

## Cognitive Integration and Referential Communication: Effects of Information Quality and Quantity in Message Decoding

MELVIN MANIS, MARK FICHMAN, AND MARJORIE B. PLATT

*University of Michigan and Ann Arbor VA Hospital*

Four experiments are reported in which respondents were presented with one or more written passages, describing the face of an actor who was portraying a particular emotional expression (e.g., surprise). After reading each message ensemble, the respondents attempted to select the appropriate target-referent from an array of 24 photographs showing the same actor in different emotional poses. The results indicated that performance was consistently affected by the *quality* of the message ensembles (as quantified by the percentage of relatively accurate descriptions in the set) and by the *quantity* of information that was available (the number of different descriptive passages in the set). When a message-set was expanded while its overall quality was reduced (by including more and more low-quality descriptions), performance deteriorated. Finally, there was no convincing evidence of information overload; despite substantial expansions in the size of the message-sets that ultimately included as many as 24 different passages, when the quality of the ensembles was held constant, performance did not decline.

In a recent series of experiments, Manis and his colleagues (Manis & van Rooijen, 1973; Manis & Platt, 1975, 1976) explored the process of information integration. In contrast to many studies in this domain, these experiments included a criterion of correctness, which made it possible to assess the *effectiveness* with which discrete informational cues were combined.

In this work (and in the experiments described below), the respondent's task was to decode a number of message-ensembles, each of which contained one or more written descriptions; our goal was to explore the performance levels (hit-rates) that would result when respondents were given several descriptions, all based on the same referent-object, and were asked to select the appropriate referent from a large array of similar stimuli. The procedure devised for these studies is an extension of the

Marjorie B. Platt is now at the Radcliffe Data Resource and Research Center, Cambridge, MA 02138. All statements are those of the authors and do not necessarily represent the opinions or policies of the US Veterans Administration. The authors express their appreciation to the Research Center for Group Dynamics at the Institute for Social Research, University of Michigan, for making their facilities available for this project. Requests for reprints should be sent to Melvin Manis, Department of Psychology, University of Michigan, Ann Arbor, MI 48109.

referential communication paradigm, and yielded an experimental situation that seemed representative of many cognitive integration tasks.

Related problems of cognitive integration have previously been considered in the research on Brunswik's lens model, and in the clinical literature, where the effectiveness of diagnostic prediction has been studied under varying conditions of information input. Research in the clinical tradition has not been particularly encouraging, however; for example, Goldberg (1968) found little evidence that diagnostic performance (accuracy) could be systematically improved by increasing the amount of information that was available to the judge.

A distressing aspect of our own earlier work on this problem was the inconsistency of some effects that were observed. Intriguing patterns that seemed relatively clear and comprehensible in one or two experiments were not always replicated in subsequent studies. The present report offers a remedy to this state of affairs; it is based on a reanalysis of three previously completed experiments, plus a fourth (newly-collected) data set, and outlines a strikingly consistent pattern of results that had previously been overlooked.

### *Additivity*

Some discussions of the integration process imply a form of "additivity," in that performance is expected to rise as more and more information is made available. In the present experiments, for example, it seemed reasonable to expect, other things being equal, that accuracy would improve in a roughly additive (monotonic) fashion as the respondents were given more and more descriptions on which to base their choices. And as we shall see, our data are generally consistent with this formulation.

A strictly additive theory has further ramifications, however. Most importantly, it implies that performance will invariably be enhanced when additional input that possesses *some* validity, no matter how little, is combined with an existing information ensemble. Note in particular that such a theory implies that performance gains will normally occur even when relatively inaccurate (or low-validity) descriptions are added to an ensemble whose other constituents are substantially higher in validity. This counterintuitive prediction can be derived from the fact that the incorrect referent-choices that are suggested by different descriptive messages are usually quite distinctive (inconstant) as we move from one message to the next. Hence, when the respondent aggregates the information implied by several passages, he/she might detect the inconsistent support for each of the incorrect response alternatives, in contrast to the convergence of evidence in support of the correct alternative (which would be consistent with *all* of the messages in the set).

### *Overload*

While a strict additive formulation implies that performance will improve as more and more information is made available, there is some reason to believe that complications may develop once the size of the message-set exceeds the respondent's processing capacity (Miller, 1960; Barendregt, 1961; Jacoby, Speller, & Kohn, 1974; Norman & Bobrow, 1975). Under these conditions, performance might deteriorate if still more information was made available ("overload"). Suppose, for example, that a subject could respond to no more than 40 semantic elements at a time, and that the average message contained four such elements. Under these assumptions an ensemble of 20 messages could not be fully utilized, since it would require the processing of 80 elements. To cope with this information overload, the respondent might divide his attention between the different passages in the message-set, processing about two elements from each of the 20 passages ( $2 \times 20 = 40$  semantic elements = our proposed maximum). A performance decrement might then be observed. This prediction is based on the assumption that performance will normally decline, if the *amount* of information that is processed remains fixed (at the respondent's maximum capacity), despite the continuing expansion of the overall message-set.

Consider a simplified illustration of this principle: in processing a message ensemble that consisted of just two descriptive passages, information integration and subsequent performance in the referential communication paradigm would be relatively difficult, if the respondent could for some reason recall only the *setting* from one passage, and the *action* from the second. By contrast, integration would be relatively simple if the setting and the action were successfully recalled from a single passage. Writing in an analogous vein, William James (1893) contended that a man who had heard a 12-word sentence would somehow know more than a dozen men, each of whom had heard a single word.

### *Information Quality*

Up to this point, much of our discussion has been implicitly focused on a situation in which the individual passages in a message ensemble are drawn from a common pool, containing many descriptions of comparable quality. Now consider what might happen if the messages were randomly selected from *two* message-pools, one containing a set of *high validity* descriptions (passages that have a relatively good probability of being accurately decoded when presented singly), and the other containing descriptions of *lower validity*.

We have previously suggested that, other things being equal, performance might be systematically influenced by the sheer *size* of the message

ensemble. It also seems reasonable to expect that for any given ensemble size, performance will be directly related to the proportion of high-valid passages in the set. Thus, given a message-set that contained, say, 12 different passages, each of which was relatively accurate (high in validity), we would expect a more effective decoding performance than we would given an ensemble comprised of six high- and six low-validity passages.

## METHOD

To explore the issues that are discussed above, four experiments were conducted; these studies were parallel in design, but differed in a number of procedural details (see below).

### *Message Presentation*

In Experiment I, the descriptive passages were presented in a booklet. Each respondent was also provided with an  $8 \times 10$  sheet containing 24 different photographs that showed the head and shoulders of an actor, portraying a wide range of posed emotions. For each message ensemble, the respondents were to choose the appropriate referent (photograph), working at their own pace.

In Experiments II, III, and IV, the individual descriptions were presented via a slide projector, with a fixed exposure time of 30 sec (in Experiment II), or 15 sec (in Experiments III and IV). After the constituent passages in an ensemble had all been shown, a "picture-slide" was presented, showing the same 24 referent-photographs as in Experiment I. In all three of these timed experiments, the picture-slide was exposed for a total of 60 sec, during which the respondents were to indicate the photograph that had been described in the preceding passage(s).

### *Experimental Design*

All four experiments were based on a complex design that involved the following variables: (a) the number of low-validity descriptions that were presented; (b) the number of high-validity descriptions; and (c) the particular photographs that were the correct choices for the different message ensembles (see Fig. 1).

Figure 1 is a simplified representation of Experiment II; the other experiments followed a similar design scheme. The various cells in Fig. 1 indicate the correct responses for the different message ensembles that the design required. For each combination of high- and low-validity passages, the experiment was arranged so that for some respondents the correct choice was Picture A, while for others it was B (or C or D). This means that by collapsing across Subgroups i-iv it was possible to arrive at an average hit-rate for each combination of messages that was mini-

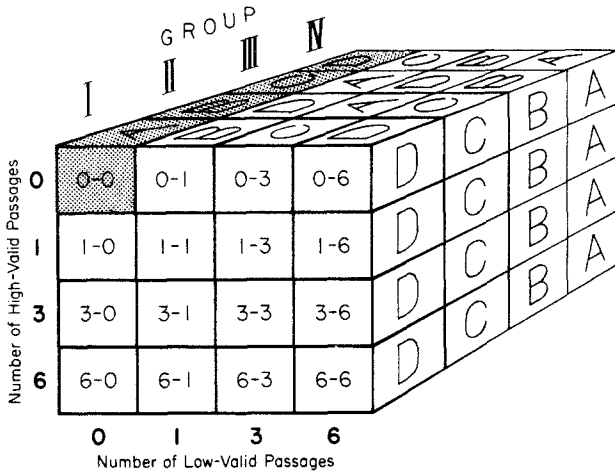


FIG. 1. A graphical representation of Experiment II. The numbers that appear on the front surface of this cube indicate the number of high- and low-validity messages that were included in the different message ensembles. The capital letters (A, B, C, . . .) signify the target-referents (correct choices) for the various message-sets. The shaded cells were not included in the experiment, since they represented void ensembles consisting of zero high-validity and zero low-validity passages.

mally affected by unusual characteristics of the individual target photographs, or by idiosyncratic features of the descriptions.<sup>1</sup>

In Experiment I the number of high-valid passages in the different ensembles constituted a *within*-subjects variable that was changed from one "trial" to the next, while the number of low-valid passages was varied *between* subjects; in Experiments II, III, and IV these assignments were reversed (i.e., the high-valid passages were varied between subjects, and the number of low-valid passages was introduced as a within-subject variable).

*Messages*

The descriptions selected for these experiments were prepared by college students in an earlier study (Manis & Armstrong, 1971), and were

<sup>1</sup> Figure 1 does not include reference to the fact that in Experiments I and IV, two different message-sets were included in each cell of the overall data cube. The counterbalancing scheme that is outlined above was applied to these pairs of message-sets. Another detail that is omitted from Fig. 1 is the fact that Experiments II and III included a set of delayed response trials, in which (unpredictably) the respondents were required to work for 2 min on a set of complex arithmetic problems, before being shown the picture-slide from which their choices were to be made. Results from these delay trials are not included in the present report.

based on an actor's portrayal of various emotional states. Here is a reasonably typical description: "This is an expression of disbelief. He is utterly amazed at something that is happening, and his face involuntarily shows his dumbfounded look."

Message ensembles were constructed by orthogonally varying the number of high- and low-validity descriptions. The validity of the individual descriptions was determined by presenting each one (singly) to a group of undergraduates similar to those in the present experiments. The high-validity descriptions yielded hit-rates between 40% and 64%. The low-validity descriptions generated hit-rates that varied between 10% and 32% (chance performance based on blind guessing would yield a hit-rate just over 4%); low-validity passages normally led to a wide variety of referent-choices and did not consistently elicit the same incorrect choice. Indeed, for each of the passages that was ultimately included in these experiments (both high- and low-validity), the correct choice was the modal selection in our norm group. Within each ensemble, the component passages were irregularly ordered, with some attempt to avoid excessive "bunching" of high- (or low-) validity passages.

In all four experiments the high- and low-validity passages were selected in accordance with a "telescope" design. Suppose, for the sake of illustration, that Description 1 was a high-valid passage that was to be included in all of the ensembles that were (a) based on Photograph A, and that (b) included just one high-validity description. The telescoping principle dictated that this same passage (1) must also be included in all ensembles that called for *more than one* high-validity description of photograph A. Similarly, if Passages 1, 2, and 3 were to be presented in ensembles that called for three high-validity descriptions of Photograph A, the telescoping scheme required that these same three passages must be included in all ensembles that contained more than three high-validity descriptions of Photograph A. The telescoping scheme was introduced to guard against the possibility that an ensemble that included, say, three high-valid passages (1, 2, and 3) might in fact contain *less* information than a single high-valid passage (4) of unusual length and inclusiveness.

Table 1 shows the various "levels" of high- and low-valid passages that were included in each experiment (e.g., in Experiment I, each message-

TABLE 1  
NUMBER OF HIGH- AND LOW-VALIDITY PASSAGES IN THE FOUR EXPERIMENTS

| Experiment | N high-valid passages | N low-valid passages |
|------------|-----------------------|----------------------|
| I          | 0, 1, 3, or 6         | 0, 1, 2, 4, or 6     |
| II         | 0, 1, 3, or 6         | 0, 1, 3, or 6        |
| III        | 0, 3, 6, or 12        | 0, 3, 6, or 12       |
| IV         | 0, 1, 3, or 11        | 0, 1, 3, or 11       |

set contained 0, 1, 3, or 6 high-valid passages, along with 0, 1, 2, 4, or 6 low-valid passages). The results presented below were based on ensembles that included at least one descriptive message (i.e., void ensembles, comprised of zero HV and zero LV components were not considered).

### *Respondents*

The respondents in these experiments were native born undergraduates at the University of Michigan ( $N_1 = 80$ ;  $N_2 = N_3 = N_4 = 160$ ). Respondents were recruited from several campus gathering spots and were paid for their efforts.

## RESULTS

The introduction to this paper focusses on two determinants of decoding performance: the *quantity* and the *quality* of the available information. These variables were operationalized as follows. (a) *Quantity* of information =  $T$  = number of high-valid (HV) passages in a given ensemble + number of low-valid (LV) passages in that ensemble. (b) *Quality* of information =  $HV/T = n$  high-valid passages in an ensemble / total passages in the ensemble.

To assess the impact of these predictors, the data from each experiment were first aggregated by collapsing across subgroups (see Fig. 1). This enabled us to determine the overall performance levels obtained with the different types of ensembles. For example, we learned that Experiment I yielded an average hit-rate of 28% for ensembles with just two low-valid passages (and no high-valid descriptions), and a hit-rate of 44% for ensembles with four low-valid passages. Correlations were calculated between aggregated hit-rates of this type and our two predictor variables, HV/T and T. The results are shown in Table 2; they indicate that performance in all four experiments was strongly affected by the *quality*, and to a lesser extent, by the *quantity* of available information. A crude evaluation of these results, using a one-tailed sign test (Siegel, 1956), yields a  $p$ -value of .06 for each predictor.<sup>2</sup> Table 2 also shows the multiple correlation between our two predictors and the observed hit-rates, and indicates that considered together, HV/T and T account for over 50% of the variance in each of our four data sets, with  $R$ s ranging from .73 to .86.

Despite this pattern of clearcut, replicated results, it is difficult to go beyond a simple sign test in evaluating the reliability of these results, since all four experiments involved a "repeated measures" design that complicated our attempts to assess the significance of the individual correlations. Thus, while Fig. 1 outlines an experimental design involving 15 types of

<sup>2</sup> The two predictors, T and HV/T, were uncorrelated in Experiments II, III, and IV. They were very modestly correlated in Experiment I ( $r = .09$ ); however, the results shown in Tables 2 and 3 are unaffected when this mild "contamination" is removed by partial correlation.

TABLE 2  
CORRELATIONS BETWEEN AGGREGATED HIT-RATES AND TWO PREDICTORS

| Experiment   | Predictor            |                            | Multiple R |
|--------------|----------------------|----------------------------|------------|
|              | Ensemble size<br>(T) | Ensemble quality<br>(HV/T) |            |
| I (N = 19)   | .34                  | .66                        | .74        |
| II (N = 15)  | .29                  | .67                        | .73        |
| III (N = 15) | .28                  | .81                        | .86        |
| IV (N = 15)  | .44                  | .69                        | .82        |

Note. N = number of different ensemble types.

message ensembles, the aggregated results for the different types of ensembles were not fully independent of one another, since each subject responded to four different message-sets.

To overcome this problem each experiment was reanalyzed, focussing separately on subsets of the data that did *not* involve repeated measurements. Consider, for example, the 15 different message ensembles for which Picture A was the correct response (see Fig. 1). These 15 ensembles were presented to different subgroups of respondents, and hence the observed hit-rates were independent of one another. Photograph B was also the target for 15 different message-sets which were presented to different subgroups of respondents; and similarly for Pictures C and D. By calculating separate correlations for the different target-referents in each experiment, we were able to obtain four separate estimates of the effectiveness of our predictors. These results are shown in Table 3, together with (a) the average of the  $r$ 's obtained in each experiment, based on the  $r$  to  $z'$  transformation (McNemar, 1969), (b) the one-tailed  $p$ -values associated with each of these averaged results, and (c) the combined  $p$ -value for the results across all four experiments (Rosenthal, 1978).

Note first that both predictor variables continue to show consistent positive relationships with the observed hit-rates. Thus,  $T$ , the total number of elements in the ensemble, was positively related to performance in 13 out of 16 correlations.  $HV/T$ , the proportion of high-valid passages, was positively related to performance in 14 of the 16 analyses.

To assess the reliability of these results, for each predictor we first determined the  $p$ -value for the mean of the correlations in the individual experiments, making the conservative assumption that the degrees of freedom associated with these results were equal to  $N - 2$ , where  $N$  was the number of different ensemble types represented in each experiment (i.e., 19 in Experiment I, and 15 in Experiments II, III, and IV). The obtained  $p$ -values were then combined across the four experiments.



TABLE 3  
CORRELATIONS BETWEEN PREDICTORS AND PERFORMANCE

| Variable: Ensemble size (T) |                 |                  |                   |                  |
|-----------------------------|-----------------|------------------|-------------------|------------------|
|                             | Exp. I (N = 19) | Exp. II (N = 15) | Exp. III (N = 15) | Exp. IV (N = 15) |
|                             | .40             | .16              | .50               | .35              |
|                             | .08             | .50              | .49               | .10              |
|                             | .08             | .05              | -.12              | .62              |
|                             | .22             | -.08             | -.06              | .10              |
| Mean <i>r</i>               | .20             | .17              | .23               | .32              |
| <i>p</i>                    | .21             | .27              | .26               | .12              |

*Note.* Combined *p* across studies: .02 (mean *p* method); .02 (adding *p* method); .05 (adding *Z* method)

| Variable: Ensemble quality (HV/T) |                 |                  |                   |                  |
|-----------------------------------|-----------------|------------------|-------------------|------------------|
|                                   | Exp. I (N = 19) | Exp. II (N = 15) | Exp. III (N = 15) | Exp. IV (N = 15) |
|                                   | .50             | -.03             | .25               | .73              |
|                                   | .64             | .48              | -.20              | .62              |
|                                   | .43             | .83              | .78               | .46              |
|                                   | .10             | .47              | .92               | .14              |
| Mean <i>r</i>                     | .44             | .50              | .58               | .52              |
| <i>p</i>                          | .03             | .045             | .01               | .025             |

*Note.* Combined *p* across studies: .001 (mean *p* method); .0001 (adding *p* method); .001 (adding *Z* method).

The results were quite clear and consistent with the pattern previously shown in Table 2. Briefly stated, the *quality* of the various message ensembles (HV/T) was strongly and reliably related to the observed hit-rates; the *size* of the message sets (T) also proved to be a significant predictor, although the results here were less striking. Note however, that the results in Table 3 are somewhat weaker (lower *rs*) than the corresponding entries in Table 2. In large part this is due to the fact that the criterion measures (hit-rates) were more reliable when they were based on a single aggregation of *all* the data in a given study, as in Table 2. The aggregation of responses across subgroups boosts reliability in a fashion that is analogous to the results obtained when a psychological test is lengthened, thereby reducing its error variance and enhancing its correlation with other variables.

While these results indicate that decoding performance was positively related to both the quality and the quantity of the information in the different ensembles, the analyses presented above are *not* sensitive to the interactive effects that these variables might produce (Anderson & Shan-

teau, 1977). Figure 2 focusses on this possibility. The figure was constructed by calculating the difference in performance between ensembles of a given size that were exclusively composed of low-valid passages ( $HV/T = 0.00$ ), and those of that same size that included only high-valid descriptions ( $HV/T = 1.00$ ).

If our two independent variables,  $T$  and  $HV/T$ , had *no* interactive effects, Fig. 2 would consist of four horizontal lines, indicating that the effects produced by changes in ensemble quality ( $HV/T$ ) were constant, regardless of the ensemble size ( $T$ ). While Fig. 2 does not exhibit this pattern, due perhaps to random perturbations in the data, it is difficult to see any consistent trend from one experiment to the next. Thus, while the results from Experiments I and II suggest that changes in message quantity ( $HV/T$ ) have their most striking effect on performance when  $T$  is small, Experiments III and IV show very different patterns. Because of the inconsistent character of these results we conclude that the present experiments provide scant support for the proposition that  $T$  and  $HV/T$  have a replicable, interactive effect on performance.

### DISCUSSION

These results are consistent with a model of the integration process that treats performance as a linear function of the quality and the quantity of the available information, with primary emphasis on quality. To appreciate the full significance of this conclusion it may be instructive to consider two models that are to some degree *inconsistent* with the present data.

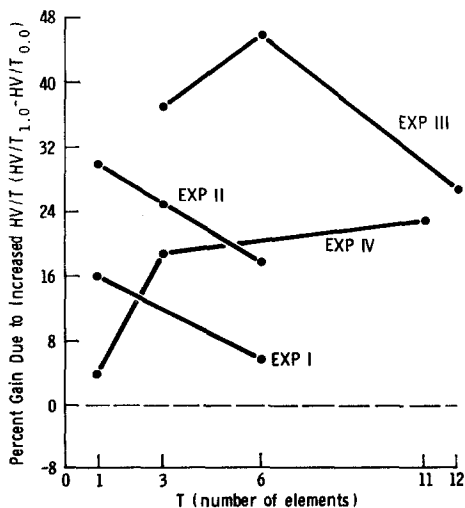


FIG. 2. Performance gains due to changes in ensemble-quality ( $HV/T$ ), as a function of ensemble-size ( $T$ ). See text.

(1) *Additive models*. All four experiments showed a modest tendency for performance to improve as a function of T, the number of descriptions in the message-set. In part then, the present results are consistent with an additive model, for they suggest that the respondents had some success in their attempts to aggregate the information from the different passages they received.

On the other hand, further analysis indicated at least one important respect in which these data are inconsistent with a *strict* additive formulation. A strict additive model suggests that performance will inevitably improve whenever additional messages that possess *some* validity (no matter how little) are introduced into an existing ensemble. Our results are in severe conflict with this formulation when we consider instances where the addition of low-validity information reduces the overall quality of an existing ensemble. (Note: Ensemble-quality, as indexed by HV/T, will always decline when low-valid material is introduced into a message-set that includes at least one high-valid component). Thus, in ensembles that included one or more high-valid passages, all four experiments yielded a *negative* relationship between performance, and the number of low-valid constituents in the different message-sets (*r*s ranged from  $-.19$  to  $-.52$  in the four experiments).<sup>3</sup> This result, showing the disruptive impact of low-validity cues when combined with cues of higher validity seems to be a robust phenomenon, for a similar pattern had previously been reported by Dudycha and Naylor in a study of multiple cue probability-learning (1966, see Fig. 3).

The deleterious effects produced by low-valid information may derive from the fact that any passage is likely to be a more salient and influential factor in the choice process when it appears in a small ensemble, rather than a large one. This implies that as more and more low-valid messages are introduced into a message-set that initially included at least one high-valid description, there should be an enhanced likelihood that the respondent might forget (or ignore) some of the high-validity material when making his choice. High-validity passages that were not considered might plausibly be "replaced" by some of the low-validity messages that had been added to the ensemble, and performance would decline as a result.<sup>4</sup>

<sup>3</sup> These results have a one-tailed *p*-value of .06. By contrast, in the message-sets that were devoid of high-valid passages, where inclusion of new LV passages simply affected the ensemble size (T), but did not affect ensemble *quality* (HV/T = 0.0 for all of these ensembles), three of the four data sets showed a positive relationship between the observed hit-rates and the number of low-valid descriptions.

<sup>4</sup> Unpublished research by Platt and Manis (Note 1) suggests that the salience of a given passage is also affected by its serial position within the overall message-set. Thus, as might be anticipated from previous work in verbal learning, messages that are presented in the middle of a series are less influential than those presented either early or late in the series.

The failure of a strictly additive model has practical implications for situations in which a selection officer or clinical diagnostician is faced with the task of evaluating a series of individuals, using a variety of informational inputs. Traditionally, such decision-makers have welcomed both high- and low-validity inputs, in the belief that information that was low in validity might nonetheless be particularly revealing in *individual* cases, and would receive its appropriate (negligible) weighting in other instances.

The present data, together with the results of Dudycha and Naylor (1966), argue against this assumption, for both studies indicate that performance normally deteriorates when low-validity cues are combined with cues of high validity. Note, however, that in evaluation contexts where the integration process can be executed through a formal regression procedure (e.g., in predicting scholastic performance), there is every reason to expect that predictive accuracy can be enhanced by incorporating relatively low-validity predictors into an informational pool of higher validity. The contrast here between intuitive human judgment and the results that can be attained through formal statistical procedures is quite marked.

(2) *Overload*. Some theorists suggest that performance will deteriorate if the cognitive system is "overloaded" with information. The present results are, however, inconsistent with this formulation when we consider message sets that vary in size (T), but are comparable in quality (HV/T). For example, in "homogeneous" ensembles that included either all high-valid or all low-valid passages, there was no convincing evidence of a performance decrement with increases in set-size, even in ensembles that included as many as 12 different passages. Similarly, the overload formulation was not supported when we considered ensembles that included equal amounts of high- and low-validity information, but varied in overall size (for example, message-sets that included three high- and three low-validity passages, vs those that included 12 high- and 12 low-validity descriptions). Despite substantial increases in the size of these ensembles, which included up to 24 different passages in Experiment III, the observed performance pattern was *not* consistent with the overload model.

Although it has long been recognized as an intriguing possibility, the overload concept does not have much empirical support in experiments on information integration.<sup>5</sup> Briefly stated, we are unaware of a single

<sup>5</sup> The overload concept has been interpreted in several distinct ways. We have looked for overload effects by varying the *amount* of information (T) presented in a cognitive integration task. James G. Miller (1960) conducted a series of experiments that varied the *rate* of information presentation, to study the common modes of response to this form of overload. Norman and Bobrow (1975) theorized about the performance patterns that might be anticipated when two tasks that utilize the same processing resources are performed simultane-

study which has convincingly shown that simple increases in the *amount* of available information (T) will ultimately depress performance in a cognitive integration task. While some recent experiments by Jacoby and his colleagues suggested the possibility of such an effect in the realm of consumer choices (1974), subsequent analysis casts considerable doubt on this interpretation (Russo, 1974). Similarly, a paper by Barendregt (1961) provides doubtful evidence for the overload concept. In this study, a clinical psychologist attempted to predict the MMPI responses of 10 psychiatric patients after having been provided with varying amounts of other test data. As more and more information was made available, performance first improved, and then eventually declined. Unfortunately, however, the largest data sets included test information that was poorer, on the average, than the smaller data sets. It is thus quite possible that the performance decrement in Barendregt's study derived from a reduction in information quality, rather than from cognitive overload.

Finally, a study with the interesting title, "Clinical prediction: does one sometimes know too much?" (Bartlett & Green, 1966) yielded results very similar to those reported by Barendregt. In this experiment, six experienced counseling psychologists predicted grade point averages based on input information that consisted of either four or 22 different predictor variables. Every judge performed more effectively when presented with the four variable predictor set. This small set of predictors was, however, substantially more valid, when considered as a separate block, than the 18 additional predictors with which it was combined to form the large (22 variable) predictor set; on the average then, the small (four variable) set contained information that was better in quality (more valid) than the information in the large (22 variable) set. The performance decrement that was observed when all 22 variables were displayed may derive from this reduction in the quality of the available information.

Overall then, the results of the present investigation closely parallel those reported in a number of related research traditions, by failing to find evidence in support of the overload concept.

## REFERENCES

- Anderson, M. H., & Shanteau, J. Weak inferences with linear models. *Psychological Bulletin*, 1977, **84**, 1155-1170.
- Barendregt, J. *Research in psychodiagnosics*. The Hague: Morton, 1961.
- Bartlett, C. J., & Green, C. G. Clinical prediction: Does one sometimes know too much? *Journal of Counseling Psychology*, 1966, **13**, 267-270.

---

ously. Under some (but not all) arrangements of this type, their model predicts a degradation in "primary" task performance as more and more resources are expended on the "interfering" task.

- Dudycha, L. W., & Naylor, J. C. Characteristics of the human inference process in complex choice behavior situations. *Organizational Behavior and Human Performance*, 1966, **1**, 110–128.
- Goldberg, L. R. Simple judgments or simple processes? Some research on clinical judgments. *American Psychologist*, 1968, **23**, 483–496.
- Jacoby, J., Speller, D. E., & Kohn, C. A. Brand choice behavior as a function of information load: Replication and extension. *Journal of Consumer Research*, 1 (June 1974).
- James, W. *Psychology, a briefer course*. New York: Holt & Company, 1893.
- Manis, M., & Armstrong, G. W. Contrast effects in verbal output. *Journal of Experimental Social Psychology*, 1971, **7**, 381–388.
- Manis, M., & Platt, M. B. Integrating referential information. In M. Kaplan & S. Schwartz (Eds.), *Human judgment and decision processes*. New York: Academic Press, 1975. Pp. 173–195.
- Manis, M., & Platt, M. B. Referential redundancy and the integration of verbal information. *Journal of Experimental Social Psychology*, 1976, **12**, 26–37.
- Manis, M., & van Rooijen, L. Integrating the information in referential messages. In R. L. Solso (Ed.), *Contemporary issues in cognitive psychology: The Loyola symposium*. Washington, DC: V. H. Winston & Sons, 1973.
- McNemar, Q. *Psychological statistics* (4th ed.). New York: John Wiley & Sons, 1969.
- Miller, J. G. Information input overload and psychopathology. *American Journal of Psychiatry*, 1960, **116**, 695–704.
- Norman, D. A., & Bobrow, D. G. On data-limited and resource-limited processes. *Cognitive Psychology*, 1975, **7**, 44–64.
- Rosenthal, R. Combining results of independent studies. *Psychological Bulletin*, 1978, **85**, 185–193.
- Russo, J. E. More information is better: A reevaluation of Jacoby, Speller, and Kohn. *Journal of Consumer Research*, 1974, **1**, 68–72.
- Siegel, S. *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill Book Co., 1956.

## REFERENCE NOTE

1. Platt, M. B., & Manis, M. *The effect of serial position on the integration of verbal descriptions*. Midwestern Psychological Association Meeting, Chicago, 1976.

RECEIVED: April 20, 1978