Connectionism and Probability Judgment: Suggestions on Biases

Pedro L. Cobos
Dept. de Psicología Básica, Social y Organizacional,
Universidad de La Laguna
Campus de Guajara, La Laguna (Tenerife) 38200

Francisco J. López, Miguel A. Rando, Pablo Fernández & Julián Almaraz
Universidad de Málaga

Abstract

In the present paper we deal with several violations of normative rules in probability judgement: the inverse-base-rate and the conjunction fallacy, among others. To reproduce these failures, a sample of subjects was asked to judge the probability of several items according to what they had learnt in a previous learning task on medical diagnosis. Attempts are made to explain the results within the connectionist framework. We based our approach in a simple network, designed by Gluck and Bower (1988), which updates its weights using the LMS rule.

Introduction

In this work we have attempted to explain some of the typical errors produced during probability judgment. Among others, we studied here the conjunction fallacy and the base-rate-neglect.

We used the task designed by Gluck and Bower (1988a) to study these errors. In this task, the subjects had to learn to diagnose two diseases, a rare and a common disease, given the presence or absence of four symptoms. The task consisted of two phases. During Phase 1, the subjects had to diagnose some hypothetical patients according to a given set of symptoms. After each diagnosis, they received some feedback on the actual disease the patient suffered from. Table 1 shows all the probabilities that were programmed. The probabilities of suffering from the disease given the symptoms in isolation were fixed according to Shanks (1990).

Table 1. Programmed Probabilities of Each Symptom Given Each Disease, and Probabilities of Each Disease Given Each Symptom

	Symptom			
	1	2	3	4
P(symptom/common disease)	.2	.3	.4	.6
P(symptom/rare disease)	.6	.4	.3	.2
P(common disease/symptom)	.5	.69	.8	.9
P(rare disease/symptom)	.5	.31	.2	.1
P(common disease/symp. isolated)	.5	.66	.82	.95
P(rare disease/symp. isolated)	.5	.34	.18	.05

During Phase 2, the subjects had to use what they had learnt in the previous phase to make some probability judgments.

The model and the hypotheses

The connectionist model designed by Gluck and Bower (1988) has been used to explain the subjects' responses to the task (see figure 1). The model has two layers of units. The first level includes four input units, one for each symptom, and the second one includes an output unit taking different values according to the disease suffered. All the input units are connected to the output unit by links that can take different values or weights. Each input unit can take two values, 1 or 0, depending on whether the symptom that it represents is present or absent, respectively. The output unit takes an activation value according to the activation values of the input units, measured by the weights of the links between the output and input units. This is the internal

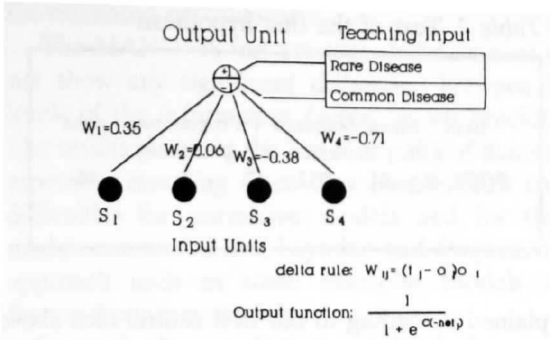

Figure 1. Simulation with the model by Gluck & Bower (1988a).

input. Therefore, the spreading direction of the activation always feeds forward. The output unit also can receive an external input that will tell the system the disease that the patient actually suffers from. If the input is equal to +1, the disease suffered is the rare one and if the input is -1 the disease is the common one.

The weights are updated after each feed-back using the Delta rule (Widrow & Hoff, 1960), equivalent mathematically to the Rescorla-Wagner rule (Rescorla & Wagner, 1972; Sutton & Barto, 1981). This rule is applied to the difference between the internal and external input in such a way that expected errors are minimized.

We postulate that, as the task progresses, this cognitive architecture appears as a base for the knowledge acquired by the subjects. It is important to explain that the processes underlying the subjects' predictions and probability judgments, based on such an architecture, can be described as single matching of patterns. We have an input pattern, formed by the activation state of the set of input units, and an output pattern, the value of the output unit, standing for the diagnosis made by the system. Our hypothesis derives from three central ideas:

1. Given the nature of the learning algorithm of this model, the final value of each weight depends on the relative validity of each symptom. Symptoms may be seen as cues that compete between them to predict each disease. Thus, when symptoms jointly appear in a hipothetical patient, the more valid symptom subtracts importance to the less valid ones, and this competition affects the magnitude of weight changes.
2. The architecture described earlier imposes on the subjects a series of limitations, such as the inability of representing the absence of information, since the input units take discrete activation values (1 for the presence of a symptom and 0 for its absence).
3. We also postulate that, given the limitation of working memory, the subjects will try to economize resources when solving an item. This means that the subjects would try to respond to the items of our tests with a single matching pair.

From these general hypotheses we elaborated the following predictions or specific hypotheses about the subjects' answers:

1. Subjects will judge the probability of suffering the rare disease greater than the programmed probability in patients with symptom S_1 given in isolation. This will be the case because symptom S_1 lacks better competitors to predict the rare disease and, on the contrary, is the worst competitor to predict the common disease. This result could be taken as analogous to the base-rate-neglect phenomenon because these probabilities were programmed to be the same (see table 1).
2. Subjects will judge the probabilities of suffering from the disease given the symptom the same as the probabilities of having the symptom given the disease. Since the system only works in one direction, from the input units to the output one, subjects will not be able to answer the items according to the probability of a symptom given the disease. Instead, they will respond to the probability of the disease given the symptom.
3. As the architecture does not allow for the representation of the absence of information, by simple matching of patterns, when we ask the subjects for the probability of suffering a disease given a symptom, and with no information about the other symptoms, they will be forced to treat the absence of information about the rest of the symptoms as the absence of the symptoms themselves.
4. We expect the presence of a conjunction fallacy in item $P(S_4S_1/R)$ in relation to $P(S_4/R)$ since the final weights obtained by the links between the units representing the symptoms and the output unit are those shown in figure 1. From this figure we can deduce that the activation of the output unit is higher when S_1 and S_4 are present than when only S_4 is present. On the contrary, the conjunction fallacy will not occur between items $P(S_4S_2/R)$ and $P(S_4/R)$ nor between items $P(S_4S_3/R)$ and $P(S_4/R)$.

Method

We carried out an experimental study to test these hypotheses. The experiment was a 2 x 2 within-subject factorial design, the sample size being 32

Four dependent variables according to the four symptoms

	Disease given symptom	Inverse probability
Absence of information condition	Example: $P(R/S_1)$	Example: $P(S_1/R)$
Complete information condition	Example: $P(R/S_1$ In Is.)	Example: $P(S_1$ In Is./R)

Figure 2. Experimental design: 2x2 within-subject factors.

subjects (see figure 2). The first factor (called the inverse factor) had two levels. In the first level, the subjects had to judge the probability of suffering from the rare disease given the presence of a symptom. In the second level, they had to make a probability judgment of having certain symptoms given the presence of the rare disease (this was called the inverse probability condition). The second factor included a complete information condition and an absence of information condition, depending on whether the subjects received information on the presence or absence of all four symptoms or only on the presence of one symptom (the symptom involved in an item).

Concerning the first hypothesis, we only have to compare the subjects' response to item $P(R/S_1$ isolated) with the relevant programmed probability (see Table 1). According to the second and third hypotheses, neither main effects nor interaction between factors are expected. Concerning the conjunction fallacy hypothesis, the same subjects were administered two kinds of items: 1) items referring to the conjunctions S_4-S_1, S_4-S_2 and S_4-S_3 in patients with the rare disease; 2) an item referring to the constituent S_4 in patients with the rare disease.

Results

Subjects' mean estimation on the probability of the rare disease in patients with only S_1 is shown in table 2. Other relevant statistics are also shown in this table.

Subjects judged the probability of suffering from the rare disease in patients with only S_1 greater than 0.5, i.e., the programmed probability. This result replicates that obtained by Shanks (1990), and is

Table 2. Test of the first hypothesis

Item	Mean	Standard error	t	Programmed probability	alpha
$P(R/S_1$ is.)	.61	.052	.52	.5	.05

explained according to our first central idea above mentioned. In figure 1 we can see that, as a result of our simulation with the conectionist model, the link between symptom S_1 and the output unit has a positive value. So, if only S_1 is activated, the output unit activation will approach a value of 1, which represents the rare disease.

According to our hypothesis about how subjects represent the absence of information about symptoms, we expected the subjects would judge the items referring to the presence of a symptom in isolation to be as likely as the items referring to the presence of that symptom, without specifying the presence or absence of the rest of the symptoms. In other words, we expected to find no main effect asociated to the information factor.

In figure 3, on the X-axis, we have represented eight items. Each of them is represented by two bars: A) the one on the left, refers to the presence of a symptom in isolation and B) the one on the right, refers to the presence of the symptom and the absence of information about the rest. The first four show the probabilities of patients with symptoms S_1, S_2, S_3, S_4 suffering the rare disease. The last four are related to the probabilities of suffering symptoms S_1, S_2, S_3, S_4 in those patients with the rare disease. The Y-axis represents the means of

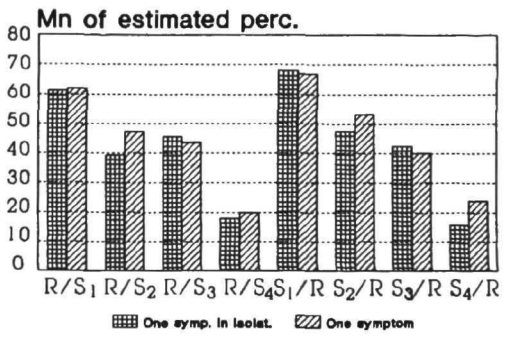

Figure 3. Hypothesis of the Representation of Absence of Information. MANOVA statistic test (Inf. effect: F = 1.28, Sig. = .302)

the estimated percentages.

The MANOVA test for repeated measures did not show any significant difference between the levels of the information factor, as we predicted. The results shown in the last four pairs of items are especially revealing since they clearly show some difficulties for normative models and for those models committed to a bayesian and frequentialist approach such as some exemplar models and feature frequency models.

One of the predictions derived from our hypotheses was that subjects would judge the probability of suffering the rare disease given a symptom as being the same as the inverse probability, that is, the probability of suffering the same symptom given the rare disease. In figure 4 we see again 8 pairs of bars. The left bar of each pair shows subjects' estimation of the probability of suffering the rare disease given the symptom below. The right bar of each pair shows the same subjects' estimations of the probability of having the symptom below in patients suffering the rare disease, that is, the inverse probability. The first 4 pairs of items refer to isolated symptoms, whereas the last four refer to items in which no information is included about the other symptoms. As our hypothesis maintained, no significant differences were found between the two levels of the inverse probability factor. In this case, the four first pairs are those which provide suggestive information, since they are hardly consistent with other alternative models, such as exemplar models or any other models engaged with the calculus of mathematical probability from relative frequencies.

Finally, we also predicted that no interaction effects between factors would take place. In figure 5 we show the differences, for each symptom, between the levels of the inverse probability factor in the complete information condition and the same difference in the absence of information condition. As we can see, the differences are very small. The MANOVA test for repeated measures did not show any interaction effect.

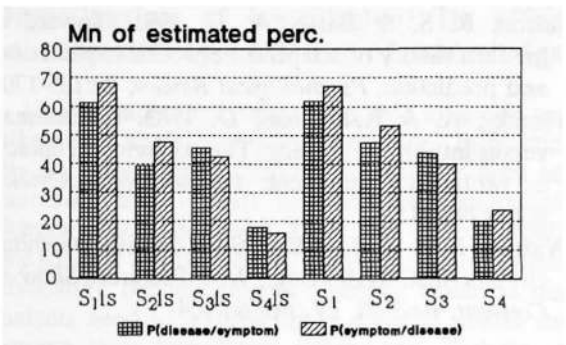

Figure 4. Hypothesis of the Inverse Probability. MANOVA statistic test (F = 1.65, Sig. = .188)

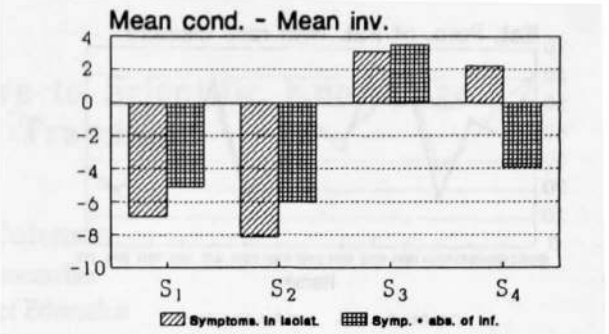

Figure 5. Information By Inverse Effect. MANOVA statistic test (F = .56, Sig. = .691).

Figure 6 illustrates the results of the conjunction fallacy test. The graph shows the results corresponding to the comparisons between judging the probability of suffering symptom S_4, in patients with the rare disease, and judgments made regarding to the different possible conjunctions between symptom S_4 and one of the other symptoms in patients with the rare disease.

The T-test for related samples revealed that only the probability of the conjunction S_4S_1 was judged significantly higher than the probability of S_4 - with

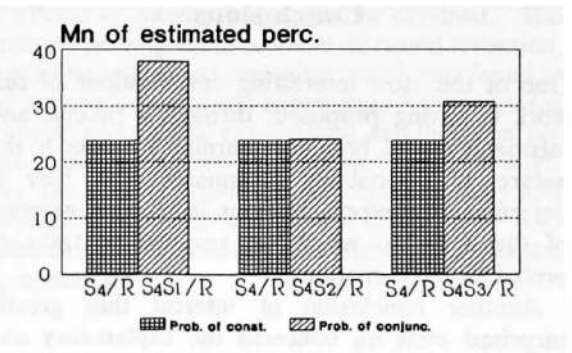

Figure 6. Test of the Conjunction Fallacy. T-paired test. Statistics, in the same order, are (t = 2.6; 1-Tail Prob. = .007), (t = .04; 1-Tail Prob. = .48), (t = 1.34; 1-Tail Prob. = .094).

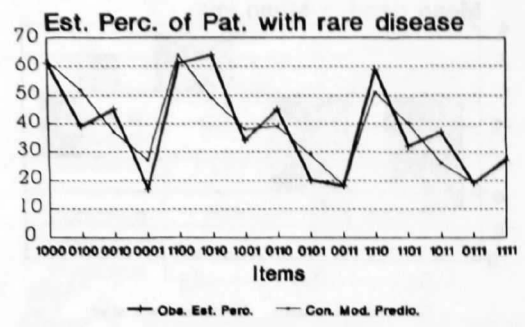

Figure 7. Connextionist Model Outputs and Subjects' Responses. Free parameter C = 1.4; SEd = 7.86; rp = .8748.

a significance level of .05. 67% of the subjects committed the conjunction fallacy when S_1 was in conjunction with S_4. On the other hand, when other symptoms where in conjunction with S_4, approximatly 50% of the subjects committed the conjunction fallacy. Again, the results in our work corroborate our hypothesis.

Finally, in figure 7 we show the results of our simulation. As can be appreciated, the connectionist model designed by Gluck and Bower fits very well the judments made by the subjects concerning the probability of suffering from the rare disease given all posible symptom configurations. Both the outputs of the model and the subjects' probability estimations were obtained at an asymptotic level of learning.

Conclusions

One of the most interesting contributions of this work is having proposed, through a precise and falsifiable model, both how learning is related to the nature of probability judgments and how it determines our representations, in working memory, of the items to which we respond in tasks of probabilistic estimations.

Another conclusion of interest that greatly surprised even us, concerns the explanatory and predictive power of such a simple model as that by Gluck and Bower. Something specially remarkable about this model is the fact that, without making substantial variations, it is able to accurately account for the results from different fields of study, such as biases in probability judgments , contingency judgments (Chapman & Robins, 1990), associative learning (Shanks, 1991), concept learning (Shanks, 1990; Gluck & Bower, 1988b), etc.

Finally, we think that this is a suggestive work since it shows a way of thinking about biases in probability judgment. Perhaps this approach does not cover the whole phenomena of biases in probability judgment, but it allows for accurate predictions and solutions to the problem of how content domains affect probability judgment.

References

Chapman, G. B. & Robins, S. J. 1990. Cue interaction in human contingency judgment. *Memory & Cognition,* 18(5):537-545

Gluck, M. A. & Bower, G. H. 1988b. Evaluating an adaptive network model of human learning. *Journal of Memory and Language,* 27:166-195.

Gluck, M. A. & Bower, G. H. 1988a. From conditioning to category learning: an adaptive network model. *Journal of Experimental Psychology: General,* 117(3):227-247.

Rescorla, R. A. & Wagner, A. R. 1972. A theory of pavlovian conditioning: variations in the effectiveness of reinforcement and non-reinforcement. In A.H. Black & W.F. Prokasy (Eds.), *Classical conditioning II: current research and theory.* New York: Appleton.

Shanks, D. 1990. Connectionism and the learning of probabilistic concepts. *The Quarterly Journal of Experimental Psychology,* 42A(2):209-237.

Shanks, D. 1991. Categorization by a connectionist network. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 17(3):433-443.

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

Tversky, A. & Kahneman, D. 1983. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychologycal Review,* 90(4):293-315.

Widrow, B. & Hoff, M.E. 1960. Adaptive switching circuits. Inst. Radio Eng., West Electron. *Show & Convent.* Rec. Pt. IV. pp. 66-104