Acta Psychologica 30 *Attention and Performance II (W. G. Koster, ed.)* 1969, 276–315

THE DISCOVERY OF PROCESSING STAGES: EXTENSIONS OF DONDERS' METHOD

SAUL STERNBERG [1]
Bell Telephone Laboratories, Murray Hill, N.J., U.S.A.

ABSTRACT

A new method is proposed for using reaction-time (RT) measurements to study stages of information processing. It overcomes limitations of Donders' and more recent methods, and permits the discovery of stages, assessment of their properties, and separate testing of the additivity and stochastic independence of stage durations. The main feature of the *additive-factor method* is the search for non-interacting effects of experimental factors on mean RT. The method is applied to several binary-classification experiments, where it leads to a four-stage model, and to an identification experiment, where it distinguishes two stages. The sets of stages inferred from both these and other data are shown to carry substantive implications. It is demonstrated that stage-durations may be additive without being stochastically independent, a result that is relevant to the formulation of mathematical models of RT.

1. INTRODUCTION

The work of DONDERS (1868) that we have been commemorating was based on the idea that the time between stimulus and response is occupied by a train of successive processes, or stages: each component process begins only when the preceding one has ended. Donders developed the *subtraction method* to measure the durations of some of these stages, and thereby study their properties; mean reaction-times (RTs) from two different tasks are compared, where one task is thought to require all the stages of the first, plus an additional stage. The difference between mean RTs is taken to be an estimate of the mean duration of the interpolated stage. The method was popular for several decades (see JASTROW, 1890) and then came into disfavor (see KÜLPE, 1895).

Although it has seen something of a revival in the last few years (e.g., NEISSER, 1963; STERNBERG, 1966; TAYLOR, 1966; POSNER and

[1] I am indebted to J. Krauskopf for several helpful suggestions. I also thank P. D. Bricker, C. S. Harris, R. S. Nickerson, and G. Sperling for criticisms of the manuscript, B. Barkow and B. A. Nasto for laboratory assistance, and D. S. Hougak, P. L. Moore, B. A. Nasto, and A. M. Pope for serving as subjects in exp. V.

MITCHELL, 1967; SNODGRASS et al., 1967; SMITH, 1968), little is known about how to test the validity of any particular application of the subtraction method. The underlying conception of the RT as a sum of durations of a series of stages is now a popular one, but there is remarkably little strong supporting evidence. And there is even less evidence that stage durations are stochactically independent, an assumption often incorporated with the idea of additivity (e.g., MCGILL, 1963; TAYLOR, 1966; HOHLE, 1967).

The early applications of Donders' idea were criticized partly because introspective data suggested that it might be difficult to devise experimental tasks that would add or delete one of the stages between stimulus and response without also altering other stages. In this paper I propose a simple method of testing for additive RT-components that opens up new possibilities for inferring the organization of mental operations from RT data without requiring procedures that add or delete stages. Unlike Donders' method it does not lead to the measurement of stage durations, but like his method it can be used to help establish the existence and properties of stages, and the relations among tnem. The method is applied to data from two kinds of choice-reaction experiment, a binary-classification ('partial-identification', or many-one) task, and a 'complete-identification' (one-one) task (BUSH et al., 1963). A generalization of the method is used to test the idea that RT components are stochastically independent. And it is shown how these methods also permit localizing the effect of a new experimental factor among a set of stages already established.

2. ASSUMPTIONS ABOUT STAGE DURATIONS IN RECENT STUDIES

2.1. *Three types of assumption*

In recent years three main propositions have been considered in the analysis of RT into components; they are listed in fig. 1. The proposition that is of main interest, and the one that reflects Donders' idea, is that there are successive functional stages between stimulus and response, whose durations are additive components of the RT. Here T_a and T_b are random variables representing the durations of two different stages, and T_w is a wastebasket category representing the total duration of all other events between stimulus and response. The first proposition implies that the mean RT is the sum of the means of the components.

A supplementary assumption sometimes treated as inseparable from that of additivity is that the RT-components are stochastically independ-

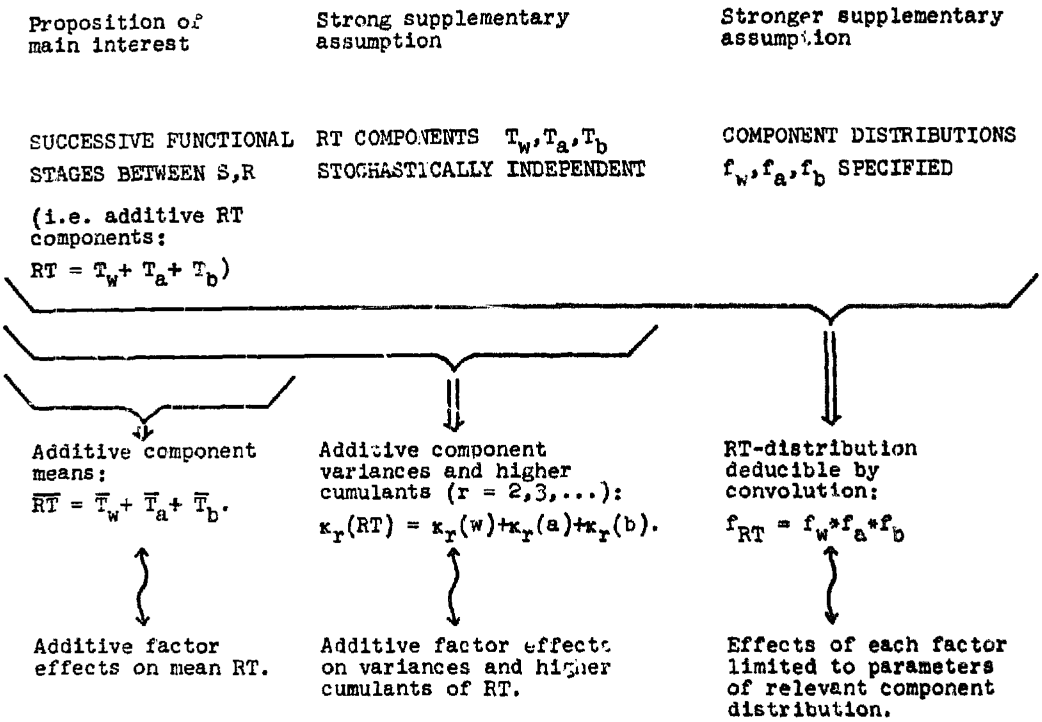

Fig. 1. Three types of assumption in the decomposition of RT, and their implications. Wavy arrows show loose implications (see section 3.1); statements in bottom row apply only to experimental factors that influence no stages in common.

ent. (I will show in section 5.4 why these assumptions should be examined separately and why a definition of stage that involves additivity without independence might be a useful one.) Taken together with the first proposition, the assumption of independence has strong implications: not only are the component variances additive – since the variance of the sum of independent quantities is the sum of their variances – but all the higher cumulants are additive as well. (Cumulants are statistics of a distribution that are closely related to its moments and that are estimated without bias by k-statistics; see Kendall and Stuart, 1958.)

An even stronger supplementary assumption is one that specifies the forms of the components' distributions. (Exponentially-distributed stage durations, for example, have sometimes been assumed.) Given the distributions and the other assumptions, one can deduce the RT distribution itself. (Intermediate cases arise when a feature of the RT distribution is inferred from assumptions in which the forms of components are only partially specified: Luce and Green (1969) provide an example.)

2.2. *Tests of the assumptions*

Almost always, a 'strong' model has been tested, in which the proposition of main interest is combined with both of the supplementary assumptions. This approach is exemplified by the work of CHRISTIE and LUCE (1956), AUDLEY (1960), RESTLE and DAVIS (1962), MCGILL (1963), MCGILL and GIBBON (1965), and HOHLE (1967). A central notion in such work is that the form of the RT distribution (which depends on the postulated distributions of the components and on their independence, as well as on their additivity) is a key to the underlying process.

According to Hohle, for example, the RT is the sum of an exponentially-distributed 'decision' component and an independent, normally-distributed 'residual' component (representing the summed durations of all other processes). He has successfully fitted the resulting theoretical distribution to RTs from several experiments. To examine the model further he obtains parameter estimates for the theoretical distribution from RTs at two or more levels of an experimental factor. One consequence of Hohle's choice of hypothetical component distributions is that these estimates reveal the extent to which each of the components is responsible for the effect of the factor on RT. If the model is correct, changes in some factors should influence only the decision component, while changes in others should influence only the residual component. But overall findings from a series of studies by HOHLE (1967) and GHOLSON and HOHLE (1968a,b) are not entirely consistent with this expectation.

Unlike Donders' method, in which experimental manipulations are required to add or delete entire stages, Hohle's requires only that the amount of processing required of a stage, and hence its duration, be manipulated. This feature, which extends the range of situations in which analysis of RTs can be performed, also characterizes the new method to be proposed in section 3.

One problem for an approach such as Hohle's, in which a strong model is invoked, is that when the model fails it is of course difficult to decide which of its several assumptions is at fault. A second problem is that rather different sets of components may give rise to RT distributions that have approximately the same form. There are several advantages, therefore, in testing relatively weaker models, or examining assumptions one at a time.

TAYLOR (1966) has tried to test the main proposition together with only the first supplementary assumption. As in the early applications of Donders' idea, he attempted to construct tasks that would add or delete entire stages. An important innovation in Taylor's work is the inclusion of a test of the additivity of stage durations: if one change in task adds stage *a* and an increment $\overline{T}_a$ in mean RT, and a different change in task adds stage *b* and an increment $\overline{T}_b$, then the two changes together should add both stages and a combined increment $\overline{T}_{ab} = \overline{T}_a + \overline{T}_b$. Such a test can validate applications of the subtraction method, and protect it from the early criticisms according to which the change in task that adds a particular stage may also alter other stages. Additivity tests also characterize the method to be proposed in section 3. Unfortunately, Taylor's test cannot be said to have succeeded, mainly because his experiment was insufficiently precise.

Whereas Taylor felt that the assumptions of additivity and independence should be tested jointly, additivity has been examined alone in a series of studies on sentence verification (McMahon, 1963; Gough, 1965, 1966). The hypothesis being tested was that negative and passive transformations in stimulus sentences add separate stages to the process of verification; if they do, the transformations should have additive effects on mean RT. This was found by McMahon, but Gough found a systematic tendency for the combined increment, $\overline{T}_{ab}$, to be less than $\overline{T}_a + \overline{T}_b$, a deviation from additivity in the same direction as the trend in Taylor's data.

Experimental operations like these, which might be thought of as deleting entire stages without altering the functions of other stages, are probably very rare; they should be considered special cases. Another example of this kind of special case arises in certain memory-search tasks (e.g., Sternberg, 1966) where it can be assumed or inferred that the number of elements scanned, and therefore the number of similar stages, is under experimental control. (Some visual-search tasks, as in Neisser, 1963, are similar.) Here the desired additivity test of the main proposition is accomplished by evaluating the linearity of the function relating mean RT to the number of elements scanned. The slope of this function represents the mean time to scan one item; its zero-intercept represents the combined durations of all events other than scanning. (Whereas Donders might have attempted to measure the zero-intercept directly, by devising an experiment in which no elements are scanned, here one can estimate its value by extrapolation.) In such search tasks, moreover, the independence assumption can be tested separately by examining the relations between each of the cumulants of the RT distribution and the number of elements scanned: if the assumption of independence is justified, these relations are also linear, and the slopes and zero-intercepts of the linear functions can be used as estimates of the cumulants of the components. These estimates in turn can be used to determine the forms of the component distributions (STERNBERG, 1964).

3. SUCCESSIVE STAGES AND ADDITIVE FACTOR-EFFECTS: A NEW METHOD

3.1. *Implications of factor-stage relations*

The method I shall propose seems to apply to a wide variety of RT situations, rather than only to those special cases where experimental manipulations can add entire stages or produce known changes in the number of identical stages. Yet it seems to permit the proposition of main interest to be tested by itself. Suppose that stages *a*, *b*, and *c* shown in fig. 2 are among a series of stages between stimulus and

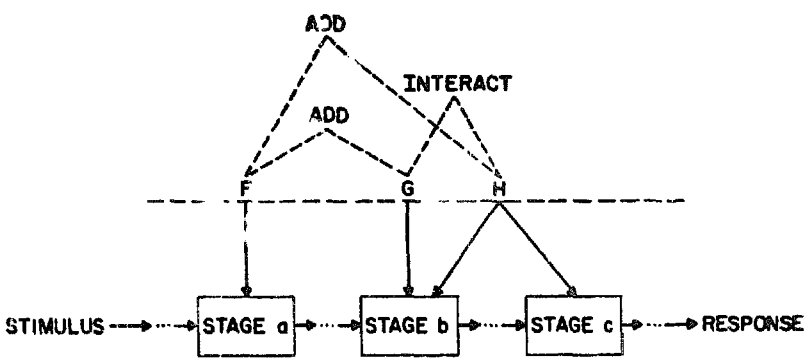

Fig. 2. Example of an arrangement of stages (*a*, *b*, and *c*) and factors (F, G, and H). Below the horizontal line are shown three hypothetical stages between stimulus and response. Horizontal arrows represent inputs and outputs of stages; time proceeds from left to right. Dots indicate the possibility of other stages in a string in which the hypothetical stages are embedded. Arrows are drawn from factors to the stages assumed to be influenced by those factors. Above the line are indicated the relations among effects of the factors on mean RT that are expected from the arrangement.

response. Suppose further that there are three experimental factors, F, G, and H, such that factor F influences only the duration of stage *a*, factor G influences only the duration of stage *b*, and factor H influences stages *b* and *c*, but not *a*. (By a 'factor' here is meant an experimentally manipulated variable, or a set of two or more related treatments called 'levels'; the 'effect' of a factor is the change in the response measure induced by a change in the level of that factor.) What are the most likely relations among the effects of the three factors on mean RT? These relations are shown above the broken line. The general idea is that when factors influence no stages in common, their effects on mean RT will be independent and additive, because stage durations are additive. That is, the effect of one factor will not depend on the levels of the others. Thus, factors F and G should have additive effects on mean RT. On the other hand, when two factors, G and H, influence at least one stage in common (stage *b*) there is no reason to expect their effects on RT to add; the most likely relation is some sort of interaction.

One can imagine exceptions to both of these rules, of course. Factors G and H might just happen to influence stage *b* additively, and their effects on RT would then also be additive, even though they influenced a stage in common. (This notion would gain strength if, for example, *other* factors either interacted with *both* G and H or with *neither.)* Alternatively, if factor F influenced the output of stage *a* as well as its duration, then it might indirectly influence the duration of stage *b*. This could lead to an interaction between the effects of factors F and G even though they did not *directly* influence any stages in common. (An example is given in section 4.4, footnote 4.) But by and large, factors that influence different stages will have additive effects on mean RT, whereas factors that influence stages in common will interact.

3.2. *The additive-factor method and the meaning of 'stage'*

The direction of these inferences is reversed in the 'additive-factor method', in which one searches for pairs of factors, like F and G, that have additive nonzero effects. Whenever such 'additive factors' are discovered, and given no stronger arguments to the contrary, it is reasonable to believe that there exists a corresponding pair of stages, *a* and *b*, between stimulus and response. (Conversely, if one cannot find a pair of additive factors that correspond to a pair of hypothesized stages, this may be taken as evidence against the hypothesis; but see section 3.4 for one exception.) Furthermore, if a third factor, H, is found to interact with G but not with F, this implies that H influences RT at least in part because of its effect on stage *b*, but not because of any effect on stage *a*.

I have deliberately avoided a precise definition of 'stage', which should await further research. The basic idea is that a stage is one of a series of successive processes that operates on an input to produce an output, and contributes an additive component to the RT. The concept of 'additivity' here entails a property of independence for mean stage-durations: the mean duration of a stage depends only on its input and the levels of factors that influence it, and not directly on the mean durations of other stages. The fundamental importance of the relations among factor effects follows from this basic idea. It is not only additivity among factors that is useful, but also the patterns of interaction: the subsets of interacting factors associated with a stage and the ways they interact allow one to infer the operations performed by that stage, and possibly also its location in a series of stages and its internal structure.

Other features that might be incorporated in a formal definition of 'stage' are: (1) Given its input, the output of a stage should be independent of factors influencing its duration. This requirement would preclude indirect factor effects, such as that of factor F on stage *b* mentioned above. (2) The stages in a series should be functionally interesting and qualitatively different and should 'make sense' in terms of other knowledge. (3) A stage should be able to process no more than one 'signal' at a time (as in WELFORD, 1960). (4) Stage durations should be stochastically independent (see section 5.4). It remains to be seen whether the stages defined by additive components have these properties.

3.3. *Additivity in two-factor experiments*

In table 1 the relation between additive RT-components and additive factors is shown in the context of a complete (2 × 2) two-factor experiment.

TABLE 1

Additive RT-components and additive factor-effects in a 2 × 2 experiment.

factor level	F_0	F_1	G_0	G_1
stage influenced	*a*	*a*	*b*	*b*
duration of stage	$T_a(0)$	$T_a(1)$	$T_b(0)$	$T_b(1)$

experimental conditions	reaction time
F_0, G_0	$RT(00) = T_w + T_a(0) + T_b(0)$
F_0, G_1	$RT(01) = T_w + T_a(0) + T_b(1)$
F_1, G_0	$RT(10) = T_w + T_a(1) + T_b(0)$
F_1, G_1	$RT(11) = T_w + T_a(1) + T_b(1)$

periment with two levels per factor. Suppose we have found a pair of factors, F and G, that influence different stages, a and b, as in fig. 2, and we study each factor at two levels, labeled zero and one. At the top of the table are given the durations of each stage as a function of the factor influencing that stage. The four pairs of experimental conditions are shown below, with corresponding RTs, T_w again represents the durations of all processes other than stages a and b. Given only the proposition of successive stages, it follows from the equations in table 1 that the means of the four RT distributions should be related by eq. (1), which is an expression of the additivity of factor effects on means:

$$\mu'(00) + \mu'(11) = \mu'(01) + \mu'(10). \tag{1}$$

With the supplementary assumption of the stochastic independence of RT components, a series of similar equations must hold for all the cumulants:

$$\varkappa_r(00) + \varkappa_r(11) = \varkappa_r(01) + \varkappa_r(10), r = 1,2,\ldots \tag{2}$$

For example, our two factors must show the same kind of additivity in their effects on the RT variance ($\varkappa_2$) as on its mean:

$$\sigma^2(00) + \sigma^2(11) = \sigma^2(01) + \sigma^2(10). \tag{3}$$

(The relations shown in eqs. (1)-(3) describe properties of population distributions, of course; because of sampling error – which grows with the order of the cumulant – it is a statistical question whether a set of empirical distributions has these properties.) Finally, if the forms, f_w, f_a, and f_b, of the component distributions are known, then the effects of a factor, say F, on parameters of the resulting RT distribution are limited to those that correspond to changes in parameters of the relevant component, in this instance, f_a. (Hohle's method, described in section 2.2, makes use of this last implication.) These ideas are summarized at the bottom of fig. 1.

Generalization of the above analysis to an experiment with p levels of factor F and q levels of factor G is straightforward (see, e.g., SCHEFFÉ, 1959). If F and G influence different stages, then for the mean RTs from the pq conditions there must exist constants $\overline{T}_w$, $\overline{T}_a(i)$, $i=1,\ldots,p$, and $\overline{T}_b(j)$, $j=1,\ldots,q$ such that

$$\overline{RT}\,(i,j) = \overline{T}_w + \overline{T}_a(i) + \overline{T}_b(j),\ i=1,\ldots,p,\ j=1,\ldots,q. \tag{4}$$

Similar conditions exist for all cumulants if independence is assumed. Equivalently, eq. (1), (2), or (3) is replaced by equating to zero each of $(p\text{-}1)(q\text{-}1)$ linearly-independent contrasts in the RT means or cumulants. Each of these contrasts represents one of the degrees of freedom associated with deviations from the additive model of eq. (4). Additivity may be evaluated either by examining deviations from an explicitly-fitted additive model (eq. (4)), by evaluating the appropriate contrasts, or by testing the interaction term in an analysis of variance.

Evidence that RT components are stochastically independent adds strength to the proposition that they represent the durations of different stages. But a failure to confirm the assumption of independence does not necessarily weaker this proposition. In section 5.4 I shall show why stage durations might be additive but not independent.

3.4. *Generalization of the method to multiple-factor experiments*

Generalization of the additive-factor method to experiments with more than two factors is not only direct, but has at least two distinct virtues. Let us consider the case of three experimental factors, F, C, and H. If any pair of these factors, such as F and G in fig. 2, influence no stages in common, then their effects should be additive not only when averaged over levels of a third factor, H (*overall interaction* of F and G zero), but also at each level of H (all *simple interactions* of F and G zero). This is true whether factor H interacts with one or both of the other factors (fig. 2) or not (fig. 3i). The fact that simple interactions of F and G are all zero implies that the three-factor interactions of F, G, and H must also be zero, and provides a more demanding test of a theory of successive stages.

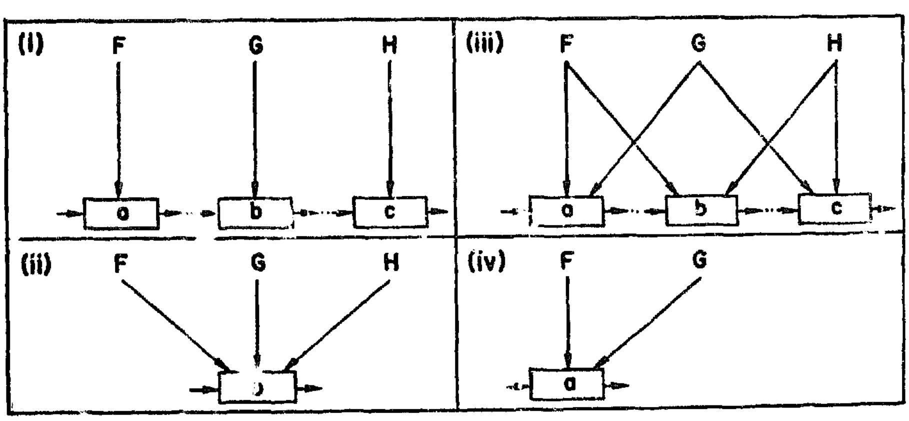

Fig. 3. Some possible arrangements of factors and stages.

In fact, of the various possible relations between three factors and a series of stages, the only ones that are expected to produce a nonzero three-factor interaction are those in which at least one of the stages is influenced by all three factors (fig. 3ii). This underlies the second virtue of multiple-factor experiments, exemplified by the contrast between the arrangements shown in figs. 3ii and 3iii. In terms of the three two-factor interactions the arrangements are equivalent: with neither do we expect any of these interactions to be zero. But for fig. 3iii, unlike 3ii, it can be shown that the three-factor interaction must be zero, even though the factors in each pair influence a stage in common. A three-factor experiment is therefore capable of discriminating between interesting alternative arrangements of stages that would be indistinguishable in a series of two-factor experiments, and is capable of leading to the discovery of successive stages even when no pair of additive factors can be found.

When all the members of a set of factors are found to influence a stage in common, a limitation of the additive-factor method is revealed: without studying additional factors there is no way to determine whether there are also other stages influenced by less than the full set. Thus, for the arrangement in fig. 2, although the data from a three-factor experiment can show that there is a stage *a* influenced by F, but not G or H, it cannot show whether there is a stage *c* influenced by H, but not by G. Similarly, for the arrangement in fig. 3iii, if one examined factors F and G only, one would be able to infer only that there was a stage *a* influenced by both F and G (fig. 3iv), and could not also discover the stage *b* influenced by F but not G. Such a stage *b* could be identified, however, in an experiment where factor H was studied at the same time as F and G (fig. 3iii).

3.5. *On applying the method and interpreting interactions*

Before we turn to experimental data, several comments about applying the method are in order.

(1) Procedures that test (or assume) the idea of stages use RT itself as the basic measure by which to assess additivity, and not any transformation of it. Additivity will in general be destroyed by nonlinear transformation of measurements. One consequence, for example, is that arithmetic means rather than harmonic or geometric means are appropriate. Furthermore, the median is inappropriate for our purpose because it is not, in general, additive. (For example, the median of a

sum of components need not be the sum of the component medians.) If stage durations are assumed to be stochastically independent, then the population minimum and maximum are suitable quantities, even though they are not, in general, linearly related to the mean (see DONDERS, 1868, Note II; TAYLOR, 1965). The difficulty here is in finding appropriate estimating statistics.

Whereas in some other domains, interactions that are removable by transformation can be regarded as arising from observations having been recorded on an inappropriate scale (COX, 1958, p. 105; KRUSKAL, 1965), here even such failures of additivity are of interest, since the basic measurement scale is specified.

(2) Experimental artifacts are more likely to obscure true additivity of factor effects than true interaction. Given the general idea of additive components, therefore, one test of an experimental procedure is the additivity of certain factor pairs that *ought* to influence different stages, such as stimulus intensity and responding limb (HOHLE, 1967, section III.D). For the same reason I would place highest credence in those two-factor interactions that are discovered in experiments in which the effects of a different pair of factors are found to add.

(3) In using the additive-factor method to test hypotheses about stages with specified functions, one cannot avoid also testing subsidiary hypotheses about the relations between the factors studied and the hypothesized stages. Suppose, for example, that we wish to test the following hypothesis, H1: stimulus encoding and response selection are accomplished by different stages, *a* and *b*. This can be tested only jointly with an additional hypothesis, H2: a particular factor, F, influences stage *a* and not *b*, and a particular factor, G, influences stage *b* and not *a*. If F and G are found to be additive, both hypotheses gain in strength. But the falsity of either H1 or H2 could produce a failure of additivity. To conclude from an observed interaction that H1 is false without assessing the validity of H2 (as in RABBITT, 1967) might therefore be an error.

(4) Certain interesting views of human information-processing are antithetical to the idea of stages whose mean durations are independent and additive. One such viewpoint has been expressed by MORAY (1967), TAYLOR et al. (1967), and POSNER and ROSSMAN (1965), for example, who propose a limited information-processing *capacity* that can be allocated to different functions in accordance with task demands. If 'capacity' is interpreted as a rate of processing, this viewpoint leads one

to expect that in a task where process a has to accomplish more, less capacity is made available for process b. If more is then also demanded of process b, its duration will increase more than if the available capacity had been greater.

In considering the implications of this discussion for patterns of factor interactions, two definitions are useful. Let an 'increase' in factor level be a change in level that increases the mean RT. And let an interaction of factors be 'positive' ('negative') if the effect of combined increases in level is greater (less) than the sum of effects of separate increases. Then even though two factors influenced different members of a pair of capacity-sharing serial processes, their effects on mean RT, rather than being additive, would interact positively. If two such processes were embedded in a series of stages, they would be identified as a single stage by the additive-factor method, and the positive interaction of the corresponding factors might indicate a capacity-sharing relation.

(5) Other forms of interaction are also of considerable interest. Suppose, for example, that two independent processes occur in parallel and that both must be completed before the next stage can begin. Then two factors that influenced them separately would interact negatively, and the processes would be identified as a single stage by the additive-factor method. A further instance of inferences from the form of an interaction – in this case, its linearity – can be found in STERNBERG, 1967.

4. APPLICATION OF THE ADDITIVE-FACTOR METHOD TO MEAN RTS IN A BINARY-CLASSIFICATION TASK

4.1. *The factors in four experiments*

I shall describe the application of these ideas first to four experiments on binary classification of numerals (three of them reported in STERNBERG, 1966 and 1967).[2]

The task in these experiments has the following paradigm: on each

[2] These experiments were designed to study the effects of factors on mean RTs. Various design features appropriate for the analysis of higher cumulants were not included, such as the use of well-practised subjects and the possibility of within-subject and within-stimulus comparisons. Where higher cumulants have been examined in these experiments (e.g. STERNBERG, 1964) some tests appeared to support the independence assumption and others did not. These analyses will be described in other reports; in the present paper only the question of additivity will be considered for binary-classification data, and not the issue of stochastic independence.

of a sequence of trials a digit is presented visually as a test stimulus. The ensemble of possible test stimuli consists of the digits from 0 to 9. The subject makes a *positive response* if the test digit is a member of a small memorized set of digits, called the *positive set*, and makes a *negative response* otherwise. By the use of payoffs that weigh accuracy heavily relative to speed, errors are held to 1 or 2 percent.

The factors to be considered here that were varied in the experiments are as follows: (1) *Stimulus quality*. The digit was presented normally ('intact') in some trial blocks, and with a superimposed checkerboard pattern ('degraded') in others. Luminance of the pattern was chosen so that degradation would increase RT without greatly increasing the error rate. (2) *Size of positive set*. In exp. I the positive set was varied from trial to trial and contained from one to six digits. In exps. II, III, and IV the positive set was fixed throughout a series of trials and contained, one, two, or four digits, each subject having a series with each set size. It was the linear increase of mean RT with this factor that was the focus of previous reports of this work. (3) *Response type (positive or negative)*. Analyses are based on correct responses only. The level of this factor is therefore determined by which response was required, that is, by whether the test stimulus was a member of the positive set. (4) *Relative frequency of response type*. The relative frequency with which positive and negative responses were required was varied between subjects by manipulation of the proportion of trials on which the test stimulus was contained in the positive set.

Factors 2 and 3 were studied in exps. I and II (STERNBERG, 1966), whose results will be used here only to provide supplementary information concerning the effect of factor 2 and the interaction of factors 2 and 3. Factors 1, 2, and 3 were examined in exp. III with twelve subjects in two sessions (STERNBERG, 1967); data from session 2 only are presented here. Factors 2, 3, and 4 were examined in exp. IV with 36 subjects in one session (unpublished). Although not designed with the additive-factor method in mind, exps. III and IV provide evidence about five of the six possible two-factor interactions among the four factors, and two of the four possible three-factor interactions.

4.2. *Results: additivity among effects of the factors on mean RT*

Evidence concerning the overall two-factor interactions in exps. III and IV is shown in the five panels of fig. 4. In each panel is shown the mean response 'profile' over subjects for levels of one factor at each

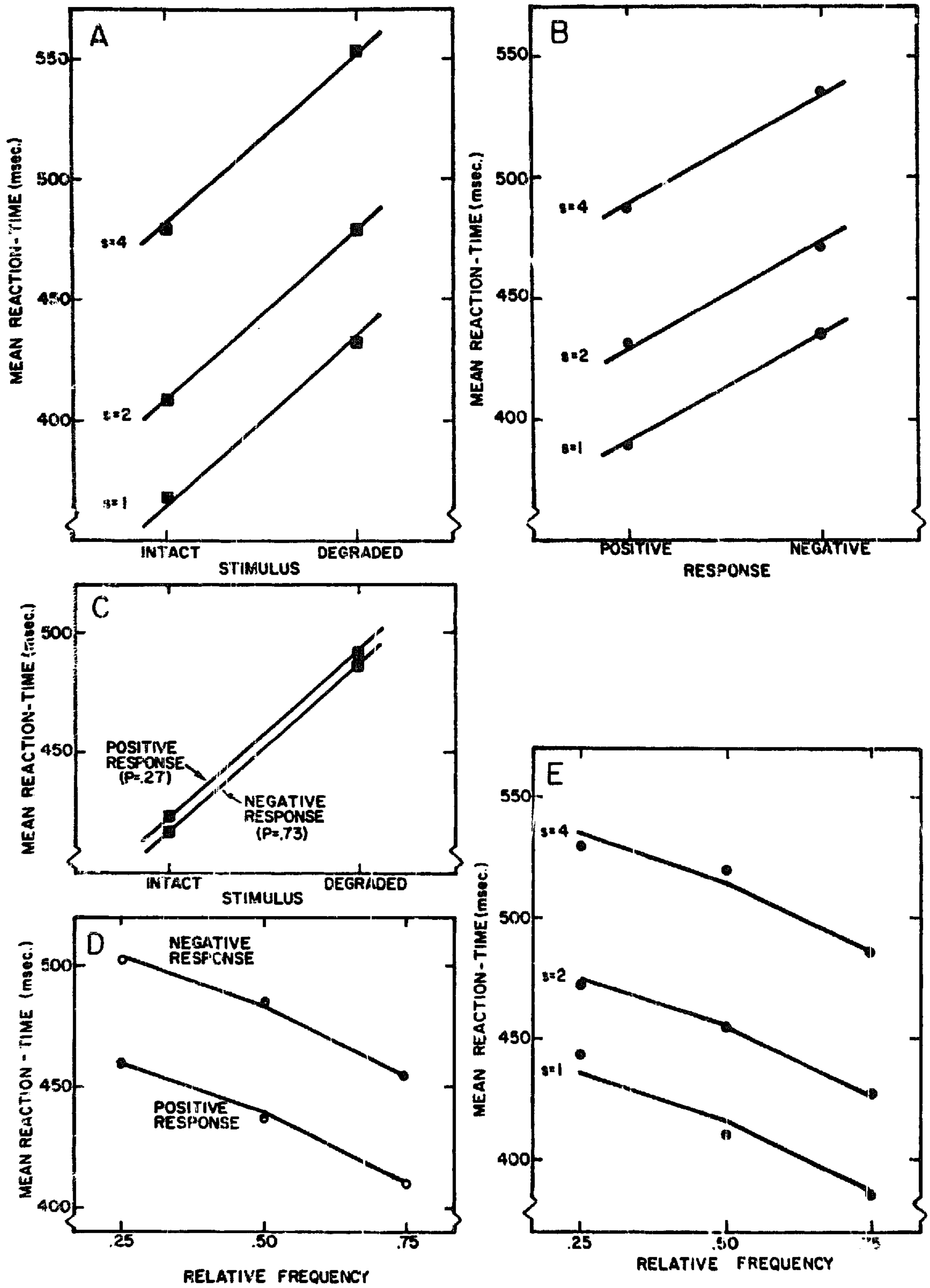

Fig. 4. Observed and best-fitting additive relations between factor effects in exp. III (squares) and exp. IV (circles). A: Effect of stimulus quality for three sizes (*s*) of positive set. RMSD (root mean squared deviation of points from parallel lines) is 1.8 msec, with 2 df. B: Effect of response type for three sizes of positive set. RMSD is 1.7 msec, with 2 df. C: Effect of stimulus quality for two response types. RMSD is 0.5 msec, with 1 df. Relative frequency (P) favored

level of another, averaged over levels of the third. For additive factors these profiles should be parallel; the lines in each panel represent the parallel profiles that best fit the data, in the sense of least squares. (Profiles are parallel if and only if equations such as (1) or (4) in section 3.3 obtain.) Goodness of fit is indicated by the square root of the mean squared deviation (RMSD) of the observed means from the best-fitting profiles, which is a quantity minimized by the fitting procedure. The RMSD should be evaluated in light of the number of degrees of freedom (df) associated with it (see section 3.3), and the size of the effect of that factor whose effect is smallest. (Where results are less clearcut, tests based on an evaluation of sampling error, as in analysis in variance, would be most appropriate.)

In all these cases the data are fitted remarkably well by an additive model. Thus, the increase in mean RT resulting from degradation was about 70 msec, regardless of the size of the positive set (fig. 4A) and regardless of the response type (fig. 4C); after averaging over the three response frequencies in exp. IV, one discovers that negative responses were about 45 msec slower than positive responses for each set size (fig. 4B);[3] as relative frequency of response was varied from 0.25 to 0.75, mean RT was shortened by about 50 msec regardless of response type (fig. 4D); and the decrease in mean RT of a response as its frequency was increased was about the same for all set sizes (fig 4E). An additive model fits worst in this last case, but with an RMSD of 3.9 msec it is satisfactory. As a very good approximation, then, these five pairs of factors are additive. (Note that without other findings the instance of additivity in exp. III shown in fig 4C would not be strong evidence for separate stages because the main effect of response type – about 6 msec – is so small. The effect of response type shown in fig. 4D was hidden in exp. III by the opposing effects of relative frequency.)

Each of these five analyses has been concerned with the overall interaction of two factors (averaging over levels of the third factor). To

[3] The additive relation between size of set (s) and response type has already been documented in the reports of exps. I, II, and III, where it was described in terms of the equality of slopes of the linear functions relating mean RT to s for the two response types.

the negative response. D: Effect of relative frequency of response for two response types. RMSD is 1.9 msec with 2 df. Filled, half-filled, and open circles each represent a different group of twelve subjects. E: Effect of relative frequency of response for three sizes of positive set. RMSD is 3.9 msec, with 4 df.

examine the simple interactions (section 3.4), and thereby assess the three-factor interaction, I fitted three-factor additive models analogous to eq. (4) to the data of exp. III (viewed as a 2 × 3 × 3 experiment) and exp. IV (viewed as a 2 × 3 × 3 experiment). The resulting RMSDs were 4.9 msec with 7 df for exp. III, and 5.8 msec with 12 df for exp. IV, both representing good agreement.

4.3. *Linearity as additivity, and its implications*

Another form of additivity revealed by the data from these experiments is shown in fig. 5. The effect of size of positive set on mean RT

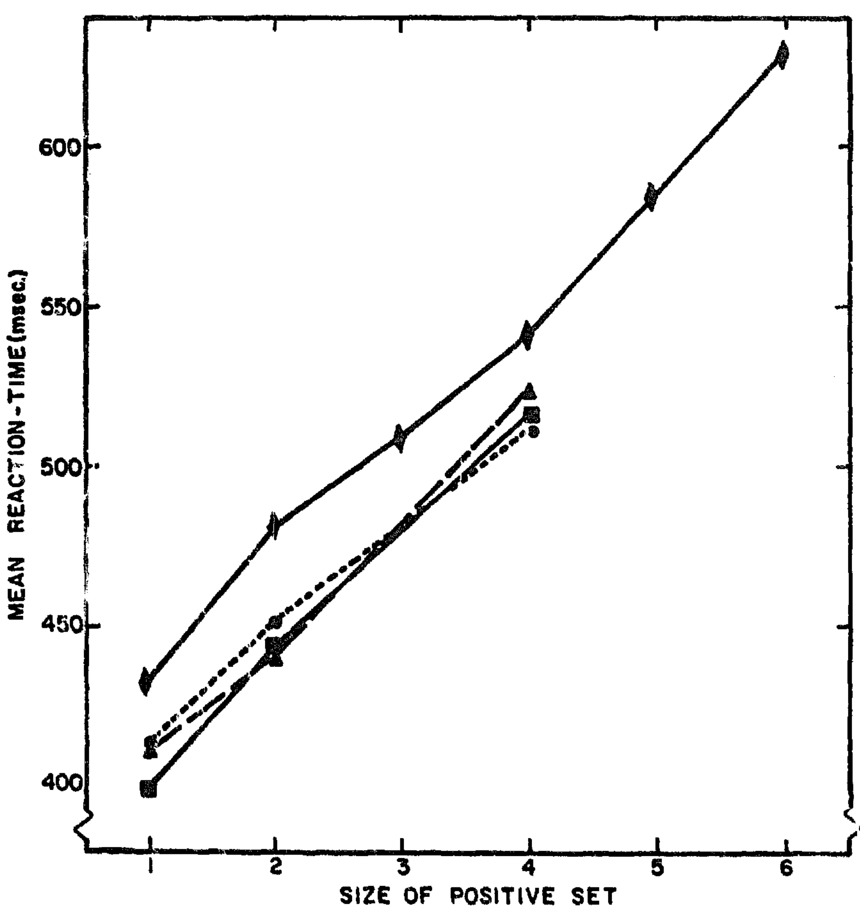

Fig. 5. Effect of size of positive set on mean RT in exp. I (diamonds), exp. II (triangles), exp. III (squares), and exp. IV (circles).

is linear. Another way of saying this is that the addition of an item to the positive set has the same effect, regardless of the size of the set. This kind of additivity suggests that each item in the positive set corresponds to a (sub)stage between stimulus and response, and that the durations of these substages represent additive components with equal means. The slope of the linear function then represents the mean duration of a substage. Various lines of evidence (STERNBERG, 1964, 1966, 1967) suggest that in each substage the test stimulus is compared to one of the items in the memorized positive set. Given this interpretation, the binary-classification experiment is one of the special cases mentioned in section 2.2, where appropriate changes in the subject's task can change the number of stages between stimulus and response. This, of course, was the original aim of Donders. Varying the number of stages is to be contrasted with merely controlling the durations of stages that are always present, as in the additive-factor method. The finding that the effect of size of positive set adds to the effects of other factors suggests that the number of stages can be varied without in this case influencing other stages; it thereby protects this particular application of Donders' idea from one source of criticism.

4.4. *Interpretation: four stages in binary classification*

In summary, then, of the six possible two-factor interactions among the four factors, all but one (stimulus quality with relative frequency of response) have been tested and found to be zero. These findings imply that at least three distinct stages are required to account for the effects of the factors studied. One stage is influenced by size of positive set and a second by response type. Whether one or two more stages are required to account for the influence of stimulus quality and relative frequency of response could be answered by the additive-factor method only in an experiment involving both of these factors. (To distinguish n stages on the basis of experiments involving n factors, using exclusively the additive relations among factor effects, *all* two-factor interactions must be shown to be zero.) The analysis is given added support by the absence of three-factor interactions in the cases examined. The linearity of the effect of size of positive set indicates that the stage associated with this factor includes a series of substages, one for each member of the set.

If we combine these inferences from the additive-factor method with supplementary arguments and plausible conjectures we are led to the more detailed picture shown in fig. 6. The additional features that have been incorporated are four stages rather than three, the functions assigned to these stages, and their order. The reasoning is as follows:

(1) The stage influenced by stimulus quality is most simply interpreted as a preprocessing or encoding stage which prepares a stimulus re-

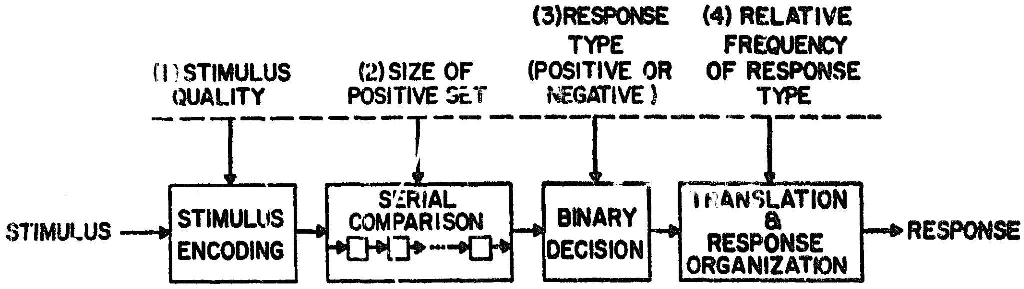

Fig. 6. Processing stages in binary classification. Above the broken line are shown the four factors examined. Below the line is shown the analysis of RT inferred from additive relations between factor pairs 1&2, 1&3, 2&3, 2&4, and 3&4, the linear effect of factor 2, and other considerations described in the text. The quality of the test stimulus influences the duration of an encoding stage in which a stimulus representation is formed. This representation is then used in a serial-comparison stage, whose duration depends linearly on size of positive set; in each of its substages the representation is compared to a memory representation of one member of the set. In the third stage a binary decision is made that depends on whether a match has occurred during the serial-comparison stage that precedes it; its mean duration is greater for negative than for positive decisions. The selection of a response based on the decision is accomplished in the final stage whose duration depends on the relative frequency with which a response of that type is required.

presentation to be used in the serial-comparison process. Otherwise it would be hard to understand how stimulus quality could influence RT without (fig. 4A) also affecting the time per comparison.[4] Any other arrangement is less plausible; this is argued in more detail in STERNBERG, 1967.

(2) The purpose of the serial-comparison stage must be to provide information for response selection. Hence any stage that depends on such information – in particular, the stages influenced by factor 3 (response type) and factor 4 (relative frequency of response type) – must follow the serial-comparison stage. (As an alternative, one might be tempted to describe the response-frequency factor in terms of the cor-

---

[4] Although no interaction was observed between stimulus quality and size of set during the second session in exp. III, an interaction was observed during the first. The form of the interaction was linear: degrading the stimulus increased the slope of the function relating mean RT to set size. This interaction is best viewed as resulting from an *indirect* influence of a factor on a stage, of the kind described in section 3.1. In this instance it is attributed to the influence of stimulus quality on the *output* of the encoding stage as well as on its duration.

related stimulus-frequencies, but the observed effects on RT would then have to be described as resulting from complicated interactions among stimulus frequency, size of set, and response type. Furthermore, relative frequency *per se* of the stimulus need have no effect on RT, as implied by results in STERNBERG, 1966.)

(3) Since factor 1 influences a stage that precedes serial comparison, and factor 4 influences a stage that follows serial comparison, these two factors must influence different stages; for this reason four stages, rather than three, are shown in the RT-analysis of fig. 6.

Included in the above argument is the rationale for all of the decisions about ordering the stages as shown in fig. 6 except the order of the stages influenced by factors 3 and 4. The functions assigned to stages, as indicated by their labels, are plausible, but should be regarded as hypotheses rather than conclusions, particularly for the last two stages. The order shown for the stages influenced by factors 3 and 4 then follows from their hypothesized functions. But confirmation of these hypotheses requires further research.

4.5. *Further comments on the method*

Given the above example, the general comments on applying the additive-factor method that were made in section 3.5 can be supplemented by four others. First, although instances of additivity of factor effects lead to the postulation of separate stages, other considerations must be used to determine the order in which these postulated stages occur. Second, any analysis produced by the method must be tentative. On the one hand, if a new factor is discovered that interacts with none of the others, then one is led to postulate an additional stage. On the other hand, new factors that interact with one or more of the others may lead to the redefinition of the functions of particular stages. Third, analyses produced by the method may be testable by new experiments, in this instance, for example, by one in which the effects of factors 1 and 4 are tested for additivity. And finally, the additive-factor method tells little about the total duration of a stage; in that sense it does not completely fulfill Donders' aims. But, as urged by some of the early investigators (e.g., JASTROW, 1890, p. 30), overall duration of a stage is more difficult to study and of less interest than whether there is such a stage, what influences it, what it accomplishes, and what its relation is to other stages.

5. Application of the additive-factor method to numeral-naming and related tasks

5.1. *Rationale and procedure for experiment V*

Let us turn now to a more traditional experiment with a one-one stimulus-response mapping, that is, one using a complete-identification task. Unlike the binary-classification experiments, exp. V was explicitly designed to permit tests of the additivity of stage durations and also of their stochastic independence. Thus, I used practised subjects, collected more data per subject, balanced the design so that linear trends in time would not destroy additivity and so that additivity could be evaluated separately for each subject, and took pains to reduce variability in both performance and measurement so that estimates of variances as well as means would be stable.

As shown in fig. 7, three factors were examined, each at two levels. The stimuli were numerals and the responses were spoken digits. The

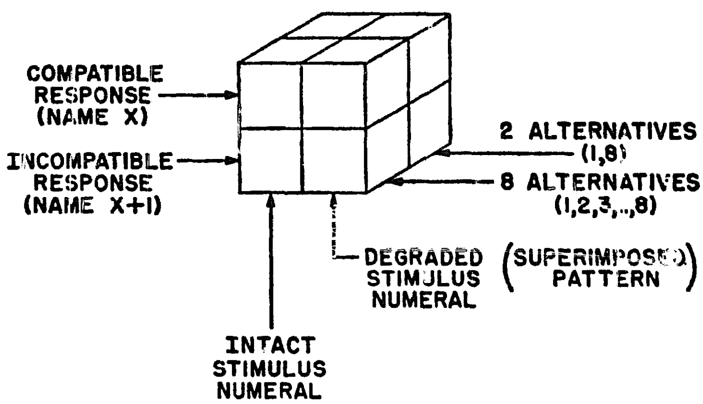

Fig. 7. Design of exp. V. Factors, each studied at two levels, were S-R compatibility, stimulus quality, and number of S-R alternatives.

number of equally-likely stimulus-response alternatives could be two or eight; *stimulus quality* was varied by presenting the numeral either intact, or degraded by a superimposed checkerboard pattern; and *S-R compatibility* was varied by making the correct response either the name of the numeral, or the name of the numeral plus one. (For example, if the numeral '1' was the stimulus, the compatible response was the spoken word 'one'; the less compatible response the spoken word 'two'.) This method of studying the relation between S-R compatibility and the effect of number of alternatives, by rearranging a compatible mapping, has the virtue of leaving stimulus and response ensembles essentially invariant; it was used earlier by Alluisi et al. (1964) and

BROADBENT and GREGORY (1965). I chose 'adding one' rather than a random assignment as the less compatible mapping so as to minimize errors and effects of practice.

In a subsequent report, I shall present full accounts of the rationale and procedure of the present experiment, with several related experiments. In brief, the choice of stimulus quality and compatibility as factors was based on the popular supposition that they must influence widely separated parts of the processes between stimulus and response in this task. Hence if these processes were organized to any extent as stages, then two different stages should be influenced by these factors, so that their effects on mean RT should be additive. If they were additive, one could go further and test the assumption of stochastic independence of stage durations by assessing the additivity of factor effects on variances and higher cumulants of the RT distribution. (Only the results concerning means and variances will be presented here.)

An incidental reason for varying number of alternatives was that the tests of additivity could be carried out more than once. But the primary reason was that a number of other experiments (e.g., BRAINARD et al., 1962; ALLUISI et al., 1964; BROADBENT and GREGORY, 1965; RABBITT, 1963) had indicated that this factor would interact with at least one of the others. I hoped to confirm this interaction in conditions under which an instance of additivity could also be found. More important, the locus of interaction might show which of the two stages was influenced by number of alternatives. The purpose of exp. V, then, was to identify a pair of stages by additivity, and simultaneously by a failure of additivity, to locate an interesting effect in one or both of them.

Conditions and payoffs were arranged to reduce errors; the average error-rate was about 2 %. There were five subjects, whose previous experience in RT experiments ranged from none to seven years. All were given practice in the task for six sessions before data were taken; the additional six sessions on which analyses are based yielded for each of four conditions 256 observations per subject for the large ensemble and 128 for the small.

5.2. *Results: additivity and interaction among effects of the factors on mean RT*

Mean RTs are shown in fig. 8, together with the pairs of best-fitting parallel lines that represent perfect additivity of the effects of stimulus quality and S-R compatibility. Since the design of exp. V permitted

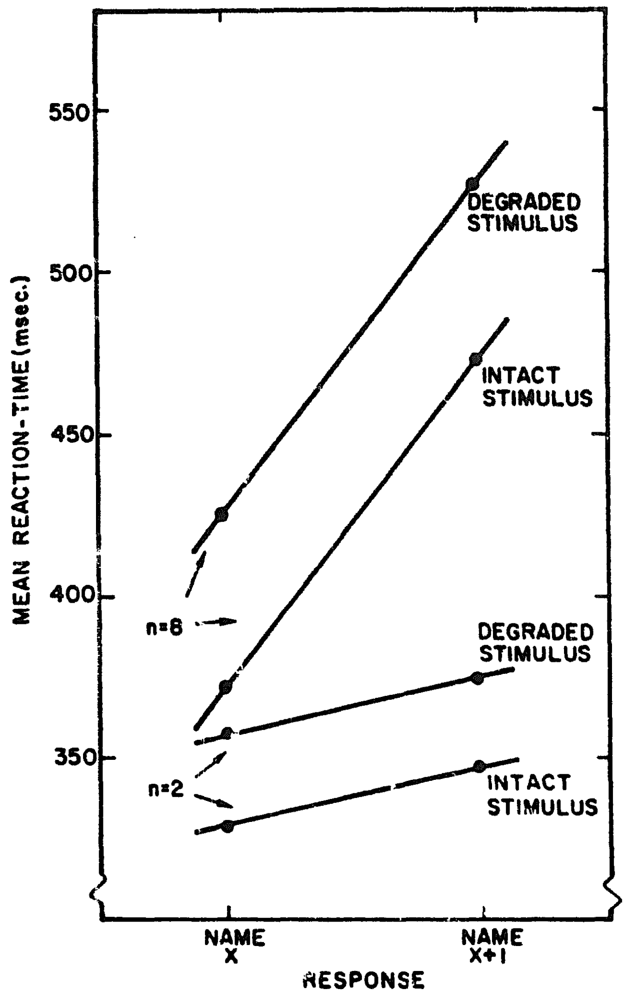

Fig. 8. Mean RTs for the eight conditions in exp. V. On the abscissa are indicated the two levels of compatibility; parameters are number of alternatives (n) and stimulus quality. For each value of n is also shown the best-fitting pair of parallel lines, which represent perfect additivity of the effects on mean RT of stimulus quality and compatibility. For $n=2$, the mean interaction-contrast is -0.9 msec with an SE of 0.9 msec Fof $n=8$ the mean interaction-contrast is $+0.4$ msec with an SE of 1.0 msec.

additivity for each number of alternatives to be assessed separately for each subject, two 1-df interaction contrasts, one for each n, were formed for each subject. This was done by taking the difference between the left and right sides of eq. (1), and normalizing (dividing by 4) so that its absolute value would equal the RMSD. The magnitude of these contrasts reflects the extent of deviation from additivity, their signs indicate whether the interaction, if any, is positive or negative (defined in section 3.5), and the standard error (SE) of their means can be evaluated by using the 4 df associated with intersubject variation. Analysis of a signed interaction contrast has the virtue of being sensitive to deviations from additivity that are systematic. Nonetheless, the analysis indicates excellent agreement with the additive model. Since these tests revealed additivity at both levels of the third factor (absence of simple interactions) there is, *a fortiori*, no three-factor interaction. In short, the effects of compatibility and stimulus quality on mean RT are perfectly additive.

But both of these factors interact with the third factor, number of alternatives. These interactions are shown more clearly in fig. 9, where

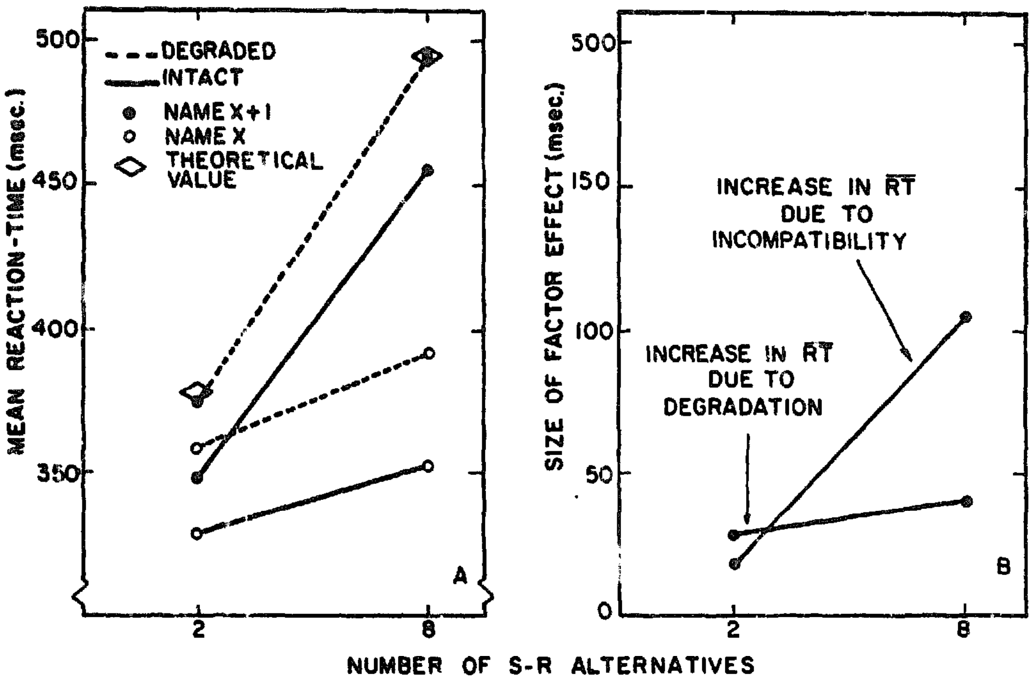

Fig. 9. Interactions between number of alternatives (n) and the other factors in exp. V. A: Mean RT to stimuli '1' and '8' as a function of n under four conditions. Theoretical values for the topmost points in each set of four are the values expected from the other three points, combined with the assumption that effects of S-R compatibility and stimulus quality are additive for each value of n. B: Effects of stimulus quality (averaged over compatibility levels) and of compatibility (averaged over quality levels) as a function of n; derived from values in panel A. Lines would be horizontal if these factors did not interact with n.

number of alternatives is indicated on the abscissa. Because we are interested in comparisons between levels of this factor without contamination from differences between stimuli or between responses, the data from the $n = 8$ conditions shown in fig. 9 are derived from the subset of trials on which the stimuli presented were the same as in the $n = 2$ conditions. The lowest pair of points in fig. 9A shows the well-known but poorly-understood fact that in the naming of highly discriminable numerals, number of alternatives has very litle effect (see SMITH, 1968); the increase in mean RT from $n = 2$ to $n = 8$ is about 20 msec. The increase is slightly greater when degraded numerals are named, substantially greater with less compatible responses, and greatest with both degraded stimuli and less compatible responses.

Fig. 9B summarizes the interactions of each of the two additive factors with number of alternatives. Number of alternatives interacts weakly with stimulus quality and strongly with S-R compatibility; expressed as a percentage of the main effect, the difference between simple effects was 35 % and 141 % respectively.[5] Such interactions are not new, of course; they have been found in other studies mentioned above, although not with the sets of stimuli and responses used here. But the use I make of these findings is, I think, new; it is shown in fig. 10.

5.3. *Interpretation: stages and factor-stage relations in the complete-identification task*

The relations found among the three factors are summarized above the horizontal line in fig. 10; the inferred analysis of RT into processing stages is shown below. Additivity of the effects of stimulus quality and S-R compatibility implies that the task is accomplished by means of at least two separate stages, designated (1) stimulus encoding (transformation of the visual stimulus into some representation of the numeral or its identity) and (2) translation and response organization. The idea of these independent subprocesses underlay Donders' work, of course, and has been discussed widely since; it is now given strong support by the additive-factor method.

Since number of alternatives interacted with both of the other factors, we must conclude that it influences the operations of both stages. (Note that one cannot justify such a conclusion simply on the grounds that RT is influenced by separate variation of both stimulus and response en-

[5] Superficially the interaction between stimulus quality and number of alternatives (n) in exp. V may appear to conflict with the absence of interaction found in exp. III between stimulus quality and size of positive set (s). But this conflict is not a real one, since variation of n in exp. V involved changing the ensemble of possible test stimuli, whereas variation of s in exp. III was accomplished without changing that ensemble. Indeed, this difference between the effects of ensemble size (n) and size of positive set (s) in relation to stimulus quality supports the view that they depend on radically different mechanisms, a view that one is also led to from the fact that, whereas mean RT usually increases logarithmically with n, it increases linearly with s.

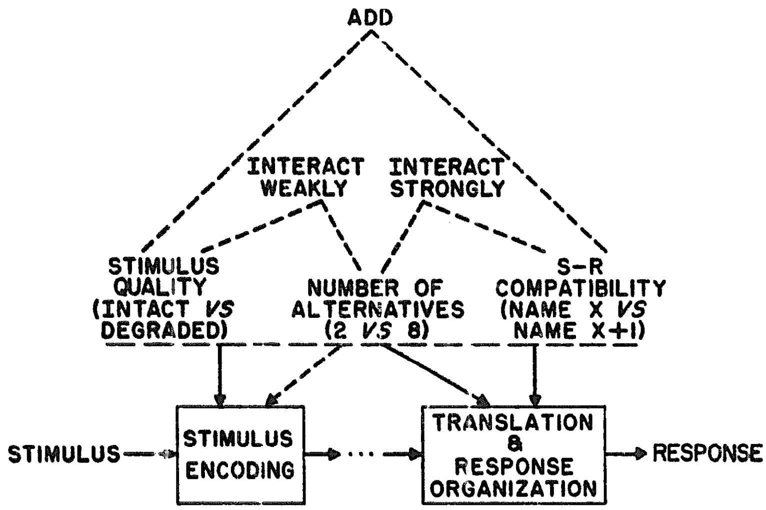

Fig. 10. Relations among factors in exp. V and the inferred stages of processing.

tropy, as in SCHLESINGER and MELKMAN, 1966; those two effects on RT could result from the influence of both factors on a single translation stage.) One might want to argue from the relative weakness of the interaction with discriminability to the relative weakness of the effect of number of alternatives on the first stage; this is why the arrow between these two is broken. These conclusions agree with those of early workers who used Donders' method; in his summary of their findings, JASTROW (1890, p. 35) concluded that number of alternatives influences the durations of both 'distinction' and 'choice', and that 'with an increase in number, the difficulty of choice increases more rapidly than the difficulty of distinction'.

As discussed in section 4, analyses of this kind are tentative. Future experiments might show, for example, that the functions assigned to the second stage in fig. 10 are accomplished by two separate stages, as in WELFORD, 1960, a 'translation' stage which was influenced by the S-R mapping, and a 'response organization' stage which was not; it would then be an open question whether number of alternatives influenced the response-organization stage as well as the other two.

The present experiment used that variety of S-R compatibility that depends on the mapping of stimuli in a fixed ensemble onto responses in a fixed ensemble. A second variety is the compatibility between entire stimulus and response ensembles, which one might call 'SE-RE compatibility'. This kind of compatibility is, of course, an interaction between factors SE and RE. In the light of our ideas about implications of interaction, a major implication of SE-RE compatibility is the existence of a nontrivial translation stage (i.e., a stage that is influenced by both SE and RE factors and that is more than merely a 'rewiring') as well as possible stimulus-encoding and response-organizing stages influenced separately by SE and RE. Indeed, the three-factor interaction found by BRAINARD et al. (1962) among SE, RE, and number of alternatives may now be interpreted as indicating the existence of a stage influenced by all three of these factors, just as the present results imply the existence of a stage influenced by both the S-R mapping and number of alternatives.

The existence of an effect of number of alternatives on both of the inferred stages in fig. 10 has at least two substantive implications. First, it suggests that the relative insensitivity of numeral-naming to this factor is a consequence of special properties of both stages. That is, both the encoding of numerals, and their translation into spoken digits, are unusual. Second, the finding conflicts with the attribution of the effect of number of alternatives to a single process, such as the single statistical decision of STONE (1960) or LAMING (1968), the single series of dichotomous choices of HICK (1952), or the single parallel process of LAMING (1966).

5.4. *On additivity of means without stochastic independence*

Let us now turn to the issue of the stochastic independence of stage durations. As I mentioned earlier, the assumption that stage-durations are additive has often been incorporated with the idea that they are independent. We have already seen that such independence has powerful consequences: for factors that influence no stages in common, it implies additivity of factor effects not only on variances, but on all the higher cumulants of the RT distribution. And a good deal of theorizing about RT-distributions is practicable only if stage durations are independent.

It is quite conceivable, however, that in some situations stage durations might be additive but not independent. As one example, consider what would happen in exp. V if a subject were 'prepared' on particular trials for a particular stimulus-response pair. Some of the implications of this idea have been explored in detail by FALMAGNE (1965). Suppose that if the stimulus presented is the one for which the subject is prepared, then both the encoding and translation stages are shorter than they would otherwise be. The result of such preparation would be a positive correlation of the durations of the two stages, rather than independence. But this effect would not alter the additivity of their mean durations.

If one thinks of 'preparation for the stimulus that is to appear' as a factor, one can place this source of correlation in the context of experimental factors and their relations to stages. Preparation can be viewed as a factor that influences both stages (like number of alternatives), but is under the subject's control rather than the experimenter's, and therefore can vary in level from trial to trial. Any such factor that was permitted to vary would induce a correlation of stage durations, whose sign would depend on whether the varying factor influenced the durations of the two stages in the same or opposite directions. So long as the variation of the subject-controlled factor (here, preparation) was not influenced by levels of the experimenter-controlled factors (here, stimulus quality and S-R compatibility) its variation would not disrupt the independence of stage-duration means or the additivity of factors that influenced the two stages separately. (If the level of a subject-controlled factor could be measured from trial to trial, even though it could not be controlled by the experimenter, and if subsets of trials on which the factor assumed the same levels were examined separately, then stage durations would presumably appear independent within those subsets.)

As a second example, in which there is a *negative* correlation between stage durations, suppose that the duration of a stage is shorter if its input is of higher quality. Furthermore, suppose that this input *is* of higher quality if the preceding stage, which produces it, has operated for a longer time. On trials on which the first stage happened to take longer, the second would be shorter, and so on. The result would be a *negative* correlation between stage durations, although, again, factor effects might remain additive.

Should the ultimate definition of 'stage', then, include a requirement of stochastic independence? In the preparation example the correlation could be viewed as a result of poor experimental control or inappropriate analysis. One might decide to retain the requirement of independence, but to conceive of this property as easily camouflaged, revealing itself only in highly refined experiments. But the second example is harder to view in this way; to retain the independence requirement one would have to identify the two processes in that example as a single stage.

The above discussion shows that the assumptions of addtivity and independence should be examined separately, because the latter may not hold, even when the former does. On the other hand, finding evidence that the durations of a set of hypothesized stages were stochastically independent would contribute to one's belief in the hypothesis, while a persistent failure to find independence might cast suspicion on either the existence of stages, or the appropriateness of the experimental methods used to study them.

5.5. *Further results: effects of the factors on RT variance*

In fig. 11 are shown the average RT-variances for each of the eight conditions in exp. V. Variances were calculated for each stimulus digit

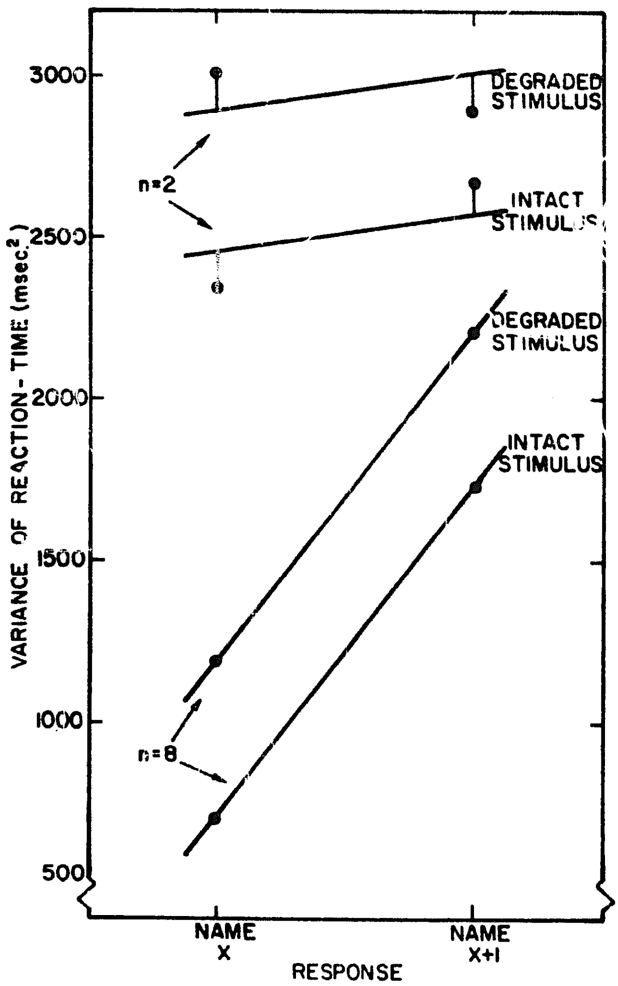

Fig. 11. Average RT-variances for the eight conditions in exp. V. On the abscissa are indicated the two levels of compatibility; parameters are number of alternatives (n) and stimulus quality. For each value of n is also shown the best-fitting pair of parallel lines, which represent perfect additivity of the effects on the RT variance of stimulus quality and compatibility. For $n=2$ the mean interaction-contrast is −111 msec² with an SE of 45 msec². For $n=8$ the mean interaction-contrast is −7 msec² with an SE of 41 msec².

in each session and were then averaged over stimuli and sessions. These averages were used to evaluate additivity for each subject separately, and for each number of alternatives, as in the case of mean RTs. Variances shown in the figure, which are averages over subjects, range from about 700 to 3000 msec².

For conditions with $n = 8$, shown is the lower part of the figure, the results agree extremely well with the additive model, lending considerable support to the assumption of independence. On the other hand, there is a definite failure of additivity for the conditions with $n = 2$ shown in the upper part of the figure. The two-choice situation seems to provide an instance where RT components are additive but not stochastically independent. In the $n = 2$ conditions subjects reported considerable variation from trial to trial in their preparedness for the stimulus that was presented. In the $n = 8$ conditions they were less aware of preparation for any particular stimulus. At present, trial-to-trial variation in the appropriateness of preparation, as in the theory of FALMAGNE (1965), appears to be the most likely source of dependence in the two-alternative conditions.

We have an instance, then, of violation of the independence assumption for RT components that we associate, on other grounds, with different stages. One implication is that methods of RT-decomposition that require the independence assumption, such as some of those described in section 2.2, may be of limited use.

An incidental finding shown in fig. 11 that is of considerable substantive interest is the marked increase in variance as the number of alternatives is reduced. This change in variance is, of course, in the opposite direction from the change in mean. It was shown by every subject and has since been replicated. For example, in the naming of intact numerals, whereas mean RT decreased by about 20 msec as number of alternatives was changed from eight to two, the variance increased by a factor of more than three. Although a change of variance in this direction may appear only under some experimental conditions, that is *can* occur is relevant to selecting among competing explanations for the well-known effect of number of alternatives on RT.

6. New interpretations of some existing instances of additivity

Clearcut instances of additive effects are relatively rare among published results, partly because factorial experiments are not popular, and perhaps also because additivity is easily destroyed by experimental artifacts and inappropriate design. But a few cases of additivity that do exist become provocative when seen from the viewpoint of the additive-factor method, and illustrate its potential to reveal not only stages of processing but also some of their properties.

6.1. *Expectancy and the stimulus-detection process*

In a simple-reaction task RAAB et al. (1961) found the effects of foreperiod and stimulus luminance to be additive. Three values of luminance were varied between sessions; mean RT decreased by 39 msec from lowest to highest. Three values of foreperiod were varied randomly from trial to trial; mean RT decreased by 14 msec from shortest to longest. An additive model fits the 3 × 3 matrix of means extremely well, giving an RMSD of 0.5 msec with 4 df.

Suppose that luminance influences the duration of a stimulus-detection stage. Then Raab's finding leads to the surprising conclusion that 'expectancy' (THOMAS, 1967) influences a different stage, rather than, for example, governing criteria in a statistical decision performed during the detection stage (MCGILL, 1963, section 2.3; FITTS, 1966), or controlling an independent 'anticipation' process that operates in parallel with the detection process (OLLMAN, 1968).

One conjecture prompted by Raab's finding is that expectancy influences response organization rather than stimulus processing. To validate such a conjecture by the additive-factor method one would need to show also that the effect of an experimental factor that was clearly associated with response organization (and not with detection) interacted with factors such as the foreperiod in a simple-reaction task, or the relative response-frequency in a choice-reaction task.

6.2. *Expectancy and the psychological refractory period*

A second example of an additive effect of time uncertainty is found in an experiment on the psychological refractory period by BERNSTEIN et al. (1968, exp. 2). Presence of a warning signal (factor 1) two seconds before the first of two stimuli S_1 and S_2 decreased the means of both RT_1 and RT_2 by about 50 msec. Varying the interval between S_1 and S_2 (factor 2) from 0 to 100 msec changed the mean of RT_1 by about 20 msec and of RT_2 by about 50 msec. The effects of these two factors were approximately additive for both responses. The implication is that the stage influenced by time uncertainty is different from the stage that displays refractoriness; such a conclusion further weakens an expectancy theory of the psychological refractory period (reviewed in SMITH, 1967).

6.3. *Sensory transmission and discrimination*

Donders and his followers believed that there exist separate stages for the transmission of sensory information from the periphery, and for its discrimination. A choice-reaction experiment of HOWELL and DONALDSON (1962) provides an instance of additivity that supports this belief. Under three stimulus-modality conditions (factor 1), one visual, one tactual, and one auditory, each of three responses was correct for one of three stimulus intensities (factor 2). To make the design orthogonal, levels of factor 2 were adjusted by cross-modality matching to be the 'same' in the three modalities. Mean RT decreased by 106 msec from visual (slowest) to auditory (fastest) modality, and by 80 msec from low to high intensity. An additive model fits the 3×3 matrix of means reasonably well, giving an RMSD of 4.5 msec with 4 df. Findings such as these would support the old idea of a transmission stage influenced by modality but not intensity, and a discrimination stage influenced by intensity but not modality.

6.4. *The selective influence of practice*

From the viewpoint of the additive-factor method, experiments occasionally suggest that practice influences some but not all of the stages in a task. In exp. III (section 4.1), for example, subjects had the same conditions in each of two sessions. For intact stimuli, mean RT declined by about 36 msec from session to session. This effect was almost perfectly additive with the effect of size of positive set, s, which increased the mean RT by 110 msec as it was varied from $s = 1$ to $s = 4$. The RMSD in this instance is 1.0 msec with 2 df. The finding of additivity suggests that practice had its effect on stages other than serial comparison.

7. USE OF INDIVIDUAL DIFFERENCES IN INFERENCES ABOUT STAGES

7.1. *'Subjects' as an additive factor*

Individual differences are often thought of as little more than an ubiquitous nuisance. Yet they seem to have the potential of providing at least supplementary information about processing stages. The simplest instance would be one in which an effect of 'subjects' on mean RT combined additively with the effect of some experimental factor, factor

E. That is, the effect of E would be invariant over subjects, despite the existence of individual differences in mean RT. Such additivity (the absence of a treatment-by-subject interaction) could be taken to imply the existence of at least two stages, one influenced by 'subjects' and not factor E, and the other influenced by E and not 'subjects'.

7.2. *'Subjects' as an interacting factor*

Interactions between 'subjects' and experimental factors can also be useful. Suppose that we have found a pair of factors, F and G, whose effects on mean RT, averaged over a set of subjects, are additive. We wish to infer that they influence no stages in common. An additional test of this inference is available if there is reliable variation from subject to subject in the sizes of the effects, that is, if 'subjects' interacts with each of the other factors. 'Subjects' would then be thought of as influencing a stage in common with F, and a stage in common with G. But if F and G influence no stages in common with each other, then the three-factor interaction of F, G, and 'subjects' should be zero (section 3.4). In other words, the effects of F and G should be additive for each subject separately and not merely for the group means, even though sizes of the effects vary over subjects. It is for this reason that the two-factor interaction contrasts of sections 5.2 and 5.5 were evaluated separately for each subject and used to obtain the SE of the mean interaction-contrast; if the mean contrast is small, such an SE would also be small only insofar as all subjects show additivity.

7.3. *Stage sensitivity and correlations of factor effects*

Finally, let us consider individual differences in the effects of factors that interact and are therefore thought to influence the same stage. Individual differences in the size of a factor's effect may be described as differences among subjects' 'sensitivity' to that factor. It is plausible that for factors that influence the same stage, sensitivities will be more highly correlated (over subjects) than for factors that influence no stages in common. (A stage can be thought of as associated with an ability or capacity, an increase in the level of any factor that influences that stage as a test of that ability, and the resulting increase in RT as the score on the test.) This idea corresponds to two properties that may be useful additions to the conception (section 3.2) of stage: (1) A stage itself is more or less sensitive, varying from subject to subject in its sensitivity to *all* the factors that influence it,[6] and (2) the sensitivities of different stages are less than perfectly correlated with one another.

If stages had these two properties, the pattern of correlations among factor effects would then supplement the pattern of interactions in providing information about stages and factor-stage relations; moreover, tentative evidence about stages could even be derived from the correlations in a set of one-factor experiments performed with the same group of subjects.

Consider as an example the analysis shown in fig. 10 of performance in exp. V. If sensitivities of the encoding and translation stages varied somewhat independently over subjects, then of the three pairs of factors, that pair whose effects should be most highly correlated are compatibility and number of alternatives (n), and the pair with the lowest correlation should be compatibility and stimulus quality. This was observed, confirming the analysis: product-moment coefficients over the five subjects for pairs of main effects are 0.88 ($p = 0.05$) for compatibility and n, 0.16 for compatibility and stimulus quality, and the intermediate value 0.42 for effects of quality and n. If this kind of finding appeared in larger experiments it would tend to confirm the usefulness of the idea of a stage having a sensitivity that varies over subjects and is uncorrelated from stage to stage.

8. Concluding remarks

The additive-factor method cannot distinguish *processes*, but only *processing stages*. This distinction bears on the interpretation, for example, of the interaction of time uncertainty and relative signal-frequency found by BERTELSON and BARZEELE (1965) and correctly felt by them to be important in the understanding of preparation. The interaction does allow one to reject the idea of *separate stages* (i.e., no stage influenced by both factors). But it does not allow one to reject the more general proposition of *separate processes;* a pair of independent processes influenced separately by the factors could conceivably operate

[6] Let α and β be parameters each influenced by the level of a different factor, and Θ_i and λ_i be individual-difference parameters that are uncorrelated over subjects. A model for the mean duration of a stage with interacting factor effects and also with the sensitivity property is $\bar{T} = \Theta_i \alpha \beta$; here the effects of both factors depend on the subject's value of Θ_i. Factor interactions could occur without the sensitivity property, however, as in a stage whose duration is $\bar{T} = (\alpha + \Theta_i)(\beta + \lambda_i)$.

in parallel, for example, and thereby produce the interaction (section 3.5).

A second proviso about the additive-factor method is related to techniques of data analysis. The usual significance tests performed in conjunction with analysis of variance are asymmetric: one is forced to assume that effects are additive (null hypothesis) unless the contrary can be proved. Given the strong implications of additivity, this assymmetry seems particularly inappropriate. To avoid this difficulty one might present findings in terms of mean interaction-contrasts and their SEs, or choose alternative hypotheses that specify interaction contrasts of theoretically-interesting magnitudes, and adjust tests so that errors of types 1 and 2 have equal probabilities with respect to such alternatives.

The idea of a processing stage that I have presented should be thought of as tentative and subject to refinement by future research. Some of the properties one might want to consider incorporating in a definition of 'stage' to make it most useful were discussed in section 3.2. As the additive-factor method is applied in various experimental situations we may want to impose an additional requirement of *stage invariance:* only those stages that emerge from the study of several different situations are of interest. (Thus, it would be desirable if the 'stimulus-encoding' stages inferred in sections 4 and 5 could be shown to have similar properties.)

Most previous attempts to use RT measurements for studying stages of processing between stimulus and response fall into two classes. Donders' subtraction method required task changes that inserted or deleted entire stages. Its range of application was limited by the difficulty of finding operations that did this but also left other stages invariant, and by the absence of tests other than introspection for determining whether the requirement of invariance was met. The second approach involves the application of precise stochastic models – 'strong' models that embody several assumptions simultaneously. This approach is limited because it does not permit the assumption of main interest – that of the existence of additive components – to be examined in isolation from assumptions about the stochastic independence of components and the forms of their distributions.

The additive-factor method is proposed as a third approach to the study of processing stages, one which avoids the limitations of both Donders' method and the strong-model approach. With it one can test the interesting assumptions in isolation, yet do so by means of operations that do not insert or delete hypothesized stages, but merely change their durations. This method had limitations too, however. Although it leads to the decomposition of a set of stages, it cannot decompose the RT itself: the absolute durations of the stages discovered are not determined. And like the other approaches, the method does not give the order of a set of stages it distinguishes. It can, however, establish the existence of processing stages and, by exposing their relations to experimental factors, help in ascertaining their properties. Its power stems from the fundamental significance of additivity, which in turn depends on the existence in RT experiments of a basic measure – that of physical time.

The concept of interaction in multifactor experiments and the associated focus on additive models was originated in the 1920's by R. A. Fischer (SCHEFFÉ, 1959, p. 90). It is perhaps not surprising that Donders and others of his time, working without this theoretical appartus, were not stimulated to perform factorial experiments or examine interactions, and that only recently (e.g., BERTELSON and BARZEELE, 1965; BROADBENT and GREGORY, 1965) has the important role of factor interactions in the study of RT been hinted at.

REFERENCES

ALLUISI, E. A., G. S. STRAIN and J. B. THURMOND, 1964. Stimulus-response compatibility and the rate of gain of information. Psychon. Sci. **1**, 111—112.

AUDLEY, R. J., 1960. A stochastic model for individual choice behavior Psychol. Rev. **67**, 1—15.

BERNSTEIN, I. H., R. R. BLAKE and M. H. HUGHES, 1968. Effects of time and event uncertainty upon sequential information processing. Percept. & Psychophys. **3**, 177—184.

BERTELSON, P. and J. BARZEELE, 1965. Interaction of time-uncertainty and relative signal frequency in determining choice reaction time. J. exp. Psychol. **70**, 448—451.

BRAINARD, R. W., T. S. IRBY, P. M. FITTS and E. A. ALLUISI, 1962. Some variables influencing the rate of gain of information. J. exp. Psychol. **63**, 105—110.

BROADBENT, D. R. and M. GREGORY, 1965. On the interaction of S-R compatibility with other variables affecting reaction time. Brit. J. Psychol. **56**, 61—67.

BUSH, R. R., E. GALANTER and R. D. LUCE, 1963. Characterization and classification of choice experiments. In: R. D. Luce, R. R. Bush and E. Galanter (eds.), Handbook of mathematical psychology, vol. 1. New York; Wiley, 77—102.

CHRISTIE, L. S. and R. D. LUCE, 1956. Decision structure and time relations in simple choice behavior. Bull. math. Biophys. **18**, 89—112.

COX, D. R., 1958. Planning of experiments. New York: Wiley.

DONDERS, F. C., 1868. Over de snelheid van psychische processen. Onderzoekingen gedaan in het Physiologisch Laboratorium der Utrechtsche Hoogeschool, 1868—1869, Tweede reeks, II, 92—120. Transl. by W. G. Koster, this volume, 412—431.

FALMAGNE, J. C., 1965. Stochastic models for choice reaction time with applications to experimental results. J. math. Psychol. **2**, 77—124.

FITTS, P. M., 1966. Cognitive aspects of information processing: III. Set for speed versus accuracy. J. exp. Psychol. **71**, 849—857.

GHOLSON, B. and R. H. HOHLE, 1968a. Verbal reaction times to hues vs. hue names and forms vs. form names. Percept. & Psychophys. **3**, 191—196.

——— and ———, 1968b. Choice reaction times to hues printed in conflicting hue names and nonsense words. J. exp. Psychol. **76**, 413—418.

GOUGH, P. B., 1965. Grammatical transformations and speed of understanding. J. verb. Learn. verb. Behav. **4**, 107—111.

———, 1966. The verification of sentences: the effects of delay of evidence and sentence length. J. verb. Learn. verb. Behav. **5**, 492—496.

HICK, W. E., 1952. On the rate of gain of information. Quart. J. exp. Psychol. **4**, 11—26.

HOHLE, R. H., 1967. Component process latencies in reaction times of children and adults. In: L. P. Lipsett and C. C. Spiker (eds.), Advances in child development and behavior, vol. 3. New York: Academic Press, 225—261.

HOWELL, W. C. and J. E. DONALDSON, 1962. Human choice reaction time within and among sense modalities. Science **135**, 429—430.

JASTROW, J., 1890. The time-relations of mental phenomena. New York: N. D. C. Hodges.

KENDALL, M. G. and A. STUART, 1958. The advanced theory of statistics, vol. 1. London: Griffin.

KRUSKAL, J. B., 1965. Analysis of factorial experiments by estimating monotone transformations of the data. J. roy. statist. Soc. B. **27**, 251—263.

KÜLPE, O., 1895. Outlines of psychology. New York: MacMillan, Secs. 69, 70.

LAMING, D. R. J., 1966. A new interpretation of the relation between choice-reaction time and the number of equiprobable alternatives. Brit. J. math. stat. Psychol. **19**, 139—149.

———, 1968. Information theory of choice-reaction times. London: Academic Press.

LUCE, R. D. and D. M. GREEN, 1969. Detection of auditory signals presented at random times, II. Percept. & Psychophys. In press.

MCGILL, W. J., 1963. Stochastic latency mechanisms. In: R. D. Luce, R. R. Bush and E. Galanter (eds.), Handbook of mathematical psychology, vol. 1. New York: Wiley, 1—120.

——— and J. GIBBON, 1965. The general-gamma distribution and reaction times. J. math. Psychol. **2**, 1—18.

McMahon, L. E., 1963. Grammatical analysis as part of understanding a sentence. Unpublished doctoral dissertation, Harvard University.

Moray, N., 1967. Where is capacity limited? A survey and a model. In: Attention and performance, A. F. Sanders e(d.), Acta Psychol. 27, 84—92.

Neisser, U., 1963. Decision-time without reaction-time: experiments in visual scanning. Amer. J. Psychol. **76**, 376—385.

Ollman, R. T., 1968. Anticipation in simple-reaction time. Manuscript in preparation.

Posner, M. I. and R. F. Mitchell, 1967. Chronometric analysis of classification. Psychol. Rev. **74**, 392—409.

——— and E. Rossman, 1965. Effect of size and location of informational transforms upon short-term retention. J. exp. Psychol. **70**, 496—505.

Raab, D., E. Fehrer and M. Hershenson, 1961. Visual reaction time and the Broca-Sulzer phenomenon. J. exp. Psychol. **61**, 193—199.

Rabbitt, P. M. A., 1963. Information-load and discriminability. Nature, **197**, 726.

———, 1967. Signal discriminability, S-R compatibility and choice reaction time. Psychon. Sci. **7**, 419—420.

Restle, F. and J. H. Davis, 1962. Success and speed of problem solving by individuals and groups. Psychol. Rev. **69**, 520—536.

Scheffé, H., 1959. The analysis of variance. New York: Wiley.

Schlesinger, I. M. and R. Melkman, 1966. The effect on choice-reaction time of stimulus information varied independently of transmitted information. J. genet. Psychol. **74**, 165—172.

Smith, E. E., 1968. Choice reaction time: an analysis of the major theoretical positions. Psychol. Bull. **69**, 77—110.

Smith, M. C., 1967. Theories of the psychological refractory period. Psychol. Bull. **67**, 202—213.

Snodgrass, J. G., R. D. Luce and E. Galanter, 1967. Some experiments on simple and choice reaction time. J. exp. Psychol. **75**, 1—17.

Sternberg, S., 1964. Estimating the distribution of additive reaction-time components. Paper presented at the meeting of the Psychometric Society, Niagara Falls, Ont., October, 1964.

———, 1966. High-speed scanning in human memory. Science, **153**, 652—654.

———, 1967. Two operations in character recognition: some evidence from reaction-time measurements. Percept. & Psychophys. **2**, 45—53.

Stone, M., 1960. Models for choice-reaction time. Psychometrika, **25**, 251—260.

Taylor, D. H., 1965. Two dimensions for reaction time distributions. Nature, **206**, 219—220.

———, 1966. Latency components in two-choice responding. J. exp. Psychol. **72**, 481—487.

Taylor, M. M., P. H. Lindsay and S. M. Forbes, 1967. Quantification of shared capacity processing in auditory and visual discrimination. In: Attention and performance, A. F. Sanders (ed.), Acta Psychol. **27**, 223—229.

THOMAS, E. A. C., 1967. Reaction-time studies: the anticipation and interaction of responses. Brit. J. math. stat. Psychol. 20, 1—29.

WELFORD, A. T., 1960. The measurement of sensory-motor performance: survey and reappraisal of twelve years' progress. Ergonomics, 3, 189—230.

DISCUSSION

Peacock: You have considered a set of factors and a set of stages; have you also mapped a set of theories for example, for expectancy and single channel, on to the former sets.

Sternberg: I would like to believe that this kind of analysis might be relevant to understanding mechanisms. I have tried to indicate how it might apply in explaining what causes choice RT to increase with the number of S-R alternatives. The fact that there appear to be two separate stages that are influenced by the number of S-R alternatives should have a great influence on the kind of theories one develops.

Nickerson: I am very interested in the reversal of the variances with the number of alternatives. Have you any speculation on it?

Sternberg: Subjects reported qualitative changes in their experience as number of alternatives was changed. In the two-alternative case, subjects reported experiencing preparation, i.e. on some trials the stimulus that appeared was the one they expected, whereas on other trials it was not. They claimed that whereas appropriate preparation speeded their responses, they were slowed down by having been prepared for the wrong stimulus. In the 8-alternative case, on the other hand, they claimed that they did not prepare. They just sat and waited for the stimulus. That could reduce the variance. My feeling is that the variance finding supports very strongly Falmagne's approach to the explanation of the effect of number of alternatives. His kind of theory, in which there are states of preparation and of non-preparation, could lead to either an increase or a decrease in variance with number of alternatives. Most other theories could not. For example, serial dichotomous choices could not easily be made to produce an increase in variance when the number of alternatives is reduced.

Nickerson: Have you found any other situation where this is true?

Sternberg: We are doing a literature search, but in fact variances are not reported in most papers. If any of you has some information on this, that would be extremely useful.

Mowbray: If you mean in a situation where it is not true, I have some. We published in about 1960 a study in which the variances were almost identical for a two-choice and a four-choice alternative in the same situation.

Hyman: I am wondering if independence and additivity could change as a function of practice. I have a notion that they would.

Sternberg: We started with already-trained subjects.

Schouten: Sanders just remarked to me how much Donders would have loved to hear this rejection of the growing scepticism by the end of the last century.

In our own work we grew a little sceptical for very circumstantial reasons. We were doing all sorts of time-study movements relating to the time-study systems in industry. These systems work in terms of allotting a time for separate elementary movements and then hoping that the sum of all these separate times will be equal to the total time. This is a case where it definitely does not hold because the whole thing becomes one gradual movement. I would like to say that your lecture and your ideas seem to me extremely inspiring for further research, just to make out where there is this additivity and where this interaction.

Kornblum: How do you expect the probabilities to work on these stages?

Sternberg: In fact your data provided an interesting case where you reported three factors, two of which interact and the third of which is additive with the others.

Falmagne: The more I hear on the problem of additivity the more I have the impression that we are getting very close to some problem that people have in measurement theory. I should like to have your comments on this.

Sternberg: There is a point here that is related to Mick's paper. It has to do with question whether you should permit yourself to transform the measurement scale from the physical scale in RT experiments. We know that in some cases one can eliminate interactions by transformation of variables. Should that be permitted in reaction time? My contention is that if you are concerned with the idea of serial processes or component stages, then you should not permit it, because what is important is whether you have additivity or not on the scale of physical time and not on any transformed scale.