CALIBRATION AND THE DECISION VARIABLE PARTITION MODEL

by

Mariam Smith

A Thesis Submitted to the Faculty of the

DEPARTMENT OF SYSTEMS AND INDUSTRIAL ENGINEERING

In Partial Fulfillment of the Requirements
For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

1 9 8 1

STATEMENT BY AUTHOR

This thesis has been submitted in partial fulfillment of requirements for a graduate degree at The University of Arizona. It is to be made available to borrowers of the University Library.

Brief quotations from the thesis are permissible, provided that the source is accurately acknowledged. Requests for permission for extended quotation from or reproduction of the thesis may be granted by the head of the department or the Dean of the Graduate College.

SIGNED: Mariam Smith

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

William R. Ferrell
W. R. FERRELL
Professor of Systems Engineering

April 13, 1981
Date

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

ABSTRACT

The thesis concerns the effect of base rates on calibration in true-false questions within the framework of a quantitative model for calibration. Two tasks were examined: one where the subjective probability of a statement being true was assessed, and a second where the statement was judged as either being true or false and the subjective probability of having made a correct decision was given. The mathematical predictions of the decision variable partition model state that in the former task an increase in the base rate yields an increase in the calibration probability. In the latter task no effect due to base rate is predicted. The results of the experiments confirmed the predictions. An unexpected effect of change in difficulty was observed even though an attempt had been made to control for difficulty across differing base rates. The decision variable partition model was found to be accurate in terms of prediction and statistically indistinguishable from the data in terms of fit.

INTRODUCTION

Decision making within uncertainty is a common phenomenon dealt with in every day life. One possible expression of such a decision process is by way of subjective probabilities. In some circumstances, a subjective probability, expressed numerically, may not have a frequency interpretation in the real world. An example of this occurs when assessing the probability of a unique event taking place. However, there are many instances where the aggregate accuracy of a collection of subjective probabilities may be corroborated. For example, one might assess the probability of an event occurring in the future such as a drought next year. The occurrence of such an event can be readily verified. When the subjective probability of an event is given as r, and such an estimate has been exercised over a large number of events, the ideal outcome would be the occurrence of the event exactly r per cent of the time.

The desirability of accurately stating the frequency of an event by means of a subjective probability can hardly be underestimated. This faculty is termed calibration. An individual is defined as being perfectly calibrated if, in the long term, for all events accorded a particular subjective probability, the frequency of occurrence equals that number accorded. Many papers concerning the topics of subjective probability and calibration have been written; an exhaustive review of the literature is presented by Lichtenstein, Fischhoff and Phillips (1977). Perhaps the most interesting point which the authors raised, apart from the general findings of calibration experiments, is that there exists as yet no proper theoretical framework into which the data fit. Psychological theories useful in explaining the underlying cognitive processes of probabilistic assessment are scant, and inadequate. There is at present only one quantitative model, developed by Ferrell and McGoey (1980), which attempts to explain the experimental data observed using general probabilistic axioms and statistical concepts.

The model, named the decision variable partition model, is not a representation of human probabilistic thinking in that it does not emanate from the psychological theory of cognitive processes. However, if the outcome of the model proves useful, and a sound model of the task is established, then the model must be said to provide a substantial contribution to the study of calibration.

The purpose of the present paper is to examine a particular aspect of calibration hitherto not reported, within the framework of the decision variable partition model. This is the effect of base rates in true-false tests. To this end, an experiment was designed and the results compared with the effects predicted by the model.

The significance of finding an adequate quantitative model is reflected largely in terms of practical implications. An ability to specify conditions under which calibration may approach optimal is both useful and important. In view of the fact that there exists a wide range of tasks which use probabilistic assessment, and that the range of possible responses is equally vast, the study was directed at two task types only. One of these was the probabilistic assessment of the veracity of a number of statements. The other was a decision of whether the statements were true, followed by giving a subjective probability of the correctness of the decision. A different discrete response set was used in each case.

BACKGROUND

In the following discussion the decision variable partition model is regarded only as applicable to true-false questions. Consider discrete propositions such as "Tucson is more populous than London", that fall into one of two sets: T, being the set of true propositions and F, being the set of false propositions.The sets, by definition, are mutually exclusive.

These propositions can be used in true-false questions eliciting subjective probability responses in two ways. One may ask for the probability that the statement is true (or false) or one may ask that the statement be classified as either true or false and for the probability that such classification is correct. The first type of task is called PT as the purpose is to assess the probability that the proposition is true. The second task is called PC since the probability of a correct response is assessed. The decision variable partition model assumes that consideration of a proposition, in conjunction with prior knowledge, yields a value y of a random variable Y corresponding to "apparent truth" or seeming veridicality of the proposition. The random variable is subsequently transformed into a decision variable X, appropriate for decision making with respect to the assessment of the subjective probability of a given proposition. In the PT case a true statement is assessed, in the PC case a correct choice of true or false response is assessed. Assignment of a numerical probability is then assumed to be made by associating values of the decision variable with acceptable probabilities in order to produce subjective probability responses, r.

The decision variable partition model of calibration is based on a normative model. Here, the responses r take on any value on [0,1], and should be the posterior probability of the event in question given the resulting value of Y. For the PT case, this is $p(T|y)$ and

$$r = \{L(y)p(T)\}/\{L(y)p(T) + [1-p(T)]\} ,$$

where the likelihood ratio $L(y) = f(y|T)/f(y|F)$. It is usually assumed that the distributions of Y conditional on T and F are equal variance normal. The appropriate decision variable to decide the veracity of a proposition is Y itself and so X=Y.

In practice, the response set $\{r_i\}$ is always discrete, be it for the respondents' convenience or experimentally predetermined. The model assumes that X is partitioned such that one value of r corresponds to each partition in such a way that larger values of r are assigned to larger values of X.

In the PC case, the normative model assumes that the initial decision of true or false is formulated depending on whether $p(T|y)$ is greater or less than 0.5. Assuming the equal variance normal distributions as above, Y is the appropriate variable with a cut-off value y_c such that $L(y) = p(F)/p(T)$. The decision variable for deciding the correctness C, of the resulting response is the distance between the obtained value y and the cut-off value y_c, thus $X = |Y - y_c|$. According to the normative model r is given as $r = p(C|x)$. Thus the expression for r is given as

$$r = \{L(x)p(C)\}/\{L(x)p(C) + [1-p(C)]\} .$$

The decision variable partition model however, assumes that X is partitioned as for the PT case, given discrete responses.

The calibration probability, $p(C|r_i)$ is the proportion of correct answers when the response r_i is given, p(C) being the proportion of correct response. Calibration itself is best represented as a graph of $p(C|r_i)$ and $p(r_i)$. Thus, perfect calibration occurs when $p(C|r_i) = p(r_i)$, yielding a straight line calibration curve.

A calibration, denoted by $\mathcal{C}$, can be described by six components and is in general represented as

$$\mathcal{C} = \{task, \{r_i\}, X, \{x^i\}, f(y|T), f(y|F), p(T)\} \quad . \quad . \quad (1)$$

Here, task represents task type, in this case either PT or

PC, and establishes the relation between 'true', T and 'correct', C. X is the decision variable, $\{x^i\}$ is the partition, $f(y|T)$ and $f(y|F)$ are the probability density functions of Y when the proposition is true and false respectively. Finally, p(T) is the probability that a proposition is associated with the true set. For simplicity, $f(y|T)$ and $f(y|F)$ are assumed to be distributed normally with means d'/2 and -d'/2 respectively and with unit variance.

The calibration probability is then found to be

$$p(C|r_i)=\frac{p(C)[F(x^i|C)-F(x^{i-1}|C)]}{p(C)[F(x^i|C)-F(x^{i-1}|C)]+p(\bar{C})[F(x^i|\bar{C})-F(x^{i-1}|\bar{C})]} \quad (2)$$

where $F(x|C)$ and $F(x|\bar{C})$ are the cumulative distribution functions of X when the response is correct and incorrect respectively.

There is evidence to suggest that the decision variable partition model fits experimentally obtained data rather well (Ferrell and McGoey, 1980). The stringent test is whether predictions made by the model are borne out by experimental data. The present paper reports the predicted effects of base rates, the proportion of true statements p(T), in the PT and PC tasks.

PREDICTIONS

The decision variable partition model accommodates a number of task types. Two in particular were addressed here, both dealing with discrete propositions. One of the tasks, PT, is to assess the probability that a given statement is true. The other task, PC, is to decide whether a proposition is true or false and subsequently to give the subjective probability that the decision is correct. The model predicts that there will be an effect on calibration when the proportion of true propositions, the base rate, alters in the PT task but not in the PC task. Typically, overconfidence is exhibited during a calibration task (Lichtenstein and Fischhoff, 1977). The prediction specifies that in the former task there will be a shift from overconfidence towards underconfidence as the proportion of true statements increases.

Mathematical view of the model prediction

The first task, PT, is the case where the probability of a statement being true is assessed. In general, calibration is denoted by C, and defined as in equation (1). It is desired to show that $p(C|r_i)$, the calibration probability, increases with an increase in $p(T)$, the base rate, independent of the type of distribution or the value of $p(C|r_i)$. The model design for the task is best represented graphically as in Figure 1.

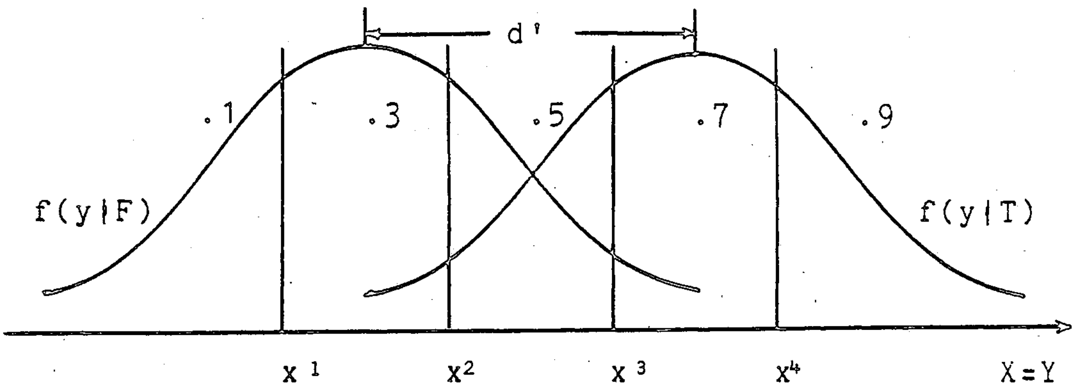

Figure 1. Model Design for the PT Task

Calibration for the PT task is denoted C_{PT} and is given by

C_{PT}= {PT,{.1,.3,.5,.7,.9},X=Y,{x^i},N(d'/2,1),N(-d'/2),p(T)},

where d' is estimated from the data and {x^i} is assumed to be fitted for p(T)=0.5. Essentially, if it can be shown that when d' remains the same but p(T) increases, then p(T|r) also increases, then the prediction can be proved. The assumptions made by the proof are illustrated in Figure 2.

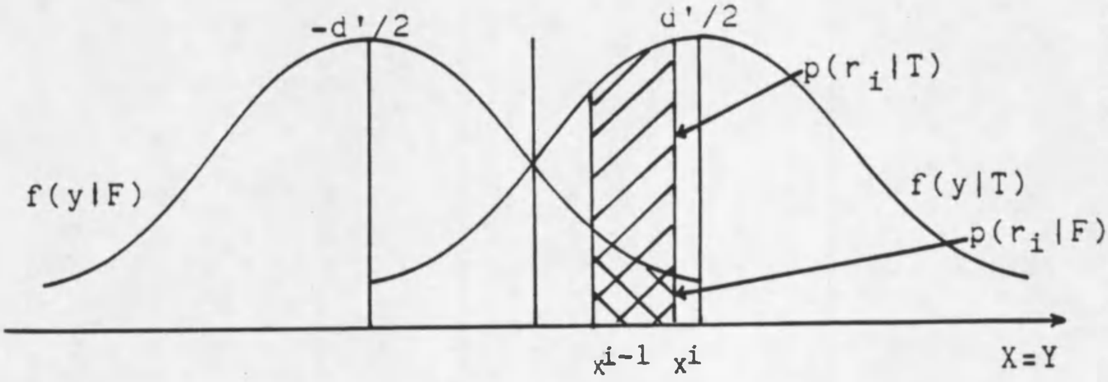

Figure 2. Assumptions implicit in Proof for PT Task

For a fixed value of d', $p(r_i|T)$ and $p(r_i|F)$ are also fixed. The calibration probability $p(T|r_i)$ may be expressed as

$$p(T|r_i) = \frac{p(r_i|T)\ p(T)}{p(r_i|F)\ 1-p(T)} \Bigg/ \frac{p(r_i|T)\ p(T)}{p(r_i|F)\ 1-p(T)} + 1 \qquad (3)$$

Now, $$p(r_i|T) = \int_{x^{i-1}}^{x^i} f(x|T)\ dx$$ and, $$p(r_i|F) = \int_{x^{i-1}}^{x^i} f(x|F)\ dx.$$

Let $L = p(r_i|T)/p(r_i|F)$ and $p = p(T)$.

Now equation (3) may be rewritten as

$$p(T|r_i) = \{Lp/(1-p)\}/\{1 + Lp/(1-p)\}$$

$$= Lp/\{(1-p) + Lp\} \qquad (4)$$

Further, let the ratios of the two values of $p(T|r_i)$ for different p(T) be given as $p = p_2/p_1$. Substitution into equation (4) then gives the expression

$$p_2(T|r_i)/p_1(T|r_i) = \{Lp_2/[(1-p_2)+Lp_2]\}/\{Lp_1/[(1-p_1)+Lp_1]\}$$

$$= \{Lp_2(1-p_1)+L^2 p_1p_2\}/\{Lp_1(1-p_2)+L^2 p_1p_2\}.$$

If $p_2 > p_1$, then $(1-p_1) > (1-p_2)$. The implication being that the numerator is larger than the denominator. That is,

$$Lp_2(1-p_1) + L^2 p_1 p_2 > Lp_1(1-p_2) + L^2 p_1 p_2 .$$

Therefore, $p_2(T|r_i) > p_1(T|r_i)$. Thus, as $p(T)$ increases, then so does $p(T|r_i)$ and the proof is complete.

The second task, PC, consists of giving a response of true or false to a proposition, and supplying the subjective probability of having given the right answer. As described in the previous section, the preliminary response decision is assumed to be made on Y using a cut-off value y_c. The respondent is assumed to have no knowledge whatsoever about the proportion of true statements. Since no feedback is given to the repondents about calibration performance, the respondent is assumed to use the cut-off value $y_c = 0$, which is the corresponding value for $p(T) = 0.5$.

The calibration for the PC task is denoted C_{PC} and is given by

$$C_{PC} = \{PC, \{.5, .6, .7, .8, .9\}, X = |Y|, \{x^i\}, N(d'/2, 1), N(-d'/2, 1), p(T)\}$$

where both d' and $\{x^i\}$ are to be determined from experimental data.

The distributions of X conditional on being correct and not correct, C and $\overline{C}$, are derived from those of Y conditional on true and false, T and F, with $X = |Y|$. Correctness is then defined as the conditions of $y < 0$ and which is indepedent of p(T). It should be noted that the result depends on two assumptions: first, that the two distributions on Y are identical in form, and second, that $y_c = 0$. In effect, the latter is assuming that the respondent is using p(T)=0.5. If the respondent were trained under a condition of p(T) not equal to 0.5 to be well calibrated, subsequent changes of p(T) without providing feedback would be expected to affect the calibration curve.

The decision variable partition model also takes into account the difficulty of the questions through the discriminability parameter d'. Let Y conditional on T and F be distributed as N(d'/2,1) and N(-d'/2,1) respectively. The parameter d' is then the distance between the means of the conditional distributions in units of their common standard deviation. This is the detectability of true from false. As illustrated in Figures 2 and 3, changes in d' will affect the calibration and response probabilities independently of the effect of p(T). If d' were kept constant under different values of p(T), it would be possible to estimate d' and $\{x^i\}$ from the p(T) = 0.5 case and from those parameters alone make quantitative predictions of the calibration curves and response frequencies for other values of p(T). If the difficulty is not constant over different p(T) conditions, then d' must be estimated from the data from those conditions in order to make the appropriate predictions. In the PC task, the ability to distinguish truth from falsehood is indicated by d'. The value of d' may be obtained using the experimentally determined values of p(C) from the relation

$$p(C) = \Phi(d'/2)$$

where Φ is the cumulative distribution function of the standard normal. For the PT task the value of d' was estimated from the ROC (receiver operator characteristic) curve obtained by treating the responses r as rating responses in a signal detection experiment (Sheridan and Ferrell, 1974, pp 369).

The model predictions are best illustrated by Figures 4 and 5 representing the PT and PC tasks respectively.

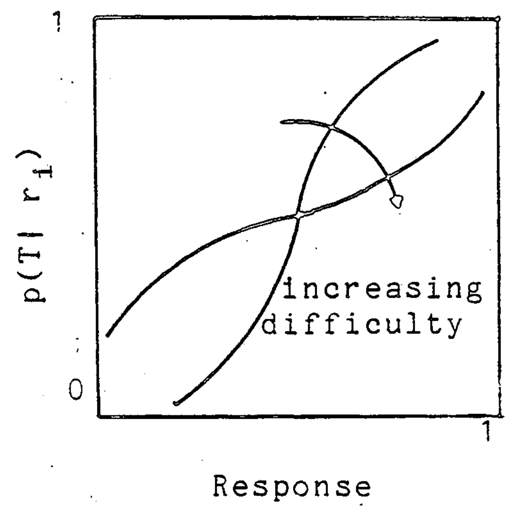

Figure 4. d' Prediction for PT Task

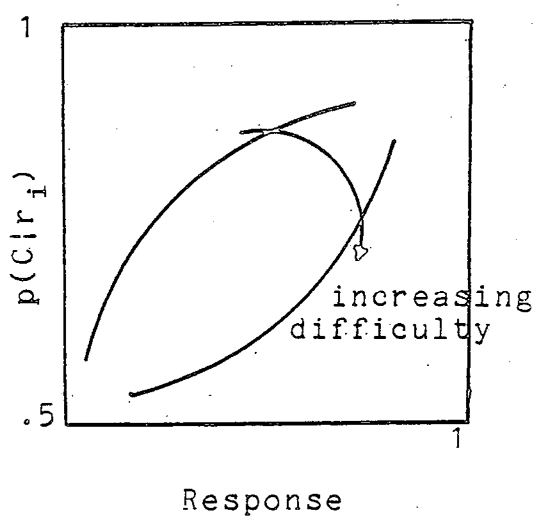

Figure 5. d' Prediction for PC Task

DESIGN

A total of twenty subjects, all volunteers, participated in the experiment. Fifteen of these were graduate students at the University of Arizona from various disciplines such as engineering, education and nursing. One subject was an undergraduate and the remainder were working adults between the ages of 25 and 35. There were about as many males as females. No specific criterion was used to select subjects. Ten subjects were randomly assigned to each of two tasks. The tasks were undertaken at each subject's convenience in terms of time and location. Each subject was given a booklet containing 600 propositions. The propositions related to population, distances between cities and historical events. Examples of each type are:

(1) "Europe is more populous than Asia";
(2) "Paris is further from Vienna than is New York";
(3) "Republican party formed before Monroe Doctrine declared".

Five hundred of the items were those used by Lichtenstein and Fischhoff (1980). The remaining hundred were constructed along the same principles. All items had an a priori classification of being either easy or hard, depending on the size of the difference on population, distance or time involved. The propositions were divided equally into three sections, with the same proportion of each type of item and difficulty level in each section. The base rate varied so that there were approximately 25, 50 and 75 per cent true statements in each section respectively. Subjects were not aware of the separate sections; indeed the test booklet was comprised of four parts. The intention was to keep difficulty constant over differing values of p(T). In order to ensure that word order had no effect on calibration performance, the entire set of propositions was reversed in one task. Thus, the first example given above would read "Asia is more populous than Europe" in one of the tasks.

Procedure

For the PT task, subjects were asked to give their subjective probability that each statement was true. The allowable response set was {.1, .3, .5, .7, .9}. For each proposition then, subjects were asked to mark one of the possible responses on an answer sheet in reply to the premise that the proposition is true. Each subject gave a total of 600 responses.

Subjects allocated to the PC task gave two responses per statement. They were required to give a yes/no answer to the question "Is the proposition true?", then give their subjective probability that their response was correct. The response set for this task was {.5, .6, .7, .8, .9}. Again, all responses were given on a separate answer sheet.

Measures were obtained for each section separately. In the PT task, the frequency of each response was obtained for true statements and for false statments. In the PC task, the frequency of each task was obtained for correctly and incorrectly answered questions. This gave the calibration probabilities and the response frequencies. Calibration curves were obtained for data summed over all subjects from each task. As indicated by the proportion correct for the three sections of the PC task, and contrary to intention, difficulty varied systematically with section. As a result, the discriminability parameter d' was calculated for each p(T) for both tasks as previously described.

RESULTS

The results of the experiment are given as graphs of calibration probabilities and proportion of responses against response categories. The difficulty associated with each base rate within a task, as indicated by the value of d', as well as the proportion correct are given in Table 1.

	pooled data			typical individual's data		
	p(T)	d'	p(C)	p(T)	d'	p(C)
task						
PT						
	.25	1.05	.25	.25	1.19	.25
	.56	.78	.56	.56	1.05	.56
	.68	.51	.68	.68	.62	.68
PC						
	.28	.54	.61	.28	.74	.61
	.45	.93	.68	.45	1.11	.71
	.74	1.20	.73	.74	1.54	.78

Table 1. Experimentally obtained values of d' and p(C)

It is clear that the difficulty associated with each value of P(T) is markedly different. The calibration curves therefore reflect these differences in addition to those due to p(T).

The data obtained for each task were pooled so that fairly smooth calibration curves resulted. Individual subjects need answer a very large set of questions before such smoothness is attained. Data obtained from individuals showed the patterns inherent in the pooled data, though some subjects were much better calibrated than others throughout the task. A "typical" subject was chosen from each task to show that there was a similarity in the calibration curves plotted. Given sufficient data for any one subject, the results should be the same as those obtained from the pooled data.

PT Task Results

Figure 6 illustrates the different calibration curves obtained for pooled data with changes in d' and in base rate. There is a marked separation between the curves, with p(T) = 0.25 exhibiting considerable overconfidence. As p(T) increases, there is a corresponding increase of the calibration probabilities, but differences attributable to d' are also included.

The case of p(T) = 0.56 was used to fit the model to the data. This was achieved on an iterative trial and error

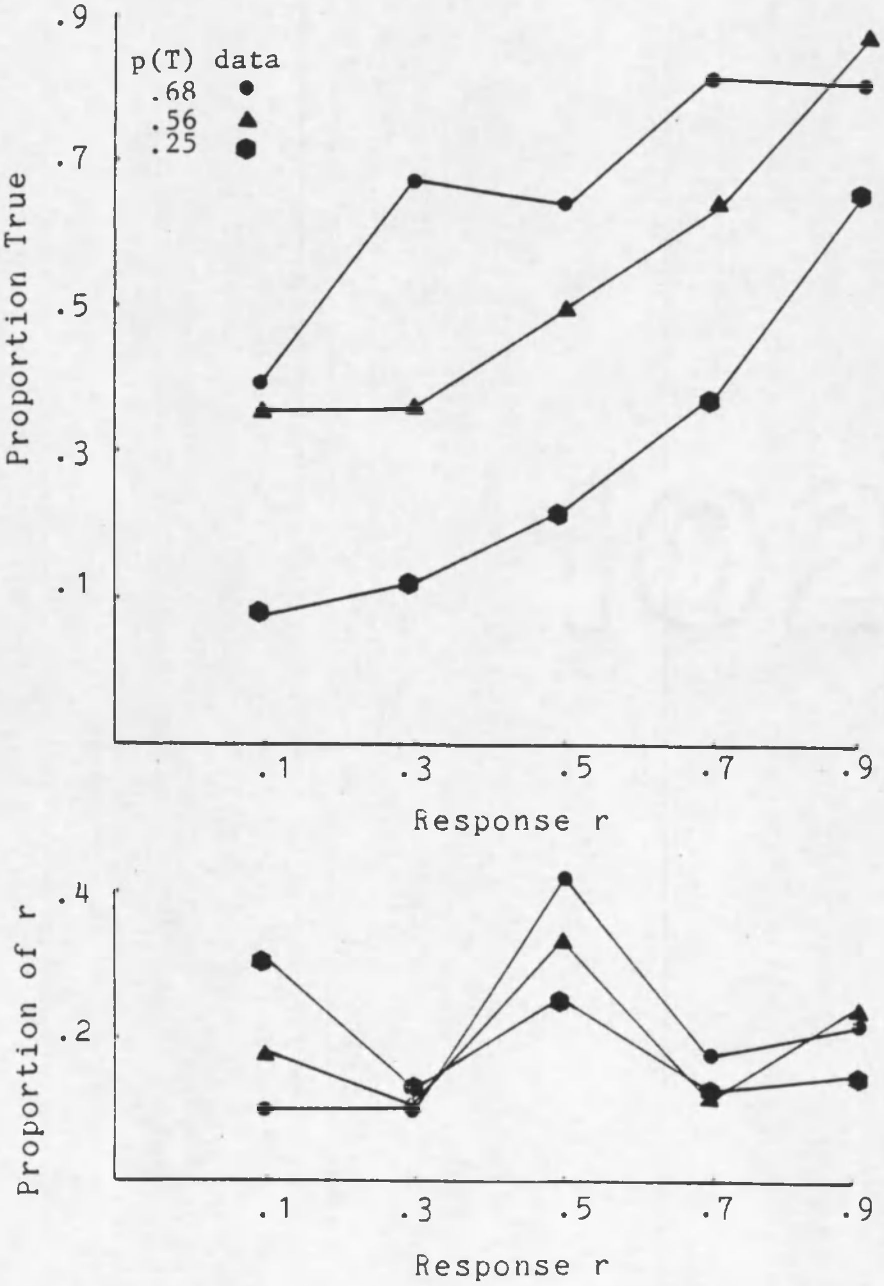

Figure 6. Obtained Data for PT Task (pooled data)

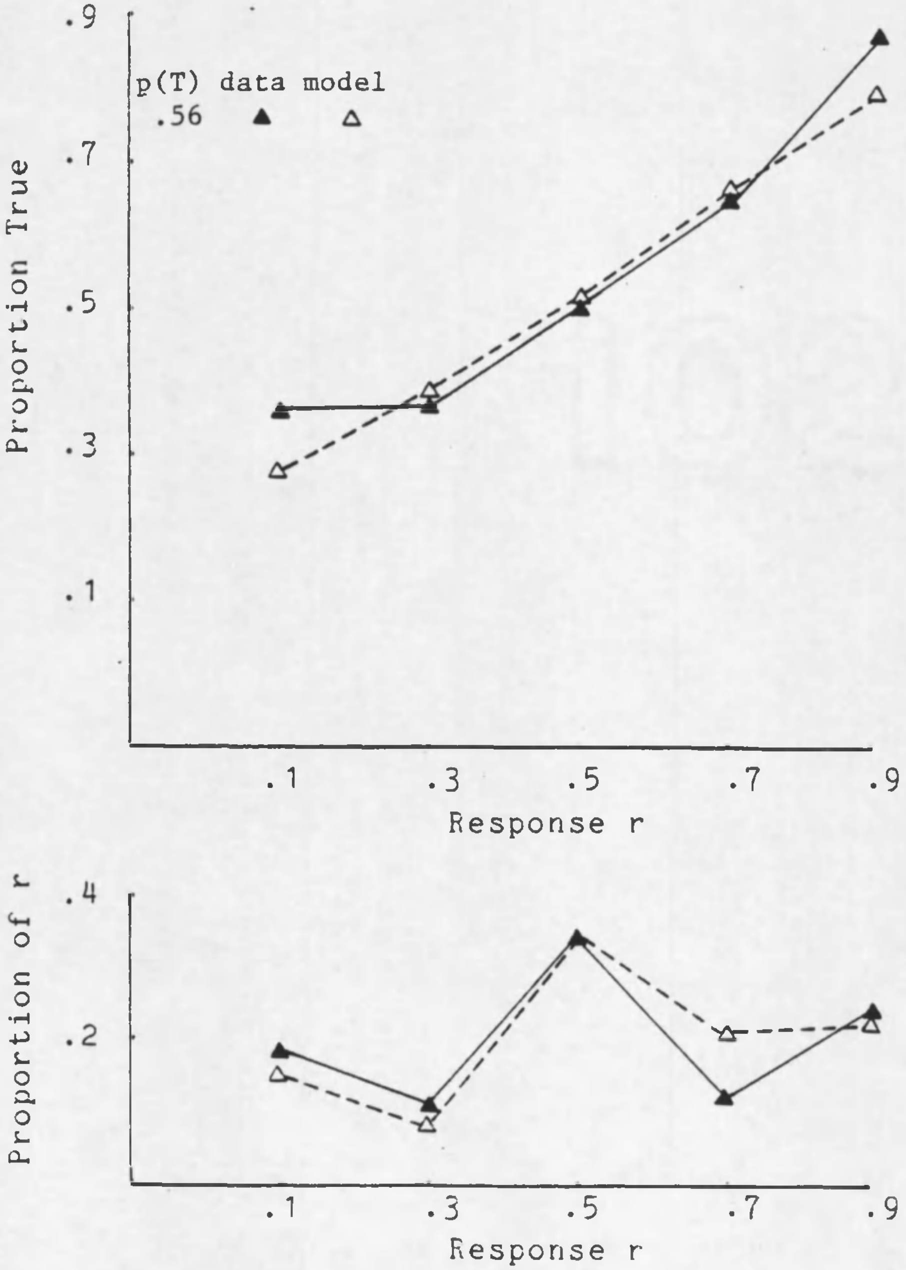

Figure 7 Fitted Curve for PT Task (pooled data)

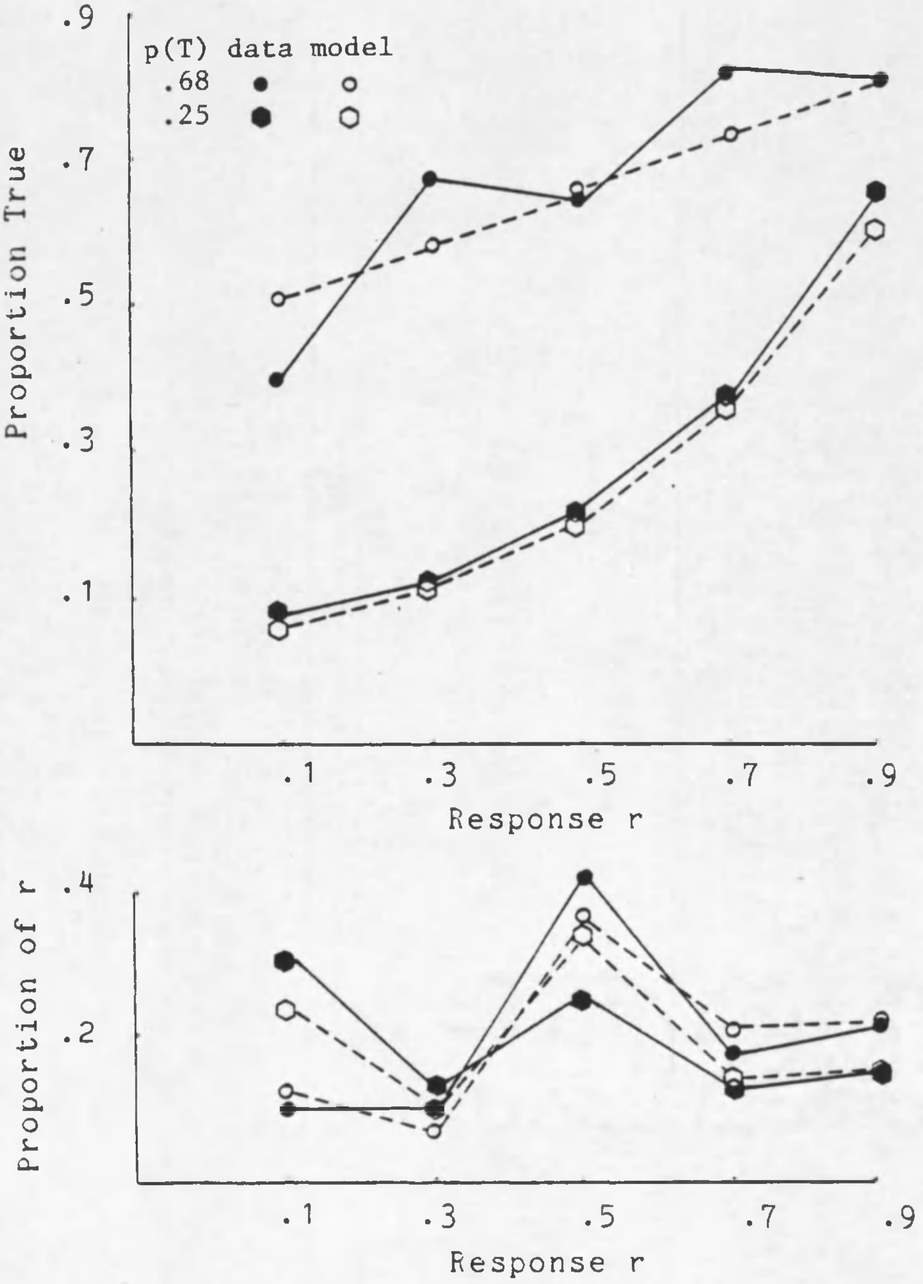

Figure 8. Predicted Curves for PT Task (pooled data)

basis to yield cut-off values of $\{x^i\}$ which would give appropriate values of $p(T|r_i)$ and $p(r_i)$ corresponding to each response. The fitted model is compared to the data in Figure 7. It was possible to obtain a reasonably good match. A Chi-square goodness-of-fit test indicated no difference between the data and the fitted model at a significance level of 0.01.

Having achieved the cut-off values $\{x^i\}$ which provided a good fit to the data, the predictions for two other base rate values, $p(T) = 0.25$ and $p(T) = 0.68$ and their corresponding values of d' were made. The predictions are compared to the experimental data in Figure 8. The Chi-square goodness- of-fit test indicated no difference between the two lines for both base rate values at the 1% significance level as expected from visual examination of the calibration curves. It should be noted that the model will always predict a smoother calibration curve since it assumes asymptotic distributions.

The above results were essentially replicated in the case of a typical subject in Figures 9, 10 and 11. As previously mentioned however, individual calibrations are given to considerable fluctuation unless a very large number of responses are made. From the results shown here, this number exceeds 200 responses. Suffice it to demonstrate that the same trends as for the pooled are apparent for each

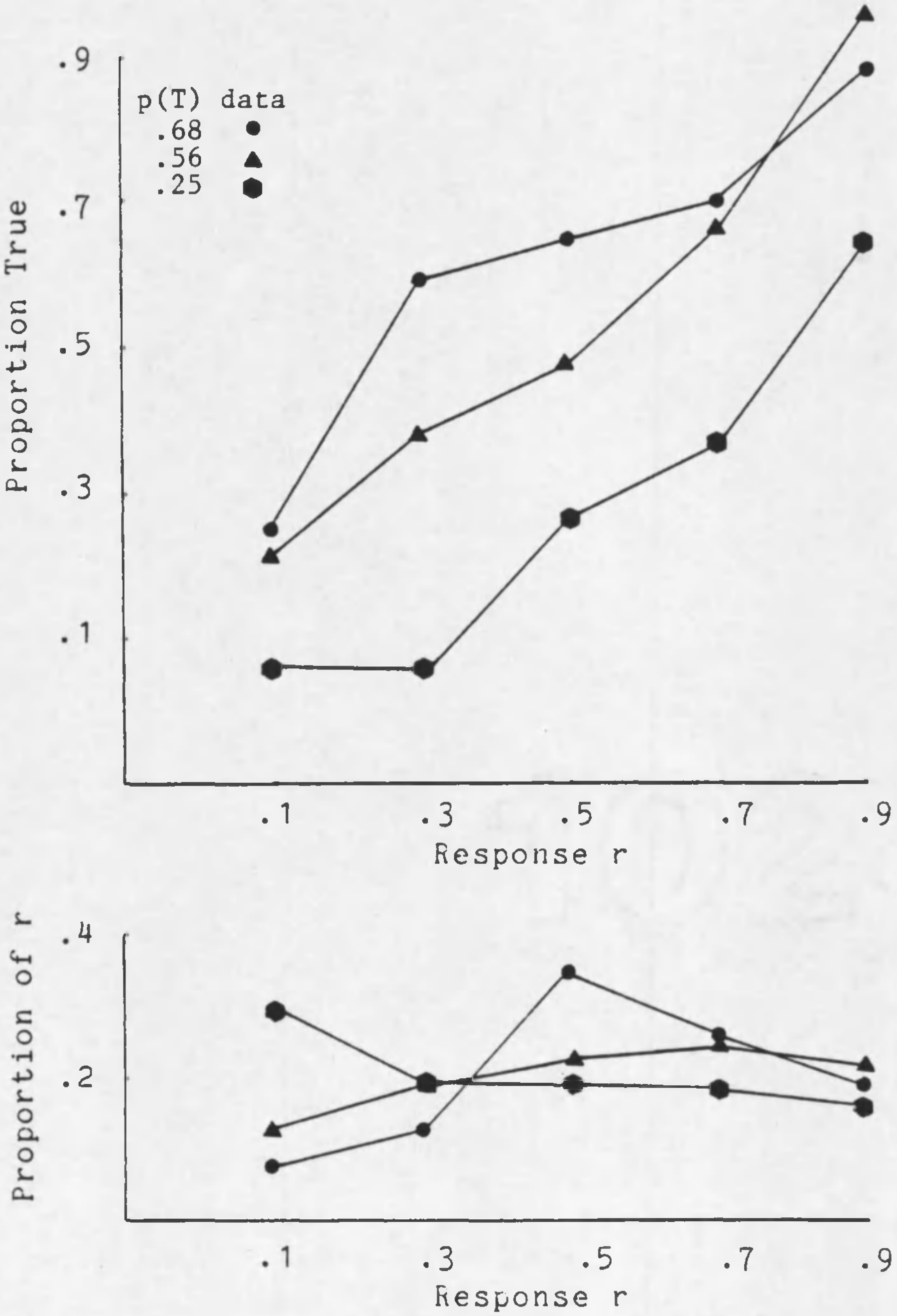

Figure 9. Obtained Data for PT Task (individual)

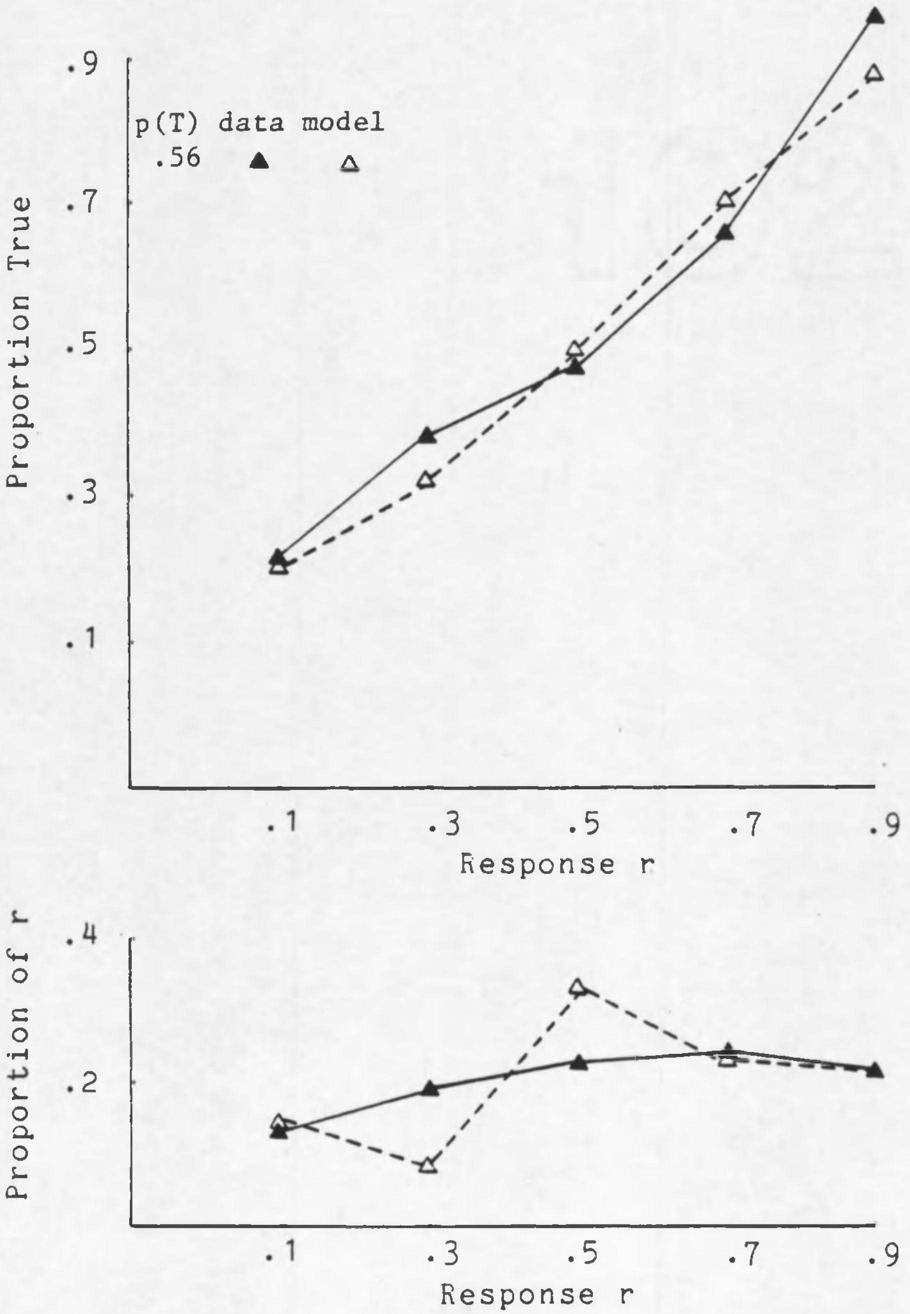

Figure 10. Fitted Curve for PT Task (individual)

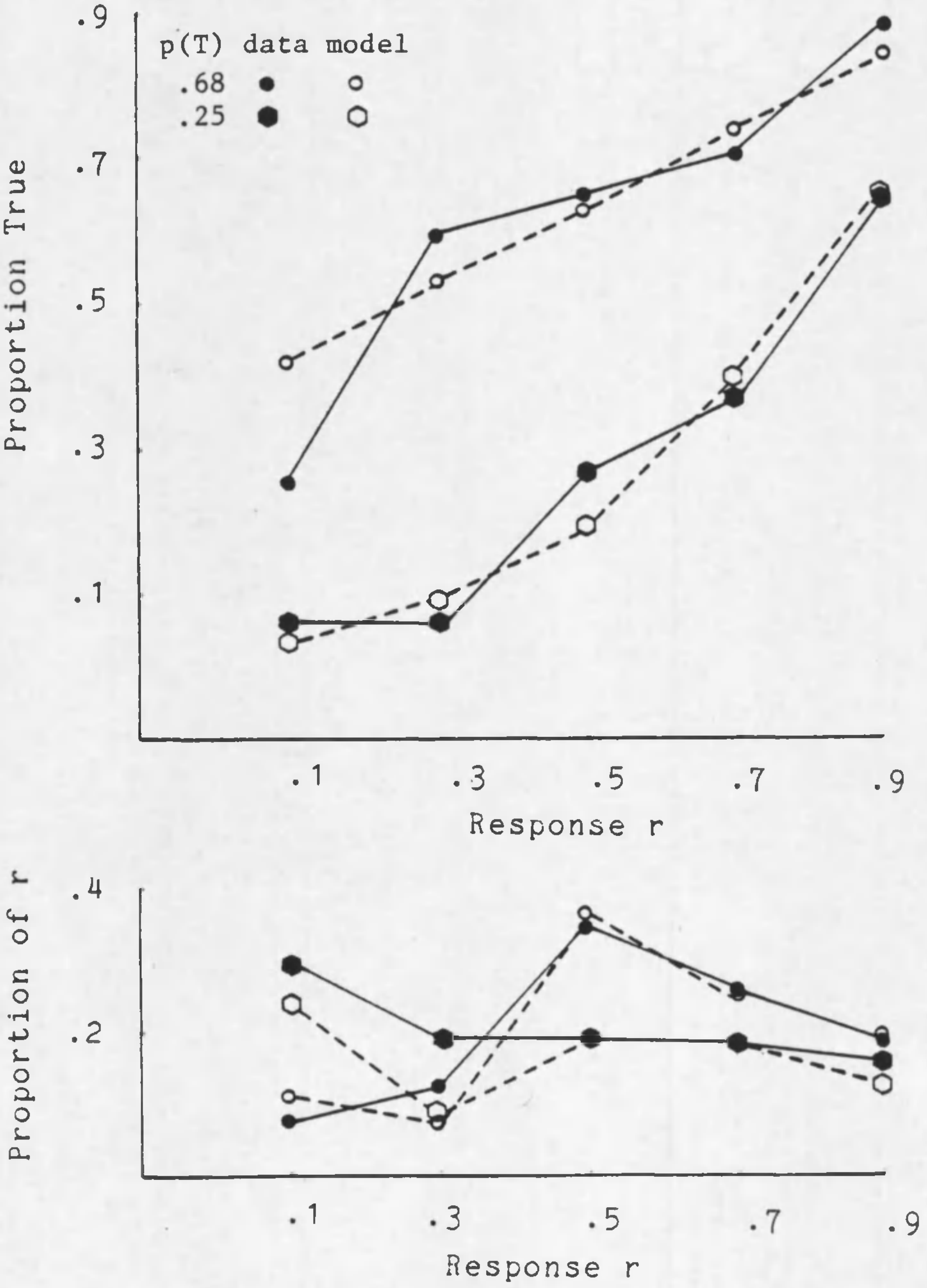

Figure 11. Predicted Curves for PY Task (individual)

base rate calibration, and, although neither the fit nor the predicted data lie as close to the experimental data as was the case for the pooled data, they are nevertheless impressively close.

The model was also used to demonstrate the effect of base rate alone, without any variation in d'. Given the fitted model for the pooled data for p(T) = 0.56, predictions were made for p(T) = 0.25 and p(T) = 0.68. Figure 12 shows the differences between the model predictions for constant d' and for the obtained data. A clear separation of calibration curves for the three base rates is appparent. There is no question that the predicted effect of p(T) is represented in the data.

PC Task Results

By comparison to the PT task calibration curves, those of the PC task show very little separation. The existing separation can be attributed entirely to the discrepancy in the value of d' across the different base rates. The calibration curves for the pooled data are given in Figure 13. As before, the model was fitted on a trial and error basis for the median base rate value and once the cut-offs were obtained predictions were made for two other curves. The fit and the predictions are illustrated in Figures 14 and 15 respectively. The Chi-square goodness-of- fit test showed no difference between the model data and

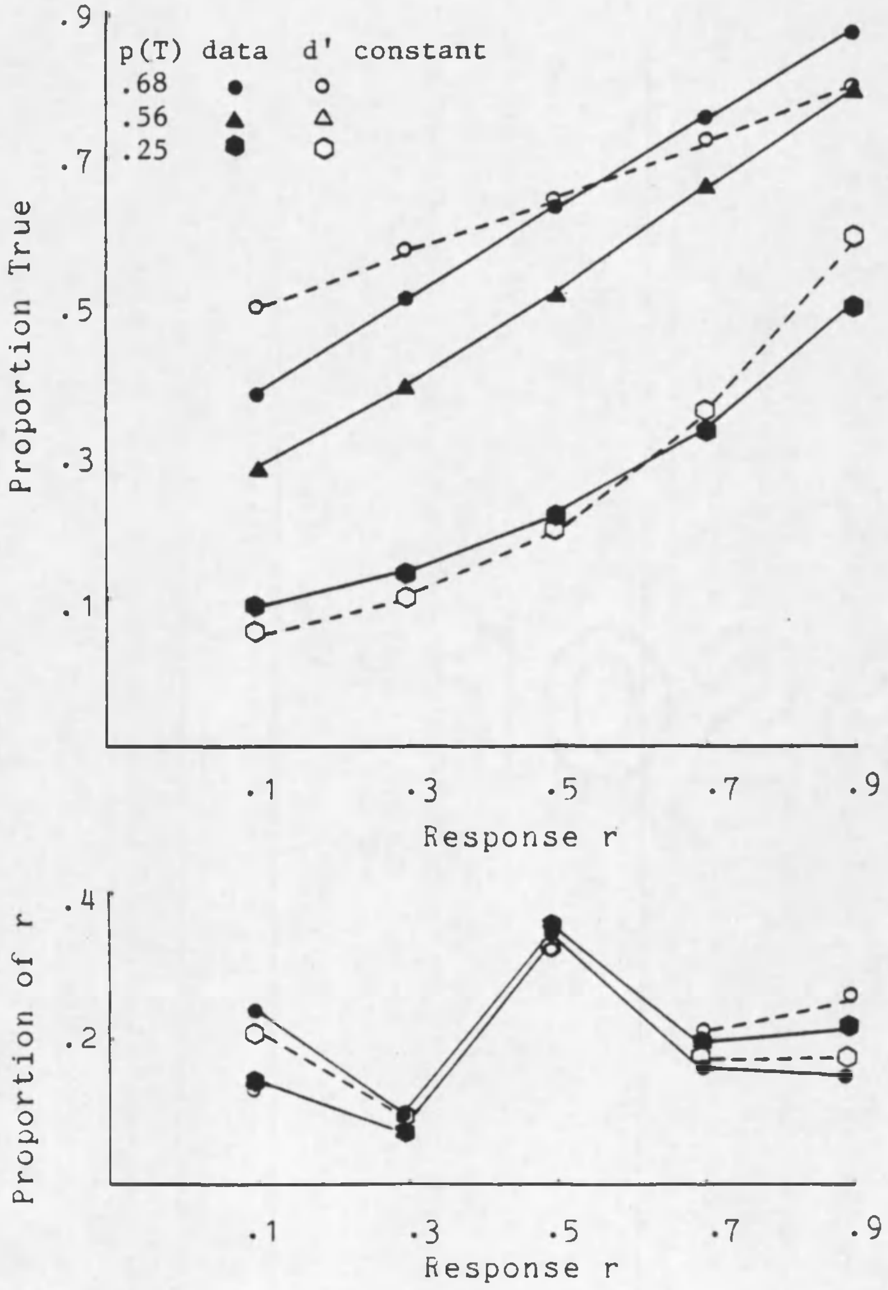

Figure 12. Model Predictions for Observed and Constant d'

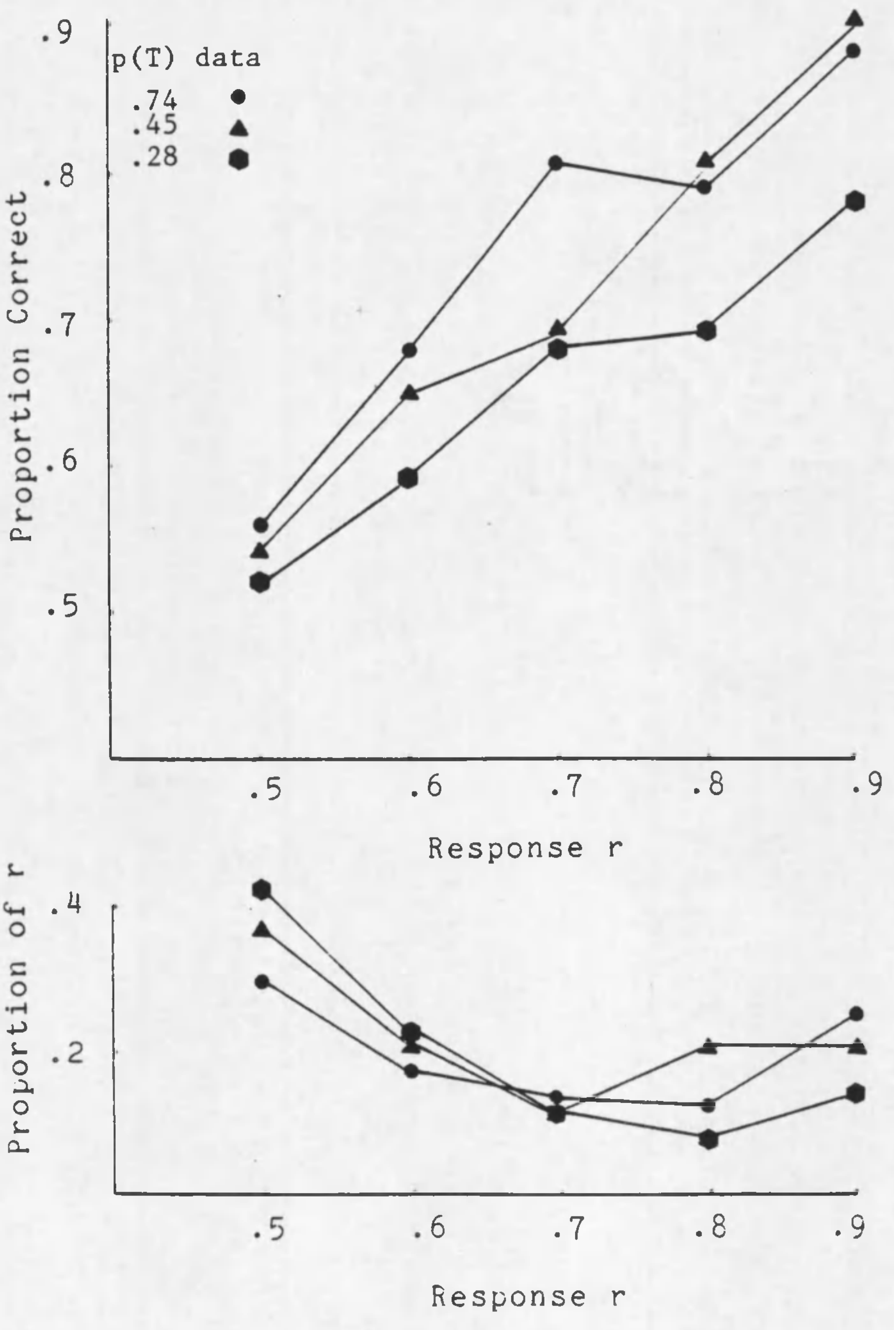

Figure 13. Obtained Data for PC Task (pooled data)

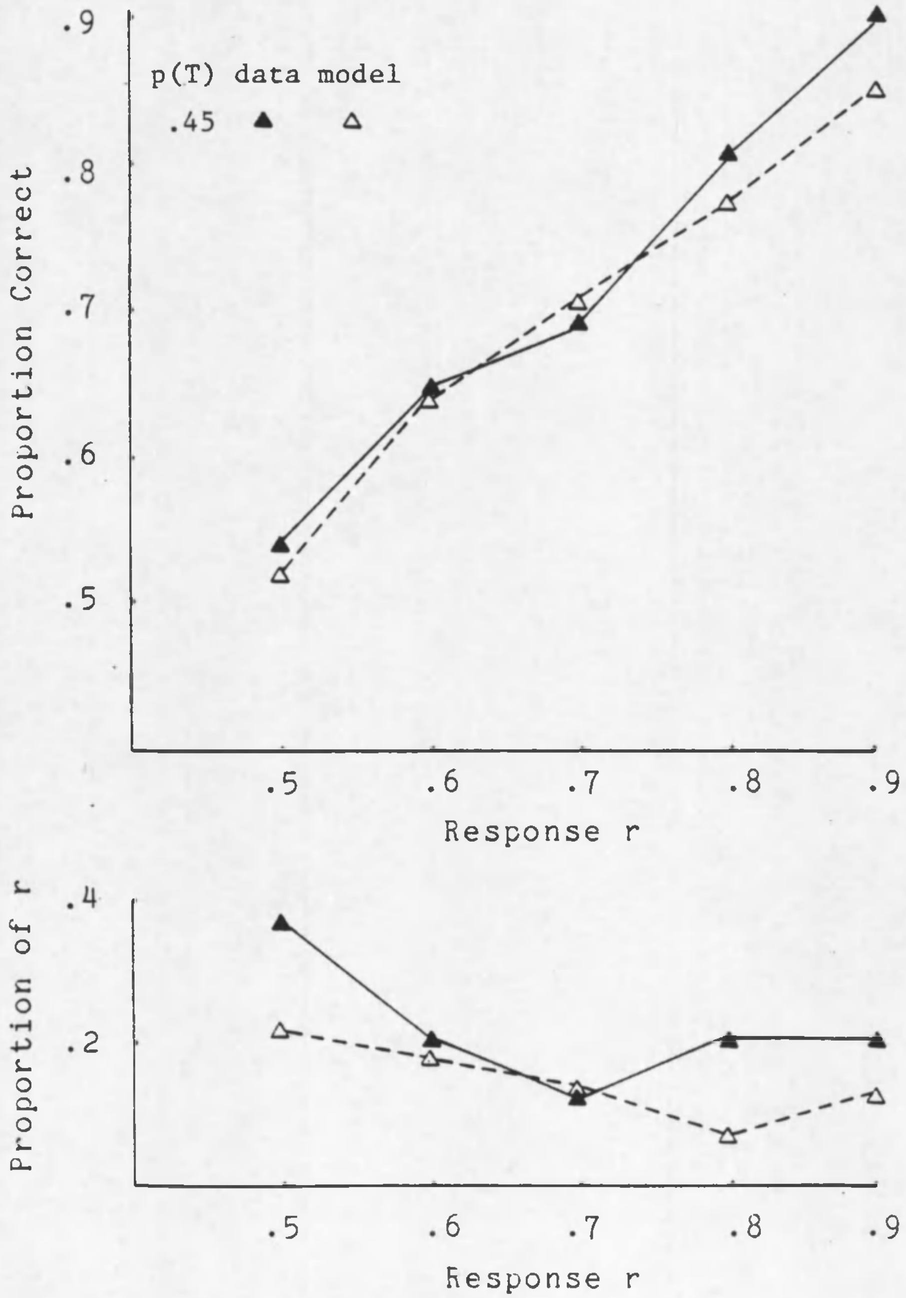

Figure 14. Fitted Curve for PC Task (pooled data)

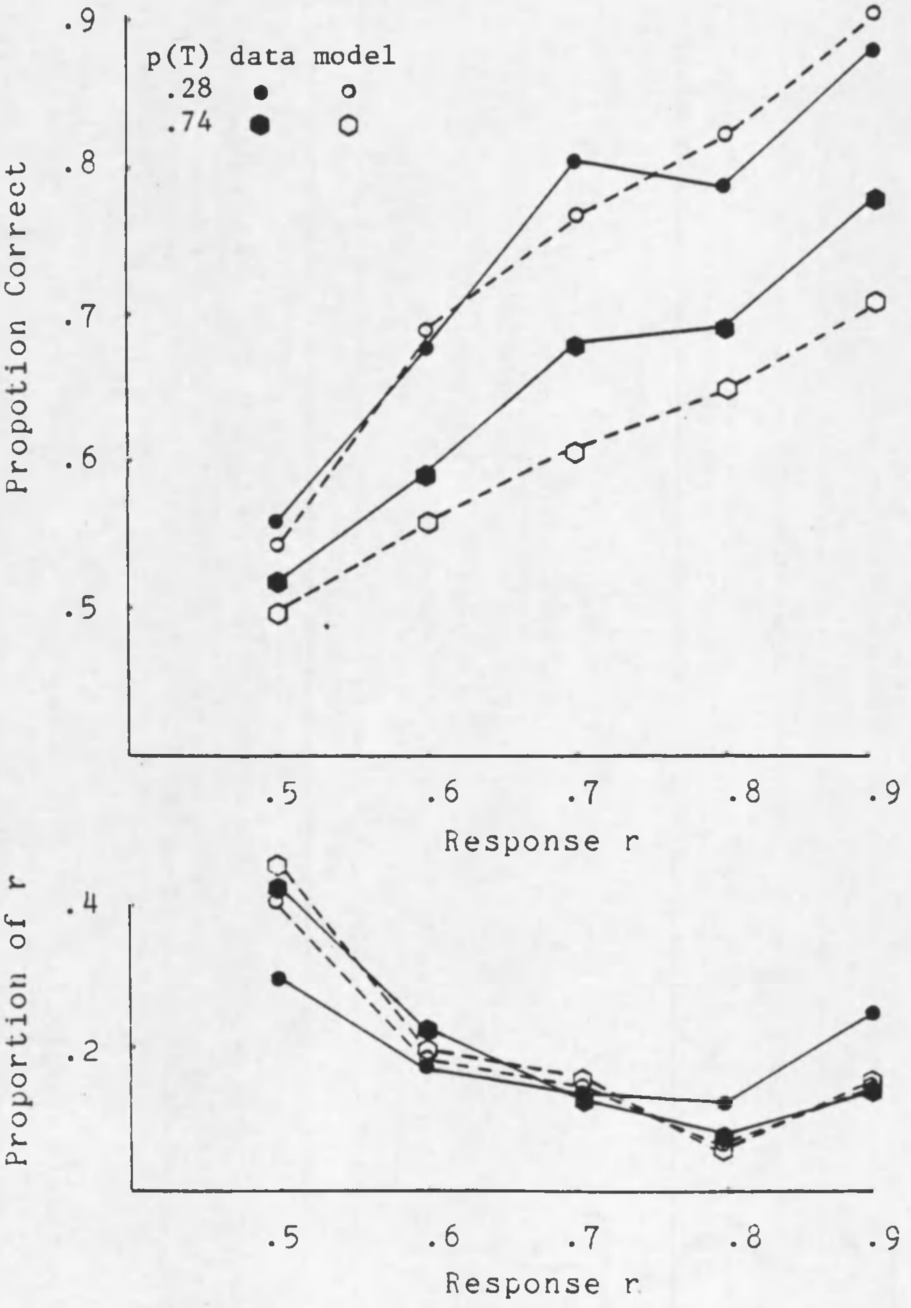

Figure 15. Predicted Curves for PC Task (pooled data)

the experimental data at the 0.01 significance level.

The corresponding data for a typical individual are given in Figures 16, 17 and 18. The presumably random fluctuation of the calibration curve is marked, though the general behaviour is the same as that exhibited by the pooled data. The fit and predictions of the model are not as good as was the case for the pooled data, nevertheless, they seem reasonably close to the observed data.

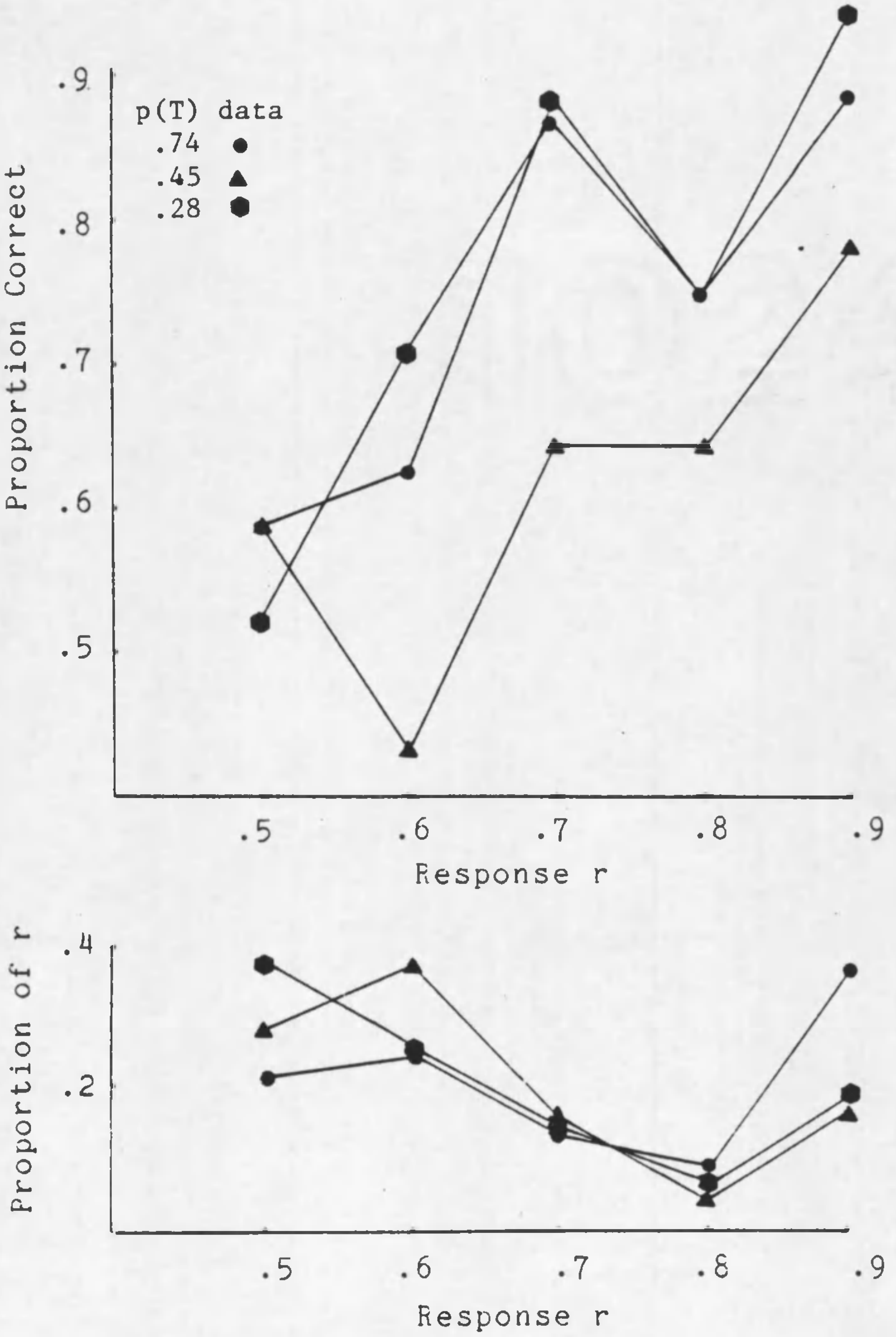

Figure 16. Obtained Data for PC Task (individual)

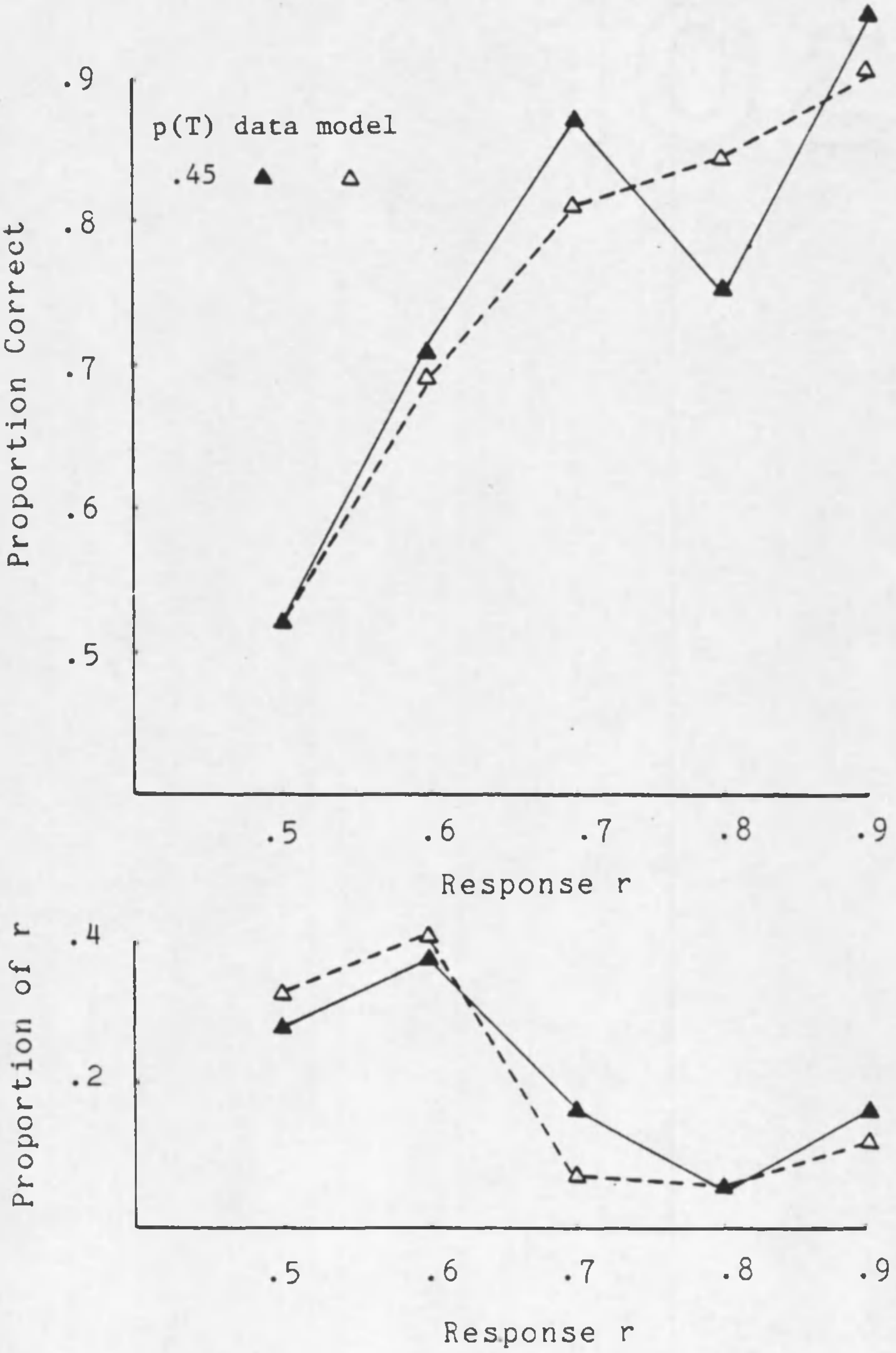

Figure 17. Fitted Curve for PC Task (individual)

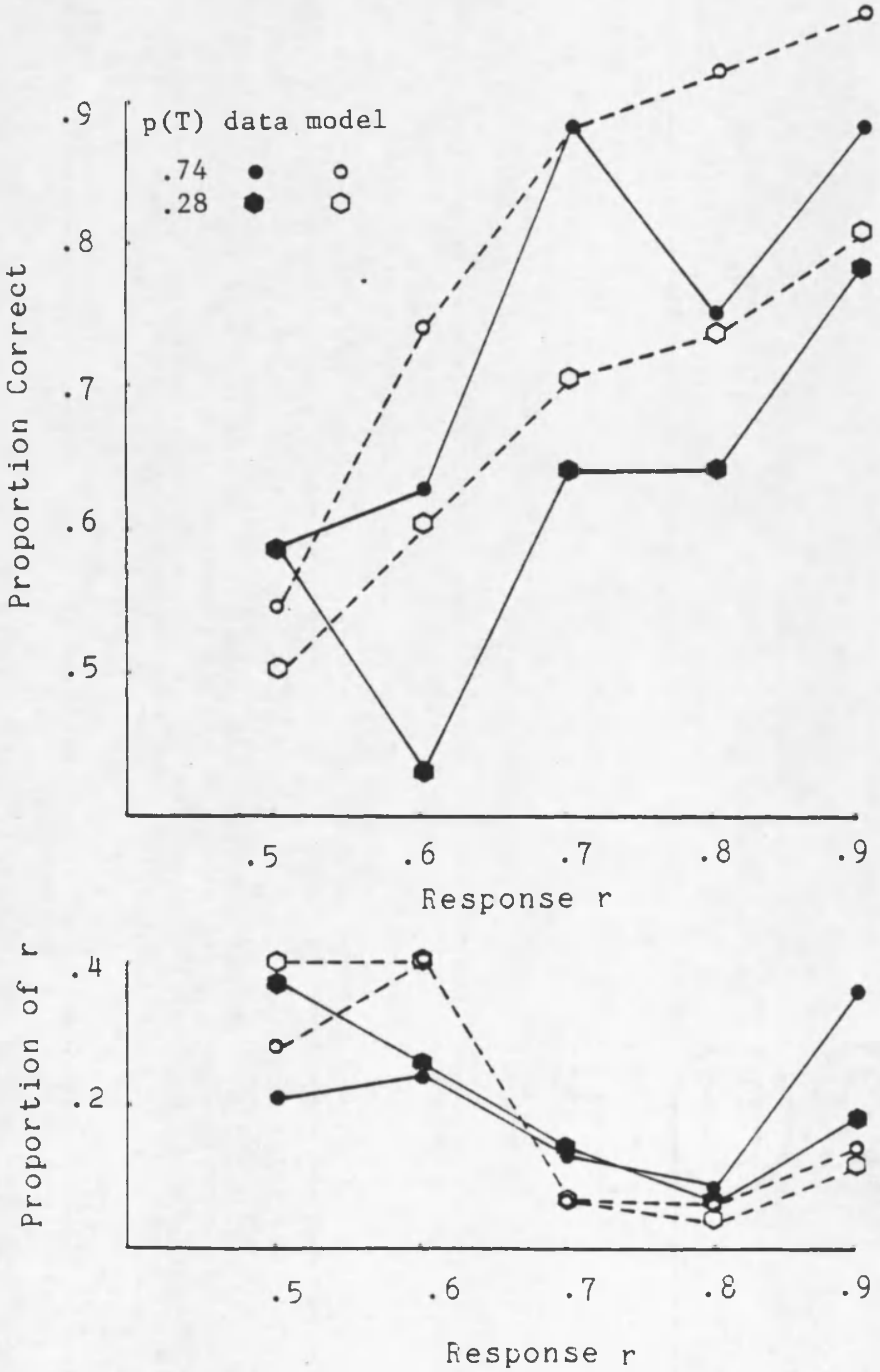

Figure 18. Predicted Curves for PC Task (individual)

DISCUSSION

The results of the experiments have shown a number of important features of calibration tasks and strongly support the decision variable partition model as a quantitative representation of calibration. In general, the results are consistent with the theoretical predictions. There are a number of particular conclusions which may be drawn:

(1) the degree of detectability affects calibration in both tasks;

(2) the base rate affects calibration in the PT but not the PC task;

(3) an increase in the proportion of true statements increases the calibration probabilities in the PT task; and,

(4) an increase in detectability increases the calibration probabilities in the PC task.

The principal contributions of the study are twofold. First, the experimental results have demonstrated how task type and base rates are related, as well as showing the effect of different detectability rates. Second, the quantitative model can be seen not only to provide theoretical predictions about calibration, but to make accurate empirical predictions as well. The latter is perhaps the more impressive bearing in mind the fact that predictions were made from one parameter estimate alone. That is, given the cut-off values $\{x^i\}$ estimated from the data of the p(T) = 0.5 condition, then with only the detectability d' estimated from the data, it was possible to predict with remarkable accuracy the calibration proportions for different base rates.

The base rate effect for the true-false tests was predicted but not tested by Ferrell and McGoey (1980). Unpublished data of their own has been presented by Lichtenstein, Fischhoff and Phillips (unpublished) which, they speculate, supports the decision variable partition model. Here the effect and the conditions for it have been stated in detail, and calculated mathematically.

The theoretical predictions made by the model are clearly supported by the observed data. It was seen in the PT task that the effects of base rate and difficulty were as predicted. Also, in the PC task there was no effect of base rate, and the only separation which occurred was because of difference in d'. The implications of these findings for cognitive processing in subjective probability assessment is yet to be determined.

Conclusion

The decision variable partition model must now take an unequivocal place in the quantitative modelling of calibration, being the only existing model of its kind. The theoretical predictions are sound and have already been shown to be supported by data (Ferrell and McGoey,1980). The empirical predictions leave no room for doubt about the adequacy of the model in encompassing the variables of subjective probability assessment in calibration.

There is much more ground to be explored in the area of calibration probabilities. It is to be hoped that the present study makes it more likely that the topic of calibration will emerge from esotericism to take a more prominent place in the literature of decision making.

REFERENCES

Ferrell, W.R. and McGoey, P.J. A Model of Calibration for Subjective Probabilities. Organizational Behavior and Human Performance, 1980, 26:32-53.

Lichtenstein, S. and Fischhoff, B. Do Those Who Know More Also Know More About How Much They Know? The Calibration of Probability Judgements. Organizational Behavior and Human Performance, 1977, 20:159-183.

Lichtenstein, S. and Fischhoff, B. How Well do Probability Experts Assess Probability? Decision Research Report 80-5, 1980.

Lichtenstein, S., Fischhoff, B. and Phillips, L.D. Calibration of Probabilities: The State of the Art. In Jungerman and G. de Zeeuw (Eds), Proceedings of the Fifth Research Conference on Subjective Probability, Utility and Decision Making, 1976.

Lichtenstein, S., Fischhoff, B. and Phillips, L.D. Calibration of Probabilities: The State of the Art. (Unpublished).

Sheridan, T.B. and Ferrell, W.R. Man-Machine Systems: Information, Control, and Decision Models of Human Performance. The MIT Press, 1974.