coding model rests on the postulate that a concurrent model cannot account for observed performance better than, or at least as well as, the covariation coding model. Indeed, Richardson deals with two other models, based on the notions of prototype representation and independent feature frequency, respectively. He shows that these concurrent models fail to explain a substantial part of the performance variance. However, Richardson does not deal seriously with another class of models that have potential relevance here: the 'instance-based' or 'exemplar-similarity' models, initially propounded by Brooks (1978) and Medin & Shaffer (1978) in the field of categorization. Briefly, the idea underlying these models is that individuals store each of the studied exemplars in memory, without making any condensation into a small amount of information, such as the occurrence frequency of individual or composed features. In the test phase, participants would perform recognition or categorization judgments of test items on the basis of their degree of similarity with specific individual instances (e.g. Brooks, 1978), or with a weighted average of all instances stored in memory (e.g. Medin & Shaffer, 1978). These judgments turn out to be sensitive to correlated features as a by-product of the use of analogies with instances in which the correlations of interest are embedded.

Reanalysing Richardson's data, Perruchet *et al.* (1993) demonstrated that a very simple exemplar-similarity model was as powerful as the Richardson covariation coding model in accounting for recognition or typicality judgments. However, they also demonstrated, through the running of hierarchical multiple regression analyses, that the covariation coding model still accounts for a reliable part of variance in performance when predictions of the exemplar model are partialed out, hence suggesting that the apparent sensitivity to covariation cannot be considered only as the by-product of the memory for specific exemplars.

Before concluding that covariation knowledge is abstracted during the experimental session, as did Richardson and collaborators, it is worth examining yet another counter-argument. This counter-argument stems from the fact that, in most of Richardson's studies, the to-be-abstracted relations bear on the size of the different components of the study items. More precisely, the best exemplars of the concepts are those in which the different components—arms, head and body for the robots—are of the same relative size (all small, all medium, or all large). The point is that this kind of dimensional covariation is ubiquitous in everyday subjective experience. Using a principle which is a part of the participant's background knowledge to structure study items makes it unclear whether he or she learns anything from the study exemplars. Indeed, there is overwhelming evidence that participants with strong expectancies for covariations between events overestimate the contingency between them, a phenomenon known as *covariation bias*. A special case of covariation bias consists in the report of a correlation even though data are in fact statistically independent, as illustrated in the classic Chapman & Chapman (e.g. 1969) experiments on illusory correlations in clinical judgment, and a number of other studies (e.g. Murphy & Wisniewski, 1989). In the cases under scrutiny here, it is possible that judgments were biased by an *a priori* expectancy, regardless of the effective relationships between the displayed events.

The experiments reported here were undertaken to test the hypothesis that participants would no longer exhibit sensitivity to covariation in a conceptual replication of Richardson's experiments using variables covarying along dimensions other than that of size. Operationally, this hypothesis states that when other variables are used, a covariation coding model would have no additional predictive value beyond exemplar model predictions, if both models are entered together as predictors in hierarchical multiple regression analyses. However, some unexpected results led us to enlarge the scope of this initial framework. Briefly, Expt 1 showed that the contrast between size and other variables is a rough and misleading approximation of a more general contrast between variables whose intercorrelations may be respectively qualified as 'pre-primed' and 'truly arbitrary'. Experiment 2 was designed to explore the relevance of this distinction when investigating the participants' ability to abstract covariations between experimental variables in incidental learning conditions.

EXPERIMENT 1

The first experiment aimed to verify that variables differing along a size dimension convey an *a priori* knowledge about their covariations, while other variables are not similarly biased. This experiment involved pictures of robots differing from one another along six ordinal, three-level variables. Half of these were size variables: head size, leg size and arm size. The other variables were the orientation of the antenna (slightly, moderately or very bent), the density of the body's texture (high, medium and low density) and the gap between the eyes (small, medium and large). Figure 1 shows all of the possible variations.

A first possibility would have been to ask participants for a direct assessment of the expected correlations between all possible pairs of features. The experiment would have tested, in fact, whether their intuitions matched experimenters' intuitions about natural covariations. However, this *direct* measure appeared ill suited for our final objective, namely the *indirect* assessment of covariation knowledge induced through the exposure to correlated events. We thus devised an original method allowing background knowledge about covariations to be indirectly evaluated.

The rationale of the method is introduced here by examining the case for arms and legs. Let us consider a reference robot, and two test robots with longer arms than the reference robot, but differing in terms of leg size; specifically, one has longer legs, and the other shorter legs, than the reference robot. Participants are asked to judge which of the test robots looks more like the reference robot using a continuous scale

10 B

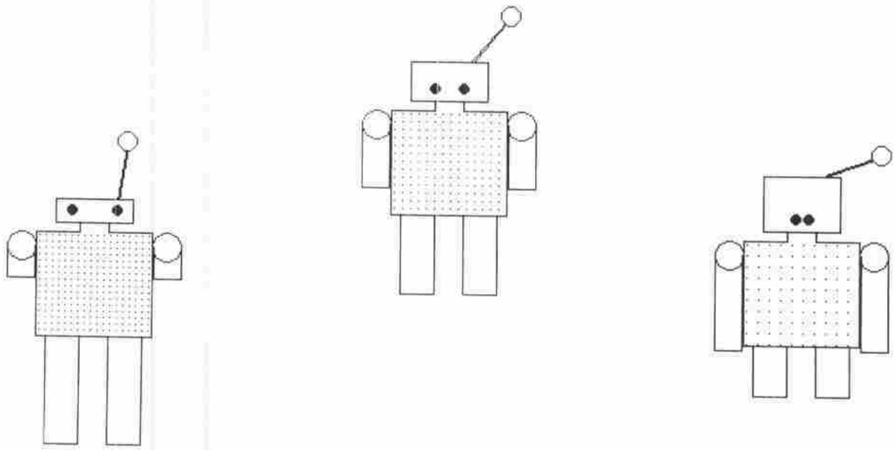

Figure 1. This figure illustrates the variations introduced in Expt 1. The robot on the left has a small head, large legs, small arms, a slightly bent antenna, high density texture and a large gap between the eyes. The robot on the right has the opposite features, and the robot in the centre possesses the intermediate values for each feature.

445n/

on which the end-points represent, at one end, the most marked choice of the short-legged robot (arbitrarily coded 0) and at the other end, the most marked choice of the long-legged robot (coded 100). The participant may give a score of 60, for instance, hence indicating that the test robot with long legs tends to be judged as more similar to the reference robot than the test robot with short legs, when both test robots have long arms. This rating in itself is uninformative about covariation knowledge, because the choice of long rather than short legs may stem from reasons unrelated to the arm length. The proper evaluation involves a comparison between this first rating and a second one, performed in similar conditions, except that both test robots now have shorter arms than the reference robot. Let us suppose that the participant rates 40 on the similarity scale in this condition, hence indicating that the test robot with long legs now tends to be judged as less similar to the reference robot than the test robot with short legs. A genuine quantitative estimate of the *a priori* knowledge about relationships between the two variables at hand is given by the difference between the two evaluations. A null difference would attest to a lack of association. Here, the 20-point difference, pending statistical confirmation, indicates that our fictitious participant tends to be sensitive to a positive relation between arm size and leg size.

Experiment 1 applied this procedure for all possible ordered pairs of variables. Participants were presented with a succession of frames comprising one robot of reference and two test robots. The robot of reference had intermediate values for the six variables (and was hence constant throughout the frames). On each frame, the test robots differed from the reference robot on two of the six variables. One of these (hereafter referred to as the context variable) was common to the two test robots, and the other (henceforth the target variable) differed. Participants had to judge on a continuous scale which of the test robots 'looks more like', or 'fits better with', the reference robot. For each context/target variable pair, the measure of interest was the difference between ratings on the target variable induced by the change in the context variable.

Method

Participants

Twenty-seven third-year university students majoring in psychology served as participants, in partial fulfilment of a course requirement.

Materials

The robots were displayed on a black and white CRT screen with a resolution of 480×640 pixels, in the spatial arrangement shown in Fig. 1. The reference robot was at the centre, and the test robots were located alongside and slightly downwards. The robots varied in height from 4.4 to 6.5 cm, and all were 4 cm in width. For all variables except body texture, the intermediate values (hereafter: 2) were set at the centre of the end values (hereafter: 1 and 3), as measured in pixels for head, arms, legs and eye lags, and in angle units for the antenna. For the body texture, the intermediate value was chosen on the basis of participants' reports in informal pilot studies. Short head, arms, and legs, very bent antenna, high density texture, and a small gap between the eyes were coded 1, and their opposite, 3.

The participants made their similarity ratings through the use of a joy-stick device. The on-line joy-stick position was reported at the bottom of the CRT screen by a marker (a 0.8 cm long horizontal rectangle) moving on a horizontal scale (length: 9 cm). The participants pressed one of the joy-stick buttons when the marker was in the desired location.

Procedure

The participants were seated in front of the CRT screen of an IBM microcomputer. The frame shown in Fig. 1 was displayed on the screen. This frame comprises the robot of reference and two test robots that show all the variations introduced during the subsequent experiments, although these specific robots are not used afterwards. Participants were told that they would be exposed to a series of similar frames throughout the experiment. They were asked to judge which of the two side robots looked more like, or fitted better with, the central robot. They were instructed to express their judgment through the joy-stick device, by moving the marker on the scale towards the chosen robot, more or less far from the centre of the scale according to the strength of their preference (they were informed that they could leave the marker at the centre of the scale if they had no preference).

Participants were first given six practice trials, in which test robots differed from those used afterwards. Sixty frames were then displayed in succession. Indeed, assessing covariation between any of the 30 (6×5) ordered pairs of context/target variables involved two frames. On each frame, test robots differed from the reference robot by their value with respect to the two variables at hand. The value for the context variable was the same for the two test robots within a given frame (either 1 or 3), and differed between frames, whereas the value for the target variable differed within a frame (1 and 3 on each frame).

The location of the test robots (i.e. the assignment of values 1 or 3 of the target variable to the right and left side of the reference robot) was randomized for each trial and each participant. The 60 frames were presented in a different randomized order for each participant.

Data analysis

The scores were initially collected on a 220-pixel scale. For the sake of convenience, all scores are reported below after being linearly transformed on a 100-point scale, with 0 indicating the choice of short head, arms, and legs, very bent antenna, high density texture, and a small gap between the eyes.

Results

The first column of Table 1 shows the mean rating by participants for each of the target variables after averaging across the context variables. Overall, participants found that robots with high heads, long arms, long legs, slightly bent antenna, low density texture and a large gap between the eyes fitted better with the reference robot than robots with the opposite features. In all cases, the mean rating departed significantly from the centre of the scale ($ts(26) > 3.63$; $p < .0012$). Although not directly relevant to our concern, this departure has the damaging consequence of lowering the scores' sensitivity to variations due to context variables by strengthening the possibility of ceiling or floor effects.

Table 1. For each of the target variables, the table indicates the mean rating score, and the algebraic difference between the rating scores obtained for the two opposite values of each context variable

Target variables	Mean rating	Context variables					
		Head	Legs	Arms	Antenna	Texture	Eyes
Head	72.7	—	4.1	5.9	−6.4	−16.4*	8.2
Legs	66.4	−1.8	—	22.3*	−2.7	−5.9	−2.7
Arms	76.4	−1.4	20.4*	—	−5.4	−3.6	10.0
Antenna	63.6	−11.8	−9.5	0.0	—	3.6	4.1
Texture	66.4	−11.8	−16.8*	7.3	−8.6	—	−8.6
Eyes	84.5	9.6	3.2	9.5*	8.6	−0.4	—

$^*p < .05$.

The other entries in Table 1 show, for each target variable, the algebraic difference between the judgments collected for the two levels of each context variable (the difference was computed as level 3 minus level 1). For a given pair of variables, two values are provided, according to the function assigned to each variable (target and context respectively, or the reverse). These two values occupy positions within the matrix that are symmetrical with regard to the main diagonal. In order to obtain an assessment of the overall correspondence between the two values obtained for each pair of variables, it is possible to compute their correlation over the 15 (6*5/2) pairs of variables. When computed on Table 1, the correlation is .734. However, the order in which variables appear in Table 1 is one of 6! (720) possible orderings. The assignment of particular differences to the upper and lower triangles of the matrix, and the consequent value of the correlation, is an accident of that particular ordering. We therefore computed a product moment correlation for each of the 720 possible arrangements of the matrix, and the resulting coefficients were averaged after r to z Fisher transformations. The mean correlation was .750 (13), $p < .002$, hence indicating that pairwise relations tended to be reciprocal. It is worth adding that this result also testifies that data are satisfactorily reliable. Indeed, because the values of each pair were collected in different conditions, the resulting correlation is an underestimation of the reliability coefficient that could have been computed if the same measures were simply repeated.

Statistical significance of the differences reported in Table 1 was computed by paired t tests. Where significant, the difference indicates that the judgment about the target variable depended on the level of the context variable, hence evidencing subjective relationship between the two variables. As expected, a reciprocal relation between arm size and leg size was clearly apparent. As expected also, the orientation of the antenna was not related to the other variables. However, several results were not anticipated. Judgment on eye spacing depended on arm size, texture density was found to be related with head and leg size, and the expected relation between head size and other size variables did not occur.

Discussion

Despite a probable lack of sensitivity due to a departure of the mean scores from the centre of the scale, our procedure revealed some consistent associations between variables. However, these associations matched only partially with the expected ones.

On the one hand, some associations did occur among some variables not involving component size. It is possible that some of these variables were in fact processed by participants as size variables, although they were not thought of as such while planning the experiment. For instance, the relations of eye gap with head and arm size become meaningful if one considers that eye gap has been processed as a size variable. The relations between some size variables and texture density are more difficult to account for along the same lines. However, these relations are not very surprising. In the real world, such relations arguably exist, although they tend to be either positive (e.g. getting closer to a picture generally increases both its size and its density) or negative (e.g. moving away the slide projector increases the size of the picture while decreasing its density).

On the other hand, associations between size variables, although strong in some cases, were not found to be a ubiquitous phenomenon. It must be realized that the material was made up of drawings of robots, that is, of artificial objects whose appearance is highly variable. For instance, the shape and the relative proportion of the different components are far more flexible than in representations of humans or animals. In this context, our initial expectations regarding the covariations of the variables defined *a priori* as size variables, may be judged retrospectively as exaggeratedly simplistic.

At first glance, the fact that the relations between size variables suffer exceptions questions our reinterpretation of Richardson's studies. Indeed, this reinterpretation is grounded in the postulate that all size variables involved in Richardson's experiments were related in *a priori* participant representations. Upon further examination, however, it appears that our reinterpretation holds true even if *a priori* associations exist only among a subset of the variables used in a given experiment. The reason for this is that the correspondence between the observations and the predictions from the theoretical models was assessed on a global basis, after averaging across variables. There is no need for participants to be sensitive to covariations between *all* pairs of variables to obtain a good correspondence in

10C

averaged performance. Note that Richardson's counterbalanced allocation of component variables to formal labels (e.g. in Richardson & Carthy, 1990, Expt 1, variable I corresponded either to head, arms or body of robots according to participants) hampered any *a posteriori* analysis aimed at examining whether participants were differentially sensitive to covariations as a function of the variables at hand.

Although most of the pairwise relations appear to be interpretable, at least on *a posteriori* grounds, there are still certain inconsistencies in the whole picture. For instance, considering that eye gap has been processed as a size variable accounts for its relation with arm size, but raises the question of why there was no relation with leg size. This outcome leads us to suspect that the pattern resulting from this kind of analysis is highly dependent on specific, superficial features of the material. This limitation is not really detrimental to our approach, but it demands that we alter our starting framework. Instead of contrasting covariations involving size and other variables, the objective of Expt 2 is to compare some of the 'pre-primed' with some of the 'arbitrary' covariations revealed in Expt 1. The distinction refers to the fact that some variables are related to each other for participants before any exposure to contingency relationships during the experimental session, whereas other variables are not similarly related. This objective does not require that the association of a given pair of variables with a particular category should be generalizable, or clearly justified on rational grounds. The only prerequisite is that the result pattern should be replicable for the same material, and our assessment of reliability suggests that this prerequisite is indeed fulfilled in our experimental set-up.

EXPERIMENT 2

The general paradigm of Expt 2 is taken from that of Richardson's studies, and more particularly Expt 1 in Richardson & Carthy (1990). Briefly, participants were instructed to study and remember a series of robots exhibiting moderately strong covariations between three ordinal variables. They were then exposed to another series of robots comprising both study and new robots, and asked for recognition judgments.

The main new factor we introduced was the nature of the variables at hand. For a first group of participants, the study material displayed covariations matching some of the 'pre-primed' covariations revealed in Expt 1 (hereafter, the pre-primed group). For a second group, the variables were not similarly related in participants' *a priori* representations (hereafter: the arbitrary group). We hypothesized that performance of the pre-primed group would show evidence of covariation abstraction, hence replicating the general outcome of Richardson's studies, whereas performance of the arbitrary group would provide no such evidence.

A second new factor was introduced, which concerned the time constraints imposed during the test phase. Participants were alternatively asked to make judgments under speeded and non-speeded conditions (with order counterbalanced). The rationale for this manipulation stemmed from the claim that performance testifying to the presence of abstraction may be due to processes taking place during the test phase while participants are coping with test items (e.g. Perruchet & Pacteau, 1991). In keeping with this claim, one may anticipate that participants faced with a test robot engage themselves in the intentional search for covariation rules (through the scanning of memorized exemplars, for instance) and then succeed in abstracting even arbitrary covariations. The speeded test condition was devised to prevent them from engaging in such a strategy. Indeed, the intentional searching for rules during the test phase is presumably time consuming, and pressing participants to respond as quickly as possible appears to be a procedure capable of impeding this mode of processing. In a more general way, manipulating time constraints in the test phase allows the part of the performance pattern which may be attributed to 'late' (Estes, 1986) or 'in-line' (Smith, 1989) processes to be at least partially separated from the part resulting from the encoding processes at work during the study phase.

Method

Participants

Thirty-two people from the same pool as that of Expt 1 served as participants.

Materials

As evidenced by Table 1, the result pattern of Expt 1 makes it impossible to draw three covarying variables (out of the six manipulated ones), and likewise, three variables without any *a priori* relations. However, this did not raise actual problems for the purpose of the present study, because in the paradigm of Richardson & Carthy 1990) we intended to replicate here, significant experimentally induced covariations involved only a subset of the variables. Labelling variables I, J and K, only the 'I and J' and the 'J and K' associations were significant. Owing to this restriction, the allocation of 'I' to legs, 'J' to arms and 'K' to eye gap allowed for satisfactory matching between pre-primed and experimentally induced covariations for the pre-primed group. Similarly, allocating 'I' to legs, 'J' to eye gap and 'K' to antenna allowed for satisfactory independence between pre-primed and experimentally induced covariations for the arbitrary group. (Incidentally, the fact that both groups shared two out of the three variables may seem paradoxical. It must be realized that changing only one variable changes in fact two out of the three pairwise relations between three variables, namely the 'I and J' and 'J and K' relations, which are those of relevance in our paradigm.) The choice of these variables has the additional advantage of leaving aside those whose pattern of interrelations observed in Expt 1 was found to be difficult to account for. These variables, namely head size and body texture, were set to their intermediate values within each group.

Procedure

Participants were randomly assigned to pre-primed and arbitrary groups until the number in a group reached 16. The groups were differentiated only by the nature of the covarying variables. For all participants, the acquisition phase comprised six repetitions of the same set (within a group) of 18 robots in different randomized orders. Each set consisted of nine different exemplars, each of which appeared one, two or three times, as shown in Table 2 (frequencies of occurrence were taken from Richardson & Carthy, 1990, Table 1). Each robot was displayed 5 s, with a 0.8 s interval between successive robots. A short, self-paced break was introduced between successive sets of 18 robots.

Participants were told that they would be exposed to schematic drawings of robots that they had seen previously, in order to recognize them later among similar robots. They were informed that they would be exposed to six series with short breaks between series, and that the same robots would be displayed on each set.

The test phase immediately followed the study phase. All 27 robots were presented twice in different random order. Half of the participants within each group were asked to give their responses as quickly

Table 2. Predicted and observed performances for each of the 27 test items. Predicted values are computed for the two theoretical models tested in the paper (arbitrary units). Observed values (percentage of old judgments) are separated according to the nature, naturally primed or truly arbitrary, of the experimentally induced correlations

Items			Predicted values		Observed values	
	IJK	Number of presentation	Covariation coding	Exemplar-similarity	Natural	Arbitrary
1.	111	3	10.49	3.5	87.50	78.13
2.	121	—	3.42	1.5	53.13	40.63
3.	131	—	1.11	0.5	15.63	28.13
4.	112	1	7.34	3.5	75.00	78.13
5.	122	—	3.42	0.5	53.13	43.75
6.	132	—	1.59	1.0	28.13	43.75
7.	113	2	5.14	2.5	62.50	75.00
8.	123	—	3.42	1.0	43.75	34.38
9.	133	—	2.27	1.5	21.88	40.63
10.	211	—	4.88	1.5	31.25	62.50
11.	221	—	3.42	1.5	53.13	68.75
12.	231	1	2.39	2.0	62.50	68.75
13.	212	—	3.42	0.5	43.75	65.63
14.	222	—	3.42	2.5	68.75	53.13
15.	232	2	3.42	4.0	75.00	68.75
16.	213	—	2.39	1.0	25.00	56.25
17.	223	—	3.42	2.0	71.88	50.00
18.	233	3	4.88	4.0	78.13	53.13
19.	311	—	2.27	1.0	21.88	31.25
20.	321	2	3.42	3.5	56.25	65.63
21.	331	—	5.14	1.5	40.63	40.63
22.	312	—	1.59	1.5	28.13	37.50
23.	322	3	3.42	4.5	56.25	75.00
24.	332	—	7.34	2.5	84.38	43.75
25.	313	—	1.11	0.5	12.50	43.75
26.	323	1	3.42	2.5	50.00	71.88
27.	333	—	10.49	2.0	78.13	53.13

as possible during the first part of the test, and allowed to complete the last part without time constraints; the other half performed the tasks in the reverse order. A signal was given in the middle of the test session to prompt them to change tasks. No mention was made of the fact that the two test parts consisted of the same set of robots.

For each robot, participants had to respond by pressing one out of the first three keys on the numeric keypad of the keyboard. They were told to press '1' if the target robot had not been viewed earlier and looked very different from the robots displayed in the study phase; '2' if the target robot had not been viewed earlier, but looked like the robots displayed in the study phase; and '3' if the target robot had been viewed in the study phase. An abridged version of these options corresponding to their numeric labels was printed at the bottom of the screen. In fact, due to the strong similarity between all of the robots, only 16.9 per cent of responses fell into '1', and 30.3 per cent into '2'. Analyses were performed after pooling responses '1' and '2', so that the final scores were akin to those issued from a conventional yes/no recognition test.

Model predictions

Model predictions are shown in Table 2. They were computed as in Perruchet *et al.* (1993). The predictions from the covariation coding model were derived from log-linear statistics, as in the studies by Richardson. The predictions from the exemplar-similarity model were the sum of two components, which were respectively proportional to (1) the frequency of occurrence of the target item. Proportionality coefficient was set to one, so that this component value was in fact equal to the number of presentation of the item; and (2) the frequency of occurrences of items similar to the target item, the similarity between two items being defined by the fact that they differ only by one value on one variable. Proportionality coefficient was arbitrarily set to 1/2. Note that the predictions of the two models are set in arbitrary units. No care was taken to fit the predicted values with realistic recognition scores, because the models allow only relative predictions to be made and in keeping with this fact, the subsequent analyses related to the models rely only upon relative information (e.g. linear correlations are insensitive to any linear transformation of the data).

For the sake of illustration, let us consider items 26 and 27 in Table 2. The value predicted by the covariation coding model is far higher for item 27 than for item 26, because item 27 is much more representative of the correlations between variables exhibited during the study phase than is item 26. On the contrary, the exemplar similarity model predicts more positive responses for item 26 than for item 27, essentially because item 26 was displayed during the study phase, whereas item 27 was not. (More precisely, item 26 was credited with one unit, because it was shown once, and with 3/2, i.e. 1.5 supplementary units because it was similar to item 23, which was shown three times; item 27 was credited with $1.5 + 0.5$ units because of its similarity with item 18 and item 26, respectively.) The overall correlation between the two sets of predictions over the 27 test robots was .44.

Results

Instructions given in the test phase to control the speed of responding were effective: the mean latency of responses was notably shorter under speeded ($M = 2.035$ s) than unspeeded ($M = 5.500$ s) conditions. For the unspeeded condition, the latencies were shorter when this condition was given after the speeded condition than when this condition was given first ($M = 4.604$ vs. 6.395 s; $F(1,28) = 4.49$, $p = .043$). The same trend was observed for the speeded condition, but the difference failed to reach significance ($M = 1901$ vs. 2169 s, $F(1,28) = 1.12$, $p > .10$). The former effect presumably reflects the difficulty that participants had slowing down their speed of responding after some practice in the speeded condition.

However, the speed of responding factor had no reliable effect upon recognition scores. Participants judged as old 69 per cent of old robots and 48 per cent of new robots under speeded conditions, and likewise, 68 per cent of old robots and 42 per cent of new robots under unspeeded conditions. A 2×2 repeated measures analysis of variance showed that the item status (old vs. new) had an effect ($F(1,28) = 22.43$, $p < .001$), hence confirming better than chance recognition of old robots. But there was no main effect of test conditions (fast vs. slow) ($F(1,28) = 1.19$, $p > .10$), and no interaction between test conditions and item status ($F < 1$).

The lack of any effect on the mean performance does not imply that speeded conditions had no effect on the pattern of responding, which is of primary importance here. To investigate this hypothesis, the percentage of old responses over all the participants was computed for each test robot separately for speeded and non-speeded conditions, and a correlation between the two sets of measures was computed. The resulting Pearson product correlation, computed over the 27 test robots, was a substantial .836.

On the basis of these results, the analyses contrasting the performances of pre-primed and arbitrary groups were run after averaging data across the two test conditions. Participants judged as old 66.9 per cent of old robots and 43.0 per cent

107

of new robots in the pre-primed group, and 70.4 per cent of old robots and 46.5 per cent of new robots in the arbitrary group. An analysis of variance performed with groups as a between-subject factor and status of items (old vs. new) as a repeated measure factor indicated no main effect for group and no interaction between group and item status (all Fs < 1). Thus the two categories of variables were of equal efficiency in promoting learning.

The percentage of old judgments was computed for each test robot, separately for the pre-primed and the arbitrary groups. The results are shown in Table 2. Scanning individual values offers some preliminary insight into the pattern of results. For instance, the performances for items 26 and 27, which were taken as privileged instances above (see Model prediction section), showed a drastic inversion for the two groups. While those perceiving pre-primed covariations gave more old responses for item 27, as predicted by the covariation coding model, the reverse trend was observed in the group perceiving arbitrary correlations, as predicted by the exemplar-similarity model. Virtually the same pattern appeared for items 23 and 24. However, such a piecemeal analysis is admittedly limited.

To obtain an overall picture of the pattern of the results, we performed correlational analyses over the 27 test robots. The correlation between groups was moderately high: .539. Thus, although not differing on mean performance, the between-groups correlation suggests that the groups displayed partially different patterns of response. For the pre-primed group, Pearson's r between recognition scores and the predictions generated by the covariation coding and the exemplar-similarity models were respectively .726 and .710. Pooled together, predictions of the two models gave a multiple R of .846. The running of hierarchical multiple regression analyses with different entrance orders demonstrated that the addition of one model significantly improved predictions made from the other, and conversely. The F to enter exemplar-similarity after covariation coding model predictions was 15.94 (1, 24), $p < .001$, and the F to enter covariation coding after exemplar-similarity model was 17.87 (1, 24), $p < .001$.

The pattern was strikingly different for the arbitrary group. Pearson's r between recognition scores and the predictions generated by the covariation coding and the exemplar-similarity models were respectively .401 and .638. Hierarchical multiple regression analyses demonstrated that exemplar similarity added significant predictive power to the covariation coding model (F to enter: 11.09 (1, 24), $p < .001$). However, the covariation coding model did not improve upon predictions made from the exemplar-similarity model (F to enter < 1); indeed, the multiple R (.652) scarcely differed from the Pearson's r involving the exemplar-similarity model alone.

Discussion

The data collected on participants faced with pre-primed covariations closely replicated those of Richardson's studies. Covariation coding and exemplar-similarity models accounted for substantial and mutually complementary parts of performance variance, as Perruchet *et al.* (1993) showed to be the case in most of Richardson's experiments. Our data fit especially well with those collected in the Richardson &

Carthy (1990, Expt 1) study, whose paradigm served as a direct model in designing the present experiment.

However, a strikingly different finding emerged when participants were faced with unprimed covariations. The exemplar-similarity model still accounted for a substantial part of performance variance, but the covariation coding model did not in any way improve the predictions made from the exemplar-similarity model alone. These results clearly support our initial hypothesis, which accounts for the predictive power of the covariation coding model in Richardson's studies by the match between experimentally induced and pre-primed covariations. When original variables were replaced by variables not related *a priori* in the participant's mind, the covariation coding model lost any specific predictive power. Inferring the psychological processes at work from the correspondence of the data with the models, these results support the contention that participants did not abstract the covariations embedded in the study material while studying under memory-oriented instructions, insofar as experimentally induced covariations are not pre-primed.

One could argue that this pattern of results is linked to the specific mode of computation of the values predicted by the two models. A large range of other computational procedures would have been possible, and the selected ones may be criticized in several ways. For instance, the exemplar-similarity model is overly simple, and the determination of its parameters is largely arbitrary. To address this problem, we performed exploratory analyses after having introduced some variations into the parameters of the exemplar model. These changes did not substantially affect the results. Of course, these analyses did not entirely refute the argument. However, the selected modes of computation have at least one advantage over other, possibly more sophisticated alternatives: they can not be suspected of being *ad hoc*. They were propounded before the present experiment was run, and had previously been applied to other data, with the very same set of parameters (Perruchet *et al.*, 1993).

Another argument contesting the validity of our conclusion is that differences between the pre-primed and the arbitrary conditions could be linked to the differential saliency or distinctiveness of the individual variables at hand, rather than to the factor of interest here. In this line of reasoning, participants faced with unprimed covariations failed to discover these covariations whereas those faced with pre-primed covariations succeeded in doing so, because the variables involved in the former case were less easily identifiable and codeable than the variables involved in the latter case. This argument finds apparent support in the fact that performances correlated slightly better with exemplar-similarity model predictions in the pre-primed group than in the arbitrary group. Although not significant, this difference (.710 vs. .638, $Z = 0.458$, $p > .10$) suggests that the study material of both groups differed with regard to characteristics other than the presence or absence of pre-primed covariations.

This objection can not be directly dismissed from available empirical data, because individual variable characteristics were not properly controlled. Although care was taken to equalize saliency between all variables in preparing material for this set of studies, we proceeded only through informal pilot experiments, which were not devised to provide data suitable for conventional statistical analyses.

However, a couple of remarks are in order. First, the fact that the correlation between performance and exemplar-similarity model predictions tends to differ for the pre-primed and arbitrary groups cannot be construed as valid support for the argument under examination. It must be realized that the correlation between performance and exemplar-similarity model predictions does not reflect only that part of performance which is imputable to the memory for exemplars. Indeed, this correlation may be due in part to genuine covariation processing, because predictions from exemplar-similarity and covariation coding models correlate. As a consequence, a drop in the predictive power of the exemplar-similarity model (assessed through the simple correlation between observed and predicted performance) may be expected when participants no longer process covariations. It would be faulty to consider this decrement as evidence that participants experienced increased difficulty in memorizing exemplars.

A second remark is that the non-difference between groups in mean recognition scores runs counter to the idea that the variables at hand were less salient in the arbitrary than in the pre-primed group. Moreover, the fact that participants discriminated between old and new robots to the same extent in both groups suggests that the variables involved in the arbitrary group could be *more* salient than the variables involved in the pre-primed group, in order to compensate for the impairment in performance due to the failure to use the relational information embedded in the study material. Whatever the case may be, a conservative conclusion is that there is no empirical evidence favouring the claim that the difficulty of abstracting arbitrary covariations is due to the lack of saliency or distinctiveness of individual variables.

Experiment 2 was also devised to examine whether participants could abstract even unprimed covariations when given sufficient time to respond in the test phase. With this aim in mind, they were successively tested under speeded and unspeeded conditions. Although this variable was found to be effective in some situations (e.g. Turner & Fischler, 1993), it introduced no reliable differences in the pattern of the result. It is likely that the number of different robots, their close similarity, and the limited amount of repetitions made it difficult to efficiently scan the stored items in order to draw abstract organizing rules, even when response time was unlimited. Wattenmaker, McQuaid & Schwert (1995) report similar difficulty for detecting feature co-occurrence through the analysis of memorized exemplars across a variety of category structures and stimulus materials, although participants were told that they could take as long as they wished before responding.

GENERAL DISCUSSION

The present set of studies was designed to investigate participants' sensitivity to the correlational structure of the environment. Our aim was to disentangle the influences of two candidate mechanisms used to account for this sensitivity, namely the genuine abstraction of empirical relations, and the establishment of analogies with stored, specific exemplars. This latter mechanism is *a priori* effective, because new items which respect the correlations tend to be similar to a larger number of old items than those which violate the correlations. The respective weights of the two possible influences were assessed through regression procedures. Our results emphasize the utmost importance of determining whether the experimentally induced correlations between variables are representative of participants' background knowledge, or rather are fully arbitrary. Participants' performance appears sensitive to both forms of correlations, as attested to by the substantial predictive value of the covariation coding model. However, hierarchical regression analyses suggest that they actually used their correlational knowledge with pre-primed correlations, whereas they failed to abstract arbitrary correlations. In the latter case, empirical sensitivity to correlational structure was a by-product of judgments of similarity with specific, old exemplars.

The strong empirical influence of background knowledge leads us to emphasize the necessity of assessing this knowledge through experimental manipulations before investigating covariation learning. This knowledge ought to be assessed *indirectly*, because it can not be easily derived from reasoning and intuition, regardless of whether reasoning and intuition proceed from the experimenter or the participants. Our first experiment provides an original method to assess knowledge indirectly. The validity of this method was confirmed *a posteriori* by the results of the second experiment, insofar as the variables dissociated on the basis of Expt 1's results revealed themselves to be processed in quite distinctive ways in Expt 2.

Of primary importance here are the consequences of the covariation bias when studying incidental rule learning. In a previous paper, Perruchet *et al.* (1993) concluded that the apparent sensitivity to covariations demonstrated in the Richardson studies could not be considered *only* as the by-product of the memory for specific exemplars. The present results show that this contention is in fact dependent upon the use of a correlational structure that partly matches the structure participants have in mind before they come into the laboratory. No similar evidence emerges when experimental variables are chosen to be *a priori* unrelated. This finding replicates and extends, with quite an original methodology, prior evidence for the failure to observe incidental covariation abstraction (e.g. Perruchet, 1994; Wattenmaker, 1991, 1993). Overall, these results run counter to Richardson's claim regarding people's ability to abstract the covariations embedded in structured environments, and consequently, the role he confers to this process in the development of knowledge in children (Richardson, 1992; see also Younger & Cohen, 1985).

More generally, incidental covariation abstraction is a case in point in the current literature pertaining to the phenomenon of implicit learning (for a review, see Berry & Dienes, 1993). Although Richardson and Wattenmaker make only allusive reference to this literature at best, their situations exhibit a striking parallelism with those involved in implicit learning research. In both cases, participants are faced with a complex rule-governed situation with memory-oriented instructions, and their adaptation to the situation is assessed indirectly through performance modifications. In the implicit learning area, a major issue focuses on the question of whether their adaptation testifies to the implicit abstraction of the rules governing the experimental situations, or may be accounted for by more elementary mechanisms, such as the memory of specific study items or simple associative processes (Perruchet & Gallego, in press). Although early contributions lent favour to the first position (e.g. Reber, 1967), subsequent studies provide growing evidence for the second position (e.g.

Perruchet & Pacteau, 1990; see review in Shanks & St John, 1994). Available evidence led us to suggest that in order to acquire the abstract rules underlying complex experimental arrangements, humans have to engage in explicit hypothesis testing, logical inference, and more generally, the various controlled processes recruited in intentional rule discovery situations (Perruchet, 1994). The evidence collected in this paper supports this general view in the particular context of covariation abstraction.

Acknowledgments

Preparation of this article was supported by the Centre National de la Recherche Scientifique (URA 1838) and Université de Bourgogne, France. We thank Dr E. Gaffan, Dr K. Richardson, and two anonymous reviewers for their helpful comments on an earlier version of the manuscript.

References

Alloy, L. B. & Tabachnik, N. (1984). Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review*, **91**, 112–149.

Ahn, W.-K., Kalish, C. W., Medin, D. L. & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition*, **54**, 299–352.

Berry, D. C. & Broadbent, D. E. (1988). Interactive tasks and the implicit-explicit distinction. *British Journal of Psychology*, **79**, 251–272.

Berry, D. C. & Dienes, Z. (1993). *Implicit Learning: Theoretical and Empirical Issues*, p. 197. Hillsdale, NJ: Erlbaum.

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds), *Cognition and Categorization*, pp. 169–215. Hillsdale, NJ: Erlbaum.

Chapman, L. J. & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid diagnostic signs. *Journal of Abnormal Psychology*, **74**, 271–280.

Estes, W. K. (1986). Array models for category learning. *Cognitive Psychology*, **18**, 500–549.

Kushner, M., Cleeremans, A. & Reber, A. (1991). Implicit detection of event interdependencies, and a PDP model of the process. In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*, Chicago, IL.

Lewicki, P. (1986). Processing information about covariation that cannot be articulated. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **12**, 135–146.

Medin, D. L. & Shaffer, M. M. (1978). A context theory of classification learning. *Psychological Review*, **85**, 207–238.

Murphy, G. L. & Wisniewski, E. J. (1989). Feature correlations in conceptual representations. In G. Tiberghien (Ed.), *Advances in Cognitive Science*, vol. 2: *Theory and Applications*, pp. 23–45. Chichester: Ellis Horwood.

Perruchet, P. (1994). Learning from complex rule-governed environments: On the proper functions of nonconscious and conscious processes. In C. Umilta & M. Moscovitch (Eds), *Attention and Performance XV: Conscious and Nonconscious Information Processing*, pp. 811–835. Cambridge, MA: MIT Press.

Perruchet, P. & Gallego, J. (in press). A subjective unit formation account of implicit learning. In D. Berry (Ed.), *How Implicit is Implicit Learning*. Oxford: Oxford University Press.

Perruchet, P. & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General*, **119**, 264–275.

Perruchet, P. & Pacteau, C. (1991). The implicit acquisition of abstract knowledge about artificial grammar: Some methodological and conceptual issues. *Journal of Experimental Psychology: General*, **120**, 112–116.

Perruchet, P., Pacteau, C. & Gallego, J. (1993). Abstraction of covariation in category learning: A critical note on K. Richardson's studies. *British Journal of Psychology*, **84**, 289–295.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, **6**, 855–863.

Richardson, K. (1986). Evidence for relational-coding in concept formation, and a note on the effects of feedback. *International Journal of Psychology*, **21**, 641–657.
Richardson, K. (1987). The coding of relations versus the coding of independent cues in concept formation. *British Journal of Psychology*, **78**, 519–544.
Richardson, K. (1992). Covariation analysis of knowledge representation: Some developmental studies. *Journal of Experimental Child Psychology*, **53**, 129–150.
Richardson, K. & Carthy, T. (1989). Concept models and concept functions: Inference from partial information. *Acta Psychologica*, **72**, 81–102.
Richardson, K. & Carthy, T. (1990). The abstraction of covariation in conceptual representation. *British Journal of Psychology*, **81**, 415–438.
Shanks, D. R. & St John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, **17**, 367–447.
Smith, E. E. (1989). Concepts and induction. In M. I. Posner (Ed.), *Foundations of Cognitive Science*, pp. 501–526. Hillsdale, NJ: Erlbaum.
Turner, C. W. & Fischler, I.S. (1993). Speeded tests of implicit knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **19**, 1165–1177.
Wattenmaker, W. D. (1991). Learning modes, feature correlations, and memory-based categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **17**, 908–923.
Wattenmaker, W. D. (1993). Incidental concept learning, feature frequency, and correlated properties. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **19**, 203–222.
Wattenmaker, W. D., McQuaid, H. L. & Schwert, S. J. (1995). Analogical versus rule-based classification. *Memory and Cognition*, **23**, 495–509.
White, P. A. (1995). Use of prior beliefs in the assignment of causal roles: Causal powers versus regularity-based accounts. *Memory and Cognition*, **23**, 243–254.
Younger, B. A. & Cohen, L. B. (1985). How infants form categories. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation*, vol 19, pp. 211–247. New York: Academic Press.

Received 11 July 1995; revised version received 15 April 1996