Consensus on Semiotic Models of Alphabetic Systems[1]

KIMBERLY JAMESON AND A. KIMBALL ROMNEY

University of California, Irvine

ABSTRACT: In this paper we report the results of an experiment that tests two variants of a generative semiotic model of the English capital or "majuscule" letters. Consensus analysis is used for estimating the "correct" *alphabetic extensions* from two different informant-groups we call "novices" and "experts." These data are then taken as a standard "cultural model" for distinguishing the relative merits of the two generative semiotic models of the alphabet. The general analytic procedure should be applicable in a wide variety of situations in which two different theories predict different measurable cultural models of the same data.

KEY WORDS: consensus, informant accuracy, semiotic models, expert knowledge, alphabetic systems

1. INTRODUCTION

In this paper we present a general procedure for testing the relative adequacy of two variants of a generative semiotic model of the English capital letters or majuscules. The models formally characterize the set of 26 majuscule letters, but since the characterizations are "generative" (algorithmic) they also determine, beyond the set of 26 existent letters, a larger set of well-formed possible letters. (Some of these have existed in ancestral alphabets, such as "Γ" in the Greek; others have never existed.) The models constitute an attempt to represent the cognitive model of the alphabet that people "have in their heads."

The two generative semiotic grammars that attempt to represent the cognitive model can be briefly characterized as consisting of *generalizations* over the set of letters (Watt 1975, 1988). An instance of such a generalization is: "letters face rightwards," which holds for all the majuscules that are asymmetric on the vertical axis except "J", which faces leftwards, and the set "N", "S", "Z", which face neither way. There are a number of such generalizations, or rules of grammar, and taken together they are characterized by the following properties: (1) they specify a set of well-formed or *grammatical* letter-forms that include both the set of 26 letter-forms that constitute the present English alphabet, and a larger subset of "pseudo-letters", equally well-formed, that our alphabet does not include; (2) the generalizations also specify, beyond the *grammatical* set, a much larger set of pseudo-letters of less grammaticality that break various

Journal of Quantitative Anthropology 2: 289—303, 1990.

rules of the grammar; and (3) the pseudo-letters can be ordered by the theory in terms of their relative ungrammaticality, using as criteria the importance and number of rules broken.

Three levels of description make up the generative formal model that characterizes the 26 English majuscules and all other highly letter-like extensions of that set. The derivative syntactic descriptions of the individual letters and pseudo-letters are, as one might expect, quite lengthy and intricate in detail and therefore will not be reproduced here. Readers interested in specific examples of syntactic descriptions may consult Watt (1988).

It is worth noting that both the types of acquisition "errors" commonly made by young children learning to write the capital letters and historical evidence from ancient Greek archaeological artifacts (see Watt and Jacobs 1975; Watt 1975) provide indirect support for some aspects of the system described above. Jameson (1989), however, was the first to present direct empirical support for the grammatical models.

Empirically testing *acceptability* of the pseudo-letters is a direct way of assessing the degree to which a particular semiotic grammar has successfully modeled the putative cognitive grammar. The implicit assumption of this approach is that informants share a common cognitive model of the alphabet. One of the goals of the present research is to test this assumption.

To measure the degree of agreement among our informants we use consensus analysis (Batchelder and Romney 1988; Romney, Weller and Batchelder 1986), a method designed to estimate both how much each informant "knows" about a given subject matter, as well as, the correct answers to the questions. Consensus analysis also provides measures of how well the data fit the assumption that the informants are reporting on a "single" culture (i.e. cognitive model) and confidence estimates on the inferred answer to each question.

This paper attempts to answer the following questions: (1) Can informants make subtle judgments about "implicit" rules when working with thematic forms like uppercase alphabetic characters? (2) Do both "novice" and "expert" informants share equally in such knowledge? And (3) How well can the judgments of informants, as measured by consensus analysis, serve to distinguish between two alphabetic theories?

The experiment presented here was conducted as part of a series of investigations (Jameson 1989) designed to determine the cognitive relevance of two generative semiotic grammars of the English alphabet. It should be emphasized that while many "psychological" models of the alphabet have been published over the years (Gibson *et al.* 1963; Künnapas 1966; Rumelhart and Siple 1974; and Townsend *et al.* 1984), the models have dealt with the perceptual aspects of letter recognition, such as how likely letters are to be confused with one another under poor viewing conditions, rather than with the cognitive aspects. None of the models have attempted to directly account for the grammaticality of the letters, that is with deliberate judgments of how letter-like new letter forms are. Further details of the investigations, including the theory and rationale underlying the alphabetic models, and the psychological implications of the results, are given in Jameson (1989).

THE EXPERIMENT

Jameson (1989) demonstrated that subjects are capable of making set-membership judgments for alphabet typefaces in an experimental situation, and that consensus analysis is useful for analyzing the resulting experimental data. The design for the present Experiment arose from designs employed in pilot experiments.

The experiment involves testing the 53 newly generated letter-forms shown in Table I as possible candidates for a sample alphabet. These 53 letter-forms fall along two alternative "wellformedness" continua, derived from theoretical models 1 and 2. Each model, for ease of experimental presentation and discussion, divides the letterforms into three subsets: *Grammatical*, *Semi-Grammatical*, and *Nongrammatical*, where *Semi-Grammatical* items have a higher degree of "grammaticality" than *Nongrammatical* items.

Informants were asked to determine whether the pseudo-letters were acceptable members of a specified variant of the English alphabet. A sample question is shown in Figure 1. In the experiment informants completed 53 forced-choice questions to determine whether or not each letter-form was acceptable as a "new letter" of the alphabet.

The purpose of the Experiment was to apply consensus analysis to determine the relative adequacy of the two models (or grammars) of the English majuscules. Both are variants of the generative semiotic grammars developed by Watt in a series of papers (1975, 1980, 1981, 1988). The two variants differ in a number of respects but their key difference lies in how they relate to the specific letter-forms used by Jameson in her experiments. One grammatical model (hereafter Model 1) made predictions about informant acceptances of letter-forms on the assumption that informants would mostly ignore relatively minor departures from the conventional forms. (For instance, Model 1 assumed that informants would accept form 12, in Table I, as an ordinary "R", ignoring the angularity of its cusp.) In contrast Model 2, was developed after pilot experiments demonstrated that informants were in fact attending to what had in advance been viewed as very minor departures from conventional letter-forms. Model 2 took that behavior into account, thus predicting that angular "R" would prove relatively less grammatical. In brief, Model 1 was lax and Model 2 was strict.

TABLE I
Experimental Booklet Items:
Question Order, Grammatical Model, and Answer Key Estimate Information

Question	Stimulus item	Model 1	Model 2	Undergrad Key	Bayesian Prob.	Expert Key	Bayesian Prob.
1		G	G	Yes	1.0000	Yes	1.0000
2		N	N	No	1.0000	No	1.0000
3		N	G	Yes	0.9941	Yes	0.9883
4		N	N	No	1.0000	No	0.9967
5		N	N	No	0.6493	No	0.9985
6		G	G	Yes	1.0000	Yes	1.0000
7		G	N	No	1.0000	No	1.0000
8		N	N	No	0.9992	No	1.0000
9		S	S	No	0.9958	No	1.0000
10		N	N	Yes	0.9841	No	0.9906
11		G	S	Yes	0.8678	Yes	0.5828
12		G	N	No	0.9953	No	1.0000
13		N	G	Yes	0.9957	Yes	1.0000
14		G	G	Yes	1.0000	Yes	0.9936
15		G	N	Yes	0.9880	No	0.9976
16		N	N	No	1.0000	No	1.0000
17		N	G	Yes	0.9741	Yes	0.9948
18		N	N	No	1.0000	No	0.9771
19		N	N	No	1.0000	No	1.0000
20		N	S	No	0.9072	No	0.8949
21		G	G	Yes	0.9985	Yes	1.0000
22		N	G	Yes	1.0000	Yes	0.9945
23		G	G	Yes	1.0000	Yes	0.9989
24		N	N	No	1.0000	No	1.0000
25		N	G	Yes	0.9706	Yes	0.8655
26		G	S	No	0.9734	Yes	0.9073
27		N	N	No	1.0000	No	1.0000
28		N	G	Yes	1.0000	Yes	1.0000
29		N	N	No	1.0000	No	0.9997
30		G	G	Yes	0.8019	Yes	1.0000
31		N	N	No	1.0000	No	1.0000
32		G	N	Yes	0.7791	No	1.0000
33		G	G	Yes	0.9998	Yes	0.8584
34		G	G	Yes	0.9963	Yes	0.9997
35		N	N	No	1.0000	No	1.0000
36		N	N	No	1.0000	Yes	0.8797
37		G	G	Yes	1.0000	Yes	1.0000
38		N	N	No	1.0000	No	1.0000

Table I (continued)

Question	Stimulus item	Model 1	Model 2	Undergrad Key	Bayesian Prob.	Expert Key	Bayesian Prob.
39		N	N	No	1.0000	No	1.0000
40		N	G	Yes	1.0000	Yes	0.9809
41		N	N	No	1.0000	No	1.0000
42		G	G	Yes	1.0000	Yes	1.0000
43		S	N	No	1.0000	No	1.0000
44		N	S	Yes	0.9697	No	0.9996
45		S	G	Yes	1.0000	Yes	0.9841
46		G	N	No	1.0000	No	1.0000
47		G	G	Yes	0.9999	Yes	0.9998
48		N	N	No	1.0000	No	1.0000
49		G	N	Yes	0.9997	No	1.0000
50		N	S	Yes	0.9440	No	0.9993
51		N	S	No	1.0000	No	0.9999
52		S	S	No	0.9049	No	0.9985
53		G	G	Yes	0.9927	Yes	0.9594

Table I Notes. Cell values in Columns 3 & 4 are as follows: G = Grammatical, N = Nongrammatical, S = Semi-grammatical, in each grammatical model. Cell values in Columns 5 & 7: "Yes" indicates that the estimated "correct" answer is "yes, this is an acceptable new letterform", and "No" indicates that the estimated "correct" answer is "no, this is not an acceptable new letterform". Columns 6 & 8 contain the Bayesian probability values from the "covariance" consensus analysis.

Model 1 is a formal iconic grammar that employs distinctive feature matrices and generative rules in its formulation. The model incorporates aspects of alphabetic evolution as well as the learning processes involved in acquisition. The model posits that the individual characters of the alphabet are morpheme-like entities that are made-up of phoneme-like entities (line-segments) which in turn are the distinctive features of parts of letters. This model is equivalent to Watt's Unified Characterization (Watt 1988) which is primarily based upon the visual aspects of the majuscules, specifying the motoric aspects of the majuscules as derived.

Model 2, discussed in Jameson (1989), is in essence a more "specific" variant of Model 1. For Model 2 the basic description given for Model 1 above is still appropriate; however, Model 2 further incorporates new rules which influence its functional output. The most salient of these new rules involve (1) provision for more variability in line-length and line-orientation, to distinguish angular from conventional "R" for instance; (2) provision for differential weighting of the distinctive features of letters; (3) re-evaluation of feature combinations which involve opposing values (i.e., curvilinearization and angularization); and (4) greater utilization of "redundancy measures" as indicators of "well-formedness".

Here are all the "new letter" candidates:

Question #1:

ABCDEFGHIJKLMNOPQRSTUVWXYZ

The "new letter" candidate:

1. Yes, this is an acceptable "new letter".
2. No, this is not an acceptable "new letter".

Fig. 1. An example of the experimental design.

The formal rules of Models 1 and 2 produce marked differences in grammatical classification as seen in Table I. For instance, item #3 is classified *Nongrammatical* in Model 1 and *Grammatical* in Model 2. Such variation across models, when viewed in terms of the empirical data analyzed by consensus analysis, permits evaluation of the models (see Jameson 1989).

SUBJECTS AND METHOD

Experimental booklets were administered to two different groups of informants. The first was a group of 30 college undergraduates who participated in the experiment for partial course credit. The second group consisted of 13 skilled professionals who participated in the experiment for cash payment. These "expert" industry informants worked in the area around Stanford University and had been employed for a duration of no less than 18 months working in the environment of a typesetting establishment or in a company which manufactured typeface designs (average duration of employment was approximately 8 years). No additional requirements (i.e., regarding age, gender, education level, native language, handedness, or any other) were used to sample the industry informants (although data on gender, cultural background, language capabilities, handedness, and so on were collected). Thus 30 undergraduate "novices" and 13 industry "experts" were sampled.

RESULTS

The results are presented separately for the two samples. Consensus

analysis (Batchelder and Romney 1988) was used to determine the informant competencies of the subjects and the estimated "correct" answer-key for the task. The consensus model results presented here were obtained using the computer package *ANTHROPAC* (Borgatti 1989). *ANTHROPAC* is written especially for the collection and analysis of systematic data in anthropological field situations. It provides a large range of statistical and plotting resources. It also produces randomized questionnaires in a number of formats. In addition to elementary statistical summaries of data it performs a number of sophisticated scaling and analysis tasks. These include correspondence analysis, principal components, consensus analysis, multidimensional scaling, and quadratic analysis and other randomization tests.

Table II and Figure 2 present summary statistics for the consensus analysis, using the covariance method with $\pi = 0.325$. The results of the consensus analysis for the undergraduate "novices" produced a mean competence level of $\bar{x} = 0.230$ with $sd = 0.327$, and 10 informants with negative competence. The undergraduate sample does not fit the model very well.

The consensus analysis carried out on the industry "experts" produced a mean competence level of $\bar{x} = 0.530$ with $sd = 0.179$, and zero informants with negative competence. The industry data were well-explained by a single factor in the consensus analysis. This indicates that the industry informants were responding to the experimental task as though they shared the same cultural knowledge.

A comparison of the "covariance" method analysis reported above with a "matches" method analysis shows that the two methods yield similar results. The undergraduate matches answer-key is correlated with the covariance answer-key $r = 0.734$; a similar measure for industry data is $r = 0.889$. Bias would lead to discrepancy between the two methods. The relatively high observed correlations suggest that bias is fairly small.

EXPERIMENTAL FINDINGS

The measures obtained for the undergraduate novices are marginal and should be regarded as somewhat unreliable. However, the consensus

TABLE II
Consensus analysis data for two samples of informants

Sample	n	Mean competence	sd	Negative competences
Undergraduate *novice*	30	0.23	0.33	10
Industry *experts*	13	0.53	0.18	0

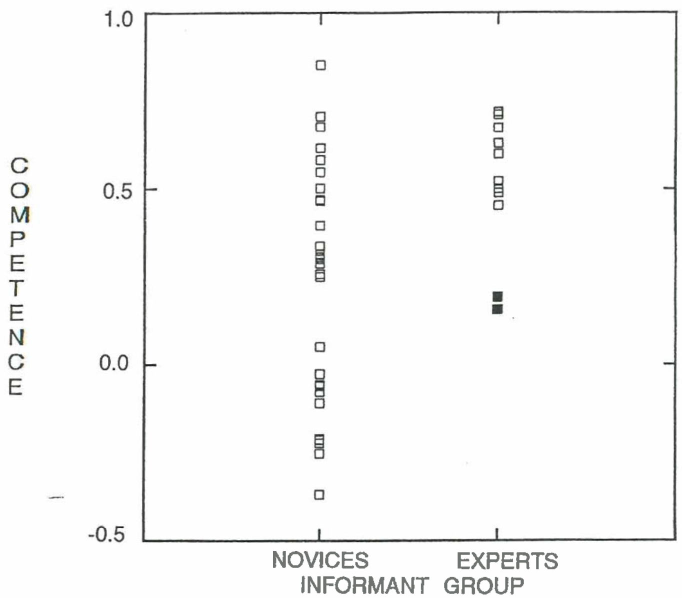

Fig. 2. Plot of the competences of undergraduate novices and industry experts.

measures for the industry experts indicate good consensus among the informants with no negative competencies. This finding is consistent with the intuitive notion that the industry "experts" share specialized "knowledge" of the generative alphabetic grammar.

It is worth noting that the mean level of competence obtained for the industry sample is depressed by two informants with low competence. These two informants (indicated in Figure 2 as solid squares) bear the singular characteristic of being specialists in the typeface design of oriental Kanji scripts (Chinese/Japanese "characters" or morphographs), differing in this respect from the other informants, who specialized in Roman alphabetic scripts. Note that the "Kanji" informants do not cluster with the other expert informants displayed in Figure 2.

CONSENSUS MODEL ANSWER-KEYS AS A TEST OF THE THEORETICAL ALPHABETIC MODELS

The answer-key estimated by the consensus analysis provides an indication of how pseudo-letters are classified ("acceptable" or "non-acceptable") by the two groups. Table I shows the classification for the undergraduate novices and the industry experts for each pseudo-letter form. Table I also shows the Bayesian probabilities that indicate a confidence level for each form. Given the validity of the assumptions of the consensus model, each Bayesian probability can be taken as the probability that the pseudo-letter was correctly classified according to the cultural model.

Estimated answer-keys represent the shared knowledge of the informants and can be compared to the predictions of the two generative alphabetic models to provide an answer as to which alphabetic model is most representative of the observed "cultural" knowledge. Thus it is assumed that the "cultural" model generally represents the "cognitive" model shared by informants.

The undergraduate novices differed on eight question items from the industry experts. For a general indication of what these estimated keys suggest for the alphabetic models, we first examined the frequency with which a given answer-key estimate supported a given alphabetic model. Table III shows the percentage of new-letter candidates accepted in each theoretical grammatical category for models 1 and 2 by novices and experts.

TABLE III
Percentage of new-letter candidates accepted in each theoretical grammatical category for models 1 and 2 by novices and experts

Answer-Key	Grammatical category	Model 1 % accept	Model 2 % accept
Novices	Grammatical	80	100
	Semi-grammatical	25	37
	Non-grammatical	34	16
Experts	Grammatical	70	100
	Semi-grammatical	25	25
	Non-grammatical	28	4

Two clear trends are discernable in the data presented in Table III. First, note that both novices and experts correctly classify a higher percentage of new-letter candidates for Model 2 than for Model 1. This illustrates the usefulness of consensus analysis in deciding between alternative theories. Second, note that the experts are somewhat better than the novices in the classification of new candidates in Model 2 but not for Model 1.

It is to be expected that as a group the industry sample should have a special kind of shared "expertise" concerning alphabetic forms, and that this should influence the observed industry-key estimates. In contrast the undergraduate sample does not have access to this "specialized" knowledge. The descriptive measures just presented bear these expectations out with respect to both Model 1 and Model 2, but not very strongly, showing only a slight difference between industry-key estimates and undergraduate-key estimates. Clearer insight into the nature of this difference may be achieved by way of a correlational analysis.

Goodman and Kruskal Gamma coefficients between the novice and expert estimated answer-keys and the two theoretical alphabetic models were calculated. Gamma (Goodman and Kruskal 1954; Freeman 1986) is a nonparametric measure of association that makes no scaling assumptions beyond the ordinal-level, which makes it an appropriate model for the present data. Many of the more familiar measures of association (e.g., Pearson's "r") represent *strong monotone models* (see Freeman 1986), and are thus inappropriate for measuring the association between a 3-valued variable and a 2-valued variable, as is the case here. The Gamma statistic is easily interpreted as the proportion of hits between two variables by the transformation given by:

$$p = (1 + \gamma)/2,$$

where p is the proportion of cases in which the two variables are in agreement, and γ is the observed Gamma statistic between the two variables.

The Gamma coefficient between the estimated answer-keys and the two theoretical models were as follows: for the undergraduate novices $\gamma = 0.67$ for Model 1 and $\gamma = 0.93$ for Model 2, for the industry experts $\gamma = 0.64$ for Model 1 and $\gamma = 0.98$ for Model 2. Note that while both Model 1 and Model 2 are correlated with the estimated answer-keys, the Gammas for Model 2 are higher than those for Model 1. The industry experts are about the same as the undergraduates for Model 1 while they are somewhat higher for Model 2.

In general, the Gammas show both estimated answer-keys to be well correlated with the theoretical models of the alphabet. However, for both groups of informants, Model 2 is better than Model 1 at predicting the answer-key estimates arising from informants' preference choices for new-letter forms. Thus, the answer-key estimates derived from the consensus analysis clearly distinguish between the two alphabetic models.

EXAMINING THE CONFIDENCE LEVELS OF THE ANSWER-KEY ESTIMATES

The confidence levels of an estimated answer-key refer to the certainty with which any given answer estimate is deemed a "correct" answer. That is, for any given question included in the experimental task the consensus analysis provides a Bayesian probability that indicates the probability that the estimated answer has been "correctly" classified. Thus the primary reason the answer-keys produced by the consensus analysis are referred to as "estimates" is that for each and every item there potentially exists some probability less than one that the given consensual "answer" to that item is not actually the "correct answer".

If we set a cut-off criterion of 0.99 or greater for a reliable answer, then the undergraduate-key contained 12 estimates and the industry-key contained 10 estimates that fail the cut-off criterion. Those items with a Bayesian probability above 0.99 are decisively classified while items with a lower probability are less reliable. As researchers we place more confidence in those answers that have a higher Bayesian probability.

One would expect informants' judgments to vary more for *Semi-grammatical* items while judgments should be more decisive, or "reliable", for those "clear-cut" new-letter forms occupying the extremes of the grammatical continuum (i.e., *Grammatical* or *Nongrammatical* items). For simplicity, however, in the following discussion we consider that the theoretical "correct" answer for *Semi-grammatical* items is "No", i.e., such items are not grammatical. It seems reasonable to ask whether an examination of the estimated answer-keys can support our overall interpretation. That is, can we gain insight into the difference between models 1 and 2 through an examination of the answer-key probabilities?

An examination of the Bayesian probabilities contained in Table I reveals the specific items that failed that the 0.99 probability criterion. It is interesting to note that for the industry-key all the estimates that "disagreed" with Model 2, (i.e., 1 *Nongrammatical* item and 2 *Semi-grammatical* items), are unreliable estimates by the criterion. That is, for the industry-key, all the estimated answers that were not in accord with Model 2 (i.e., 3 out of 53) were among those 10 items that failed to satisfy the cut-off criterion. However, this effect was not observed for Model 1: 15 of the 53 responses conflicted with Model 1, and only 4 of these 15 exhibited a probability less-than 0.99. That is, 11 of these responses conflicted decisively with Model 1.

The fact that in the industry-key all the estimates in conflict with Model 2 are also unreliable estimates indicates that the industry sample did not decisively disagree with any of Model 2 classifications. The industry-key is in conflict with Model 2 only when the estimated answers receive a Bayesian probability less-than 0.99. In contrast, the industry-key often reliably "disagrees" with Model 1.

To parallel the above discussion for the undergraduate-key, it is not the case that all the estimates that "disagree" with the grammatical categories of Model 2 are among those 12 unreliable estimates from the undergraduate-key. In contrast to the industry informants who always "disagreed" with Model 2 with some uncertainty, in the undergraduate sample only 5 of the 12 unreliable answer estimates observed were in conflict with Model 2. Among the 12 unreliable estimates observed in the undergraduate-key, 7 are estimates which are in accord with the grammatical categories suggested by Model 2. In terms of Model 1, 15 of the 53 undergraduate-key estimates conflicted with the model, and only 6 of these exhibited a probability less than 0.99.

Comparing the reliability of answer-keys between the industry and undergraduate groups is somewhat problematic due to the fact that as a rule fewer "expert" informants will give rise to a more reliable answer-key estimate than many more "novices". This issue, discussed at length elsewhere (Maher 1987), directly bears upon the reliability of the estimates discussed here. In the present analysis no strong conclusions are suggested from the comparison of the reliability of answer-key estimates from an industry sample of size 13 to that of a undergraduate sample of size 30, (especially since the fit to the consensus model seems less than ideal for the undergraduate sample).

Ideally, to make such a comparison one could estimate the reliability coefficient for the type of informant population sampled (i.e., expert or novice) and then determine the size of sample needed in order to produce equally reliable answer-key estimates from the two different samples.

In the present experiment, according to the Spearman-Brown Prophecy formula computations, one would need to sample a minimum 59 undergraduate informants in order to obtain a key-estimate as reliable as that observed from the industry sample of size 13. Even then a sample of 59 undergraduates may not yield a key-estimate which is the equal of the industry sample since the observed fit of the model to undergraduates has been marginal at best.

The above comparison suggests that when the industry informants disagree with Model 2 they give unreliable estimates. Also, the industry sample yields far fewer unreliable items in their respective answer-key estimate than would the answer-key estimate from an undergraduate sample of equal size.

Such a finding might be seen as suggesting that in the future one could empirically identify "expert" informants simply by virtue of the fact that when the sample size is small they yield fewer unreliable answer-key estimates than another comparison sample. Although the possibility of identifying "expert" informants by virtue of competence measures is built into the consensus analysis, examining the properties of answer-key estimate reliability to identify "experts" may be an additional method to utilize, whether concordance is obtained for a sample or not. In addition, when a model of informants' responses exists, examination of the unreliable answer-key estimates might suggest further improvements in the response model in question. Such an examination might point towards an improved and more predictive model of the empirically observed behaviors of groups, and therefore towards a more precise cognitive model of individuals.

GENERAL DISCUSSION

We note that the findings presented here are consistent with the findings of independently conducted experiments that employed paired-comparison methodology and Thurstonian scaling techniques to derive numerical "Performance Rating" estimates for the pseudo-letter stimuli (see Jameson 1989). Utilizing the data for all 53 items from the present experiment in the computations, the Gamma between the undergraduate-key and the Performance Ratings is associated at $\gamma = 0.551$. The same measure for the industry-key is significantly higher at $\gamma = 0.706$. This is interesting since both the undergraduate and the industry Gammas are computed using Performance Ratings derived exclusively from undergraduate paired-comparison data. These findings are further evidence for a "common understanding" of the knowledge domain being accessed by both undergraduates and industry informants.

In this analysis we observe results that indicate that both answer-key estimates obtained via the present Experiment (undergraduate-key and industry-key) are highly associated with the independently obtained performance ratings derived from paired-comparison data. These results give independent external support to the application of the consensus model in the present experimental paradigm and further suggest that the consensus model not only is capable of capturing differences in informant responses for alphabetic stimuli, but also is useful as a tool for distinguishing between theoretically and functionally different alphabetic models.

CONCLUSIONS

We have answered the questions raised at the beginning of the paper. First, it has been shown that when viewed collectively informants' responses do indeed reflect subtle judgments concerning the "implicit" rules that presumably govern thematic (set-defined) forms like the upper-case majuscules. Second, the observed results also demonstrate that "expert" and "novice" informants do not share equally in the domain of knowledge investigated. Third, it has been shown that informants' responses do indeed distinguish between two very similar alphabetic theories. The data consistently supported one cognitive model (Model 2) over another (Model 1) as a predictor of informants' responses.

Some further observations are notable. First, consensus analysis, is shown to be an appropriate and useful tool for examining a domain typically considered "psychological". Second, the consensus analysis was found to be capable of detecting subtle qualitative differences in informants' knowledge, as was seen in the difference between "Kanji" informants compared to other industry informants. This suggests that consensus analysis may be useful in the investigation of the nonhomogeneity of informant samples. Third, the answer-key estimates from the consensus analysis were found to be interpretable in terms of detailed aspects of the alphabetic models (especially Model 2), as was the "reliability" of those estimates. These findings could be employed to produce improved alphabetic models.

Finally, the consensus analysis accurately distinguished between informants which were *a priori* considered to be "novices" and "experts" for the domain under investigation. Specifically, consensus analysis theory was shown to be an appropriate model for the industry response data, whereas the response data of the undergraduate sample was not as well-suited to the model.

NOTE

[1] The authors would like to thank W. C. Watt who provided extensive comments on an earlier version of this paper.

REFERENCES

Batchelder, W. J. and A. K. Romney 1988 Test Theory Without an Answer Key. *Psychometrika* 53: 71—92.

Borgatti, S. 1989 Using Anthropac to Investigate a Cultural Domain. *Cultural Anthropology Methods Newsletter* 1:11. Published by the Department of Anthropology, University of Florida, Gainsville, FL.

Freeman, L. C. 1986 Order-Based Statistics and Monotonicity: A Family of Ordinal Measures of Association. *Journal of Mathematical Sociology* 12(1): 49—69.

Gibson, E. K., H. Osser, W. Schiff, and J. Smith 1963 An Analysis of Critical Features of Letters Tested by a Confusion Matrix. In *Cooperative Research Project No. 639: A Basic Research Program on Reading*. Washington: U. S. Office of Education.

Goodman, L. A. and W. H. Kruskal 1954 Measures of Association for Cross Classifications. *Journal of American Statistical Association*, 49: 733—764.

Jameson, K. 1989 *An Empirical Investigation of Semiotic Characterizations of Alphabetic Systems*. Doctoral Dissertation; University of California, Irvine.

Künnepas, T. 1966 Visual Perception of Capital Letters. *Scandinavian Journal of Psychology*, 7: 189—196.

Maher, K. 1987 *A Multiple Choice Model for Aggregating Group Knowledge and Estimating Individual Competencies*. Doctoral Dissertation, University of California, Irvine.

Romney, A. K., S. C. Weller and W. H. Batchelder 1986 Culture as Consensus : A Theory of Culture and Informant Accuracy. *American Anthropologist*, 88: 313—338.

Rumelhart, D. E. and P. Siple 1974 Process of Recognizing Tachistoscopically Presented Words. *Psychological Review*, 81: 99—118.

Townsend, J. T., G. G. Hu, and J. Evans 1984 Modelling Feature Perception on Brief Displays with Evidence for Positive Interdependence. *Perception and Psychophysics*, 36: 35—49.

Watt, W. C. 1975 What is the Proper Characterization of the Alphabet? I: Desiderata. *Visible Language*, vol. 9: 293—327.

Watt, W. C. 1980 What is the Proper Characterization of the Alphabet? II: Composition. *Ars Semiotica*, IV: 1, 3—46.

Watt, W. C. 1981 What is the Proper Characterization of the Alphabet? III: Appearance. *Ars Semiotica*, IV: 3, 269—313.

Watt, W. C. 1988 What is the Proper Characterization of the Alphabet? IV: Union. *Semiotica*, 70-3/4: 199—241.

Watt, W. C. and D. Jacobs 1975 The Child's Conception of the Alphabet. In M. P. Douglass ed. *Thirtyninth Yearbook, Claremont Reading Conference*. Claremont, CA: Claremont Graduate School.