Salience in Learning in a Probabilistic Environment

Stephen E. Edgell

University of Louisville

William P. Neace

Bellarmine University

J. Isaiah Harbison,

University of Maryland

and

Eric S. Brian

University of Louisville

The paradigm of multiple-cue probability learning (MCPL) has often been used to study behavior in a probabilistic environment. Behavior in a probabilistic environment is an important component of decision making. A subject in a MCPL experiment must learn through experience to use information, often called cues, to predict a criterion often called an event. A variant of this paradigm is one where the both the cues and the event are nonmetric. Castellan (1977) noted that in the paradigm of nonmetric multiple-cue probability learning (NMCPL) it is important to recognize the dimensionality of the cues. In the most common implementation of the paradigm of NMCPL, on each trial, a value of each cue dimension, randomly chosen according to probabilities that are parameters of the experiment, is presented to the subject. The subject then must predict what event will occur. After the subject makes a prediction of the event, the correct event is presented. Which event occurs on each trial is randomly determined according to a probabilistic relationship between the cues and the event, which is a parameter of the experiment. This procedure is repeated for many trials The subject must learn the relationship between the cues and event by this trial by trial feedback..

Often the experiment is presented to the subject as a medical decision making task. The cues are presented as symptoms or results of medical tests from a patient and the event as the disease the patient has. These may be given real names, made up names, or be totally abstract. NMCPL has been used to study many questions important to the field of decision making. These include the effects of irrelevant information (e.g., Castellan, 1973; Edgell & Hennessey, 1980; Edgell, et al., 1996), base rate neglect (e.g., Gluck & Bower, 1988; Medin & Edelson, 1988; Estes et al., 1989; Shanks, 1990; Edgell, et al., 2004)), utilization of configural information (e.g., Estes, 1972; Edgell & Castellan, 1973; Stockburger & Erickson, 1974 ;Edgell, 1978 & 1980; Gluck & Bower, 1990; Edgell & Roe, 1995)

In the typical NMCPL experiment, subjects response rates of the events given each pattern of cues quickly begins to more or less approximate that of the environment and asymptotes near that value during later trials. The exact level of that asymptote depends upon many factors. One rather surprising factor that affects the level is the physical representation of the cues. Many researchers in the field even now do not seem to be aware of this. Edgell and Morrissey (1992) discovered that the physical representation of the cues often resulted in different levels of utilization even if the cues were of the same validity. This was found even after hundreds of learning trials. This effect was found even though the cues and events in their experiment were very distinctive and easy to discriminate. The cues were also abstract. It had been previously found, in the similar experimental paradigm of concept formation (also called concept learning or concept identification) that the learning rate was affected by the physical representation of the cues (Vandierendonck, 1977). This paradigm is identical to NMCPL except that the relationship between the cues and event is deterministic rather than probabilistic.

Edgell, et al (1992) further studied this effect of salience. Those authors found a set of three binary valued cues that were equally salient. The three cue dimensions were equally salient to the degree that in an environment with one of the cue dimensions relevant and the other two irrelevant, the utilization was the same regardless of which cue dimension was the relevant one. Those authors also proposed an explanation for the salience effect and ran two experiments to test it. However, work by Kruschke and Johansen (1999) has contradicted that explanation. Later in the present paper we will further explore this issue.

A model by Castellan and Edgell (1973) that had enjoyed much success in accounting for the various findings from the NMCPL paradigm had no mechanism to account for salience effects. Kruschke and Johansen (1999) proposed a connectionist learning model, RASHNL, to account for findings in the paradigm of NMCPL including salience effects. The model is an extension of an earlier model, ALCOVE (Kruschke, 1992). RASHNL added limited-capacity attention that rapidly shifts and a learning mechanism for the distribution of these shifts. In the present paper we present three studies that further explore the effects of salience in NMCPL and test this model. We then further develop the model proposed by Castellan and Edgell (1973) to account for salience effects

.Experiment 1

Is one salience parameter for each dimension sufficient?

RASHNL, has several parameters that must be estimated from the data to obtain the best fit. Among these are one salience parameter for each cue dimension. Kruschke and Johansen (1999) noted, "For highly separable component dimensions, the salience of their configuration should, presumably, depend only on the saliences of their components. (p. 1087)" Their RASHNL model applies this. With the one salience parameter, for each cue dimension, RASHNL can account for the salience effects found by Edgell and Morrissey (1992) and by Edgell, et al (1992). We proposed to test if this is also sufficient in slightly more complex NMCPL environments.

It was noted by Kruschke and Johansen (1999) that only the relative values of the salience parameters have an effect on the model. The relative value of the salience parameter should be a function of only the physical representation of the cues. It should not depend upon the validity of the cue. RASHNL will be sensitive to the validity of the cue through trial by trial learning. To model the findings of Edgell, et al (1992), the three salience parameters, one for each of the three cue dimensions, of RASHNL must all be set equal. In the Edgell, et al (1992) study there were three cue dimensions, and which one was relevant was manipulated in three conditions. It should be noted that the cue dimensions of Edgell, et al. (1992), shape, color, and size of a geometric figure, are highly separable component dimensions. RASHNL, with equal salience parameters, would predict that in an environment with these same stimuli but with two of the cue dimensions relevant with equal validity they would be equally utilized. Further, RASHNL would predict that if any combination of two cue dimensions were relevant in one condition, the utilization would be equal to another condition where a different pair of cue dimensions was relevant.

Thus, it was proposed to test this critical prediction of RASHNL. A study like Edgell, et al (1992), using the three stimuli they found to be equally salient was designed. Each of the cue dimensions has two possible values, and the criterion also has two possible values. Two of the three cue dimensions were relevant while the third was irrelevant. The two relevant cue dimensions had a moderate validity. One value of each of the relevant dimensions predicted one of the two criterion values with probability .7 and the other with probability .3. The other value of that cue dimension predicted the same two criterion values in the opposite direction. That is it predicted the same one of the two criterion values with probability .3 and the other with probability .7. The two relevant cue dimensions functioned addively. That is when the value of both of the cues predicts a criterion value with probability .7 (.3), together they predict that criterion value with probability .9 (.1). If the value of one cue predicts a criterion value with probability .7 (.3) and the other cue predicts the same criterion value with probability .3 (.7), together they predict that criterion value with probability .5. Since the two cue dimensions function addively, they do not function configurally. Indeed there was no relevant configural information in the environments in this study.

These probabilities give a validity of .2 for each relevant cue dimension. This validity measure ranges from 0, which indicates irrelevant, to .5, which indicates a cue dimension that perfectly predicts the criterion. The details of this measure of validity are given in Edgell, 1978 & 1980, Edgell & Roe, 1995, and Kruschke & Johansen, 1999. Basically, it is a linear transformation using the same linear model as used in a factorial analysis of variance. Given that each criterion value occurs equally often (a base rate of .5), a .7/.3 cue has a validity of .2 because the two cue values add or subtract .2 from the base rate of .5.

There are three ways of choosing which two of the three cue dimensions are the relevant ones. Each of these possibilities was run in a separate condition. Condition 1 had color and shape relevant, condition 2 had shape and size relevant, and condition 3 had color and size relevant. Recall that RASHNL not only predicts equal utilization of the two relevant cue dimensions within each condition, but also predicts equal utilization across conditions for each relevant cue dimension.

Method

Subjects. The subjects were 118 college students who participated to satisfy a requirement in their introductory psychology class. They were randomly assigned to the three conditions.

Apparatus. The stimuli were presented on computer video displays, and responses were made using two button response boxes. The apparatus was the same as Edgell, et al. (1992) used, and it was in the same room.

Procedure. Instructions were read to the subjects who had copies to follow along as the instructions were being read to them. The instructions explained the task. The task was presented as a medical diagnosis task. Subjects were told that there were the results of three medical tests run on each patient who had one of two possible diseases. The outcome of each medical test and the diseases were coded abstractly to prevent using any previous medical knowledge. The test results were coded as a circle or a triangle, that was red or blue, and was larger or smaller. The diseases were called X and O. Theses are the same stimuli found to be equally salient by Edgell et al. (1992). Detailed descriptions of them may be found there. The instructions further stated "As with real world medical tasks, it is impossible to be correct 100% of the time. Patients with the same test results may not have the same disease." .

On each trial a patient's set of test results was presented on the screen and the procedure waited for a response to be entered by pressing the button labeled “X" or the button labeled “O.” The response made was immediately written on the screen using a short phrase, and 0.3 sec later the correct diagnosis was written on the screen using a short phrase. The test results, response, and feedback remained on the screen for 1.5 sec, after which the screen was erased. After an intertrial interval of 0.5 sec, the next set of cues was presented on the screen. This procedure continued for 400 trials. Because subjects were run in groups of up to four at a time, in booths screened from view of each other, subjects who finished their 400 trials before all the subjects were done were run for additional unrecorded trials until all the subjects were finished. The order of the trials was randomly permuted in blocks of 80 for each subject. Within each block of 80 trials the numbers of each type of trial (what test results were presented and what the correct disease was) gave exactly the desired conditional probabilities. Further, each of the 8 stimuli patterns occurred with equal frequency and each disease occurred equally often.

Results and Discussion

Subjects’ utilizations after learning are what is of interest. Proportions of subjects’ responses predicting one of the diseases for each of the eight patterns of cue values were calculated for each block of trials. An inspection of these revealed that subjects’ had clearly reached asymptote by the last trial block. Asymptotic response proportions from the last trial block were converted to utilization weights. This gave an asymptotic utilization weight for each subject for each of the two relevant cue dimensions. Utilization weights are calculated in the same manner as validity measures. They range from 0, which mean the subject was ignoring the cue dimension, to .5, which means the cue dimension is totally controlling the subject's responding. Mean asymptotic utilizations were calculated within conditions for each relevant cue dimension, and are shown in Table1.

As seen in Table 1, mean asymptotic utilizations were not equal as predicted by RASHNL. The best way to compare across conditions is to total the utilizations of the two relevant cue dimensions in each condition. These totals are also given in Table 1. They are not equal

Table 1

Asymptotic Mean Utilizations for Experiment 1

Cue Dimension	Condition		
	1	2	3
color	.130	-	.158
shape	.153	.132	-
size	-	.058	.083
Total	.284	.189	.241

(F(2,115)=3.080, p=.050). The difference between the two relevant cue dimensions in condition 1 was not significant (t(40)=.738, p=.465). However, the differences between the two relevant cue dimensions in condition 2 (t(39)=3.171, p=.003) and condition 3 (t(36)=2.412, p=.021) were significantly different. In an environment with two of three cue dimensions relevant, salience effects are more complex than in an environment with only one of the three cue dimensions relevant. Salience is not only a function of the physical representation of the cue dimensions, but the effect it has interacts with the validities of the cue dimensions. No model, such as RASHNL, with a salience parameter for each cue dimension will be capable of accounting for salience effects. A more complex mechanism will be required.

Experiment 2

The physical representation of the criterion also affects salience

Salience is usually thought of as a function of the physical representation of the cues. However, in a study where the cue was represented by color and the criterion was also represented by the same colors, Goodie & Fantino (1999) found that although utilization was affected by the environmental conditional probabilities of the criteria given the cues, there was a tendency for it to be affected by responding the same color. This is a new aspect to salience. Not only the physical representation of the cues but of the criterion also has an effect on utilization.

When Edgell et al (1992) were searching for three equally salient cue dimensions they started with size (a larger difference than the present), color, and shape (square and triangle), which were not equal in salience. The salience was size higher than color higher than shape. To reduce the salience of size, they reduced the size differential. To increase the salience of shape, they increased the difference in shapes to a three sided figure, the triangle, and an infinite sided figure, a circle. These changes worked and gave the stimuli used in Experiment 1. However, the criteria was coded as an X or O, which were chosen because of their familiarity from the children's game. When shape was relevant in their experiments as well as in the present Experiment 1, circle more often predicted O. The circle is about 5 to 6 cm and is colored, while the O is a regular computer O on the screen. Although there is not a direct relation as in the Goodie & Fantino (1999) study, but there could be some relation. An Experiment was run to explore the limits of the effect of the relationship between the physical representation of the criterion. The experiment was exactly like Experiment 1 but with the relationship between circle and O reversed. That is when relevant circle more often predicted X. The other two cues relationship was also reversed.

Method

Subjects. The subjects were 121 college students who participated to satisfy a requirement in their introductory psychology class. They were randomly assigned to the three conditions. These subjects were from the same pool as those in Experiment 1, and were run at the same time. Which experiment a subject was run in was random.

Apparatus. The apparatus was the same as in Experiment 1.

Procedure. The procedure was the same as in Experiment 1, except that subjects were run in groups of up to 5 at a time.

Results and Discussion

Proportions of subjects' responses predicting one of the diseases for each of the eight patterns of cue values were calculated for each block of trials. An inspection of these revealed that subjects' had clearly reached asymptote by the last trial block. Asymptotic response proportions from the last trial block were converted to utilization weights. Mean asymptotic utilizations were calculated within conditions for each relevant cue dimension, and are shown in Table 2. Comparing Table 1 with Table 2 it can be seen that the utilizations are similar except for those of shape. Ignoring condition, there was a significant difference between the total utilization in the two experiment ($t(237)=3.966$, $p<.001$). There was a significant difference for condition 1 between the two experiments ($F(2,79)=9.497$, $p<.001$). This was due to a significant difference in the utilization between the experiments due to shape ($F(1,80)=19.050$, $p<.001$), but not to color ($F(1,80)=.463$, $p=.498$). Similarly, there was a significant difference for condition 2 between the two experiments (F(2,76)=8.469, p<.001). This was due to a significant difference in the utilization between the experiments due to shape (F(1,77)=16.256, p<.001), but not to size (F(1,77)<.001, p=.983). Condition 3 showed no significance (F(2,75)=1.670, p=.195), for either color (F(1,76)=2.284, p=.135) or size (F(1,76)=.906, p=.344).

Table 2

Asymptotic Mean Utilizations for Reverse Criterion, Experiment 2

Cue Dimension	Condition		
	1	2	3
color	.149	-	.109
shape	.041	.027	-
size	-	.058	.057
Total	.190	.085	.166

Reversing the relationship between the representation of the criterion and the cues has a large effect on the utilization of shape. However, it has little or no effect on the utilization of color or size. Thus, the similarity between the physical representation of the criterion and the cues can be less than the same and still have a sizable effect on utilization. It should be noted that this effect did not go away after hundreds of learning trials.

Experiment 3

The Explanation for Salience Effects

Edgell et al (1992) postulated a reason for the effect salience has on utilization. They proposed that the physical representation affects a subject's memory for what stimuli appeared on each trial. If one misremembers the stimuli, that cue dimension will appear less valid. A full explanation of how the memory hypothesis accounts for salience effects is given in the Appendix. It is well known that a cue dimension with less validity is utilized less. Edgell et al (1992) tested the memory hypothesis and found an almost perfect negative correlation for the number of errors for remembering which cue value appeared and utilization of that cue dimension. Of course, this correlation only proves that memory errors and utilization covary. It does not confirm that one causes the other, although there is a reasonable causal mechanism.

Edgell et al (1996) did a definitive study. They noted that if the memory error hypothesis is correct then the salience of another, but irrelevant cue dimension would have no effect on the utilization of a relevant cue dimension. They first showed that an additional cue dimension will lower the utilization of a cue dimension, and that how much effect it has is a function of it's validity. If the additional cue dimension is irrelevant, its salience cannot affect its perceived validity. Regardless of memory errors the relationship between the cue and the criterion will still appear 50/50. An explanation of this is given by example in the appendix. If the memory error hypothesis is correct, the salience of an additional cue dimension would have no effect on the utilization of another cue dimension. If the memory error hypothesis is wrong and the salience effect is due to amount of shared attention allocated, then the more salient the additional cue dimension should result in less utilization of another cue dimension. They compared the utilization of a cue dimension when paired with an irrelevant cue dimension of low salience with one of high salience. They found only a slight a very nonsignificant difference as predicted by the memory error hypotheses. A replication found also a slight and nonsignificant difference, but in the opposite direction. They concluded that the level of salience of additional cue dimension did not affect the level of utilization of the other cue dimension as predicted by the memory error hypotheses and contrary to an amount of attention allocated hypotheses.

RASHNL, which uses an amount of attention mechanism to account for salience effects, predicts that a more salient irrelevant cue will have a larger effect on the utilization contrary to the findings of Edgell et al (1996). Kruschke & Johansen (1999) note that they had difficulty finding cue dimensions that would produce the effect RASHNL predicted. The ones they found were pairs of words. On each trial subjects saw two words, one from each pair. The order of the words on the screen which was varied randomly trial to trial. It is unclear if the subjects were even aware of the dimensional nature of the stimuli. However, those authors did find the effect RASHNL predicts contradicting the findings of Edgell (1996). Kruschke & Johansen (1999) argued that Edgell (1996) failed to find the effect due to low power.

RASHNL assumes that salience has an effect because of the allotment of attention between cue dimensions is affected by it. Thus, if there is only one cue dimension, it must get all of the attention. For salience to affect utilization, there must be at least two cue dimensions to compete. However, the memory error hypothesis predicts even with only one cue dimension perceived validity would be lower for a cue dimension with less salience. This would result in a lower utilization of it. We tested this prediction of RASHNL. Two conditions were run, each with only one cue dimension of moderate validity (.25). One condition used a cue dimension of high validity and the other of low validity.

Method

Subjects. The subjects were 78 college students who participated to satisfy a requirement in their introductory psychology class. They were randomly assigned to the two conditions.

Apparatus. The apparatus was the same as in Experiment 1.

Procedure. The procedure was the same as in Experiment 1, except for the following: The stimuli consisted in the high salience condition of a small and a much larger blue triangle and in the low salience condition of a blue triangle and a square. These stimuli was found to be of high and low salience, respectively, by Edgell (1992). There were 200 trials, permuted in blocks 40. The subjects were run in groups of up to 5 at a time in cubicles screened from each other.

Results and Discussion

Response proportions given the two cue values were calculated for each block of trials. An inspection revealed that subjects had reached asymptote by the third trial block. Response portions from the last two trial blocks were averaged together for each subject. Thus the asymptotic data are from 80 trials, as in the previous two experiments. Having more trials in the means, lowers the error variance. These asymptotic response proportions were converted to a utilization weight for the cue dimension. The means within the two conditions were very nearly equal: .242, for the high salience condition and .236 for the low salience condition. They were not significantly different ($t(76)=.144$, $p=.886$). This supports the prediction of RASHNL and the competition for attention account for the effect of salience that model proposes. It contradicts the memory error hypothesis.

Theory

We now take up the issue of a theory to account for the findings reported here on the effect of salience in NMCPL. Castellan and Edgell (1973) proposed a hypothesis generation model in which, on each trial, the subject forms a hypothesis of whether or not paying attention to a particular item of information will lead to a correct answer. The model proposes that the subject entertains hypotheses for paying attention to the event base rate (i.e., ignoring the cues), to each cue dimension, and to each pattern of cue dimensions. That is, on each trial the subject would hypothesize with some probability (which is a model parameter) that paying attention to the event base rate would lead to a correct answer and with one minus that probability that paying attention to the event base rate would not lead to a correct answer. In that same way, with a different probability, the subject would generate one or the other hypothesis for paying attention to the first cur dimension. This process would continue for each cue dimension and each pattern of cue dimensions. If the subject generated one and only one expectation of making a correct answer, then the subject would pay attention to the corresponding item of information on that trial. In any other case, the subject would generate a new set of hypotheses according to the same probabilities until reaching the state where one and only one expectation is positive. The model was proposed in two versions, but only the second is being considered here because it was found to fit better (Castellan & Edgell, 1973) and because the first version of the model predicts that subjects are unable to utilize relevant configural information. There is much evidence that subjects can and do utilize relevant configural information (e.g., Edgell & Castellan, 1973; Edgell, 1978; Edgell, 1980; Estes, 1972; Stockburger & Erickson, 1974).

The model further postulates that after the subject chooses which information to pay attention to, the value of that information on this trial is observed by the subject. Then, conditional on the value of the information, expectations are generated as to whether each of the two responses will be correct or not. Again, these expectations are generated with probabilities that are parameters of the model. The subject, as before, generates sets of hypotheses until one and only one response is expected to be correct. The subject then makes this response.

Castellan and Edgell (1973) proposed that the model parameter values (the probabilities of expecting to be correct if a particular item of information is used and the probabilities of expecting to be correct if a particular response is made conditional on the observed value of the item of information that was chosen to be observed) would be adjusted through learning trial by trial.

Although the mechanism of this learning adjustment was never specified, asymptotic values were hypothesized. These asymptotic values are those that would result from optimal use of the information by the subject. Because these values are functions of the environment probabilities, no parameters were estimated from subjects' data. No mechanism was included that would account for the effects of salience.

As was found in the present experiments and others, these effects are very complex. With the present level of understanding of these effects, the only possibility is to estimate the parameters from the data. The probabilities, of expecting to be correct if a particular response is made conditional on the observed of the item of information that on each trial was chosen to be observed (the s parameters in the model), we would not expect to be affected by salience. Thus we continued to use the asymptotic estimates of Castellan and Edgell (1973) for these. The probabilities, of expecting to be correct if a particular item of information is used (the z parameters in the model), should be a function of both the validity of that item of information and its salience. The probabilities of using each item of information (the w parameters of the model) are a direct function of the z parameters of the model. Because mathematically it is simpler, we chose to estimate these w parameters from the subject's data. For three cue dimensions as were used in Experiments 1 and 2 there are 8 such w parameters. However, they sum to one. Also, because the event base rate and one of the cue dimensions in each condition of those experiments was irrelevant the two w's for them are not separable. Thus there are 6 free parameters to be estimated. We fit the model to each condition of Experiments 1 and 2 using least sum of squared error to the asymptotic mean response proportions with a direct search method (Chandler, 1969). The model's predicted response proportions were then transformed to utilization weights. These model utilization weights for Experiment 1 are given in Table 3 and for Experiment 2 are given in Table 4. Comparing Tables 1 and 3 and Tables 2 and 4, it can be seen that the model fit extremely well.

Table 3
Model Predicted Utilizations for Experiment 1

Cue Dimension	Condition 1	2	3
color	.131	-	.158
shape	.153	.132	-
size	-	.058	.084

Table 4
Model Predicted Utilizations for Experiment 2

Cue Dimension	Condition 1	2	3
color	.149	-	.109
shape	.042	.027	-
size	-	.058	.057

The estimated values of using each item of information (the w parameters) are given in table 5 for Experiment 1 and Table 6 for Experiment 2. Recall that one cue dimension in each condition being irrelevant as well as the event base rate being irrelevant all the s parameters for them are .5. Thus only the sum of the two w parameters for them is reported. The values of the w parameters, as can be seen in Tables 4 and 6, show is no simple pattern that makes sense given the environments of the experiments.

Table 5

Estimated model probabilities for paying attention to each aspect of the stimuli for Experiment 1

Aspect	Condition 1	Condition 2	Condition 3
color	.046	*	.067
shape	.046	.198	*
size	*	.072	.000
pattern of color and shape	.165	.135	.149
pattern of color and size	.113	.046	.000
pattern of shape and size	.179	.046	.000
overall pattern	.146	.025	.367
combined irrelevant (*)	.036	.049	.421

Table 6

Estimated model probabilities for paying attention to each aspect of the stimuli for Experiment 2

Aspect	Condition 1	Condition 2	Condition 3
color	.121	*	.237
shape	.016	.035	*
size	*	.099	.088
pattern of color and shape	.042	.018	.046
pattern of color and size	.249	.046	.046
pattern of shape and size	.043	.034	.046
overall pattern	.046	.000	.000
combined irrelevant (*)	.485	.769	.538

One could also estimate the parameters of the RASHNL model separately for each condition. RASHNL for environments such as those in Experiments 1 and 2 has 9 parameters. Eight of these are free parameters as one of the salience parameters can be fixed and the other two estimated relative to it. However, finding separate estimates violates the philosophy underlying RASHNL. None of its parameters should vary as a function of the validly of the components of the stimuli in the environment. That is learned by the model through exposure to the environment over trials. The salience parameters are only a function of the physical representation of the cue dimensions. The w parameters of the Castellan and Edgell (1973) model, on the other hand are functions of both.

References

Castellan, N.J., Jr. (1973). Multiple-cue probability learning with irrelevant cues. *Organizational Behavior and Human Performance, 9*, 16-29.

Castellan, N. J., Jr. (1977). Decision making with multiple probabilistic cues. In N. J. Castellan, Jr., D. B. Pisoni, & G.R. Potts (eds) *Cognitive theory* Vol II (pp. 117-147) New Jersey: L. Erlbaum.

Chandler, J.P. (1969). STEPIT-Finds local minima of a smooth function of several parameters. *Behavioral Science, 14,* 81-82.

Edgell, S. E. (1978). Configural information processing in two-cue probability learning. *Organizational Behavior and Human Performance*, *22*, 406-416.

Edgell, S. E. (1980). Higher order configural processing in nonmetric multiple-cue probability learning. *Organizational Behavior and Human Performance*, *25*, 1-14.

Edgell, S.E., Bright, R.D., Ng, P.C., Noonan, T.K., & Ford, L.A. (1992). The effects of representation on the processing of probabilistic information. In B. Burns (Ed.), *Percepts, Concepts, and Categories: The representation and processing of information* (pp. 569-601). New York: Elsevier.

Edgell, S.E., & Castellan, N.J., Jr. (1973). The configural effect in multiple-cue probability learning. *Journal of Experimental Psychology, 100,* 310-314.

Edgell, S.E., Castellan, N.J., Jr., Roe, R.M., Barnes, J.M., Ng, P.C., Bright, R.D., & Ford, L.A. (1996). Irrelevant information in probabilistic categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1463-1481.

Edgell, S.E., Harbison, J.I., Neace, W.P., & Nahinsky, I.D. (2004) What is learned from experience in a probabilistic environment. *Journal of Behavioral Decision Making, 17,* 213-229.

Edgell, S.E., & Hennessey, J.E. (1980). Irrelevant information and event base rates in nonmetric multiple-cue probability learning. *Organizational Behavior and Human Performance, 26,* 1-6.

Edgell, S.E., & Morrissey, J.M. (1992). Separable and unitary stimuli in nonmetric multiple-cue probability learning. *Organizational Behavior and Human Decision Processes, 51*, 118-132.

Edgell, S.E., & Roe, R.M. (1995). Dimensional information facilitates the utilization of configural information: A test of the Castellan-Edgell and the Gluck-Bower models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1495-1508.

Estes, W.K. (1972). Elements and patterns in diagnostic discrimination learning. *Transactions of the New York Academy of Sciences, 34*, 84-95.

Estes, W. K., Campbell, J. A., Hatsopoulos, N., & Hurwitz, J. B. (1989). Base-rate effects in category learning: A comparison of parallel network and memory storage-retrieval models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 556-571.

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

Gluck, M. A. & Bower, G. H. (1990). Component and pattern information in adaptive networks. *Journal of Experimental Psychology: General, 119,* 105-109.

Goodie, A.S., & Fantino, E.(1999). Base rates versus sample accuracy: Competition for control in human matching to sample. *Journal of the Experimental Analysis of Behavior, 71*, 155-169.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22-44.

Kruschke, J.K., & Johansen, M.K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1083-1119.

Medin, D. L. and Edelson, S. M. (1988). Problem structure and the use of base-rate information from experience. *Journal of Experimental Psychology: General, 117,* 68-85.

Shanks, D. R. (1990). Connectionism and human learning: critique of Gluck and Bower. *Journal of Experimental Psychology: General, 1990,* 101-104.

Stockburger, D.W. and Erickson, J.R. (1974). Probabilistic discrimination learning with dimensionalized stimuli. *Organizational Behavior and Human Performance, 11,* 157-171.

Vandierendonck, A. (1977). The relation between cue salience and learning rate in concept identification. A test of the Trabasso-Bower model. *Acta Psychologica, 41*, 467-479.

Appendix

Memory Hypothesis

How the memory hypotheses explains salience effects can best be explained by example. Suppose we have an environment with one binary valued cue dimension (blue/red) and one binary valued event (X/O). Suppose blue and red occur with equal frequency (P(blue)=P(red)=.5).

EXAMPLE 1: A relevant cue

Suppose we have a 70/30 cue (P(X|blue)=.7, P(X|red)=.3) with perfect memory. You observe the following 4 joint occurrences with the relative frequencies as indicated :

Blue – X 35%

Blue – O 15%

Red – X 15%

Red – O 35%

The perceived validity is, as it should be, 70/30.

Now suppose there are 20 % memory errors: (60% of trials are remembered 40% are not with 20% randomly correct and 20% randomly incorrect).

Correctly remembered	misremembered	total Perceived
Blue – X 28%	Red – X 7%	Blue – X 31%
Blue – O 12%	Red – O 3%	Blue – O 19%
Red – X 12%	Blue – X 3%	Red – X 19%
Red – O 28%	Blue – O 7%	Red – O 31%

The perceived validity is only 62/38. It is easily seen that the higher the percentage of trials with memory errors, the lower will be the perceived validity.

EXAMPLE 2 An irrelevant cue

Suppose we have a 50/50 cue (irrelevant) with perfect memory. You observe the following 4 joint occurrences with the relative frequencies as indicated :

blue – X 25%

blue – O 25%

red – X 25%

red – O 25%

The perceived validity is 50/50 (irrelevant, but Ss are correct 50% of the time so they don't see it as irrelevant).

Now suppose there are 20 % memory errors: (60% of trials are remembered 40% are not with 20% randomly correct and 20% randomly incorrect).

Correctly remembered	misremembered	total Perceived
Blue – X 20%	Red – X 5%	Blue – X 25%
Blue – O 20%	Red – O 5%	Blue – O 25%
Red – X 20%	Blue – X 5%	Red – X 25%
Red – O 20%	Blue – O 5%	Red – O 25%

The perceived validity is 50/50, still the same. Regardless of what percentage of trials have memory errors; we still get the same perception.