VARIA

Juan Antonio García-Madruga
Nuria Carriedo
María José González-Labra
(Editors)

MENTAL MODELS IN REASONING

UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA

VARIA (37220PB01)
MENTAL MODELS IN REASONING

Diseño de cubierta:
Francisco Gutiérrez y equipo de diseño gráfico de la UNED

ISBN: 84-362-4034-0
Depósito legal: M. 38.490-2000

Primera edición: octubre de 2000

Impreso en España - Printed in Spain
Fotocomposición: Safekat, S. L.
Belmonte de Tajo, 55. 28019 Madrid
Imprime: Fernández Ciudad, S. L.
Catalina Suárez, 19. Madrid 28007

INDEX

LIST OF CONTRIBUTORS

BRUNO BARA. Università di Torino. Italy
MÓNICA BUCCIARELLI. Università di Torino. Italy
RUTH M.J. BYRNE. Trinity College, Dublin. Ireland
ANTONIO CAÑO. Universidad de Málaga. Spain
MANUEL CARREIRAS. Universidad de la Laguna. Spain
NURIA CARRIEDO. UNED. Spain
JEAN-PAUL CAVERNI. CREPCO. Aix en Provence. France
PEDRO L. COBOS. Universidad de Málaga. Spain
ANTONIO CORRAL. UNED. Spain
GINO DE VOOGHT. University of Ghent. Belgium
VICKY DIERCKX. University of Ghent. Belgium
ORLANDO ESPINO. Universidad de La Laguna. Spain
JONATHAN EVANS. Plymouth University. UK
PABLO FERNÁNDEZ-BERROCAL. Universidad de Málaga. Spain
M.ª JOSÉ FERRACES. Universidad de Santiago. Spain
JUAN GARCÍA-MADRUGA. UNED. Spain
M.ª JOSÉ GONZÁLEZ-LABRA. UNED. Spain
VITTORIO GIROTTO. CREPCO. Aix en Provence. France
DAVID GREEN. University College, London. UK
FRANCISCO GUTIÉRREZ. UNED. Spain
PHILIP JONHSON-LAIRD. Princeton University. USA
PAOLO LEGRENZI. Universitá di Milano. Italy
VINCENZO LOMBARDO. Università del Piemonte Orientale. Italy
FRANCISCO J. LÓPEZ. Universidad de Málaga. Spain
MONTSERRAT MARTÍN. Universidad de Santiago. Spain
ALICE MCELENEY. Trinity College, Dublin. Ireland
ENRIQUE MESEGUER. Universidad de la Laguna. Spain
SERGIO MORENO. Universidad de Granada. Spain
MARCO PALMONARI. Università di Siena. Italy
JEAN-LUC PÉRIS. CREPCO. Aix en Provence. France

10

BIASES IN PROBABILISTIC REASONING MAY BE PRODUCED BY ASSOCIATIVE LEARNING MECHANISMS

PEDRO L. COBOS,
ANTONIO CAÑO
FRANCISCO J. LÓPEZ

En los primeros estudios sobre sesgos en tareas de categorización probabilística, Kahneman y Tversky (1973) pusieron en relación tales sesgos con el modo en que las categorías se aprenden, se representan en la memoria y se usan. Desde esta perspectiva, los procesos de categorización basados en la similitud entre un objeto y los prototipos de las posibles categorías eran los responsables de sesgos como el de la desestimación de frecuencias de categorías o el de la falacia de la conjunción. Como tales procesos se concebían como rápidos, poco costosos y disparados por las propiedades de los estímulos, se asimilaron como la base constituyente del razonamiento intuitivo, proceso opuesto al razonamiento extensional (Tversky y Kahneman, 1983). Una de las principales razones por las que el enfoque de los sesgos y heurísticos de Tversky y Kahneman ha terminado por desanimar a muchos investigadores descansa en el escaso compromiso con la formulación de su teoría en términos de modelos computacionales. Nuestra propuesta, en pocas palabras, es que los mecanismos de aprendizaje asociativo, que interesantemente son rápidos, poco costosos y disparados por las propiedades de los estímulos, resultan ideales para la formalización computacional del razonamiento intuitivo (en Hinton, 1990 y Sloman, 1996 se puede encontrar una caracterización del razonamiento intuitivo basada en procesos asociativos). Más concretamente, la desestimación de frecuencias de categorías y la falacia de la conjunción en tareas de categorización probabilística se pueden interpretar, en algunas circunstancias, como ilusiones cognitivas debidas a la intervención de procesos de aprendizaje asociativo que resultan disparados por los contenidos presentes en las tareas de categorización. Cuando las personas

experimentan situaciones en las que algunos acontecimientos o propiedades (claves) son seguidos por otros (resultado), se forman lazos asociativos que permiten predecir los últimos a partir de la ocurrencia de los primeros. Los lazos asociativos formados durante tal situación de aprendizaje permiten la generación automática de un producto en una situación posterior de enjuiciamiento probabilístico si los argumentos de los juicios incluyen contenidos iguales o similares a las claves y resultados de la anterior situación de aprendizaje. Este producto, a su vez, sesga los juicios de las personas llegando a producir resultados normativamente inadecuados como la desestimación de frecuencias de categorías o la falacia de la conjunción. En el presente capítulo hemos revisado una serie de trabajos experimentales, especialmente los realizados en nuestro laboratorio, que apoyan nuestra propuesta. Asimismo, mostramos cómo una red neuronal asociativa puede dar cuenta de los fenómenos originados en los diferentes experimentos.

INTRODUCTION

In the first studies on biases in probabilistic categorisation tasks, Kahneman and Tversky (1973) explained the biases on the basis of frequently used heuristics, i. e., a sort of cognitive *shortcut* to solve daily life tasks. Some of these heuristics were related to the way in which knowledge about categories is acquired, represented in memory and used. Under this view, categorisation processes based on the similarity between an object and each of the category prototypes were responsible for biases as the base rate neglect and the conjunction fallacy. As these processes were conceived as rapid, effortless and feature-driven, they were taken as the basis for intuitive reasoning, as opposed to extensional reasoning (Tversky & Kahneman, 1983).

Under this perspective, biases in probabilistic reasoning tasks were also conceived as cognitive illusions. Two main reasons can be mentioned for this. First, heuristics, though responsible for the errors, were thought to be appropriate solutions for some current tasks out of the laboratory context. Second, heuristic processes were supposed to be primed by some elements of the laboratory task that are characteristic of those extra-laboratory contexts where heuristics are appropriate. Thus, in some sense, people are induced to take the experimental task as if it were another one (Cohen, 1981; López, Cobos, Caño & Shanks, 1998).

Unfortunately, Tversky and Kahneman's heuristics and biases approach has lacked of instantiation in terms of cognitive processing models. This, as well as other related aspects, has led the research program to two undesirable states (Dougherty, Gettys, & Ogden, 1999; Gigerenzer, 1996): a) heuristics have been invoked to explain almost every thing, but the conditions that switches on or off such heuristics remain obscure; b) there has been a tendency to base the explanation and description of behaviour on the normative principles violated. For example, the base rate neglect has served to refer to a violation of the Bayesian theorem as well as an explanation of that violation. Thus,

though initially promising and stimulating, this theoretical view has disenchanted a good number of researchers who have verified how little advance has been produced after three decades of research on biases in probability judgement.

Our claim, in a few words, is that associative learning mechanisms are good candidates to be the processes underlying the rapid, effortless and feature-driven intuitive reasoning conceived by Tversky and Kahneman (see Hinton, 1990 and Sloman, 1996 for an associative-based characterisation of intuitive reasoning). Specifically, the base rate neglect and the conjunction fallacy in probabilistic categorisation tasks can be characterised, in some circumstances, as cognitive illusions due to the associative learning processes evoked by the contents of the categorisation task. When people are repeatedly exposed to situations in which some events or features (cues) are followed by others (outcomes), associations are formed allowing to predict the second from the first ones. The associative links formed during the learning situation can automatically generate an output in a later probability judgement situation if the same or similar cues and outcomes constitute the judgements' arguments. This output, in turn, biases people's judgements and yields some normatively inadequate responses as the base rate neglect and the conjunction fallacy.

Along this chapter, we report some empirical evidence that support our claim. We have focused on two experiments made in our laboratory, though other interesting results have been published which have inspired our proposal. As will be shown, there is a common research strategy in all the experiments we will cite. Such strategy includes the following aspects: a) participants are provided with an associative learning task followed by a judgement phase where they have to estimate the probability of some statements relating the cues and outcomes of the learning phase, b) the learning and the judgement phase are arranged so as to meet the conditions under which some biases are typically obtained. Briefly speaking, a cognitive illusion is experimentally induced by providing an associative learning context. We also report the result of a simulation run with an associative neural network to assess the adequacy of associative learning principles to account for participants' judgements. Before starting with the empirical evidence, however, it is important to envisage which task associative mechanisms are aimed at and how such task can be performed. After that, we will articulate an associative explanation of biases in probability judgement tasks.

Associative mechanisms and predictive learning

Associative theories of learning have benn mainly developed in the field of animal learning. A good amount of phenomena have led researchers to think of animal conditioning as reflecting the learning of predictive relationships between events. Animal learning in classical conditioning, for example, is said to be a mechanism to acquire a sort of knowledge which reflects the causal texture of the environment (Dickinson, 1980; Tolman & Brunswick, 1935). One important phenomenon on which this assertion is based is the effect of relative validity of cues (more commonly known as selective learning effects). It is largely known that for conditioning to occur it does not suffice it to provide

individuals with a temporal contiguity relationship between a cue and an outcome. If the target cue always appears with another which, on its own, is a reliable predictor of the outcome, then the target cue will develop little or no conditioning at all. This is what happens in blocking experiments, for instance. In these experiments a given cue A is repeatedly paired with the outcome during a first stage. During the second stage, the target cue, B, and cue A are jointly paired with the outcome. At test, the target cue elicits little conditioned response compared with another condition including only AB trials (Kamin, 1968).

In these cases, the target cue is said to have a low relative validity. That is, when cue B is paired for the first time with the outcome, it is less valid as a predictor of the outcome than its accompanying cue. A similar way to think of B is as a redundant cue, for A is all what is needed to predict the outcome in AB trials. As a consequence, we cannot be sure of how reliable is B as a predictor of the outcome. But we can provide a more precise measure of unidirectional predictive relationships. Allan (1980) has proposed a contingency based measure, ΔP, which is the result of the probability of the outcome given the cue minus the probability of the outcome given the absence of the cue, holding constant everything else. A strict application of this contingency formula to the blocking design yields a value of 0 for the target cue B because the probabilities of the outcome given the presence and absence of such cue, holding constant the presence of the accompanying cue A, both equal 1. Interestingly, it has been shown that animals are impressively sensitive to variations of the DP measure of the relationship between cues and outcomes (Rescorla, 1966, 1968) as well as between responses and outcomes (Hammond & Paynter, 1983). Moreover, animals are said to be well-calibrated with respect to ΔP. Therefore, it is not unjustified to say that learning rely on a device well suited for the task of detecting predictive relationships between events.

Probably the most popular associative theory proposed to explain such phenomena is the Rescorla-Wagner theory (Rescorla & Wagner, 1972). This is a highly parsimonious theory which is able to account for an overwhelming number of data coming from animal as well as human learning experiments. The theory establishes a rule to calculate associative strength changes between the representations of the cues and the representations of the outcomes on the basis of the occurrences of such cues and outcomes. This rule, which is mathematically equivalent to the well known delta rule extensively applied in connectionist networks (Sutton & Barto, 1981), states that the change of the associative strength between a cue representation and an outcome representation is directly proportional to the surprise caused by the outcome (namely, the difference between what is expected to occur about the outcome from all the cues present and what actually occurs). It is mathematically formalised as follows:

$$\Delta V_{ij} = \alpha_i \beta_j (\lambda_j - \Sigma V)$$

where α_i and β_j refers to the salience of the cue and the outcome, respectively; λ_j is 1 when the outcome is present and 0 when it is absent; and ΣV is the sum of the associative strengths between all the cues present and the outcome (i.e., the expectation of the outcome). It has been shown that, under some conditions,

this rule yields equivalent asymptotic results as the ΔP calculus (see Cheng & Holyoak, 1995, and Cheng, 1997 for a detailed computational analysis of the RW rule). Thus, the RW rule is an algorithm that effectively detects predictive relationships between cues and outcomes. Finally, it gives a good explanation of the relative validity effect. In fact, this model was conceived to account for this phenomenon.

Since the last few years, some researchers have been interested in determining how strongly human predictive learning parallels the findings found in the animal conditioning research (Allan, 1993; Dickinson, Shanks, & Evenden, 1984; Shanks, 1993, 1995; Wasserman, 1990). Experimental designs inspired in animal learning experiments have been used in causal, contingency, and category learning experiments. The results obtained greatly resemble those obtained in animal conditioning. Two frequent aims in these experiments have been to determine whether contingency, predictive and causal judgements are affected by the ΔP measure of contingency and to verify whether contingency, predictive, causal and probability judgements are affected by the relative validity of cues. There is convincing evidence that contingency, predictive and causal judgements are nicely sensitive to ΔP. Moreover, there are some data showing that contingency judgements accurately fit ΔP when trial-by-trial learning procedures are used (López, Almaraz, Fernández, & Shanks 1998; Wasserman, Elek, Chatlosh, & Baker, 1993). Actually, there is some controversy in non-contingent situations, for some experiments show that people tend to overestimate non-contingent relationships between cues and outcomes in high outcome-density conditions, i.e., when the outcome occurs very frequently. This result also parallels animal learning findings. However, there is evidence that such overestimation disappears if the number of trials is incremented and if participants are properly instructed (López, Almaraz, Fernández and Shanks, 1998; Matute, 1996). Interestingly, there is evidence that overestimation in animal learning also tends to disappear with extended training (Benedict & Ayres, 1972).

On the other hand, the relative validity effect has been repeatedly obtained in different situations. It has been found in predictive, contingency, and causal learning situations (Baker, Mercier, Vallee-Tourangeau, Frank, & Pan, 1993; Chapman & Robbins, 1990; Dickinson & Shanks, 1985; Shanks & López, 1996; Van Hamme & Wasserman, 1993). Relative validity effects have also been found in probabilistic categorisation tasks (Cobos, López, Rando, Fernández & Almaraz, 1993; Estes, Campbell, Hatsopoulos & Hurwitz, 1989; Gluck & Bower, 1988; Kruschke, 1996; Myers, Lohmeier & Well, 1994; Nosofsky, Kruschke & McKinley, 1992; Shanks, 1990). As we will see later, the later results are relevant here because they allow to establish a link between relative validity effects and some biases in probability judgements. Specifically, it is noteworthy that relative validity effects make people to deviate from empirical conditional probabilities.

The RW model has been successfully employed to account for the above mentioned results as well as others. As will be shown, in some cases, it has been even used to give a quantitative account for the observed judgements, obtaining an impressive good fit (Gluck & Bower, 1988; Estes *et al.*, 1989; Cobos *et al.* 1993). Of course, some assumptions have to be made, though minimal, to extend the RW theory to account for judgements because such theory is only

concerned with connection strengths between representations of cues and outcomes. Normally, it is assumed a monotonically increasing function relating connection strengths and judgements. In categorisation tasks where people have to assign a series of stimuli to a set of mutually exclusive categories, a ratio rule is usually applied (Gluck & Bower, 1988; Estes *et al.*, 1989). That is, if we think of a given category representation as a given output node of a feedforward connectionist network, the probability of that category would be equal to the amount of activation of the corresponding output node divided by the summed amount of activation from all the output nodes:

$$P(i) = \frac{O_i}{\sum_j O_j}$$

Here, the amount of activation of a given output node is assumed to be equal to the summed connection strengths of the representations of the cues present[1].

Other more sophisticated connectionist models have been proposed in some predictive learning situations, such as ALCOVE (Kruschke, 1992), Pearce's configural model (Pearce, 1994), or ADIT (Kruschke, 1996). These models make different assumptions about how stimuli are represented or about attentional shifts as part of learning. However, it is noteworthy that a crucial feature of these models to account for the data is still the error driven learning rule on which they are based. Thus we will focus on the more simple RW model for several reasons: a) it instantiates the essential error-driven learning principle to account for the data, b) despite its simplicity, it can cope with a wide range of phenomena of animal and human learning, c) it contains very few free parameters, d) the other models do not improve the RW's fit to the data of the experiments we will describe.

In summary, there is consistent evidence that animal and human learning is intended to capture the predictive structure of the environment. On the other hand, there is nothing new in saying that associative and, more generally, connectionist mechanisms are ideal devices to learn predictive relationships. In fact, models belonging to this family accounts for most of the data coming from animal learning research and many human learning results.

An associationist-based explanation of biases in probability judgement

Though we will focus on the base rate neglect and the conjunction fallacy, the explanation offered here is not intended to be restricted to these fallacies. However, as our proposal is applied to these phenomena, let us start with a short description of each one.

[1] As many other researchers, we will assume some equivalence between associative and connectionist models. So, we will interchangeably use connections and associations, associative strength and connection weight, cues representations and input nodes, and outcomes representations and output nodes.

The base rate neglect refers to a fallacy obtained in Bayesian problems. Generally, participants have to estimate the probability of an object or person belonging to a series of mutually exclusive categories. Participants are usually informed of the frequency of the different categories (the base rates) and of the probabilities of finding the features which define the object they have to categorise in each category (conditional probabilities). The Bayesian formula to calculate the probability of the object belonging to one of the categories (A) is as follows:

$$P(A|O) = \frac{P(O|A) \cdot P(A)}{P(O|A) \cdot P(A) + P(O|\bar{A}) \cdot P(\bar{A})}$$

where O is the object to be classified; A and $\bar{A}$ are the target category and its complementary set, respectively; P(A|O) is the probability that object O belongs to category A; and P(O|A) and P(O|$\bar{A}$) are the probability of object O in category A and category $\bar{A}$, respectively.

The base rate neglect is obtained when participants' judgements are not affected, or are only slightly affected, by information about category frequency [P(A) and P($\bar{A}$)]. This phenomenon has been mainly studied in cases where there is some kind of conflict between the base rates and the conditional probabilities information. This happens when those features which define the object are more probable in a low-frequency category [that is, P(O|B)>P(O|A), and P(B)<P(A)] (see Figure 10.1). In such cases, participants' judgements can be better predicted from the conditional probability information rather than from the base rates.

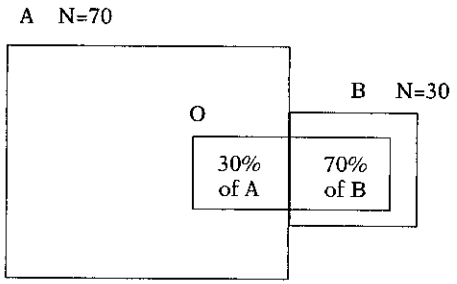

Figure 10.1. A Bayesian structure problem with a conflict between base rate information and conditional probability information.

The conjunction fallacy occurs when someone assigns a higher probability to a conjunction of events than to one of its constituents. This implies the violation of the conjunction theorem, which states that the probability of the conjunction of events is always lower or equal to the probability of any of the events that constitute the conjunction [that is, P(A∩B) ≤ P(A), and P(A∩B) ≤ P(B)]. In the Linda problem (Tversky & Kahneman, 1983), surely the most cited

in this context, a character named Linda is described to participants as an activist woman. Later, they have to sort in a decreasing order a series of statements depending on their probability. The most relevant statements are: Linda is active in the feminist movement (A), Linda is a bank teller (B), Linda is a bank teller and is active in the feminist movement (A & B). As a result, many participants estimate that statement A & B is more likely than statement B (Agnoli & Krantz, 1989; Fiedler, 1988; Politzer & Noveck, 1991; Wolford & Taylor, 1990).

According to our proposal, biases in probability judgements are a consequence, in some cases, of a cognitive illusion due to the operation of an associative learning mechanism (ALM). This type of cognitive illusion occurs when the contents of the probability judgement task are involved in a previous predictive learning situation, that is, a situation where an individual has to learn from examples to predict some outcomes from some cues. As we said before, associative learning mechanisms are specially well suited to learn predictive relationships from examples. The repeated exposure to examples favours the creation of associations between representations of the cues and representations of the outcomes so that, when a given cue is detected, the associated outcome representation is automatically activated. The amount of activation of the outcome representation can be taken as a measure of to what extent the outcome is expected.

When participants in an experiment have to make a probability judgement on the relationship between, say, two contents that have played the role of cue and outcome in a predictive learning task, the ALM automatically generates, as an output, a certain amount of expectation about the outcome occurrence. This amount of expectation is interpreted as a measure of the relationship between both contents. So, in a way, participants have a direct and automatic access to knowledge about the relationship between the contents involved in the probability judgement task. The bias in probability judgements occurs if that knowledge is accepted when the requested probability judgement is made, provided that ALM's output violates some assumption of the experimenter's normative analysis of the probabilistic reasoning task. In short, the learning context favours a way of coding and storing certain information which efficiently solves the problem faced in such context, but which generates a cost in a subsequent task if it requires solving different problems with similar contents (see López, Cobos, Caño & Shanks, 1998 for further details; see also Ratcliff & McKoon, 1996, for a similar account of some biases in implicit memory tasks).

To simplify, we can summarise our associationist explanation of biases in probability judgements in the following four steps: 1) the contents of the probability judgement have played the role of cues and outcomes in a previous associative learning situation; 2) during the learning situation, directed associations from representations of cues to representations of outcomes are formed that allow for a direct and automatic access to knowledge about the relationship between cues and outcomes; 3) during the judgement task, the contents that are made explicit for participants activates ALM which, in turn, can produce a single-step output that violates the experimenter's normative analysis of the task; 4) ALM's output can bias individual's judgement if it is accepted as the basis for the response.

How can our proposal be applied to explain the base rate neglect? As we said before, the base rate neglect has been shown many times in probabilistic

categorisation tasks in which the features that define the object to be categorise is more likely given the least frequent category. The task we have used in our experiments, which is based in Gluck & Bower (1988)'s, will serve as an example. In Gluck and Bower's task, participants had to learn to diagnose two fictitious diseases from four possible symptoms on a trial by trial basis. At test, participants had to rate the probability with which a series of hypothetical patients suffered from one or the other disease given each of the possible symptoms. 75% of the hypothetical patients suffered from a common disease (C), while the remaining 25% suffered from a rare one (R). Table 10.1 shows the conditional probabilities relating the diseases and the symptoms. One of the patients showed symptom S_1, which is more likely in disease R (p=.6) than in disease C (p=.2). Despite this, there are as many patients with S_1 who suffer from disease R as patients with S_1 who suffer from disease C. However, participants gave higher ratings to the probability of disease R given S_1 than to the probability of disease C given S_1.

To show how participants' judgements could be accounted for on the basis of the RW learning rule, Gluck and Bower run a simulation with a two layered neural network which updated its weights via the RW rule. The input layer contained four units (S_1, S_2, S_3, & S_4), one for each symptom, while the output layer consisted of two units, one for disease R (R) and another for disease C (C). Input units could take on values of 1 or 0 depending on whether the represented symptom was present or absent in a given patient, respectively. The output units were linear ones.

A key aspect for an associationist-based account of this sort of base rate neglect is that the difference between the probability of S_1 given R and the probability of S_1 given C correlates with a difference between the relative validity of S_1 as a predictor of R and as a predictor of C. That is, S_1 is the most valid symptom to predict disease R, while it is the least valid symptom to predict disease C (see Table 10.1). So, on the basis of the RW learning rule, unit S_1 will end up with a stronger association with unit R than with unit C. If we keep this in mind, the application of our four step account to this situation is straightforward. First, participants rated the probability of statements relating contents that had formed part of a previous predictive learning situation. Second, during the learning stage, associations between such contents (S_1-R and S_1-C) developed allowing to predict diseases from symptoms. Third, the explicit presence of S_1 in the judgement required automatically and rapidly triggered a sort of knowledge of the predictive value of S_1 for disease R and for disease C. Fourth, such knowledge was accepted as the basis for the probability judgement.

Now, it is clear why we said before that relative validity effects produce some deviations from empirical conditional probabilities. However, such deviations are unimportant if we take into account the overall set of training trials. Gluck and Bower's model produces a qualitatively significant deviation from empirical probabilities only in trials in which the patient presents symptom S_1 as the only symptom[2]. These trials are very unfrequent in the overall set of training trials, and at the same time, the associative weights responsible for this deviation are also responsible for a greater accuracy in the remaining training trials. On the other hand, as will be shown in the

simulation's results, the neural network is very sensitive to category base rates. It could not be otherwise because neural networks that use error driven learning methods are trying to optimise predictions in the overall training set. Therefore, it could be claimed at the same time both that ALM is responsible for the above base rate neglect, and that this mechanism essentially produces Bayesian behaviour (McClelland, 1998). We think that this is an interesting point because it allows to conciliate two empirically based claims about probability judgement: a) on the one hand, probability judgements are plagued with irrationalities; b) on the other hand, humans are well calibrated in real life situations.

Finally, it is important to realise how crucial is the role of steps 3 and 4 in accounting for the above base rate neglect. According to such steps, the information that is made explicit in the judgement requirement makes ALM to generate an output which, in turn, strongly determines people's responses. This is also a key aspect because ALM cannot generate a response to the item $P(R|S_1)$ in a single step. Note that this item is not equivalent to test the neural network with the input vector 1000, which stands for the presence of symptom S_1 in isolation. S_1 in item $P(R|S_1)$ refers to the set of patients with symptom S_1, among which we can find patients who also present S_2 or S_2 & S_3 or S_3, etc. Thus, to obtain an adequate response from the neural network we should test it with many input patterns. Using mental models theory's terms (Johnson-Laird, 1983), this should involve creating a mental model for each possibility compatible with having S_1 to feed ALM and storing the result in working memory (see Shanks, 1990). Tversky and Kahneman (1983) refers to this as a decompositional strategy. But this is very hard when we have many possible mental models. Interestingly, according to the principle of cognitive economy of the mental models theory, the first mental model people create is determined to a large extent by the information that is made explicit in the premises and such mental model strongly influences people's responses (Johnson-Laird & Byrne, 1991; see also Johnson-Laird, Legrenzi, Girotto, Sonino Legrenzi & Caverni, 1999). Thus, steps 3 and 4 can be thought of as deriving from the mental models theory.

Regarding the conjunction fallacy, it has usually been found when an unfrequent event is joined with a frequent one. In these situations people tend to judge the conjunction of the two events as more likely than the unfrequent one. We can arrange a similar situation in Gluck and Bower's task. For example, suppose that people have to judge the probability of symptom S_4 in disease R patients and the probability of symptoms S_4 & S_1 in the same set of patients. As will be shown later, S_4 ends up with a negative associative weight with R and with a large positive weight with C, while S_1 ends up with a positive associative weight with R and a negative associative weight with C. As ALM is fed by the explicit information, the unfrequent constituent item would produce an output

[2] This is only true in Shanks (1990; Experiment 3). In Gluck and Bower (1988; Experiment 1), though the probability of R and of C were the same in patients with S_1, they were not in patients with S_1 as a unique symptom. In the later case, disease R was more likely than disease C. However, Shanks arranged training trials so that the probability of R and of C were also the same in patients with S_1 as a unique symptom. Despite Shanks' procedural modification, Gluck and Bower's neural network still predicts a bias toward disease R with an input vector consisting of 1000.

which is highly incompatible with being a disease R patient compared with the conjunction item. Thus, the conjunction item would be judged as more likely than the unfrequent constituent item. Of course, this would implicitly involve an inversion in the judgements direction. That is, while the items ask for the probability of symptoms given the diseases (D-S direction), ALM's outputs are based on S-D directed associations which allow to predict diseases from symptoms. But just because of ALM's functioning, knowledge about S-D directed relationships are readily accessible, while D-S directed relationships are not. In other words, because during the learning stage some stimuli are repeatedly processed so as to learn to predict diseases, people are biased to process those stimuli in the same way even if they have to judge D-S directed relationships in a later test stage.

We are not claiming, however, that conjunction fallacies in such cases are the byproduct of participants confounding the probability of symptoms given diseases with the probability of diseases given symptoms. We state, rather, that participants are biased to process the symptoms as the inputs of ALM, whose outputs exert an influence on probability judgements. The extent to which ALM's outputs determine probability judgements could depend, among other things, on participants' opportunities to detect normative violations of such outputs.

As we have tried to show, associative learning mechanisms could well be those rapid, effortless and feature-driven (content-addressable) processes envisaged by Tversky and Kahneman as responsible for biases in probability judgement. Furthermore, associative learning mechanisms can be conceived as an instantiation of some sort of representativeness heuristic. As knowledge from the different examples is superimposed to a large extent in the same weight matrix, associative networks develop a sort of abstracted prototype. Moreover, the outputs of this kind of networks are nothing more than the computation of the similarity between input patterns and the prototypes stored in the weights (Cobos & Almaraz, 1995)[3]. Finally, it could also be argued how these mechanisms provide a way of thinking of other heuristics such as accessibility and plausibility.

Evidence supporting the associative-based explanation of biases

If biases in probability judgement can be caused by cognitive illusions due to the use of contents that have taken part in a previous predictive learning situation, then these biases should be experimentally elicited by the use of a predictive learning task before the judgement stage. In fact, this is just what Gluck & Bower did. In their experiments, participants received information about the relationship between a series of events through a predictive learning task rather than through the usual verbal descriptions or numeric presentations in terms of likelihoods. Later on, participants were asked to judge the probability of a series of statements relating the same events. As mentioned before, participants judged the less frequent disease as more likely than the more

[3] Associative neural networks can, at the same time, store specific information of exemplars. Thus, learning from exemplars does not necessarily means abstracting a prototype at the cost of learning specific information of exemplars (McClelland & Rumelhart, 1985).

frequent one despite S_1 having been paired the same number of times with each disease. This is what happens in base rate neglect experiments, and this is why the phenomenon found by Gluck and Bower is called apparent base rate neglect. As far as we are concerned, Gluck and Bower's experiments were the first to establish a link between a very known fallacy in probability judgement and associative learning processes.

The apparent base rate neglect has been replicated in a good number of experiments using Gluck and Bower's task or similar procedures (Cobos *et al.*, 1993; Estes *et al.*, 1989; Kruschke, 1996; Myers *et al.*, 1994; Nosofsky *et al.*, 1992; Shanks, 1990). Among these works, there are three empirical evidences that strengthen the associationist explanation of this phenomenon. First, in all these experiments participants are well calibrated regarding empirical probabilities. Observed judgements seem to be quite sensitive to base rates. So there seems to be no room for an explanation based on ignoring base rates as suggested by Tversky and Kahneman's representativeness heuristic. The way in which Tversky and Kahneman conceive such heuristic is equivalent, in many cases, to applying the Bayesian formula without the base rates (Gigerenzer & Murray, 1987). As the simulation's results will show, sensitivity to base rates is just what should be expected from Gluck and Bower's neural network. This model fits reasonably well Bayesian probability calculus and predicts, at the same time, the apparent base rate neglect observed in participants' judgements in patients with S_1.

Second, the associative neural network has been used to obtain quantitative fits with impressive good results. Estes *et al.* (1989) obtained learning curves from participants' performance through the training stage. The learning task they used was the same as Gluck and Bower's except for some minimal procedural changes. All participants were trained with the same sequence of hypothetical patients. They also ran a simulation with Gluck and Bower's neural network using the same training sequence participants experienced to obtain learning curves from the neural network's performance. The simulation's results nicely fitted the observed curves. The fit provided by the neural network was even better than the fit provided by an exemplar-based memory model. Estes et al. also reported the results of a series of analysis showing that both participants and the neural network produced categorisation responses well calibrated with respect to Bayesian probabilities. We have also obtained quantitative fit results in our laboratory using the same task and the same model. Specifically, Cobos *et al.* (1993) recorded participants' probability judgements for every possible combination of symptoms at the test phase. The simulation's results were almost identical to those of Figure 3, where we can appreciate how closely the network approached the observed judgements.

Third, there is evidence supporting the explanatory status of relative validity regarding the apparent base rate neglect. For example, Gluck and Bower (1988; Experiment 2) altered the relative validity of S_1 and S_4 while holding the same programmed conditional probabilities of these symptoms. To achieve this, they changed the conditional probabilities of S_2 and S_3 so that the former became the best predictor of R and the latter became the best predictor of C. As a consequence, participants gave lower ratings for the

probability of R given S_1 and for the probability of C given S_4 than participants in Experiment 1. Moreover, the apparent base rate neglect tended to disappear. Similar results have been obtained by Kruschke (1996) and Cobos (1996). Apart from these base rate neglect designs, there are other experiments showing relative validity effects in probability judgements (e.g. Price & Yates, 1995).

We will show in what follows some evidence regarding the adequacy of an associationist-based account of biases in probability judgement. These experiments constitute, to a large extent, a replication of Cobos *et al.* (1993)'s, experiments. We used Gluck and Bower's learning task with some changes. For example, there were 160 training trials rather than 240. In each trial participants received information about the symptoms present in a hypothetical patient and made a diagnostic decision. After the diagnostic response, participants were provided with corrective feedback. We used the same programmed probabilities as Gluck and Bower, which can be seen in Table 10.1. However, as Shanks (1990), we also arranged training trials so that the probability of R and of C were the same in patients with S_1 as a unique symptom. With such arrangement, any apparent base rate neglect result would be very difficult to explain from conventional exemplar-based memory models or multiple-trace ones as MINERVA2. Hypothetical patients with no symptoms were also included in the learning phase. To simulate the results we used a modified version of the network described above (see Figure 10.2), including a single node in the output layer (see also Gluck & Bower, 1988). Now, the target output for the single node is +1 for disease R and –1 for disease C. The logistic function was used to transform the output activation values in to probability judgements as indicated below:

$$P(R) = \frac{1}{1 + e^{-c \cdot net}}$$

where *net* is the net input to the output unit and *c* is a constant free parameter. The training consisted of several epochs of 160 trials as it was done for the training of participants. A small learning rate was used to obtain weight values near asymptote ($l = 5 \cdot 10^{-5}$).

TABLE 10.1
Programmed probabilities of each symptom given each disease and of each disease given each symptom in Experiment 1

	Symptoms			
	1	2	3	4
P(symptom|common)	0.2	0.3	0.4	0.6
P(symptom|rare)	0.6	0.4	0.3	0.2
P(common|symptom)	0.5	0.7	0.8	0.9
P(rare|symptom)	0.5	0.3	0.2	0.1

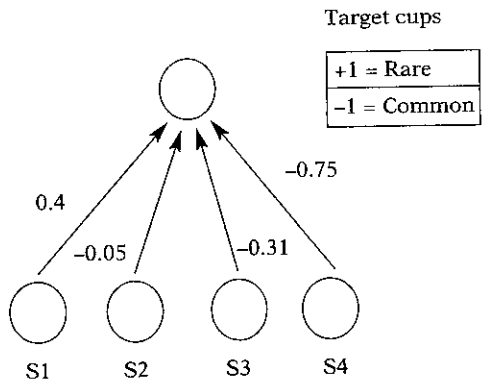

Figure 10.2. Asymptotic weight values obtained with the neural network used for the simulation.

In Experiment 1 we focused on four objectives: 1) to test whether the diagnostic learning task could induce the apparent base rate neglect bias in a later probability judgement phase; 2) to assess the extent to which category base rates affect participants' judgements; 3) to test whether the learning task situation could also induce the conjunction fallacy bias; 4) to assess the extent to which the neural network explains the whole set of data.

As explained above, we centered our attention on symptom S_1 to search for the apparent base rate neglect. We expected to replicate participants' bias toward indicating the presence of disease R in patients with S_1 found by other researchers. Figure 10.2 shows the asymptotic weights of the different connections. As can be seen, input node S_1 has a positive connection with the output node, meaning that symptom S_1 is more diagnostic of disease R than of disease C. As explained before, this result of the simulation is due to S_1 having more relative validity as a predictor of R than as a predictor of C. To test participants' sensitivity to category base rates, we recorded their ratings on the probabilities of R and of C given every possible symptom configuration so as to allow an analysis of category frequency on probability judgements. If we look at the weights in Figure 10.2, we can easily appreciate the network's sensitivity to base rates. Note that the sum of the absolute values of the negative weights is largely greater than the sum of the corresponding values of the positive weights. This means that, for the overall set of possible symptom configurations, the network will give much more disease C responses than disease R responses.

To search for evidence of conjunction fallacy biases, participants had also to judge the probability of some symptoms given each disease. Specifically, in patients suffering from R they had to judge the probability of the following items: S_4, S_4 & S_1, S_4 & S_2, and S_4 & S_3. Likewise, in patients suffering from C participants had to judge the probability of the following items: S_1, S_1 & S_2, S_1 & S_3, and S_1 & S_4. The conjunction fallacy would involve higher ratings

for the conjunction items than for the critical items S_4 and S_1 in the context of R and of C, respectively. As we mentioned in point 2, our central claim is that ALM will automatically generate outputs whenever the same cues (symptoms) and outcomes (diseases) constitute the judgements' arguments. This would be the case even if participants have to judge the probability of symptoms given the diseases, which is in the opposite direction of the associative links. To the extent that such output influences probability judgements, some conjunction fallacies should be expected. A look at Figure 10.2 again reveals that some conjunction fallacies are clearly expected, while others should not occur. For example, the input pattern 1001, which stands for symptoms S_1 & S_4, would produce an output which is much more consistent with suffering from R than the input pattern 0001. So participants should judge the conjunction S_4 & S_1 as more likely than symptom S_4 in patients suffering from R. On the other hand, the input pattern 1001 produce an output which is much more consistent with suffering from C than the input pattern 1000. So participants should judge the conjunction S_1 & S_4 as more likely than symptom S_1 in patients suffering from C. Another conjunction fallacy that could be expected consists in judging the conjunction S_1 & S_3 as more likely than symptom S_1 in patients suffering from C, though this is not as greatly expected as the others.

We went to considerable length to prevent participants from confusing the probability of a disease given a symptom with the probability of a symptom given a disease, and to prevent them from confusing the explicit absence of a symptom with the absence of information about the presence or absence of a symptom. Otherwise, the biases could be interpreted as simple linguistic misinterpretations. This was achieved via instructions, visual cues during the task that helped to discriminate, and filler items that compelled participants to discriminate between those kinds of items.

The median rating for the probability of R given S_1 was 0.63, which was statistically higher than the empirical probability 0.49. Consequently, participants biased their judgements toward the rare disease, as it was expected.

Figure 10.3 shows the mean ratings for the probability of R given each of the possible configurations except for the no symptom case. In this figure the simulation data obtained with the neural network model is also displayed as well as the judgements that should be expected if participants were absolutely insensitive to category base rates.

As can be seen, contrary to the representativeness heuristic explanation, participants' judgements, as well as the neural network's estimations, were strongly influenced by category base rates, for both were far from what could be expected on the basis of simple base rate neglect.

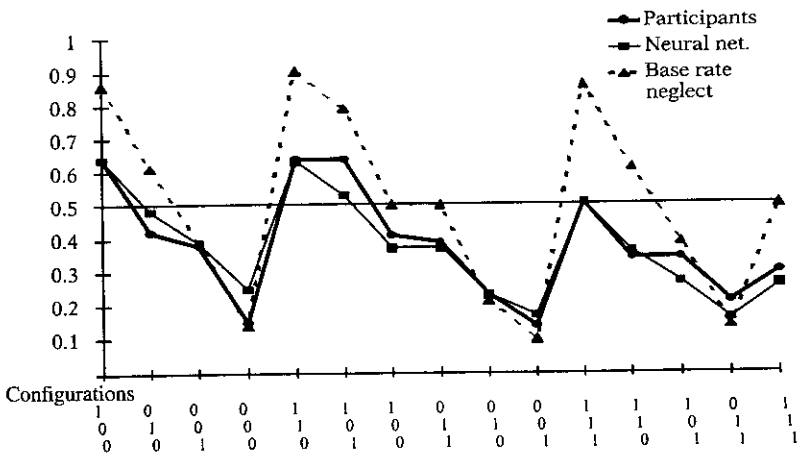

Figure 10.3. Participants' mean probability judgements, simulation data obtained from the neural network and judgements expected from neglecting category base rates in Experiment 1.

Finally, Table 10.2 shows the mean ratings regarding the test for the conjunction fallacies, as well as the net input for the output unit for each item.

TABLE 10.2
Mean ratings for the probability of the items concerning the test for the conjunction fallacy and the corresponding net inputs for the output unit in Experiment 1

Diseases					
R					
	Items	S_4	S_4 & S_1	S_4 & S_2	S_4 & S_3
	Judgements	0.26	0.45	0.26	0.32
	Net input	–0.75	–0.35	–0.80	–1.06
C					
	Items	S_1	S_1 & S_4	S_1 & S_3	S_1 & S_2
	Judgements	0.45	0.64	0.50	0.45
	Net input	0.40	–0.35	0.09	0.35

Regarding disease R patients, only the conjunction S_4 & S_1 received significantly higher ratings than the S_4 constituent. Regarding C, only the conjunction S_1 & S_4 received higher ratings than the S_1 constituent. Both conjunction fallacies were expected by the model. In the context of C, judgements for the conjunction S_1 & S_3 were not statistically different from

judgements for the S_1 item. However, it is also the case that the network does not predict the latter conjunction fallacy to the same extent as in the former cases. Specially remarkable are the differences found between the empirical probabilities and participants' probability estimates in those conjunction items that were rated as more likely than their critical constituents (S_1 & S_4 in R patients, and S_1 & S_4 in C patients). For instance, only 14% of the patients suffering from R presented both S_1 and S_4 whereas the mean probability judgement was .45. In patients suffering from C, 10% of the patients presented S_1 and S_4, while participants' mean probability judgement was .64. According to the net activation values shown in Table 10.2, we can easily explain why these estimations reached such high values. Note that the output node activation value with input pattern 1001 is –0.35. This value, which is not very extreme, is highly consistent with being a disease C patient. This explains why participants' ratings are above .5 in the case of disease C patients. On the other hand, such value is not extremely inconsistent with being a disease R patient, so the ratings obtained in disease R patients were close to .5.

Accordingly, we could say that we have answered the four objectives of Experiment 1. It has been shown that: 1) the diagnostic learning task may induce the base rate neglect bias in a later probability judgement phase (what constitutes a replication of other researchers' findings); 2) overall, participants' judgements are strongly affected by category base rates; 3) the learning task may also induce the conjunction fallacy bias; 4) the neural network used provides a reasonably good account of the whole pattern of results.

To explain the conjunction fallacies found in Experiment 1 we assumed that probability judgements in the D-S direction [P(S|D)] are based on the outputs generated by the associative learning mechanism, where links are in the S-D direction. We have carried out Experiment 2 to obtain convergent evidence supporting this assumption. We wondered to what extent judgements in the D-S direction were determined by the same mechanism as judgements in the S-D direction. Thus, we manipulated three within-subjects factors: 1) judgement direction (S-D vs. D-S), 2) category base rate (.75 vs. .25), and 3) the probability of a symptom given each disease (.2, .3, .4 and .6).

As we showed before, the associative neural network, as well as participants' judgements in the S-D direction, were influenced by category base rates. On the one hand, if the associative learning mechanism determines judgements in the D-S direction to the same extent than judgements in the S-D direction, we should expect the main effect of the conditional probability of symptoms and the main effect of category base rate, but neither the main effect of judgement direction nor the interaction between category base rate and judgement direction. On the other hand, normatively, we have no reason to expect a category base rate effect in the D-S judgements. Consequently we should expect the main effect of the conditional probability of symptoms, the main effect of judgement direction, and the interaction between category base rate and judgement direction. This interaction should be the result of having a category base rate effect in the S-D direction but not in the D-S direction.

The same learning task, apparatus and stimuli used in Experiment 1 were also used in Experiment 2. Table 10.1 shows the programmed probabilities

TABLE 10.3
Mean ratings for the probability judgements in the different conditions of Experiment 2

Disease C				
	Symptoms			
	S_1	S_2	S_3	S_4
D|S	0.30	0.57	0.58	0.91
S|D	0.33	0.62	0.58	0.92
Disease R				
	Symptoms			
	S_4	S_3	S_2	S_1
D|S	0.16	0.49	0.47	0.71
S|D	0.20	0.49	0.50	0.77

used in the present experiment. The design, as well as the results, are shown in Table 10.3.

According to the design, all participants had to judge the probability of 16 experimental items.

The statistical analysis revealed the reliable main effects of category base rate and of conditional probabilities of symptoms. Judgement direction was only marginally significant, and interestingly, the interaction between category base rate and judgement direction was far from statistical reliability. An important prediction from our proposal was that the category base rate should have an effect in the S-D and D-S levels. The analysis of simple effects confirmed our predictions. The simple effect of category base rate was significant in the S-D direction, as well as in the D-S direction. This is difficult to explain if we do not assume that the processes underlying the D-S judgements are strongly determined by a mechanism specialised in predicting diseases from symptoms. This is the case for the associative learning mechanism we have postulated, which is well suited for the predictive learning task in the S-D direction but not for the D-S judgements lately required. Therefore, these results support our claim that a predictive learning situation primes the ALM participation, which is an efficient algorithm at dealing with the predictive learning problem. But this happens at some cost because ALM is a content driven process and so can generate outputs by the contents of the predictive learning situation even in a new task where ALM is not an efficient process any more.

CONCLUSION

In Experiments 1 and 2 we have shown that predictive learning tasks can induce biases typically found with very different procedures (e.g. Kahneman and Tversky,

1973). This support our view that biases as the base rate neglect and the conjunction fallacy could, in some circumstances, be related to the way in which knowledge about categories and predictive relationships is acquired and represented in memory. We have also shown an interesting relationship between associative learning processes, which have been successfully employed to understand category and predictive learning, and biases in probabilistic categorisation tasks. Specifically, a simple neural network model has provided a good account of the whole pattern of results (see also Gluck & Bower, 1988 for similar results and theoretical interpretation). All this research shows relative validity effects and relates such effects with the phenomenon of the base rate neglect.

Finally, though our approach is indebted to Tversky & Kahneman's heuristics and biases approach, we would like to stress some differences to underline what our own contribution may be. The heuristics approach has been a general framework to account for a set of biases in probability judgements made by individuals in a wide range of situations. Within this framework, biases as the base rate neglect or the conjunction fallacy have been considered as the result of the operation of a series of intuitive and fast processes (heuristic) which enable the individual to solve efficiently and without effort certain tasks frequently found in everyday life. The presence of some features which are characteristic of these tasks in the probabilistic reasoning problems that participants must solve in the laboratory can activate such heuristics. This is a source of deviations from the standard normative theory that supports the experimenters' analysis of the reasoning task.

One of the main criticisms against this explanatory framework is its limitations regarding formal modelling and its lack of a detailed specification of the cognitive processes involved (Gigerenzer, 1996). A closely related issue is the dependency of the explanation and the description of the behaviour on the normative principles violated. This fact is manifested in different ways. For example, the characterisation of the base rate neglect and the conjunction fallacy from the heuristic approach has been mostly based on the lack of adjustment between participants' responses and the results obtained from the application of the normative theory by the experimenter. This has led the research program to a somehow paradoxical situation: to explain the non-normative behaviour we have nothing but the normative framework. Thus, though people neglect base rates, for instance, they are quasi-bayesian after all (Gigerenzer & Murray, 1987). Another example is the identification between the observed behaviour and the labels used to designate the violated principle. As a consequence, this fact has tended to conceal the diverse origins of the identified bias. The base rate neglect is a clear example of this situation (Koehler, 1996).

Unlike the heuristics approach, our explanation does not intend to cover all biases nor all the situations in which they occur. In fact, it is limited to a specific set of situations. In such situations, people make probability judgements about the relationship between a series of events that are similar or identical to those involved in a previous predictive learning situation. In this case, judgements can be affected by the intervention of an ALM. This mechanism solves efficiently the task of predicting relevant events from numerous cues that can occur at any time. However, the efficiency of these processes entails the

production of inadequate responses when different decisions have to be made about similar or identical contents.

On the other hand, our explanation shows a degree of precision and formal modelling which can neither be found in the heuristic approach nor in many other explanations of biases in probabilistic reasoning. Although this attempt to reach precision and formal modelling in our explanation of probabilistic judgements means sacrificing the range of those phenomena which can be explained, it also involves a deeper understanding of the origins of biases in probabilistic reasoning. First, because it enables us to specify more precisely the conditions in which biases are obtained in probability estimates. One of the criticisms launched by Gigerenzer against the heuristics and biases approach is the lack of precision in the specification of the conditions in which biases are obtained. Secondly, because the experimental data are much more informative to the extent that they can be interpreted in the light of theories specified as information-processing models.

Finally, another crucial difference between the heuristics and biases approach and our own lies in the underlying strategy on which our research is based. According to our account, biases such as the base rate neglect and the conjunction fallacy are, in some cases, the result of processes that do not fulfil the judgement task's goals but are effective in a different context. Such context has been identified here as the learning of predictive relationships between events and category learning may be considered as a particular case (Shanks, 1995). Thus, an adequate characterisation of those processes that induce biases requires the previous understanding of the processes involved in the learning of predictive relationships. In turn, this requires the identification of the environmental variables which are crucial to detect predictive relationships (see also López, Cobos, Caño & Shanks, 1998 for further details). Consequently, we come to the conclusion that to understand the processes that induce biases in probability judgements, the judgement task must be analysed within the framework of predictive relationship learning rather than only within the framework of the normative theory the reasoner violate (Bayes's theorem and conjunction theorem). Those processes that induce biases can only be identified from a previous analysis of the learning task for which they are designed. According to this, our explanation of the biases obtained in the judgement phase is completely determined by the analysis of the learning tasks participants had to solve previously. For instance, the direction of the predictions (S-D), as well as the relative validity of cues, are crucial features of the learning task to understand biases in the judgement phase.

To some extent, our research program falls into an *ecological approach*, which seems to be a present-day and promising view (it can be more explicit or implicitly appreciated in e. g. Anderson, 1991; Brase, Cosmides & Tooby, 1998; Fiedler, 1996; Gigerenzer & Goldstein, 1996). What is essential in this approach is that the analysis of the basic tasks that humans have to solve in their normal environment to survive is taken as a heuristic to understand the processes involved in laboratory tasks and to analyse such tasks.

Curiously, within the heuristics and biases approach, it was first thought that the processes that induce biases in probability judgements could be related

to categorisation processes and to the way in which knowledge about categories is represented in normal life (Kahneman & Tversky, 1973). However, the possibility of taking category learning as a framework was not taken any further to augment our understanding of such bias induction processes. Instead, normative theories, whose principles were violated by participants, have been considered the basis for the description of the processes responsible for the biases. This is probably one of the main reasons why the understanding of biases has advanced so slowly for such a long time, as Gigerenzer has pointed out (Gigerenzer, 1996).

ACKNOWLEDGEMENTS

This work is part of a research project supported by Junta de Andalucía (HUM0105).

REFERENCES

AGNOLI, F., & KRANTZ, D. H. (1989): «Suppressing natural heuristics by forma instructions: The case of the conjunction fallacy». *Cognitive Psychology, 21*, 515-50.

ALLAN, L. G. (1980): «A note on measurement of contingency between two binary variables in judgement tasks». *Bulletin of the Psychomic Society, 15*, 147-149.

— (1993): «Human contingency judgments: rule based or associative?» *Psychological Bulletin, 114*, 435-448.

ANDERSON, J. R. (1991): «Is human Cognition adaptive?» *Behavioral and Brain Sciences, 14*, 471-517.

BAKER, A. G., MERCIER, P., VALLÈ-TOURANGEAU, F., FRANK, R., & PAN, M. (1993): «Selective associations and causality judgments: the presence of a strong causal factor may reduce judgments of a weaker one». *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 414-432.

BENEDICT, J. O., & AYRES, J. B. (1972): «Factors affecting conditioning in the truly random control procedure in the rat». *Journal of Comparative and Physiological Psychology, 78*, 323- 330.

BRASE, G. L., COSMIDES, L., & TOOBY, J. (1998): «Individuation, counting, and statistical inference: The role of frequency and whole-object representations in judgment under uncertainty». *Journal of Experimental Psychology: General, 127*, 3-21.

CHAPMAN, G. B., & ROBBINS, S. J. (1990): «Cue interaction in human contingency judgment». *Memory and Cognition, 18*, 537-545.

CHENG, P. W. (1997): «From covariation to causation: A causal power theory». *Psychological Review, 104*, 367-405.

CHENG, P. W., & HOLYOAK, K. J. (1995): «Complex adaptive systems as intuitive statisticians: causality, contingency, and prediction». In J. A. Meyer, & H. Roitblat (Eds.), *Comparative approaches to cognition*, Cambridge, MA: The MIT Press.

COBOS, P. L. (1996): *Una aproximación conexionista al estudio de los errores de estimación de probabilidades en tareas de categorización probabilística*. Unpublished doctoral dissertation.

COBOS, P. L., & ALMARAZ, J. (1995): «Representación de conceptos y categorización: Un modelo conexionista». In M. Carretero, J. Almaraz, & P. Fernández (Eds.), *Razonamiento y Comprensión*, pp. 133-52. Trotta, Madrid.