



January 2001

# Separate modifiability, mental modules, and the use of pure and composite measures to reveal them

Saul Sternberg

*University of Pennsylvania*, [saul@psych.upenn.edu](mailto:saul@psych.upenn.edu)

Follow this and additional works at: [http://repository.upenn.edu/ircs\\_reports](http://repository.upenn.edu/ircs_reports)

---

Sternberg, Saul, "Separate modifiability, mental modules, and the use of pure and composite measures to reveal them" (2001). *IRCS Technical Reports Series*. 25.

[http://repository.upenn.edu/ircs\\_reports/25](http://repository.upenn.edu/ircs_reports/25)

University of Pennsylvania Institute for Research in Cognitive Science Technical Report No. IRCS-01-08.

This paper is posted at ScholarlyCommons. [http://repository.upenn.edu/ircs\\_reports/25](http://repository.upenn.edu/ircs_reports/25)  
For more information, please contact [libraryrepository@pobox.upenn.edu](mailto:libraryrepository@pobox.upenn.edu).

---

# Separate modifiability, mental modules, and the use of pure and composite measures to reveal them

## Abstract

How can we divide a complex mental process into meaningful parts? In this paper, I explore an approach in which processes are divided into parts that are *modular* in the sense of being *separately modifiable*. Evidence for separate modifiability is provided by an instance of *selective influence*: two factors  $F$  and  $G$  (usually experimental manipulations) such that part **A** is influenced by  $F$  but invariant with respect to  $G$ , while part **B** is influenced by  $G$  but invariant with respect to  $F$ . Such evidence also indicates that the modules are functionally distinct. If we have *pure measures*  $M_A$  and  $M_B$ , each of which reflects only one of the parts, we need to show that  $M_A$  is influenced by  $F$  but not  $G$ , while  $M_B$  is influenced by  $G$  but not  $F$ . If we have only a *composite measure*  $M_{AB}$  of the entire process, we usually also need to confirm a *combination rule* for how the parts contribute to  $M_{AB}$ .

I present a taxonomy of separate-modifiability methods, discuss their inferential logic, and describe several examples in each category. The three categories involve measures that are *derived pure* (based on different transformations of the same data; example: separation of sensory and decision processes by signal detection theory), *direct pure* (based on different data; example: selective effects of adaptation on spatial-frequency thresholds), and *composite* (examples: the multiplicative-factor method for the analysis of response rate; the additive-factor method for the analysis of reaction time). Six of the examples concern behavioral measures and functional processes, while four concern brain measures and neural processes. They have been chosen for their interest and importance; their diversity of measures, species, and combination rules; their illustration of different ways of thinking about data; the questions they suggest about possibilities and limitations of the separate-modifiability approach; and the case they make for the fruitfulness of searching for mental modules.

## Comments

University of Pennsylvania Institute for Research in Cognitive Science Technical Report No. IRCS-01-08.

# Separate modifiability, mental modules, and the use of pure and composite measures to reveal them

Saul Sternberg  
University of Pennsylvania

Reprinted from *Acta Psychologica*, 106, 147-246, 2001,  
with permission from Elsevier Science.

This pdf file contains:

Table of Contents: Text (page 143)  
Table of Contents: Appendices (page 144)  
List of Examples (page 145)  
Some Features of the Examples (page 145)  
Lists of Tables, Figures, and Equations (page 146)  
The Article (pages 147-246).

Note that page numbers for the article in the pdf file (147 - 246)  
match the page numbers for the article as printed.

**Contents (Part 1: TEXT)**

<b>148</b>	<b>1. Introduction</b>
	1.1 Antecedents
	1.2 Notation
	1.3 Twofold partition as an initial step in understanding a complex process
	1.4 Processes, measures, factors, and selective influence
	1.5 Statistical issues
	1.6 Process decomposition versus task comparison
	1.7 Relevance of brain measurement to process decomposition
	1.8 Organization of the paper
<b>155</b>	<b>2. Pure measures: Definitions and inferential logic</b>
	2.1 Selective influence of factors on processes and their measures
	2.2 Hypothesis, prediction, and inference for pure measures
	2.3 Why all four properties are important
<b>158</b>	<b>3. Composite measures: Definitions and inferential logic</b>
	3.1 Composite measures based on summation
	3.1.1. Illustration: The Additive-Factor Method.
	3.2 Composite measures based on multiplication
	3.3 Factorial experiments
<b>163</b>	<b>4. Introduction to three examples of inference based on derived pure measures</b>
<b>164</b>	<b>5. Isolation of a timing module in the rat (Example 1)</b>
	5.1 Two kinds of elaboration of the two-module analysis
	5.2 Comments
<b>167</b>	<b>6. Parallel neural modules revealed by the lateralized-readiness potential (Example 2)</b>
	6.1 Comments
<b>172</b>	<b>7. Separation of sensory and decision processes by signal detection theory (Example 3)</b>
	7.1 Why has SDT failed in this respect?
	7.2 An approximation to full modularity of sensory and decision processes when reinforcer ratio is controlled
	7.3 Comments
<b>178</b>	<b>8. Introduction to three examples of inference based on direct pure measures</b>
<b>179</b>	<b>9. Evidence for modular spatial-frequency analyzers from selective adaptation (Example 4)</b>
	9.1 Comments
<b>182</b>	<b>10. Evidence from individual neurons for temporally-delimited (serial) neural modules (Example 5)</b>
	10.1 Estimation of stage durations
	10.2 Relation of these findings to the demonstration of separate modifiability
	10.3 Comments
<b>186</b>	<b>11. Evidence from fMRI for modular neural processes implemented by anatomically delimited processors (Example 6)</b>
	11.1 Comments
<b>190</b>	<b>12. Introduction to four examples of inference based on composite measures</b>
<b>191</b>	<b>13. Evidence for modular spatial-frequency analyzers from 'probability summation' at threshold (Example 7)</b>
	13.1 Justification of a multiplicative combination rule for non-detection probability
	13.2 Three tests of the joint hypothesis
	13.3 Comments
<b>196</b>	<b>14. Evidence from amplitude of the event-related potential for modular neural processes (Example 8)</b>
	14.1 The additive-amplitude method
	14.2 Application of the additive-amplitude method to word classification
	14.3 Comments
<b>200</b>	<b>15. Multiplicative combination rule for response rate (Example 9)</b>
	15.1 A plausible systematic deviation from additivity measured by multiplicative interaction of scaled factor levels
	15.2 Comments
<b>206</b>	<b>16. Reaction time as a composite measure: Selective effects of sleep deprivation (Example 10)</b>
	16.1 Comments
<b>210</b>	<b>17. General Discussion</b>

Separate Modifiability, Mental Modules, and the Use of Pure and Composite Measures to Reveal Them  
**Contents (Part 2: APPENDICES, GLOSSARY, ACKNOWLEDGEMENTS, REFERENCES)**

<b>213</b>	<b>Appendix A1. Process decomposition versus task comparison</b>
	A1.1 Introduction
	A1.2 Qualitative task comparison
	A1.2.1 Behavioral studies of memory.
	A1.2.2 Effects of sleep deprivation
	A1.2.3 Task-specific effects of localized brain lesions.
	A1.2.4 Task-specific effects on localized brain activation.
	A1.3 Quantitative task comparison: Subtraction and division methods
	A1.3.1 Donders' subtraction method.
	A1.3.2 Jacoby's division method.
	A1.3.3 Derived pure measures from subtraction and division methods.
<b>221</b>	<b>Appendix A2. Considerations in the choice of factors</b>
	A2.1 Unitary factors and qualitative process invariance
	A2.2 Manipulated versus selective factors
	A2.3 Inferential logic when hypotheses about modules specify the factors
<b>223</b>	<b>Appendix A3. Composite measures, combination rules, and stochastic independence</b>
	A3.1 Measures, combination rules, and plausibility
	A3.2 Multiplication as combination rule: Implications for data analysis of the zero-correlation requirement
	A3.3 Stochastic independence of process contributions as further evidence of modularity
<b>226</b>	<b>Appendix A6. Evidence for neural and functional modules that correspond</b>
<b>228</b>	<b>Appendix A7. Details of the analysis of brightness discrimination by pigeons</b>
	A7.1 Support for the equal-variance Gaussian detection model
	A7.2 Evaluation of an alternative measure of the decision process
<b>229</b>	<b>Appendix A9. Factorial experiments and multiple-level factors with pure measures</b>
	A9.1 Advantages of a factorial design.
	A9.2 Advantages of multiple factor levels.
<b>231</b>	<b>Appendix A10. Inferring neural processing stages from single-unit recordings</b>
	A10.1 Classification of neurons
	A10.2 Combining activation times over neurons
	A10.3 Effects of stimulus and response factors on stage durations
<b>233</b>	<b>Appendix A11. fMRI and modular processes: Requirements and statistical issues</b>
	A11.1 Requirements for a process-decomposition study
	A11.2 Statistical issues
<b>235</b>	<b>Appendix A13. Fitting and testing one-channel and two-channel models of detection</b>
<b>236</b>	<b>Appendix A15. Numerical scaling of factor levels for multiplicative and hybrid combination rules</b>
<b>237</b>	<b>Appendix A16. Processing stages as modules</b>
	A16.1 Multiplicative combination rule for the proportion of response omissions
	A16.2 What is a 'stage'?
	A16.3 Additive effects of factors on mean reaction time: Alternative interpretations
<b>240</b>	<b>ACKNOWLEDGEMENTS</b>
<b>241</b>	<b>GLOSSARY</b>
<b>242</b>	<b>REFERENCES</b>

## Examples

(Examples 11 and 12 are not numbered in the text.)

Ex.#	Example	Pages	
		TEXT	APPX
1	Isolation of a timing module in the rat	164	—
2	Parallel neural modules revealed by the lateralized-readiness potential	167	—
3	Separation of sensory and decision processes by signal detection theory	172	228
4	Evidence for modular spatial-frequency analyzers from selective adaptation	179	229
5	Evidence from individual neurons for temporally delimited (serial) modules	182	231
6	Evidence from fMRI for modular neural processes implemented by anatomically delimited processors	186	233
7	Evidence for modular spatial-frequency analyzers from 'probability summation' at threshold	191	235
8	Evidence from amplitude of the event-related potential for modular neural processes	196	—
9	Multiplicative combination rule for response rate	200	236
10	Reaction time as a composite measure: Selective effects of sleep deprivation	206	237
(11)	Evidence for neural and functional modules that correspond	—	226
(12)	Frequency of response omissions as a composite measure	—	237

## Some Features of the Examples

Example	Modules Functional(F) Neural(N)	Measure(s)			Combination Rule
		Derived Pure	Direct Pure	Com- posite	
1. Timing (rat)	F	*			—
2. LRP & parallel processes (human)	N	*			—
3. SDT & luminance (pigeon)	F	*			—
4. Selective adaptation & spatial frequency (human)	F		*		—
5. Single units & serial processes (monkey)	N		*		—
6. fMRI (hypothetical; human)	N		*		—
7. Detection of compound gratings (human)	F			*	Multiplication
8. ERP amplitude (human)	N			*	Summation
9. Response rate (rat)	F			*	Multiplication
10. Reaction time / sleep (human)	F			*	Summation
(11) LRP & serial processes (human)	F & N		*	*	Summation
(12) Response omissions (human)	F			*	Multiplication

## Tables

Table	Page	Title
1	156	The hypothesis of selective influence of factors on processes, and its implications for pure measures.
2	157	Inferential logic for pure measures.
3	160	Inferential logic for a composite measure with summation as the combination rule.
4	161	Inferential logic for a composite measure with multiplication as the combination rule.
5	165	A 2×2 factorial design.
6	173	"Yes-No" signal-detection experiment: Data matrix and three payoff matrices.
7	222	Alternative inferential logic for pure measures.
8	223	Alternative inferential logic for a composite measure with summation as the combination rule.
9	230	A factorial design with each factor at two levels.
10	231	A factorial design with each factor at four levels.

## Figures

Figure	Page	Example
1	165	
2	166	1
3	169	
4	170	2
5	176	3
6	180	4
7	183	
8	184	5
9	188	6
10	194	7
11	198	
12	199	8
13	202	9
14	208	10
15	227	(11)

## Equations

Equation	Page
1	159
2	159
3	159
4	161
5	162
6	162
7	162
8	193
9	193
10	193
11	201
12	202
13	204
14	205
15	206
16	220
17	220
18	220
19	235
20	236
21	236
22	236
23	237
24	237



ELSEVIER

Acta Psychologica 106 (2001) 147–246

---

---

acta  
psychologica

---

---

www.elsevier.com/locate/actpsy

# Separate modifiability, mental modules, and the use of pure and composite measures to reveal them

Saul Sternberg

*University of Pennsylvania, Philadelphia, PA 19104-6196, USA*

---

## Abstract

How can we divide a complex mental process into meaningful parts? In this paper, I explore an approach in which processes are divided into parts that are *modular* in the sense of being *separately modifiable*. Evidence for separate modifiability is provided by an instance of *selective influence*: two factors  $F$  and  $G$  (usually experimental manipulations) such that part **A** is influenced by  $F$  but invariant with respect to  $G$ , while part **B** is influenced by  $G$  but invariant with respect to  $F$ . Such evidence also indicates that the modules are functionally distinct. If we have *pure measures*  $M_A$  and  $M_B$ , each of which reflects only one of the parts, we need to show that  $M_A$  is influenced by  $F$  but not  $G$ , while  $M_B$  is influenced by  $G$  but not  $F$ . If we have only a *composite measure*  $M_{AB}$  of the entire process, we usually also need to confirm a *combination rule* for how the parts contribute to  $M_{AB}$ .

I present a taxonomy of separate-modifiability methods, discuss their inferential logic, and describe several examples in each category. The three categories involve measures that are *derived pure* (based on different transformations of the same data; example: separation of sensory and decision processes by signal detection theory), *direct pure* (based on different data; example: selective effects of adaptation on spatial-frequency thresholds), and *composite* (examples: the multiplicative-factor method for the analysis of response rate; the additive-factor method for the analysis of reaction time). Six of the examples concern behavioral measures and functional processes, while four concern brain measures and neural processes. They have been chosen for their interest and importance; their diversity of measures, species, and combination rules; their illustration of different ways of thinking about data; the questions they suggest about possibilities and limitations of the separate-modifiability approach; and the case they make for the fruitfulness of searching for mental modules. © 2001 Elsevier Science B.V. All rights reserved.

---

*E-mail address:* saul@psych.upenn.edu (S. Sternberg).

0001-6918/01/\$ - see front matter © 2001 Elsevier Science B.V. All rights reserved.

PII: S 0001-6918(00)00045-7



*PsycINFO classification:* 2260; 2300; 2400; 2520; 2530

*Keywords:* Cognitive processes; Perceptual motor processes; Neurosciences; Methodology; Models

---

## 1. Introduction

How can we divide a complex process into meaningful parts, or “modules”? I start by considering a general criterion for characterizing the modules of a complex process, a criterion that reflects recent thinking within psychology. Focusing on it provides a new taxonomy that unifies and relates diverse approaches to module identification. The criterion is *separate modifiability*, a kind of independence: if two processes are modules, then it should be possible to change each process without changing the other. Thus, we can find mental modules by searching for pairs or sets of mental processes that are separately modifiable. At present, the taxonomy encompasses three kinds of experiment that use separate modifiability to search for modules; I illustrate the taxonomy with a set of 10 diverse examples.<sup>1</sup>

The principle has been applied in two situations, which correspond to the two main categories of the taxonomy. In one, we believe or hypothesize that we have a *pure measure* of each of the hypothesized modules. Within this category there are two sub-categories: we can have *direct* pure measures, each depending on different data, or we can have *derived* pure measures, which depend on different aspects or functions of the same data. Perhaps the most famous case of pure measures comes from signal-detection theory (SDT), which is claimed to “separate” a sensory process from a decision process (Ex. 3). Here, the data might consist of the proportions of “hits” and “false alarms” in a detection or discrimination experiment. These proportions can be transformed and combined in different ways to create derived pure measures of discriminability (reflecting the sensory process) and of bias (reflecting the decision process). With pure measures, the hypothesis that the measures are indeed pure must usually be tested, along with separate modifiability.

In the second main category of experiments we have a *composite measure* of two or more hypothesized modules. That is, we have a measure to which we believe or hypothesize each module makes a contribution, such as detection probability (Ex. 7), amplitude of the event-related potential (ERP) (Ex. 8), or response rate (Ex. 9). Here, to evaluate separate modifiability, we need something more: we have to know or hypothesize a *combination rule* that tells us how the modules contribute to the composite measure. For example, in the *additive-factor method* (AFM) (Sections 3.1.1, 16; Appendix A.16.3), the hypothesized combination rule is *summation*: the composite measure, reaction time (RT), is the *sum* of the contributions from the modules. With a composite measure, the combination rule must usually be tested,

---

<sup>1</sup> “Example” and “Examples” will be abbreviated “Ex.” and “Exs.”, respectively.

along with separate modifiability. However, if the combination rule is known (as in Ex. 8, where it is known from physics), separate modifiability can be evaluated on its own.

*Separate modifiability of component processes and their functional distinctness.* These two ideas are closely related. “Functional distinctness” is not well defined, so its implications are not clear, but it at least suggests separate modifiability: if two parts of a process serve different functions (e.g., representing sensory information vs making a decision based on that representation), then it seems plausible that if they can be changed at all, one can be changed without the other (e.g., by manipulating the stimulus vs the costs and benefits of alternative decisions). Conversely, if we have demonstrated that the two parts are separately modifiable by finding different factors that influence them selectively, this is evidence for their being functionally distinct; if they were not, the same factors should influence them. Moreover, as additional factors are found that do not jointly influence them, the nature of the sets of factors that do and do not influence each process should provide information that helps to further specify its distinct function. In short, modules defined by separate modifiability are plausibly processes we would regard as functionally distinct (another criterion for modularity), and tests of separate modifiability can elucidate their distinctness.

The goal of the present paper is to explicate the separate-modifiability approach to process decomposition, and to demonstrate its fruitful use in diverse areas and in the analysis of both functional and neural processes.

### 1.1. *Antecedents*

The approach to understanding complex systems and processes by dividing them into functionally distinct parts is an old one. Bechtel and Richardson (1993) discuss examples from biology dating as far back as the 18th century. Within psychology, an early example is the decomposition of RT into functionally distinct and separately modifiable *stages* of mental processing, postulated and investigated by Donders (1868) and pursued by many modern psychologists. (See, e.g., Christie & Luce, 1956; Luce, 1986; Meyer, Osman, Irwin, & Yantis, 1988; Sanders, 1998; Sternberg, 1969, 1998b; Townsend & Ashby, 1983.) Shallice suggest that “modularity is a basic principle that we often use to understand natural systems” (1988, p. 20). Arguments for modular organization in the design of artifacts and algorithms and their speculative extension to design in nature have been advanced by Simon (1962) and also, in his “principle of modular design”, by Marr (1976).<sup>2</sup>

---

<sup>2</sup> In 1983, Fodor introduced and discussed a concept of modularity (narrower than those of Simon and Marr and the one advanced here) that has had a large impact in some areas of psychology, but is outside the scope of the present paper. In the context of his argument for neuroanatomical and functional modularity, Shallice (1988, Section 2.1) argues that Fodor’s criteria are too restrictive.

Studies of artificial neural networks have suggested that an arrangement of functionally specialized sub-networks may have advantages over a non-modular architecture (Happel & Murre, 1994), and may develop during learning if short neural connections are favored (Jacobs & Jordan, 1992). Researchers have also explored other conditions under which sub-networks develop functional specialization (Jacobs, 1997; Jacobs, Jordan, & Barto, 1991; Jacobs, Jordan, Nowlan, & Hinton, 1991).

Evidence for functional localization in the real brain, from the effects of lesions (Shallice, 1988) and from single-cell studies (e.g., Van Essen, Anderson, & Felleman, 1992), has contributed to the argument for processing modules, based on the idea that “anatomically distinct subsystems exist to carry out separate micro-functions” (Shallice, 1988 p. 19), and that “computational independence is made easier by separability of physical processes in the brain” (Shallice, 1981). Cowey (1985, pp. 55–56) has linked Marr’s (1976) argument for functional modularity of *processes* to anatomical localization of neural *processors*, based on the localization of a processor facilitating the selective inhibition of the process it implements, and on the importance of the shortness of interneurons in excitatory and inhibitory interactions within a processor. Such arguments suggest the hypothesis that operations carried out by anatomically localized processors implement functionally modular neural processes.

*Roberts’ independent-measures method.* Some of the ideas in this paper reflect those of Seth Roberts. What I call the application of the separate-modifiability criterion to derived pure measures (different measures based on the same set of data; Section 2) is the same as his “independent-measures method”, which he applied in several studies starting with Roberts (1981) and has discussed extensively since (1983, pp. 351–356; 1987, pp. 165–167; 1993; and 1998, Section 2.2).

## 1.2. Notation

Boldface roman letters, **A**, **B**, . . . , will be used to name hypothesized functional modules, which we learn about from behavioral measures that may be pure (e.g.,  $M_A$ ) or composite (e.g.,  $M_{AB}$ ). If **A** and **B** are influenced by changes in the levels of factors  $F$  and  $G$ , respectively, this may be made explicit by expressions such as  $M_A(F)$  and  $M_{AB}(F, G)$ . The contributions of **A** and **B** to  $M_{AB}$  may be described, respectively, as  $u_A$  and  $u_B$ , or, making explicit their dependence on  $F$  and  $G$ , as  $u_A(F)$  and  $v_B(G)$ , sometimes abbreviated  $u(F)$  and  $v(G)$ . Neural processes, which we learn about from brain measures, will be denoted by greek letters  $\alpha, \beta, \dots$ , with measures such as  $M_\alpha$  and  $M_{\alpha\beta}$ . The result of changing the level of a factor will be described in one of two ways. One is as a *difference*, the other as a *ratio*. Consider a pure measure  $M_A$  that depends on a factor  $F$  with levels  $\{F_j\}$ , and let  $\overline{M}_A$  be its mean. Consider the result of changing from level  $F_1$  to level  $F_2$ . We define the *effect* of  $F$  on  $\overline{M}_A$  as the *difference*  $\overline{M}_A(F_2) - \overline{M}_A(F_1)$ ; the *proportional effect* (“p.effect”) as the *ratio*  $\overline{M}_A(F_2)/\overline{M}_A(F_1)$ . A quantity such as  $RT_{\bullet k}$ , in which the subscript  $j$  in  $RT_{jk}$  is replaced by a dot, represents the mean of  $RT_{jk}$  over all values of  $j$ .

### 1.3. Twofold partition as an initial step in understanding a complex process

For simplicity we shall mainly consider dividing a process into just two modules, which requires just two factors. Such a twofold partition may often be a reasonable starting point in decomposing a process.<sup>3</sup> In general, extending the argument to more than two modules is straightforward: one or both of the modules already identified may be further partitioned into sub-modules. In another kind of elaboration, the functions of the identified modules can be further specified by considering the pattern of effects of additional factors. Such elaborations of an initial analysis (which may also serve to test it) are discussed in connection with Exs. 1 and 10.

If a process consists of modular sub-processes, then it would seem important to discover what they are and to roughly characterize their distinct functions. But this is only an initial step in understanding the process, which requires knowing in greater detail what the modules do, how they are organized, and how their effects combine.

### 1.4. Processes, measures, factors, and selective influence

More formally, two parts (sub-processes) **A** and **B** of a complex (mental) process are defined as modules if and only if they are *separately modifiable*. Evidence favoring separate modifiability is provided if we find an instance of *selective influence*, that is, find two factors *F* and *G* (usually experimental manipulations – independent variables) that influence **A** and **B** selectively: a change in the level of *F* influences **A** but leaves **B** invariant, while a change in the level of *G* influences **B** but leaves **A** invariant. However, we can observe only one or more *measures* of a process, not the process itself. If we have *pure measures*  $M_A$  and  $M_B$ , each of which reflects only one of the component processes, we need to show that  $M_A$  is influenced by *F* but invariant with respect to *G*, while  $M_B$  is influenced by *G* but invariant with respect to *F*. To demonstrate separate modifiability if we have only a *composite measure*  $M_{AB}$  of the entire process usually requires us also to test a *combination rule* for how the modules contribute to  $M_{AB}$ .

*The generality of selective influence.* In the absence of a compelling argument for why it should be otherwise, we would regard the property of selective influence of factors on processes or their measures as being satisfied only if the required invariances are general, within appropriate ranges of the levels of *F* and *G*. For example, the invariance of  $M_A$  with respect to *G* has to obtain at all levels of *F*.<sup>4</sup> The

<sup>3</sup> It is possible, however, for a process to contain parts whose modularity would not be revealed by a twofold partition. Suppose a process consists of three modules **A**, **B**, and **C**, and we find three factors *F*, *G*, and *H* such that *F* influences **A** and **B** but not **C**, *G* influences **B** and **C** but not **A**, and *H* influences **A** and **C** but not **B**. Separate modifiability would then have been demonstrated for each pair of sub-processes, and hence for the set of three. But it would not have been demonstrated for any twofold partition. For example, suppose a twofold partition separates **A** + **B** = **X** from **C**. Then, because all three factors influence **X**, both of the factors (*G* and *H*) that influence **C** would also influence **X**; thus **X** and **C** would not have been shown to be separately modifiable.

<sup>4</sup> With pure measures, explicit testing of such generality requires a factorial design (Section 3.3), often not used; with composite measures, factorial designs are required for other reasons.

reason for this desideratum is to avoid drawing strong conclusions based on accidents of particular factor levels. However, in the present context, selective influence can be persuasive evidence for separate modifiability even if such generality fails (Appendix A.9.1).

### 1.5. Statistical issues

The tests of statistical significance that are most familiar to psychologists are asymmetric: they force us to assume that there is no effect of a factor – invariance, the “null hypothesis” – unless the contrary can be proved. However, the demonstration of selective influence (e.g., of  $F$  on  $M_A$ ) relies as strongly on showing that  $M_B$  is invariant with respect of  $F$  as on showing that it is influenced by  $G$ . The conclusion of invariance should not be permitted to be drawn by default, simply because data are insufficiently precise. One useful guide, applied in the examples below, is obtained by expressing the magnitude of deviations from invariance as a one-dimensional quantity (easy if  $F$  is studied at only two levels), and providing an estimate of its precision such as its standard error (S.E.). We are then forced to consider the power of the test: how large the deviations would have to be in order to be detected. (A supplementary precaution is to check that the effect of  $F$  on  $M_A$  is significantly greater than the effect of  $G$ , and the converse for  $M_B$ .) In Ex. 3 (pure measures), the factors have more than two levels, which permits such one-dimensional quantities to be defined so as to be sensitive to the degree to which deviations from invariance are systematic. In Ex. 9 (composite measure), where a failure of invariance would be revealed as failure of additivity of the effects of factors with multiple levels, a similar quantity is defined.<sup>5</sup> Statistical issues are further discussed in Appendix A.11.2.

### 1.6. Process decomposition vs task comparison

The focus of the present paper is on methods for discovering the parts (sub-processes) of the complex process used to carry out a particular task. It is central in these methods to find factors that influence these sub-processes selectively. And it is therefore important that factors be selected such that their variation changes the process quantitatively but not qualitatively: that is, factors should influence the

---

<sup>5</sup> A possibly better approach, not integrated into the present paper, is to apply the kind of *equivalence tests* (Berger & Hsu, 1996) that have been developed to answer the question of the bioequivalence of drugs, tests that reverse the asymmetry of the more traditional ones. Here the null hypothesis asserts that two drugs are *not* equivalent; the risk described by the significance level is that equivalence (i.e., approximate invariance across drugs) will be asserted when it is false. To perform such a test it is necessary to define an interval around zero (the alternative hypothesis) such that if a one-dimensional measure of the true effect fell within that interval, invariance could be declared. Interactions can be treated similarly: for example, the alternative hypothesis can be an interval around zero for a one-dimensional interaction measure. Formal incorporation of such intervals into inference is rare in psychology, but we often use them informally; it seems desirable to make them explicit, recognizing that opinions about the appropriate interval may differ and may change.

execution of sub-processes, but not cause them to be added, deleted, or replaced. One feature of these separate-modifiability methods is that they provide evidence that bears on the validity of the *task theory* that guides their use – a hypothesis about the structure of the complex process used to carry out a task.

These separate-modifiability methods should be distinguished from the formally similar *task-comparison method* often used in psychological research, where the interest is in finding factors that selectively influence measures of performance in *different* tasks, each involving a *different* complex process, and the finding of greatest interest is *double dissociation* of tasks. Using such findings to identify the parts of the associated complex processes usually requires a theory for each task, theories that gain at most weak validation from the results, which in general provide no information about the separate modifiability of the hypothesized parts. It is important to distinguish the task-comparison method (Appendix A.1) from the methods discussed below, whose goal is to partition the one complex process used to carry out one particular task.

### 1.7. *Relevance of brain measurement to process decomposition*

The methods of cognitive neuroscience that use brain measurements taken during performance of a task offer great promise for helping us to understand the functional organization of mental processes, as well as the anatomical organization of the processors that implement them. For this reason I included, among the ten examples, four that involve such brain measures. As we shall see, the inferential logic that applies to *functional* processes (**A**, **B**, . . . , inferred from *behavioral* measures) can also be applied to *neural* processes ( $\alpha$ ,  $\beta$ , . . . , inferred from *brain* measures). Inference from brain measurements to functional processes raises several difficult issues (Meyer et al., 1988; Rugg & Coles, 1995; Sarter, Berntson, & Cacioppo, 1996). One is based on the observation that systematic changes in a brain measurement that are associated with task variations provide only correlational information; they do not necessarily reflect brain events that are *required* to perform the task (“task-essential”).<sup>6</sup> Caution is thus needed in their interpretation, whether in the context of task comparison (Appendix A.1.2.4) or process decomposition. A second issue is the relation between the modularity (that might be inferred from behavioral measures) of functional processes, and modularity of the neural processes assumed to implement them – modularity that might be inferred from brain measures. That these two kinds of modularity are associated seems plausible if the performance of a task depends on a set of distinct functional processes and if the neural processors that implement them are functionally specific. But does one kind of modularity necessarily imply the other (Section 11.1)? Much depends on the validity of correspondences that are assumed, e.g., that the latency of a brain measurement is temporally linked to some significant time point during a functional process (Sections 6, 10;

<sup>6</sup> Unlike brain *measures*, this problem does not arise with brain *factors* (such as lesions, or transcranial magnetic stimulation) used together with task factors and behavioral measures.

Appendix A.6); or that the level of activation of a brain region reflects the amount of “work” associated with implementing a process (Section 11; Appendix A.11).<sup>7</sup>

Different sources of reassurance about task essentiality are available for different brain measures. Examples include the following: It is helpful to look for (and find) the same patterns of selective factor effects on brain and task-performance measures (Section 11; Appendix A.6), acknowledging that these may not always be available (Section 6). In measures of single-neuron activity, where individual trial data are usable (Section 10), we can determine that such activity occurs on every trial (or on every trial when a particular stimulus or response occurs) between stimulus and response, and seldom or never between trials. (In contrast, for measures such as the ERP or functional magnetic-resonance imaging (fMRI) that often require averaging over trials, the measured events may occur on only some of the trials.) Lesion data are helpful when a processor is believed to have been localized, as in some fMRI studies (Section 11). And timing data are helpful when the measure has good time resolution, as in ERP or single-neuron measurements (Sections 6, 10, 14; Appendix A.6), data that, for example, permit us to know whether the measured event occurred during the stimulus–response interval, and when during that interval.

These and related issues are further considered in the context of particular examples.

### 1.8. *Organization of the paper*

I discuss the inferential logic associated with these methods and general considerations related to their use in Sections 2 and 3. Section 2 starts by distinguishing between the selective influence of factors on unobservable processes **A** and **B**, and on observable measures of those processes,  $M_A$  and  $M_B$ , and then explicates the inferential logic associated with the use of pure measures. Section 3 does the same for composite measures  $M_{AB}$  with summation (Section 3.1) and multiplication (Section 3.2) as the combination rules. With the latter combination rule an additional hypothesis must be included: that the contributions of **A** and **B** to  $M_{AB}$  are uncorrelated.

Ten examples are discussed in Sections 4–16, based on measures that are derived pure (Exs. 1–3), direct pure (Exs. 4–6), and composite (Exs. 7–10). Each group is preceded by an introduction (Sections 4, 8, 12); each example concludes with a section containing comments about some of its lessons and limitations. The examples, some flawed and limited, were chosen for their interest and importance, their diversity of measures, species, and combination rules, their illustration of various ways of making inferences from data in this context, the questions they suggest about the possibilities and limitations of the approach, and the extent to which they argue for the existence of modular mental processes and the value of searching for them. The measures are neural, for investigating neural processes (Exs. 2, 3, 5, 8;

---

<sup>7</sup> These assumptions are akin to “linking hypotheses” about the connections between perceptual and physiological states (Teller, 1984; Geisler & Albrecht, 2000).

Appendix A.6) and behavioral, for investigating functional processes (Exs. 1, 3, 4, 7, 9, 10); species include monkey (Ex. 5), rat (Exs. 1, 9), pigeon (Ex. 3), and human (Exs. 2, 4, 6–8, 10). For several of the examples, only mean data were available, which limited the possible analyses.

A caveat that applies to all the examples but is not reiterated for each is that the inferences from any single application of these methods have to be regarded as tentative. Among the reasons are: First, like all cases where a theory is confirmed by the success of its predictions, the degree of confirmation is moderated by the existence (which can change) of plausible alternative theories that generate the same predictions, along with other knowledge, including results of related experiments. Second, other measures of the same processes might provide different answers. And third, the arguments rest heavily on demonstrations of invariance as factor levels are changed – of the values of measures, or of the sizes of effects on measures – and persuasive evidence for invariance requires highly precise data, as well as the possibility of assessing the extent to which any deviations are systematic.

The reader is encouraged to skim Sections 2 and 3 to become familiar with terminology and logic, and then return to them after considering a few of the examples. For interested readers, the text is followed by appendices with heading numbers starting with “A” and keyed to corresponding main sections of the text. Because of their similarities, and because task comparison (Sections 1.6, 11) is more common in psychology than process-decomposition methods, the longest appendix (Appendix A.1) attempts to distinguish and relate the two approaches, arguing for the advantages of the separate-modifiability approach for the decomposition of mental processes. Other appendices are concerned with details of particular examples, and others with general issues such as the choice of factors (Appendix A.2), the plausibility of combination rules (Appendix A.3.1), the stochastic independence of contributions of modules to composite measures (Appendices A.3.2, A.3.3), advantages of factorial experiments and multiple-level factors (Appendices A.2.1, A.9), numerical scaling of factor levels (Appendix A.15), and statistical issues (Appendix A.11.2). Two further instances of process decomposition that complement the other examples are briefly described in Appendices A.6 and A.16.1.

Following the appendices is a glossary of the main abbreviations.

## 2. Pure measures: definitions and inferential logic

Suppose a process is hypothesized to consist of two modules **A** and **B** with corresponding hypothesized *pure measures*  $M_A$  and  $M_B$ . They are “pure” in the sense of being selective, or separate: for example,  $M_A$  should not vary with changes in **B**. Examples of measures of a process include its duration, a discriminability measure such as  $d'$  in SDT, an intensity measure such as a rat’s peak bar-pressing rate, and for a neural process, the amount of activity in a brain region.



Table 1

The hypothesis of selective influence of factors on processes, and its implications for pure measures

Parts of the Hypothesis	Associated Predictions
Process <b>A</b> is influenced by $F$	$F$ is likely to change some $M_A$
Process <b>A</b> is invariant with respect to $G$	$G$ changes no $M_A$
Process <b>B</b> is influenced by $G$	$G$ is likely to change some $M_B$
Process <b>B</b> is invariant with respect to $F$	$F$ changes no $M_B$

### 2.1. Selective influence of factors on processes and their measures

We cannot know about a mental process **A** as such, but only about one or more measures,  $M_A$ , of it. For this reason, we have to consider what the hypothesis of selective influence of factors  $F$  and  $G$  on processes **A** and **B** says about their influence on pure measures of **A** and **B**; the implications are spelled out in Table 1.<sup>8</sup> Thus, we test the hypothesis of selective influence of factors on processes by examining the behavior of various measures. Suppose we find that  $M_A$  varies with the level of  $F$  but not with the level of  $G$ , and that  $M_B$  varies with  $G$  but not with  $F$ . Using “ $\leftarrow$ ” for “is influenced by” and “ $\nleftrightarrow$ ” for “is invariant with respect to”, these four properties can be described as  $M_A \leftarrow F$ ,  $M_A \nleftrightarrow G$ ,  $M_B \leftarrow G$ , and  $M_B \nleftrightarrow F$ . Such a finding of selective influence of factors  $F$ ,  $G$  on measures  $M_A$ ,  $M_B$  supports the hypothesis of selective influence of  $F$ ,  $G$  on processes **A**, **B**, and thus supports separate modifiability (hence modularity) of **A**, **B**. It also supports the hypothesis that  $M_A$  and  $M_B$  are pure measures of **A** and **B**. Note that the choice of measures may influence our findings, and hence our conclusions. That is, two measures of the same process may behave differently, so a claim of invariance of a process may be limited to just the aspect of that process that is reflected by the measure we use. For example, if  $M_A$  is a particular pure measure of **A**, then, whereas  $M_A \leftarrow F$  implies  $\mathbf{A} \leftarrow F$ ,  $M_A \nleftrightarrow G$  does not imply  $\mathbf{A} \nleftrightarrow G$ . However, in the absence of arguments to the contrary it is reasonable to start with the idea that if an hypothesized measure of a process is invariant, then so is the process itself.

Suppose process **A** influences process **B**, such that a change in **A** causes a change in  $M_B$ . Then, because any factor that influences **A** should also indirectly influence  $M_B$ , we should be unable to use  $M_B$  to demonstrate separate modifiability of **A** and **B**. Similarly, for a composite measure  $M_{AB}$  we should be unable to find evidence supporting the modularity of **A** and **B** if the contribution of **B** to  $M_{AB}$  is influenced by **A**. Thus, in general, the separate modifiability of modules precludes one influencing the other. As discussed in Section 7.1, however, there may be good reason to regard **A** and **B** as modular even if one “influences” the other in the sense that **B** operates on information (the operand) provided by **A**. (After all, processes arranged sequentially

<sup>8</sup> For neural processes, change **A** to  $\alpha$ ,  $M_A$  to  $M_\alpha$ , etc., in this table and in Tables 2–4, 7, and 8.

Table 2  
Inferential logic for pure measures

<b>Joint Hypothesis</b>	
<i>H1</i> : Processes <b>A</b> and <b>B</b> are modules (separately modifiable).	
<i>H2</i> : $M_A$ , $M_B$ are pure measures of <b>A</b> , <b>B</b> .	
<b>Prediction</b>	
We may be able to find factors $F$ and $G$ that influence $M_A$ and $M_B$ selectively: $p_1: M_A \leftarrow F$ , $p_2: M_B \leftarrow F$ , $p_3: M_B \leftarrow G$ , $p_4: M_A \leftarrow G$ .	
<b>Alternative Results</b>	
We find factors $F, G$ that influence $M_A$ and $M_B$ selectively.	We fail to find such factors.
<b>Inferences</b>	
Support for joint hypothesis $H1 + H2$ .	Refutes one/both of $H1$ , $H2$ , or we didn't look enough for $F, G$ .

– as stages – have often been thought to be “data-dependent”, with one stage using data furnished by an earlier stage, yet are shown to be separately modifiable using  $M_{AB} = \overline{RT}$ .<sup>9</sup>) In testing for modularity in such a case, a measure  $M_B$  should be sought that reflects not the operand, but properties of the operator.

## 2.2. Hypothesis, prediction, and inference for pure measures

In Table 2 is outlined the two-part hypothesis associated with pure measures, the prediction to which it leads (which includes the four indicated properties,  $p_1$ ,  $p_2$ ,  $p_3$ , and  $p_4$ ) and the alternative possible findings and associated inferences.<sup>10</sup> Here and elsewhere (e.g., Tables 3, 4, 7, and 8) we would of course like to

<sup>9</sup> The requirement of “constant stage output” for such processes (e.g., Sanders, 1998; Sternberg, 1969, Section 3.2) must thus be interpreted with some subtlety.

<sup>10</sup> If the hypotheses about **A** and **B** are sufficiently detailed, then an alternative formulation to the one in Table 2 (Appendix A.2.3, Table 7) may be more appropriate, in which the joint hypothesis incorporates the specification of factors  $F$  and  $G$ .

ask separately about each part of the joint hypothesis (here *H1* and *H2*), but in a single such test we cannot. One consequence is that failure of the prediction may be less informative than success. If factors *F* and *G* are found not to influence  $M_A$  and  $M_B$  selectively, then, guided by our (revised) beliefs about **A** and **B**, we can try other factors, or abandon *H2* and consider other measures, or abandon *H1*.

### 2.3. Why all four properties are important

Suppose we tested for and found only  $p_1$  and  $p_2$  of the four properties listed under Prediction in Table 2: a factor *F* influences  $M_A$  but not  $M_B$ . Why do we need the others? One reason is that  $M_B \leftarrow F$  could occur simply because  $M_B$  is not *sensitive*, making the invariance  $M_B \leftarrow F$  less persuasive. (For example, if the measure  $M_B$  is the duration of process **B**, then *F* could modify **B** without altering its duration.) This possibility is rendered less plausible if we can show that *some* manipulation can alter  $M_B$ , that is, if we can find a factor *G* such that  $M_B \leftarrow G$  (property  $p_3$ ). We would then know that both  $M_A$  and  $M_B$  are sensitive.<sup>11</sup> The second reason is that even with  $p_3$  added, we have no support for the hypothesis that **B** can be modified without **A**; for this, we also need  $p_4$ :  $M_A \leftarrow G$ . If we added  $p_4$  but not  $p_3$ , our evidence would also be insufficient:  $p_4$  could arise simply because *G* was not a *potent* factor. This possibility is rendered less likely if we can show that *G* can influence *some* process, which we learn from  $p_3$ ,  $M_B \leftarrow G$ . Thus, the four properties together tell us that each factor is potent (can modify *some* process) and that each measure is sensitive (can be altered by *some* factor), as well as providing evidence that, for processes **A** and **B**, each can be modified without the other.<sup>12</sup>

## 3. Composite measures: definitions and inferential logic

Suppose we have one measure  $M_{AB}$  that reflects a property of the entire complex process of interest, a process that consists of component processes **A** and **B**. That is, we know only the *combined* contributions of **A** and **B** to a measure  $M_{AB}$  that depends on both of them. To make inferences from the data, we must know or hypothesize how the contributions of **A** and **B** combine in influencing  $M_{AB}$ . In this paper I consider primarily two combination rules, *summation* and *multiplication*. (Hybrid rules are also briefly considered, in Section 15 and Appendix A.15.) These rules are

<sup>11</sup> One limitation of the criterion of separate modifiability is that if a process (and hence all measures of it) is insensitive – if we cannot modify it – it will not be revealed as a module. On the other hand, convincing evidence of the non-modifiability of a process in the context of modifiable others might argue for its being functionally distinct, and modular in a different sense.

<sup>12</sup> Suppose a subset of the properties obtains, such as  $p_1$ ,  $p_3$ , and  $p_4$  (see Ex. 3). This might be interpreted as “partial modularity”, consistent with **A** influencing **B** but not the reverse; in the task-comparison method this would be analogous to something between single and double task-dissociation (Appendix A.1.2.1).

associated with the two ways mentioned in Section 1.2 in which the result of changing the level of a factor can be described, as an *effect* for one and as a *p.effect* for the other.<sup>13</sup>

### 3.1. Composite measures based on summation

Suppose we either know or hypothesize that the combination rule is *summation*: the combined contribution of processes **A** and **B** to  $M_{AB}$  is the *sum* of their separate contributions, or  $M_{AB} = u_A + v_B$ . Letting  $u \equiv u_A$  and  $v \equiv v_B$ , if factors  $F$  and  $G$  selectively influence **A** and **B**, respectively, we have

$$M_{AB}(F_j, G_k) = u(F_j) + v(G_k), \quad (1)$$

where, for example,  $u(F_j)$  is a function that describes the relation between the level of  $F$  and the contribution of **A** to  $M_{AB}$ . Now, in general we work with averages rather than individual values of  $M_{AB}$ , and we regard  $u_A$  and  $v_B$  as random rather than deterministic variables. Using  $\bar{M}$ ,  $\bar{u}$ , and  $\bar{v}$  to indicate means (expectations) of these random variables, it is pleasantly convenient that with no further assumptions (such as the stochastic independence of  $u$  and  $v$ ), Eq. (1) implies

$$\bar{M}_{AB}(F_j, G_k) = \bar{u}(F_j) + \bar{v}(G_k). \quad (2)$$

(As we shall see, matters are not so simple if the combination rule is multiplication.) From Eq. (2) it is easy to show that  $F$  and  $G$  are *additive factors*: the combined effect on  $M_{AB}$  of changing the levels of both  $F$  and  $G$  is the sum of the effect of changing only  $F$  and the effect of changing only  $G$ . For changes from the first to the second level of each factor,

$$\begin{aligned} \text{effect}(F, G) &\equiv \bar{M}_{AB}(F_2, G_2) - \bar{M}_{AB}(F_1, G_1) \\ &= [\bar{u}(F_2) + \bar{v}(G_2)] - [\bar{u}(F_1) + \bar{v}(G_1)] \\ &= [\bar{u}(F_2) - \bar{u}(F_1)] + [\bar{v}(G_2) - \bar{v}(G_1)] \equiv \text{effect}(F) + \text{effect}(G). \end{aligned} \quad (3)$$

Eq. (3) is equivalent to the effect of  $F$  on  $M_{AB}$  being invariant over levels of  $G$ , and vice versa. In Table 3 is outlined the two-part hypothesis associated with such a composite measure, the prediction to which it leads, and the alternative possible findings and associated inferences.<sup>14</sup>

#### 3.1.1. Illustration: the additive-factor method

One way in which summation can arise as the combination rule is if processes **A** and **B** are arranged as *stages* – functionally distinct operations that occur during

<sup>13</sup> Care must be taken in the estimation of the p.effect from values subject to sampling error because, in general,  $E(\mathbf{x})/E(\mathbf{y}) \neq E(\mathbf{x}/\mathbf{y})$ , where  $E(\cdot)$  is expectation, even when  $\mathbf{x}$  and  $\mathbf{y}$  are stochastically independent.

<sup>14</sup> Whereas properties  $\{p_k\}$  (Tables 2, 7) apply to observable quantities, analogous properties  $\{p'_k\}$  (Tables 3, 4, 7, 8) apply to contributions to a composite measure that are not directly observable. Also note that if **A** and **B** are hypothesized to exhaust the process (in a twofold partition), then in Tables 3 and 4 the fourth line under Prediction can be deleted, and in Table 8, the third line of  $H1^*$ .

Table 3  
Inferential logic for a composite measure with summation as the combination rule

<b>Joint Hypothesis</b>	
<i>H1</i> : Processes <b>A</b> and <b>B</b> are modules (separately modifiable).	
<i>H3</i> : Contributions $u_A, v_B$ of <b>A</b> , <b>B</b> to $M_{AB}$ combine by <i>summation</i> .	
<b>Prediction</b>	
We may be able to find factors $F$ and $G$ that influence <b>A</b> and <b>B</b> selectively:	
$p'_1: u_A \leftarrow F, p'_2: v_B \leftarrow F, p'_3: v_B \leftarrow G, p'_4: u_A \leftarrow G,$	
and jointly influence no other process.	
If so, their <i>effects</i> on $M_{AB}$ will be <i>additive</i> .	
<b>Alternative Results</b>	
We find factors $F$ and $G$ with <i>additive effects</i> on $M_{AB}$ .	We fail to find such factors.
<b>Inferences</b>	
Support for joint hypothesis $H1 + H3$ .	Refutes one/both of <i>H1</i> , <i>H3</i> , <i>or</i> we didn't look enough for $F, G$ .

non-overlapping epochs, and such that the response occurs when both operations have been completed. The measure  $M_{AB} = \overline{RT}$ , the duration of the full process, can be written as the sum of individual process durations,  $T_A + T_B$ . If factors  $F$  and  $G$  influence processes **A** and **B** selectively, then  $RT = T_A(F) + T_B(G)$ . This implies that the effect of  $F$  on  $\overline{RT}$  in an  $F \times G$  factorial experiment (the difference in  $\overline{RT}$  produced by a change in level of  $F$ ) is *invariant* over levels of  $G$ . Also implied is the complement: invariance of the effect of  $G$  on  $\overline{RT}$  over levels of  $F$ ; together, *mutual invariance*. Such invariance is equivalent to the effects on  $\overline{RT}$  of  $F$  and  $G$  being *additive*, and is one aspect of the AFM, in which the finding of such invariance is used as evidence for *H1* and *H3* of Table 3. (An example due to Sanders, Wijnen, & Van Arkel, 1982 is described in Section 16.) Indeed, in the present context, the AFM might better be called the “method of invariant factor effects”.<sup>15</sup>

<sup>15</sup> In an alternative formulation of the inferential logic associated with the AFM, a statement about the selective influence of particular factors is incorporated in the joint hypothesis (Appendix A.2.3, Table 8; Sternberg, 1998b, Section 14.6.3).

3.2. Composite measures based on multiplication

Suppose we either know or hypothesize that the combination rule is *multiplication*: the combined contribution of processes **A** and **B** is the *product* of their separate contributions, or  $M_{AB} = u_A \times v_B$ . If factors *F* and *G* selectively influence **A** and **B**, respectively, then we have

$$M_{AB}(F_j, G_k) = u(F_j) \times v(G_k). \tag{4}$$

As when summation is the combination rule, *u* and *v* are random rather than deterministic variables, and we will be working with averages rather than individual values of  $M_{AB}$ . With the multiplicative combination rule, however, an additional hypothesis is needed to permit the derivation of the equivalent to Eq. (2) from Eq. (4): we must assume (*H5* in Table 4) that the contributions from **A** and **B** to  $M_{AB}$  have

Table 4  
Inferential logic for a composite measure with multiplication as the combination rule

<b>Joint Hypothesis</b>	
<i>H1</i> : Processes <b>A</b> and <b>B</b> are modules (separately modifiable).	
<i>H4</i> : Contributions $u_A, v_B$ of <b>A</b> , <b>B</b> to $M_{AB}$ combine by <i>multiplication</i> .	
<i>H5</i> : Contributions of <b>A</b> and <b>B</b> to $M_{AB}$ are uncorrelated.	
<b>Prediction</b>	
We may be able to find factors <i>F</i> and <i>G</i> that influence <b>A</b> and <b>B</b> selectively: $p'_1: u_A \leftarrow F, p'_2: v_B \leftarrow F, p'_3: v_B \leftarrow G, p'_4: u_A \leftarrow G,$ and jointly influence no other process. If so, their <i>proportional effects</i> on $M_{AB}$ will be <i>multiplicative</i> .	
<b>Alternative Results</b>	
We find factors <i>F</i> and <i>G</i> with <i>multiplicative p.effects</i> on $M_{AB}$ .	We fail to find such factors.
<b>Inferences</b>	
Support for joint hypothesis <i>H1 + H4 + H5</i> .	Refutes one/more of <i>H1, H4, H5</i> , <i>or</i> we didn't look enough for <i>F, G</i> .

zero correlation.<sup>16</sup> (The expectation of a product of two random variables is the product of their expectations if and only if they are uncorrelated.)

Using  $\bar{M}$ ,  $\bar{u}$ , and  $\bar{v}$  to indicate means (expectations), and given the additional assumption, it follows from Eq. (4) that

$$\bar{M}_{AB}(F_j, G_k) = \bar{u}(F_j) \times \bar{v}(G_k). \quad (5)$$

From Eq. (5) it is easy to show that  $F$  and  $G$  are *multiplicative factors*: the p.effect on  $M_{AB}$  of changing the levels of both  $F$  and  $G$  is the product of the p.effect of changing only  $F$  and the p.effect of changing only  $G$ .<sup>17</sup> For changes from the first to the second level of each factor,

$$\begin{aligned} p.effect(F, G) &\equiv \frac{\bar{M}_{AB}(F_2, G_2)}{\bar{M}_{AB}(F_1, G_1)} = \frac{\bar{u}(F_2)\bar{v}(G_2)}{\bar{u}(F_1)\bar{v}(G_1)} = \frac{\bar{u}(F_2)}{\bar{u}(F_1)} \times \frac{\bar{v}(G_2)}{\bar{v}(G_1)} \\ &= \frac{\bar{M}_{AB}(F_2, G_k)}{\bar{M}_{AB}(F_1, G_k)} \times \frac{\bar{M}_{AB}(F_j, G_2)}{\bar{M}_{AB}(F_j, G_1)} \equiv p.effect(F) \times p.effect(G). \end{aligned} \quad (6)$$

Illustrations are provided by Exs. 7 and 9. Eq. (6) is equivalent to the p.effect of  $F$  on  $M_{AB}$  being invariant over levels of  $G$ , and vice versa. Note that if Eq. (4) is true for  $M_{AB}$ , it is also true for its reciprocal; simply replace  $u(F_j)$  and  $v(G_k)$  by their reciprocals in Eq. (4). Analogs to Eqs. (5) and (6) then follow, as above.<sup>18</sup>

If two modules contribute multiplicatively to  $M_{AB}$ , and their contributions are greater than zero, then they contribute additively to  $\log(M_{AB})$ :

$$\log[M_{AB}(F_j, G_k)] = \log[u(F_j)] + \log[v(G_k)].$$

Under these conditions, Eq. (6) thus implies

$$\log[p.effect(F, G)] = \log[p.effect(F)] + \log[p.effect(G)]. \quad (7)$$

By using the logarithm of the composite measure, we can thus transform the multiplicative case into the additive one, which is sometimes convenient (Section 15) and may be very useful (Appendix A.3.2).

### 3.3. Factorial experiments

We shall be considering (complete) factorial experiments in which each of the  $n_F$  levels of  $F$  is tested with each of the  $n_G$  levels of  $G$ , thus generating  $n_F \times n_G$  “conditions”. This kind of experiment permits us to determine the extent to which the effect of each factor is modulated by the level of the other. For testing selective influence with pure measures, such experiments are not required, but they do add important information, as demonstrated in Section 7.2 and discussed in Section 11

<sup>16</sup> Some of the implications for data analysis of  $H5$  being required are discussed in Appendix A.3.2.

<sup>17</sup> For any level  $G_k$  of  $G$ , the p.effect on  $M_{AB}$  of a factor  $F$  with two levels  $F_1, F_2$  is the value of the measure when  $F = F_2$  divided by its value when  $F = F_1$ .

<sup>18</sup> As shown by Anderson (1996) and Roberts (1987), analyses of such data benefit from numeric scaling of the factor levels; see Sections 13, 15, and Appendix A.15.

and Appendix A.9.1. However, factorial experiments are essential with composite measures, because the predictions take the form of statements about how factor effects combine. For example, if the combination rule is summation, as in Exs. 8 and 10, then selective influence implies that the effect of one factor should be invariant over levels of the other, which corresponds to additive factor effects; such a prediction can be tested only in a factorial experiment.

#### 4. Introduction to three examples of inference based on derived pure measures

In each of the examples of Sections 5–7, different computations on the same set of data provide hypothesized pure measures of two component processes.

**Example 1** (*Isolation of a timing module in the rat*). Roberts (1981) measured response rate as a function of time from signal onset for rats performing a time-discrimination task (the *peak procedure*); he discovered that two measures based on the response-rate function were influenced selectively by different factors. Further experiments permitted him to confirm and elaborate the inferred decomposition. This example is included because it is simple, clear, and prototypical.

**Example 2** (*Parallel neural modules revealed by the lateralized readiness potential (LRP)*). The LRP is a measure of electrical signals at the scalp that reflects the preparation of a left-hand or right-hand response. With appropriate experiments (Osman, Bashore, Coles, Donchin, & Meyer, 1992), two brain measures have been derived from it, indicators of when the response is selected and of when the decision is made to execute that response. The pattern of factor effects not only indicates the modularity of the two corresponding neural processes, but, in addition, shows that they are arranged in parallel.

**Example 3** (*Separation of sensory and decision processes by signal-detection theory (SDT)*). SDT is associated with the best known use of derived pure measures of two hypothetical processes (sensory **S** and decision **D**) that underlie performance in many tasks. Its measure  $M_S = d'$  of the sensory process is influenced by “sensory factors” (such as stimulus properties) but not “decision factors” (such as rewards and penalties). However, until recently SDT has failed to provide any  $M_D$  that is invariant with respect to sensory factors. Recent work with pigeons (Alsop, 1998; McCarthy & Davison, 1984) suggests that the problem lies not with  $M_D$  but rather with the principle factor used to influence and control **D**. By using a different factor, the approximate modularity of **S** and **D** can be demonstrated. This example, therefore, shows how critical is the choice of factors. Another feature of this example is the examination of more than two levels of each factor, which permits testing the extent to which deviations from invariance are systematic.



### 5. Isolation of a timing module in the rat (Ex. 1)

In the *peak procedure* as used by Roberts (1981),<sup>19</sup> a rat that has learned to press a lever for food is presented, randomly, with two kinds of trial, each starting with the onset of a signal such as a light. On *food trials*, a food reward is “primed” after a fixed interval (such as 20 s) called *time of food (TF)*, in the sense that the first lever press after *TF* causes food delivery and signal offset. On *empty trials*, no food is delivered, and after 2 or 3 min the signal turns off, independent of the rat’s behavior. After a rat is trained, its response rate during an empty trial rises and then falls, reaching a maximum at about the time when food would become available on food trials. Results from this procedure and others suggest that animals have a *clock process, C*, that measures such time intervals. Roberts’ goal was to *isolate* this process, that is, to show that *C* can be modified and measured separately from the remainder of the stimulus–response path, which I shall call the *response process R*; Among other functions, *R* uses information from *C* in its control of response rate.

For each condition, Roberts started with the function that relates mean response rate to time from signal onset, the *response-rate function*. (See Fig. 1 for examples of mean functions.) From such a function we can derive many properties that might be pure measures; examples are the maximum rate (*peak rate*), the time at which the peak rate is achieved (*peak time*), the *spread* (width) of the function, and the *tail rate* (rate toward the end of empty trials). Earlier results from a similar procedure (Catania, 1970) had shown in a pigeon that peak rate could be changed without changing peak time. Roberts was interested in the possibility that processes *C* and *R* are modules associated with pure measures  $M_C$  (peak time) and  $M_R$  (peak rate), respectively.

In one experiment (1981, Exp. 1) Roberts randomly mixed trials on which the signals were light and sound. He varied two factors, *TF*, defined above, and the proportion of trials on which food was delivered, *probability of food (PF)*. During Phase 1 of the experiment, *PF* was fixed at 0.8 while *TF* varied:  $TF = TF_1 = 20$  s with one of the two signals;  $TF = TF_2 = 40$  s with the other. During Phase 2, *TF* was fixed at 20 s while *PF* varied:  $PF = PF_1 = 0.2$  with one of the two signals;  $PF = PF_2 = 0.8$  with the other. In both phases, the two signals occurred randomly and equally often. The conditions studied were thus  $\{(TF_1, PF_2), (TF_2, PF_2)\}$  in Phase 1, and  $\{(TF_1, PF_1), (TF_1, PF_2)\}$  in Phase 2 (shown in Table 5, column 2 and row 1, respectively). Response-rate functions after training, averaged over the 10 rats, are shown in Fig. 1. Fig. 2 shows the effects on peak rate and peak time of changing *TF* (panel A) and of changing *PF* (panel B).

These findings of selective influence (and others: Roberts, 1993, 1998) support the hypothesis that the process that controls response rate can be partitioned into two different modules (*C* and *R*) that control peak time ( $M_C$ ) and peak rate ( $M_R$ ), respectively, and that  $M_C$  and  $M_R$  are pure measures of these modules.

<sup>19</sup> See Roberts (1998) for further discussion and access to other illuminating papers.

Table 5  
A  $2 \times 2$  factorial design<sup>a</sup>

		Probability of food, $PF$	
		$PF_1 = 0.2$	$PF_2 = 0.8$
Time of food, $TF$	$TF_1 = 20$ s	<b>(20, 0.2)</b>	<b>(20, 0.8)</b>
	$TF_2 = 40$ s	(40, 0.2)	<b>(40, 0.8)</b>

<sup>a</sup> Boldface entries show conditions in Exp. 1 of Roberts (1981).

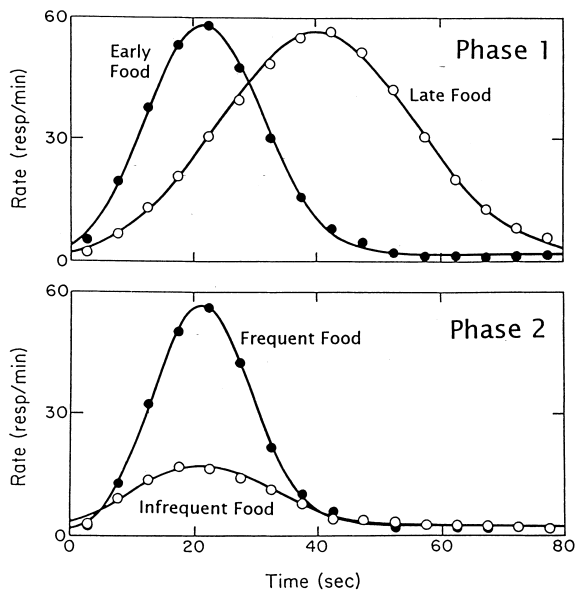


Fig. 1. Mean response rate vs time from signal onset within consecutive 5 s intervals in the two phases of Roberts' (1981) Experiment 1. During Phase 1  $PF$  was fixed at 0.8, and  $TF$  differed for the two possible signals:  $TF_1 = \text{early} = 20$  s vs  $TF_2 = \text{late} = 40$  s. During Phase 2  $TF$  was fixed at 20 s, and  $PF$  differed for the two possible signals:  $PF_1 = \text{infrequent} = 0.2$  vs  $PF_2 = \text{frequent} = 0.8$ . These functions are based on the data for both food trials and empty trials. The fitted curves are described by Roberts (1981, pp. 245–246). After Fig. 2.3 of Roberts (1998).

### 5.1. Two kinds of elaboration of the two-module analysis

Such a two-module analysis can be extended in two ways. First, each additional factor that is found that influences or does not influence  $M_C$  or  $M_R$  permits a more refined characterization of the identified modules. Much can be learned about what a (sub-)process does from which of a set of factors influence it (and how), and which do not.

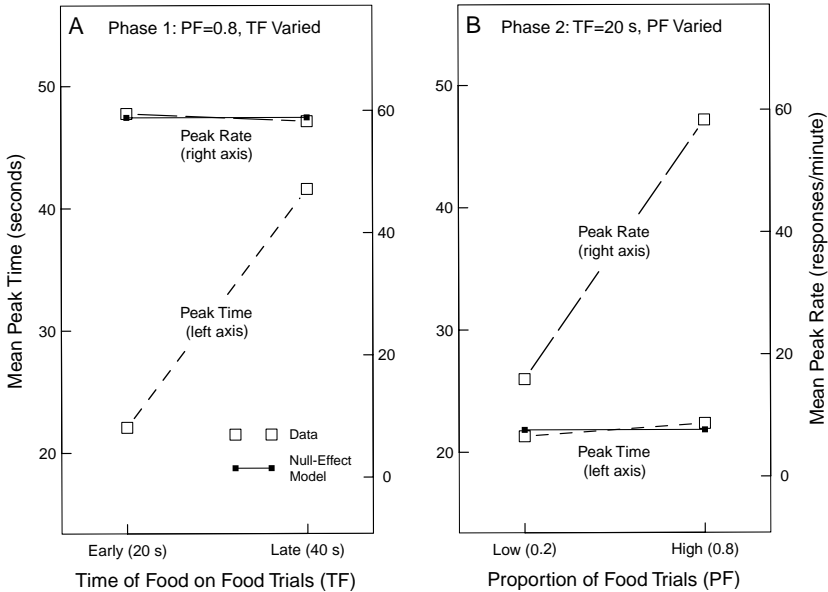


Fig. 2. Panel A: Effects of *TF* in Phase 1 on peak rate ( $-1.3 \pm 2.5$  responses/min) and on peak time ( $20 \pm 1$  s). Panel B: Effects of *PF* in Phase 2 on peak rate ( $42 \pm 5$  responses/min) and on peak time ( $1.1 \pm 1.0$  s). Also shown are null-effect models for peak rate in panel A and peak time in panel B. Note that for the conditions on the left in panel A and on the right in panel B (where the levels of *TF* and *PF* are the same) values of both peak rate and peak time are almost identical; the stability across Phases 1 and 2 is impressive. The differences between the effects of *TF* (Phase 1) and *PF* (Phase 2) on each of the two measures are highly reliable: for peak rate the difference is  $44 \pm 6$  responses/min; for peak time the difference is  $18 \pm 1$  s. Roberts (1981) used robust and resistant analyses; the values here, derived with methods more familiar to psychologists, differ slightly from his. Variability estimates are based on differences among the ten rats.

Second, the two-module process can be further partitioned, as follows: Suppose we have found that factors *TF* and *PF* influence measures  $M_C$  and  $M_R$  selectively, as above. If a third measure  $M_X$  is found that is influenced selectively by a new factor *H*, but not by *TF* or *PF*, and if *H* influences neither  $M_C$  nor  $M_R$  (adding five properties to  $p_1, \dots, p_4$  of Table 2), then this indicates the presence of a third module **X**, and a further partition is justified.<sup>20</sup>

Roberts has achieved both kinds of elaboration in his analyses of timing behavior. First, several other findings permit further characterization of processes **C** and **R** (Roberts, 1981). For example, a “blackout” (signal interruption) early during a trial changes peak time but not peak rate, while, after experience, prefeeding (before the

<sup>20</sup> For composite measures, both kinds of elaboration of a two-factor two-module analysis require factorial experiments with two or three factors. Examples of the first kind of elaboration in such a case are described in Section 16.1.

session) does the reverse. The inference is that these two manipulations selectively influence **C** and **R**, respectively. These findings also add support to the conclusion from the results described above that **C** and **R** are separate modules, and that  $M_C$  and  $M_R$  are pure measures of them. Second, other measures derived from the response-rate function, such as relative spread (width/peak-time), can be changed while peak time and peak rate remain invariant (Roberts, 1982, 1993); this indicates that the S–R path can be partitioned further.<sup>21</sup>

## 5.2. Comments

In this investigation of separate modifiability, the joint hypothesis (Table 2) includes *H1*: Processes **C** and **R** are modules (separately modifiable) and *H2*: Peak time  $M_C$  and peak rate  $M_R$  are pure measures of **C** and **R**, respectively. Roberts' discovery of factors *TF* and *PF* that influence the two measures selectively supports the joint hypothesis. The peak time and peak rate measures are different aspects of the same data (the response-rate function), rather than aspects of different data; hence these are *derived* rather than *direct* pure measures.

The boldface entries in Table 5 describe the design of Roberts (1981) experiment. As mentioned in Section 3.3, complete factorial experiments are not required when pure measures are available, as in this example, where the condition ( $TF_2$ ,  $PF_1$ ) was excluded. What these results tell us is that when  $PF = PF_2 = 0.8$ ,  $p_1 : M_C \leftarrow TF$  and  $p_2 : M_R \leftarrow TF$  (Phase 1) and that when  $TF = TF_1 = 20s$ ,  $p_3 : M_R \leftarrow PF$ , and  $p_4 : M_C \leftarrow PF$ . What would a complete experiment have added? First, two tests of *qualitative generality*: given selective influence of factors on measures, properties  $p_1$  and  $p_2$  should also be found when  $PF = PF_1 = 0.2$ , and properties  $p_3$  and  $p_4$  should also be found when  $TF = TF_2 = 40s$  (Section 1.4); such implications could be tested. And second, two tests of *quantitative invariance*: given the joint hypothesis, the effect of *TF* on  $M_C$  should be invariant with respect to *PF*, and the effect of *PF* on  $M_R$  should be invariant with respect to *TF*. Thus, there are internal consistency checks of the method, and here, a complete factorial experiment would permit them.

## 6. Parallel neural modules revealed by the lateralized readiness potential (Ex. 2)

Consider a trial in a choice–reaction experiment where two alternative responses are made by the two hands. Recall that the part of the motor cortex that controls a hand is contralateral to that hand. When enough information has been extracted from the stimulus to permit selection of the hand, but before any sign of muscle activity, the part of the motor cortex that controls that hand becomes more active than the part that controls the non-selected hand. This asymmetric activity can be

<sup>21</sup> Most of these findings and conclusions are summarized in Table 25.1 and Fig. 25.10 of Roberts (1993); see also Roberts (1981).

detected as an increase in the difference between electrical potentials at the two corresponding scalp locations. Let  $A_{MC}(t)$  (an index of *motor-cortex asymmetry*) express this difference as a function of time from stimulus onset.  $A_{MC}(t)$  is normally zero, but becomes positive when the response hand is selected; the increase of such asymmetry is called the *lateralized readiness potential* (LRP) (see Eimer, 1998 for a review).

I shall use “Event 1” to mean the onset of the LRP; the latency  $T_1$  of Event 1 can thus be used to indicate when the response has been selected. Normally (on “Go” trials)  $A_{MC}$  continues to rise until the overt response is initiated. If a “NoGo” signal tells the subject to inhibit the response before it would otherwise occur, however,  $A_{MC}$  starts falling.<sup>22</sup> The time at which  $A_{MC}(t; NoGo)$  diverges from  $A_{MC}(t; Go)$  – the latency  $T_2$  of “Event 2” – can thus be used to indicate when the NoGo signal is discriminated and response preparation ceases.

Osman et al. (1992 see also Osman, 1998) exploited these properties of  $A_{MC}(t)$  in a study in which the visual stimulus on each trial had two features, *position* and *category*. Its position (left vs right, which was rapidly discriminated) indicated which response to make, should a response be required. Its category (letter vs digit, which was discriminated more slowly) indicated whether this was a Go trial (on which the selected response should be activated) or a NoGo trial (on which any response should be inhibited).

The purpose of the study was to investigate the relation between two processes underlying performance in the task: One is *location discrimination and response selection* ( $\alpha$ ). The other is *category discrimination, and activation vs inhibition of the selected response* ( $\beta$ ). Events 1 and 2 indicate the completion of processes  $\alpha$  and  $\beta$ , respectively. Can response preparation start when the position but not the category of the stimulus has been discriminated? And, if so, can category discrimination proceed in parallel with response preparation? To answer such questions, Osman et al. examined the effects of two factors: One (in Exp. 1) is *Go–NoGo Discriminability* ( $GND$ ), which should influence  $\beta$ ; it could be *easy* (letter and digit with dissimilar shapes,  $GND_1$ ) or *hard* (similar shapes,  $GND_2$ ). The other factor (in Exp. 2) is the *stimulus-response mapping* ( $SRM$ ), which should influence  $\alpha$ ; it could be *easy* (spatially compatible  $SRM$ : respond with the hand on the same side as the stimulus,  $SRM_1$ ) or *hard* (spatially incompatible  $SRM$ : respond with the hand on the opposite side,  $SRM_2$ ).

Consider the asymmetry function for Go trials,  $A_{MC}(t; Go)$ : this was obtained by linearly combining four different  $ERP(t)$  functions, where  $ERP(t)$  is the average (over an appropriate subset of trials) of scalp potential vs time from stimulus onset. The four  $ERP(t; Go)$  functions were obtained from the left and right motor cortex for each of the two signaled responses on Go trials. They were combined in such a way

<sup>22</sup> I use the term “inhibit” informally here, without distinguishing a separate inhibitory process from the mere cessation of preparation.

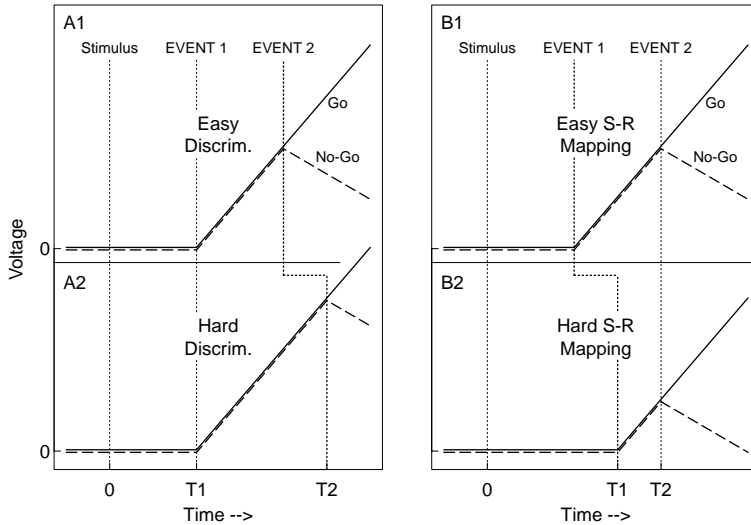


Fig. 3. Schematic idealized asymmetry functions  $A_{MC}(t)$  from Osman et al. (1992). Event 1 is the LRP onset; Event 2 is the divergence of  $A_{MC}(t; Go)$  from  $A_{MC}(t; NoGo)$ . Panels A1 and A2: Asymmetry functions from the two conditions in Exp. 1, in which  $GND$  could be *easy* ( $GND = GND_1$ ) or *hard* ( $GND = GND_2$ ). Panels B1 and B2: Asymmetry functions from the two conditions in Exp. 2, in which  $SRM$  could be *easy* ( $SRM = SRM_1$ ) or *hard* ( $SRM = SRM_2$ ).

that a positive  $A_{MC}(t; Go)$  would indicate preparation for the signaled response.<sup>23</sup>  $A_{MC}(t; NoGo)$  was obtained similarly.

Idealizations of the resulting  $A_{MC}(t)$  functions are shown in Fig. 3, and the observed values of  $\bar{T}_1$  and  $\bar{T}_2$  in Fig. 4. Each of the four panels of Fig. 3 shows the pair  $A_{MC}(t; Go)$  and  $A_{MC}(t; NoGo)$  for one condition. The two latency measures for a condition were derived in different ways from this pair of  $A_{MC}(t)$  functions: The latency  $T_1$  of Event 1 (onset of the LRP) is the time at which the *sum* of the two  $A_{MC}(t)$  functions reliably exceeds baseline. The latency  $T_2$  of Event 2 (divergence of the Go and NoGo LRPs) is the time at which their *difference* reliably exceeds zero. In

<sup>23</sup> For Go (letter) trials in Exp. 1, for example, there were four conditions, left vs right position  $\times$  easy vs hard discrimination, each providing data from about 150 correct trials per subject. On each trial on which the left response  $R_L$  was signaled, the left and right motor cortex ( $MC_L$  and  $MC_R$ ) generated potentials that were measured as a function of time from stimulus onset. For each subject, these potential functions were averaged over each such set of trials, to give  $ERP(t; Go, R_L, MC_L)$  and  $ERP(t; Go, R_L, MC_R)$ . Corresponding ERPs were obtained for Go trials on which the right response  $R_R$  was signaled. For each condition, these four ERPs were then combined to give  $A_{MC}(t; Go) = [ERP(t; Go, R_L, MC_L) - ERP(t; Go, R_L, MC_R)] + [ERP(t; Go, R_R, MC_R) - ERP(t; Go, R_R, MC_L)]$ . Because the LRP is a negative-going potential at the motor cortex contralateral to the response, it causes the  $A_{MC}(t)$  defined in this way to become positive when the signaled response is selected.

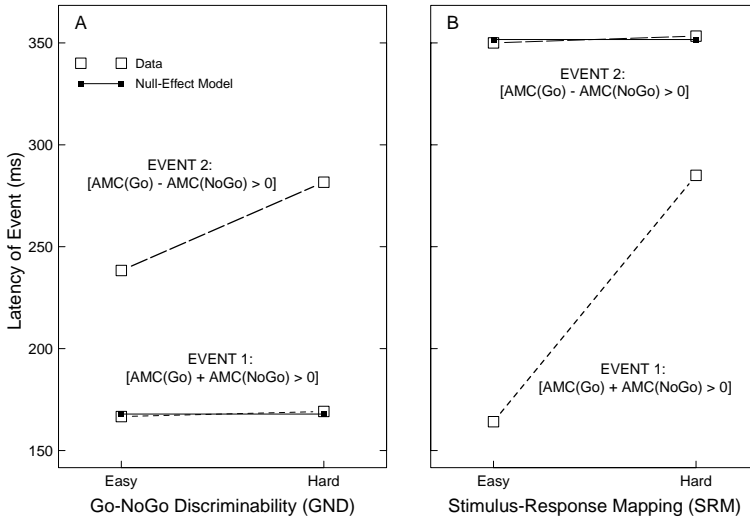


Fig. 4. Mean effects of two factors on the latencies of Events 1 and 2 in Osman et al. (1992). Panel A: Effects of *GND* (Exp. 1,  $N = 6$ ). Its effect on  $\bar{T}_1$  is  $2.5 \pm 5.0$  ms; its effect on  $\bar{T}_2$  is  $43 \pm 14$  ms; the difference between these effects is  $41 \pm 11$  ms ( $p \approx 0.01$ ). Panel B: Effects of *SRM* (Exp. 2,  $N = 6$ ). Its effect on  $\bar{T}_1$  is  $121 \pm 17$  ms; its effect on  $\bar{T}_2$  is  $3.3 \pm 8.8$  ms; the difference between these effects is  $129 \pm 27$  ms ( $p \approx 0.01$ ). Also shown are null-effect models for  $\bar{T}_1$  in panel A and  $\bar{T}_2$  in panel B. For the RT data, see Tables 1 and 5 in Osman et al. (1992).

Exp. 1, *GND* influenced  $\bar{T}_2$  (by 43 ms) but not  $\bar{T}_1$  (compare Figs. 3A1 and A2). The absence of any effect of *GND* on  $\bar{T}_1$  ( $\bar{T}_1 \approx 170$  ms on both Go and NoGo trials) shows that the position of the stimulus, but not its category, controlled the start of response preparation. More generally, because the stimulus influenced response preparation before both of its features were discriminated, these findings from Exp. 1 demonstrate the transmission of “partial information” from the perceptual process to the response process.

In one of the conditions of Exp. 2 the *SRM* was hard, which was expected to delay selection of the response. To ensure that stimulus position had an opportunity to influence response preparation on NoGo trials in both conditions, it was important to delay the Go–NoGo discrimination; Osman et al. therefore increased letter–digit shape similarity such that  $\bar{T}_2 \approx 350$  ms as shown in Fig. 4B. In this experiment, *SRM* influenced  $\bar{T}_1$  (by 121 ms) but not  $\bar{T}_2$  (compare Figs. 3B1 and B2). (Note that increasing the level of mapping difficulty from *SRM*<sub>1</sub> to *SRM*<sub>2</sub> therefore *reduced* the interval between Event 1 and Event 2.) *SRM* and *GND* thus influenced the two measures  $\bar{T}_1$  and  $\bar{T}_2$  selectively, supporting the hypothesis that they are pure measures of two different modular processes (Table 2).

The results also show how  $\alpha$  and  $\beta$  are arranged. Suppose they were arranged sequentially, at stages. Prolonging the first of two stages by  $\Delta t$  ms should delay completion of the second by the same amount: the prolongation  $\Delta t$  should be *propagated* to the completion time of the next stage. If we assume equal delays be-

tween completion of each process and its effect on  $A_{MC}(t)$ , then the order of process completions would be the same as the order of Events 1 and 2. The finding (Exp. 2) that the effect of *SRM* on  $T_1$  is *not* propagated to  $T_2$  would then be sufficient to invalidate a stage model. If we relax the equal-delays assumption, permitting us to assume the opposite order of process completions, then the propagation property requires that any effects on  $T_2$  propagate to  $T_1$ , contrary to what was found (Exp. 1) for the effect of *GND*. The alternative to a stages arrangement is that  $\alpha$  and  $\beta$  overlap in time and operate in parallel. Given a plausible linkage of  $\alpha$  and  $\beta$  with the behavioral response, the  $\overline{RT}$  (on Go trials) would then be determined by the completion time of the slower of the two. Such an arrangement is consistent with the further finding (from Exp. 2) that the effect of *SRM* on  $\overline{RT}$  (16 ms) is dramatically smaller than its effect on  $\overline{T}_1$  (121 ms).

### 6.1. Comments

As mentioned above, latencies such as those shown in Fig. 4 are determined from  $ERP(t)$  functions, which are averages over trials. It is highly likely that latencies of the corresponding brain events vary from trial to trial. Given the low thresholds used to define Events 1 and 2, quantities such as  $T_1$  and  $T_2$  would then reflect the left-hand tail, rather than the mean, of the latency distribution over trials. To the extent that the shape or spread of this distribution varies with factor level, this complicates the interpretation of changes in  $T_1$  and  $T_2$ , and of quantitative similarities and differences between effects of factors on  $\overline{RT}$  and their effects on latencies such as  $T_1$  and  $T_2$  (Meyer et al., 1988, Section 8).

At the time of the Osman et al. (1992) study, it had not been recognized that the LRP might result partly from lateralized activation of sensory neurons and from response tendencies “automatically” elicited by lateralized stimuli.<sup>24</sup> It seems unlikely, however, that either instance of selective influence in Figs. 4A and B (on which the inferences are based) could be the spurious consequence of such “direct” effects of stimulus lateralization, rather than reflecting the task-specified response. For example, the effect of *SRM* on  $T_1$  in Exp. 2 (Figs. 3B and 4B) shows that  $T_1$  is at least in part determined by the task-specified response. Hence, the invariance of  $T_1$  in Exp. 1 (Figs. 3A and 4A) is unlikely to be an artifact of stimulus position.

A contrasting example is outlined in Appendix A.6, in which pure measures based on the LRP provide evidence for a serial arrangement of two neural processes. In that case, unlike this one, a composite behavioral measure ( $\overline{RT}$ ) leads to a similar analysis of corresponding functional processes. In the present example, however,  $T_1$  indicates the time of a process (response selection) for which there may be no pure behavioral measure, and whose contribution to the composite  $\overline{RT}$  measure cannot be described by any simple combination rule.

<sup>24</sup> See Eimer (1998), but also Vallé-Inclan and Redondo (1998).



### 7. Separation of sensory and decision processes by signal-detection theory (Ex. 3)

The most influential approach to deriving pure measures of two processes underlying the performance of a task is the one associated with SDT, as described, for example, by Swets, Tanner, and Birdsall (1961).<sup>25</sup> At the heart of this approach is the recognition that even simple psychophysical tasks involve decision processes (that may be biased) as well as sensory mechanisms. Consider a psychophysical experiment in which two types of trial are randomly intermixed, each with a slightly different light intensity. On one type of trial the brighter light,  $S_T$  (the “target” stimulus) is presented; on the other type the dimmer light,  $S_{NT}$  (the non-target stimulus) is presented. The observer’s task is to respond with either  $R_T$  (“it was the target”) or  $R_{NT}$  (“it was the non-target”). On each trial, according to SDT, the observer forms a unidimensional internal representation of the stimulus; let us call these representations  $X_T$  and  $X_{NT}$ , for  $S_T$  and  $S_{NT}$ , respectively. Because  $S_T$  is brighter than  $S_{NT}$ ,  $X_T$  will tend to be larger than  $X_{NT}$ . It is also assumed, however, that because of external and internal noise,  $X_T$  and  $X_{NT}$  are random variables with distributions, rather than being fixed constants, and that because  $S_T$  and  $S_{NT}$  are similar, these distributions overlap. It is the overlap that creates the discrimination problem for the observer. In what follows, I shall assume that the distributions of  $X_T$  and  $X_{NT}$  are Gaussian with equal variances (supported in the experiment described below, but often false), and also that  $S_T$  and  $S_{NT}$  trials are equally frequent. The present treatment of SDT is simplified in other ways as well.

According to SDT, the value of  $X$  on a trial results from the operation of a sensory process **S**; a decision process **D** then uses this value to select one of the two responses, selecting  $R_T$  if  $X$  exceeds a criterion value, selecting  $R_{NT}$  otherwise. The choice of criterion value determines the direction and magnitude of response bias.

The data from such an experiment can be described by four numbers arranged in a  $2 \times 2$  data matrix as shown in Table 6, where the rows correspond to the two trial types  $S_T$  and  $S_{NT}$  and the columns correspond to the two responses  $R_T$  and  $R_{NT}$ . In the top row are the proportions of the target ( $S_T$ ) trials that elicited each response, which estimate  $Pr\{R_T|S_T\}$  (the true positive or “hit” probability), and  $Pr\{R_{NT}|S_T\}$  (the false negative or “miss” probability). In the bottom row are the proportions of the non-target ( $S_{NT}$ ) trials that elicited each response, which estimate  $Pr\{R_T|S_{NT}\}$  (the false positive or “false alarm” probability) and  $Pr\{R_{NT}|S_{NT}\}$  (the true negative or “correct rejection” probability). From such a table, two measures can be derived: One is  $d'$ , presumed to be a pure measure of the sensory process **S**, and proportional

<sup>25</sup> See also Green and Swets (1966), especially Chapter 4 on “Basic Experiments: Separation of Sensory and Decision Processes”; Macmillan and Creelman (1991), especially Chapter 2 on “The yes–no experiment: Response bias”; and Swets (1998) on “Separating Discrimination and Decision in Detection, Recognition, and Matters of Life and Death”.

Table 6  
 Yes–No signal-detection experiment: data matrix and three payoff matrices, PM1, PM2, and PM3

		Data Matrix		PM1		PM2		PM3	
		$R_T$	$R_{NT}$	$R_T$	$R_{NT}$	$R_T$	$R_{NT}$	$R_T$	$R_{NT}$
$S_T$	$Pr\{R_T S_T\}$ (True Positive)	$Pr\{R_{NT} S_T\}$ (False Negative)	1	0	1	0	$\pi_T$	0	
	$Pr\{R_T S_{NT}\}$ (False Positive)	$Pr\{R_{NT} S_{NT}\}$ (True Negative)	0	1	-3	1	0	$\pi_{NT}$	

to  $\bar{X}_T - \bar{X}_{NT}$ , which increases with discriminability. The other is a measure of response bias or criterion, presumed to be a pure measure of the decision process **D**.

Many factors have been used in attempts to influence **S** and **D**, some expected to influence just a sensory process (sensory factors or “s-factors”), and some expected to influence just a decision process (decision factors or “d-factors”). Stimulus features such as the luminance difference between  $S_T$  and  $S_{NT}$  are examples of s-factors used to influence the measure  $d' = M_S$ . In studies with human observers,  $2 \times 2$  payoff matrices (PMs), containing positive or negative values associated with the four possible outcomes on a trial, have been used as a factor *PM* to influence the response bias associated with **D**. Examples are shown by payoff matrices PM1 and PM2 in Table 6, where each cell contains the number of units of money earned for the corresponding S-R pair. PM1 provides symmetric rewards, while PM2 induces a “conservative” criterion – a bias favoring  $R_{NT}$ . In analogous studies with animals, partial reinforcement is often used: a food reward is provided for only a fraction of the correct responses. The entries in PM3 are therefore probabilities less than 1.0; one way to try to induce a conservative criterion would be to set  $\pi_{NT} > \pi_T > 0$ .

In terms of the inferential logic of Table 2, the SDT approach has met with only partial success:  $d'$  has been found to be sensitive to s-factors (such as the luminance difference mentioned above), and invariant with respect to d-factors (such as the payoff-matrix), which argues for  $d' = M_S$  being a pure measure of **S**. From the viewpoint of a psychophysicist investigating **S**, the SDT approach can thus be extraordinarily helpful. However, despite attempts to find a pure measure  $M_D$  of the decision process – in particular, a measure that reflects response bias – none has been found that is invariant with respect to s-factors (as well as being sensitive to d-factors) in experiments with either humans (Dusoir, 1975, 1983), or with animals, using payoff matrices such as PM3 in Table 6 (Alsop, 1998). The SDT approach thus seems to provide only a “partial modularity” (footnote 12) in which property  $p_2$  ( $M_D \leftrightarrow$  s-factors, Table 2) does not obtain.

### 7.1. Why has SDT failed in this respect?

One possibility is that a suitable measure  $M_D$  has yet to be found. It should be pointed out that **D** is hypothesized to operate on information (the operand) provided by **S** (Section 2.1). Whereas the operator may well not be affected by the operand, **D** must be influenced in *some* sense by the operand, and hence by s-factors as well as d-factors. This might make it especially difficult to find a measure that reflects only that aspect of **D** (the operator) that is invariant over s-factors, even if there is such an aspect. (Recall from Section 2.1 that different measures of a process may reflect different aspects of it, and that a claim of invariance of a process may be limited to just the aspect of that process reflected by the particular measure.)

A second possibility is that it may be difficult to find s-factors with respect to which to demonstrate the desired invariance of a measure of response bias. For example, if the level of discriminability is random from trial to trial (several intensities of  $S_T$ , one of  $S_{NT}$ ), it is not obvious how to define a separate measure of response bias  $M_D$  for each  $S_T$  intensity that it could permit testing such invariance. On the other hand, if the level of discriminability is fixed over a series of trials or cued on each trial, as in Dusoir's (1983) experiments, then *expected discriminability* or expected noise level or signal level (which might well be a d-factor) is confounded with *discriminability* (an s-factor).

A third possibility is that suitable d-factors have not been used. For example, when the relative frequencies of  $S_T$  and  $S_{NT}$  have been manipulated instead of being kept equal (a d-factor), it has been found (Markowitz & Swets, 1967) that  $d'$  is affected, presumably because such variation also affects "knowledge of the signal", an s-factor. The possibility considered by Alsop (1998), and by others whose work he reviews, is that the problem may be a failure of the assumption that *PM* influences **D** selectively and controls response bias. It is true that if a subject operates so as to maximize the expected value of a trial, and learns the  $X_T$  and  $X_{NT}$  distributions, then the *optimal* criterion can be computed from them and the payoff matrix. But humans have been found to select criteria that are far from optimal (e.g., Green & Swets, 1966, Chapter 4), and a few animal studies suggest that response bias is controlled, not by the conditional probabilities described by the payoff matrix, but by the obtained distribution of rewards over the two alternative responses, sometimes called the *reinforcer ratio* (*RR*).

One way to characterize *RR* is as  $Pr\{R_{NT}|Reward\}$ , the proportion of the total number of rewarded responses that are associated with  $R_{NT}$ . Given PM3 of Table 6, and assuming equal frequencies of  $S_T$  and  $S_{NT}$ ,  $Pr\{R_{NT}|Reward\} = r/(1+r)$ , where  $r$ , the expected ratio of rewarded  $R_{NT}$  responses to rewarded  $R_T$  responses, is given by  $[\pi_{NT}Pr\{R_{NT}|S_{NT}\}]/[\pi_TPr\{R_T|S_T\}]$ . In addition to depending on the values of  $\pi_T$  and  $\pi_{NT}$  specified by the payoff-matrix, the ratio  $r$  and hence *RR* also depend on discriminability and the current response bias. Controlling or manipulating the payoff matrix thus provides only partial control of a factor (*RR*) that may influence **D** selectively, but whose level is also affected by s-factors. Even when only **S** is of interest and only s-factors are varied, the failure of a constant payoff-matrix to insure a

constant response bias could be important, if the interpretation of their effects requires that s-factors do not also influence **D**. Fortunately it is possible to control *RR* itself instead of the payoff matrix.

### 7.2. An approximation to full modularity of sensory and decision processes when *RR* is controlled

In a luminance-discrimination experiment with six pigeons, McCarthy and Davison (1984) used a linked concurrent pair of variable-interval (VI) schedules to control *RR*. On each trial in a series, one of two light intensities appeared on the center key of three keys; the two trial types were equally frequent. The correct response was to peck the left key ( $R_T$ ) if the center key was “bright”, and to peck the right key ( $R_{NT}$ ) if it was “dim”. Correct responses were reinforced with food, with a mean probability of about 0.37, controlled by the VI schedules. Two factors were varied orthogonally: The *luminance ratio* (*LR*) of the two lights was varied by letting the dimmer luminance be one of five values, including, for the lowest *LR* level, a value of equal to the brighter luminance. And *RR* could be one of three values, 0.2, 0.5, or 0.8.<sup>26</sup> There were thus  $5 \times 3 = 15$  conditions. For each bird, each condition was tested for a series of consecutive daily sessions until a stability requirement was satisfied; the data analyzed came from the last seven sessions in each condition (about 1060 trials per condition per bird).

For each condition and each bird, the data can be represented in a data matrix like that in Table 6, and can be summarized by two proportions,  $Pr\{R_T|S_T\}$  and  $Pr\{R_T|S_{NT}\}$ . If the distributions of  $X_T$  and  $X_{NT}$  (see above) are Gaussian with equal variances, and  $z(\cdot)$  is the  $z$ -transform of a proportion (the inverse Gaussian distribution function), then the (“ROC”) curve traced out when  $z(Pr\{R_T|S_T\})$  is plotted against  $z(Pr\{R_T|S_{NT}\})$ , as *RR* is changed from 0.2 to 0.5 to 0.8, is expected to be linear with slope 1.0. Examination of the set of such curves (six birds  $\times$  five levels of *LR*) supports this expectation, and hence the equal-variance Gaussian model.<sup>27</sup>

Given such support for the model, suitable estimators for the discriminability and criterion measures for each condition are  $\hat{d}' = z(Pr\{R_T|S_T\}) - z(Pr\{R_T|S_{NT}\})$  and  $\hat{c} = [z(Pr\{R_T|S_T\}) + z(Pr\{R_T|S_{NT}\})]/2$ . (Just as in Ex. 2, the two measures of interest are derived from the basic data by determining the sum and difference of two initial measures.) Note that the origin for the criterion measure is the midpoint between  $\bar{X}_{NT}$  and  $\bar{X}_T$ ; the sign of the criterion thus reflects the direction of the bias. Means over birds of these two measures are shown in Fig. 5.

The left side of the figure (and associated ANOVAs) shows that while the criterion responds strongly to factor *RR* (panel A1), the often-demonstrated invariance of  $d'$  with respect to  $d$ -factors is also persuasive here (panel A2): there is neither a main effect of *RR* on  $d'$ , nor is there any modulation by *RR* of the effect of *LR*. The in-

<sup>26</sup> If  $RR = Pr\{R_{NT}|Reward\} = 0.2$ , for example, the ratio  $r$  defined above is 1/4; for each rewarded  $R_{NT}$  response there are four rewarded  $R_T$  responses, encouraging a liberal (low) criterion for  $R_T$ .

<sup>27</sup> See Appendix A.7.1 for details.

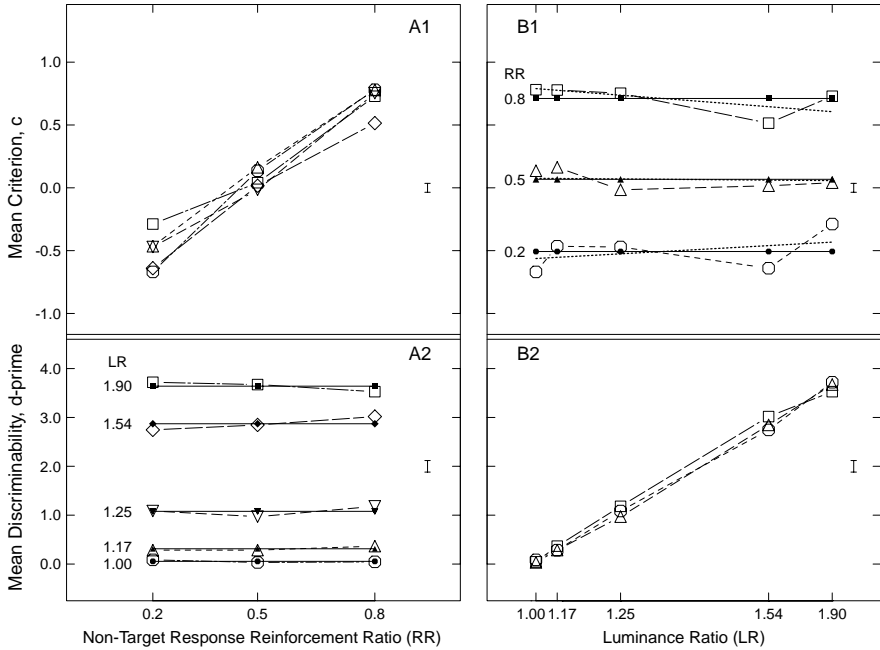


Fig. 5. Mean effects of  $RR = Pr\{R_{NT}|Reward\}$  (panels A1, A2) and  $LR$  (panels B1, B2) on criterion  $\hat{c}$  (panels A1, B1) and on discriminability  $\hat{d}'$  (panels A2, B2) are shown by unfilled points and light lines.  $RR$  and  $LR$  levels have been scaled so as to linearize their mean effects on  $\hat{c}$ -value and  $\hat{d}'$ -value (panels A1 and B2), respectively. Filled points and heavy solid lines in panels A2 and B1 represent fitted models in which  $\hat{d}'$  and  $\hat{c}$  are invariant with respect to factors  $RR$  and  $LR$ , respectively. The dotted lines in panel B1 represent a fitted model with a multiplicative interaction of the two factors. The  $\pm$ S.E. error bars reflect estimates of the variability of each plotted point after removing mean differences among birds. Plotting symbols correspond from top panels to bottom, but not from left to right; the plotted  $y$ -values are the same from left to right. From McCarthy and Davison (1984); basic data kindly provided by Alsop.

variant model fits well. We thus have evidence for the hypothesis that while  $RR$  is potent, as shown by its influence on the criterion  $c$  (hence on  $\mathbf{D}$ ), it leaves invariant our measure  $\hat{d}'$  of discriminability (and hence of  $\mathbf{S}$ ). These statements are based on standard ANOVA tests in which  $LR$  and  $RR$  are treated as factors (categorical variables). Such tests, being global rather than focused, would reveal any violation of invariance, whether erratic or orderly, but not be especially sensitive to any.

Because we are especially interested in *systematic* modulation of the effect of a factor  $F$  by a factor  $G$ , tests that focus on systematic effects rather than being global (and may thus be more sensitive to such effects) are also useful. An example of a systematic modulation is one that has a pattern that is qualitatively the same for each level of a factor  $F$ , but whose magnitude changes monotonically with the ordered levels of  $G$  (a “monotone interaction”; Section 15.1), or linearly with a numeric variable  $\mathcal{G}$  that represents the scaled levels of  $G$  (a multiplicative interaction; Section 15.1). (Roughly, if a small change in  $\mathcal{G}$  alters the effect of  $F$  a little, then a greater change should alter it more.) To permit such focused tests, each level of each factor

was given a numerical “potency value” equal to the corresponding mean (over levels of the other factor) of the measure most sensitive to it; for each factor the  $x$ -axis values in Fig. 5 have been plotted on such a scale. (This is why the means of the functions in panels A1 and B2 are linear.) Let “ $\mathcal{RR}$ ” and “ $\mathcal{LR}$ ” represent these numerically scaled versions of the two factors. No significant violations of invariance of  $d'$  with respect to  $RR$  were found from any of several such focused tests, including the multiplicative interaction of  $\mathcal{LR}$  and  $\mathcal{RR}$ .<sup>28</sup>

Effects of  $LR$  are shown on the right side of the figure. Panel B2 shows the orderly effect of  $LR$  on  $d'$ . Panel B1 shows that to a good approximation the criterion is uninfluenced by  $LR$ , while it causes discriminability to range widely, from  $d' \approx 0$  to  $d' \approx 3.6$ . In this case, however, a global test shows that the small  $LR \times RR$  interaction is significant:  $F(8, 40) = 2.25; p = 0.04$ .<sup>29</sup> Results of fitting a multiplicative interaction of  $\mathcal{LR}$  and  $\mathcal{RR}$  are shown by the dotted lines; criteria tend to be less extreme at higher levels of discriminability. One advantage of representing failures of invariance with such multiplicative interactions is that they have only one degree of freedom, and thus permit simple expression of their size and variability over subjects. Here we find the coefficient of the interaction term  $\mathcal{LR} \times \mathcal{RR}$  to be  $-0.07 \pm 0.03$ , which is not significant. In appreciating the magnitude of the interaction coefficient, it is useful to know that if the dotted lines maintained their average separations but converged fully at  $LR = 1.90$ , the coefficient would be  $-0.48$ , about seven times as large as the value observed.<sup>30</sup>

Overall, this study along with others (see Alsop, 1998) seems promising in its suggestion that by using something other than the payoff matrix as a  $d$ -factor in signal-detection experiments, the separate modifiability of **S** and **D** can be demonstrated. While these data are from pigeons, a recent study with humans (Johnstone & Alsop, 2000) also suggests that when  $RR$  is controlled, the criterion  $c$  is approximately invariant over changes in discriminability.

### 7.3. Comments

This example illustrates how important is the choice of factors. When the measures are  $M_S = d'$  and  $M_D = c$ , and the factors are the  $s$ -factor  $LR$  and the  $d$ -factor  $PM$ , the modularity of **S** and **D** is hidden:  $M_D$  depends on  $LR$  as well as  $PM$ . In contrast, if the factors are  $LR$  and  $RR$ , the pigeon data suggest that to a good ap-

<sup>28</sup> See Section 15 for a similar approach to testing additivity of factor effects on a composite measure, rather than invariance of an effect on a pure measure.

<sup>29</sup> When data for the highest  $LR$  (for which the ROCs were unusually irregular) are removed, however, the interaction becomes non-significant:  $F(6, 30) = 1.48; p = 0.22$ .

<sup>30</sup> Unfortunately, in a similar study (Alsop & Davison, 1991) with more extreme  $RR$  values (0.1, 0.5, 0.9), the multiplicative interaction is somewhat larger, in the same direction, and highly significant ( $-0.12 \pm 0.01, p = 0.0003$ ), which perhaps argues against ignoring the multiplicative interaction in the present experiment. Differences between these studies include the ordering of conditions and the occurrence in the McCarthy and Davison (1984) experiment of different outcomes for correct vs incorrect responses on non-reward trials.

proximation they influence  $M_S$  and  $M_D$  selectively, thus demonstrating the separate modifiability of **S** and **D**. One way to think of this is that the dependence of  $M_D$  on the two factors  $LR$  and  $PM$  occurs because they both affect  $RR$ ; if we control  $RR$  rather than  $PM$ , we find  $M_D$  to be invariant with respect to  $LR$ , just as  $M_S$  is invariant with respect to d-factors.

Also illustrated by this example is focused testing for a systematic deviation from invariance, the importance of factorial experiments and factors with more than two levels for doing so, and the helpfulness of numerical scaling of factor levels for permitting the systematic deviation to be easily expressed and evaluated.

Finally, this example permits me to emphasize that the purity of a measure is a hypothesis to be tested ( $H2$ , Table 2), by considering an alternative to  $M_D = c$  as a measure of **D**: the likelihood ratio  $M'_D = \beta$  at the criterion, well known in the context of SDT.<sup>31</sup> In Appendix A.7.2 I show that unlike  $c$ ,  $\beta$  is drastically influenced by  $LR$  in the present experiment, hence is not a pure measure of **D**.

## 8. Introduction to three examples of inference based on direct pure measures

In each of the examples of Sections 9–11, hypothesized pure measures of two processes are based on different data sets.

**Example 4** (*Evidence for modular band-limited spatial-frequency analyzers in the visual system*). Demonstration of the selective effects of sensory adaptation is a classic use of direct pure measures, and one in which the issue of modularity of perceptual processes has been important theoretically. Here, the hypothesized pure measures are detection thresholds for simple gratings of different frequencies. However, it may be more appropriate to regard this as an example of task comparison rather than process decomposition.

**Example 5** (*Partial evidence from single-cell recordings for temporally delimited neural modules (stages)*). Activation latencies of cortical neurons that are linked to sensory and motor events, respectively, provide pure measures of the time from stimulus to sensory-neuron activation, and the time from motor-neuron activation to overt response, candidates for the durations of modular processes  $\alpha$  and  $\gamma$ , respectively. This example raises questions about how to demonstrate that brain measures reflect events that are essential in performing a task, and (because a process  $\beta$  intervenes between  $\alpha$  and  $\gamma$ ) about what evidence is needed to demonstrate modularity of three processes rather than two.

**Example 6** (*Evidence from fMRI brain measurements for modular processes implemented by localized neural processors*). Because of its relatively good spatial resolution, fMRI provides a promising measure for the decomposition of neural processes.

<sup>31</sup> See, e.g., Macmillan and Creelman (1991, Chapter 2); Swets (1998).

As I could not find a suitable study, the data are hypothetical. The discussion contrasts the task-comparison method (in which fMRI or PET measurements have often been employed; Appendix A.1) with the method of process decomposition promoted here, outlines some of the assumptions required for both methods, and argues for the superiority of the latter for the analysis of mental processes. I suggest that inferences are considerably strengthened by finding behavioral measures that show the same pattern of selective influence as the brain measures, and again mention the advantages of factorial experiments and factors with more than two levels.

### 9. Evidence for modular spatial-frequency analyzers from selective adaptation (Ex. 4)

In the 1960s a revolution occurred in the way in which vision scientists thought about pattern perception. The visual system came to be regarded as performing something close to a spatial Fourier analysis of a pattern, that is, an analysis into spatially sinusoidal components in different overlapping spatial-frequency bands (Campbell & Robson, 1968). One discovery that favored this view was Blakemore and Campbell's (1969) finding of frequency-selective adaptation effects with sinusoidal gratings, "perhaps the single most convincing psychophysical demonstration that the visual system contains multiple channels, each of which operates more or less as a band-pass spatial frequency filter" (De Valois & De Valois, 1988, p. 183).

The data in Fig. 6, collected for a different purpose, are derived from contrast-threshold measurements of low- and high-frequency gratings for one of the subjects (DS) in Graham's (1970) Ph.D. thesis (see also Graham, 1972).<sup>32</sup> For these data, the joint hypothesis (Table 2) asserts (*H1*) that the analyzers **L** and **H** are modules that separately determine the contrast thresholds for low (1.3 cycles/deg) and high (7.5 cycles/deg) spatial frequencies, respectively, and (*H2*) that the contrast thresholds for sinusoidal gratings at these two frequencies are pure measures  $M_L$  and  $M_H$  of their respective modules.<sup>33</sup>

The factors, each with two levels, are *low-frequency adaptation* (*LA*) (with a grating of 1.0, 1.3, or 1.7 cycles/deg), at either  $LA_1 = 0\%$  or  $LA_2 = 80\%$  contrast, and *high-frequency adaptation* (*HA*) (with a grating of 5.5, 7.5, or 10.0 cycles/deg), at either  $HA_1 = 0\%$  or  $HA_2 = 80\%$  contrast. The adapting stimuli  $\{(LA_j, HA_k)\}$  were homogeneous (blank), (0%, 0%); low-frequency, (80%, 0%); or high-frequency, (0%, 80%). The goals of Graham's experiment did not require the fourth condition

<sup>32</sup> Graham's primary interest was in the shapes of the functions that relate threshold elevation to test frequency, after adaptation, for each of a set of adapting frequencies.

<sup>33</sup> The contrast of a sinusoidal grating is half the difference between the peak and trough luminances divided by their mean, a number between zero and one, often described in percent. The contrast threshold is the contrast at which the presence of a grating is just detectable. Whereas typical adaptation contrasts range from 20% to values close to 100%, threshold contrasts tend to be 1% or less.



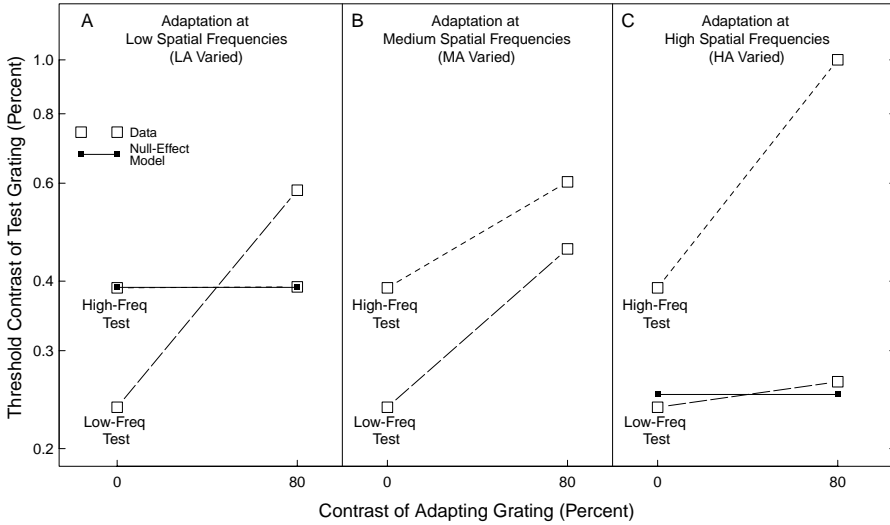


Fig. 6. Data from Graham (1972, Fig. 5). As a function of adaptation contrast, each panel shows threshold contrasts (on a logarithmic scale) for low-frequency (1.3 cycles/deg) and high-frequency (7.5 cycles/deg) test gratings. What differs across panels is the spatial-frequency range of the adapting gratings: In cycles/deg they are *low* (1.0, 1.3, 1.7) in panel A (factor *LA*); *medium* (2.25, 3.0, 4.0) in panel B (factor *MA*); and *high* (5.5, 7.5, 10.0) in panel C (factor *HA*). Each factor has two levels, 0% and 80% contrast. The values of threshold contrast for 0% adapting contrast (left-hand points) in the three panels represent the same data. In units of  $100 \times \log(\text{threshold})$ , mean threshold elevations in panels A, B, and C are, respectively, 39.0, 28.5, and 4.6 units for the low-frequency test ( $M_L$ ) and 0.2, 19.1, and 41.0 units for the high-frequency test ( $M_H$ ). The estimated S.E. for these mean elevations is 4.1 units and is based on 12 *df*; the elevations are therefore highly significant ( $p < 0.001$ ) except for the high-frequency test with low-frequency adaptation ( $0.2 \pm 4.1$  units) and the low-frequency test with high-frequency adaptation ( $4.6 \pm 4.1$  units). Also shown are null-effect models for the effects of *LA* on  $M_H$  and of *HA* on  $M_L$ .

of a complete factorial design,  $(LA_2, HA_2) = (80\%, 80\%)$ : adaptation with a compound grating consisting of both low and high frequencies.<sup>34</sup> To increase sample size and provide a measure of variability I selected the data to include adapting frequencies near (as well as at) the center frequencies of the hypothesized modules. The plotted values  $M_L$  and  $M_H$  are post-adaptation threshold contrasts of the two test gratings obtained by the method of adjustment, averaged over the three adapting frequencies in each set.

These conditions produced the data shown in Figs. 6A and C. Fig. 6A shows that changing *LA* from  $LA_1$  to  $LA_2$  increases  $M_L$  but not  $M_H$ ; Fig. 6C shows that changing *HA* from  $HA_1$  to  $HA_2$  does the opposite. Thus *LA* and *HA* influence  $M_L$  and  $M_H$  selectively, supporting *H1* and *H2* of Table 2.

Fig. 6B shows that a third factor, *medium-frequency adaptation* (*MA*) (with a grating of 2.25, 3.0, or 4.0 cycles/deg), at either  $MA_1 = 0\%$  or  $MA_2 = 80\%$  contrast,

<sup>34</sup> See Appendix A.9.1 for the likely results of such a factorial experiment.

leaves neither  $M_L$  and  $M_H$  invariant; this is interpreted to indicate that the bandwidths of analyzer modules **L** and **H**, centered at 1.3 and 7.5 cycles/deg, respectively, are both large enough to include some or all of these medium frequencies. Note that finding a factor,  $MA$ , that influences both  $M_L$  and  $M_H$  does not detract from the demonstration of separate modifiability provided by finding two factors that influence those measures selectively.

Especially when taken together with related studies, these data argue persuasively for the existence of band-limited spatial-frequency analyzers that control contrast thresholds within their bands, and whose adaptive states are separately modifiable by high-contrast simple gratings.

### 9.1. Comments

In this investigation of separate modifiability, the joint hypothesis (Table 2) includes  $H1$ : Analyzers **L** and **H** are modules (separately modifiable) and  $H2$ : Contrast thresholds  $M_L$  and  $M_H$  for low and high spatial frequencies are pure measures of the hypothesized modules. The joint hypothesis is supported by the discovery that factors  $LA$  (when  $HA = HA_1$ ) and  $HA$  (when  $LA = LA_1$ ) influence the two measures selectively. Because  $M_L$  and  $M_H$  are based on different data, they are *direct*-pure rather than *derived*-pure measures.

A different approach to the modularity of spatial-frequency analyzers is taken in Ex. 7 from Sachs, Nachmias, and Robson (1971), an experiment in which stimuli in the conditions of primary interest were compound gratings containing two frequencies best detected by two distinct analyzers. The procedure and results of the latter experiment are consistent with the idea that on every trial, both analyzers participated in the complex process that the subject used to perform the task.

In the present study, however, where thresholds were obtained by the adjustment method and the test frequency was known to the subject (who might therefore be able to adjust the detection process, dependent on frequency), it is hard to argue that  $M_L$  and  $M_H$  are measures within the same task. Strictly, we should regard this as an example of the task-comparison method (Section 1.6; Appendix A.1), with the two tasks being detection of the low- and high-frequency test gratings. Suppose, however, that  $M_L$  and  $M_H$  were obtained in a “concurrent task”, where the stimuli on different trials could contain either frequency, both, or neither, and where on each trial the subject made separate judgements about the presence of each test grating. Under such frequency-uncertainty conditions, both analyzers would participate on every trial, and the claim that a single complex process was under analysis would have greater merit. Performance in such a concurrent task appears to be no worse than under corresponding fixed-frequency conditions (Graham, Kramer, & Haber, 1985), suggesting that the pattern of adaptation effects would be similar to that in Fig. 6.

In Appendix A.9 we consider the advantages of expanding a study like the present one into a factorial design with multiple-level factors.

## 10. Evidence from individual neurons for temporally delimited (serial) neural modules (Ex. 5)

In experiments in which RT of human subjects is measured and where it is possible that component processes are arranged in *stages* (Sections 3.1.1, 16; Appendix A.16.2), we usually have no pure measures of their durations. Instead, the RT has to be treated as a composite measure that reflects the entire process. Suppose that the end of a functional processing stage corresponds to a neural process that ends with the (asynchronous) “activation” of an appropriate population of neurons (Parker & Newsome, 1998). When sub-human primates are subjects in such experiments, measurements of the activity of single neurons then provide estimates of time points that may inform us about these boundaries between one such stage and the next. Of course, such time points are useful for identifying (modular) stages only if the stages so defined can be shown to be separately modifiable. A pioneering but partial example is provided by the work of Mouret and Hasbroucq (2000).

*Task essentiality of neural activity.* Whether a brain measure reflects events that are essential to the task being performed is often problematic (Section 1.7). One virtue of Mouret and Hasbroucq’s approach is that the signal-to-noise ratio is high enough so that each trial provides an interpretable datum. For a neuron to be of interest, they can therefore require that it become active during the interval between stimulus and response on every successful trial of a particular type, and that its activation can be shown to be either contingent on or temporally coupled to a particular trial event (Appendix A.10.1).<sup>35</sup> Meeting of such criteria cannot guarantee the task essentiality of a neuron’s activity, but it does constitute strong supporting evidence.

The subject was a rhesus monkey performing in a two-choice RT experiment. His thumb (*t*) rested on one surface, his four fingers (*f*) on another. The stimulus was vibration of one of the surfaces; the response was an isometric increase in the force he applied to one of the surfaces. The factor of principal interest was *SRM*: In high-compatibility trial blocks ( $SRM = SRM_1 = \textit{easy}$ ), the digit(s) stimulated on a trial also responded; the two trial types can be designated *t-T* and *f-F*. In low-compatibility trial blocks ( $SRM = SRM_2 = \textit{hard}$ ), the mapping was reversed; the trial types were *t-F* and *f-T*. An extracellular electrode was used to record the spikes (action potentials) of individual neurons in primary somatosensory cortex (S1) and primary motor cortex (M1). During each of many sets of trials on which RT was measured, the activity at a particular electrode placement (probably arising from a single neuron) was also measured; the authors estimated the occurrence time (an *activation time*,  $T_a$ ) of an increase in its spike rate (an *activation*). Neurons were selected that regularly showed activations between stimulus ( $T_s = 0$ ) and response ( $T_r = RT$ ). Because about 25% of these neurons systematically showed two such bursts of

<sup>35</sup> Such criteria cannot easily be imposed if, because the *S/N* ratio is low, a method requires averaging over many trials, as in typical uses of ERPs (Sections 6, 14; Appendix A.6) or fMRI measurements (Section 11; Appendices A1.2.4, A.11).

activity, the neuron-activation (rather than the neuron) became the unit of analysis, whose occurrence time on each of many trials was measured. For simplicity, I write as if such activation pairs for an electrode placement were generated by two different cells, and call each unit a “cell\*”. Based on selection criteria described in Appendix A.10.1, cells\* fell into four classes, two sensory ( $SC_2$  and  $SC_4$ ) and two motor ( $MC_2$  and  $MC_4$ ); data from cells\* that satisfied none of the four sets of selection criteria were discarded.

### 10.1. Estimation of stage durations

Fig. 7 shows the way in which three hypothesized neural processing stages were defined, and how stage durations were estimated by combining mean activation times with mean occurrence times of stimulus and response. Averaging over trial types, the data in Fig. 8 show the estimated mean durations of  $\alpha$ ,  $\beta$ , and  $\gamma$  to be  $\bar{D}_\alpha = 39$ ,  $\bar{D}_\beta = 142$ , and  $\bar{D}_\gamma = 63$  ms, respectively. The mean effect of *SRM* on  $\bar{RT}$  is 33 ms. Which stage or stages are responsible for this effect? Panel A of Fig. 8 shows that whereas *SRM* influenced  $\bar{RT}$  for those trials on which sensory cells\* were measured, it had no effect on mean latency of the neuron response, hence none on  $D_\alpha$ . Panel B shows that for trials on which motor cells\* were measured, the mean

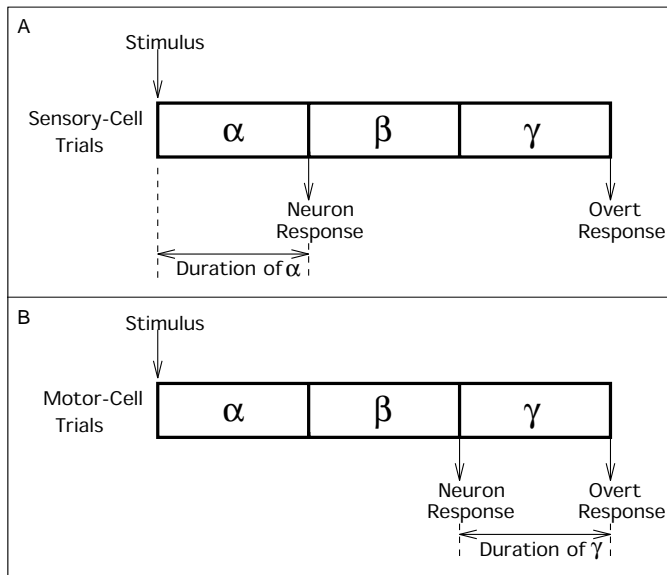


Fig. 7. Three hypothesized processing stages and their definitions in terms of the times of events on trials on which sensory cells\* were measured (panel A) and on which motor cells\* were measured (panel B), in the experiment by Mouret and Hasbroucq (2000). Stage  $\alpha$  with mean duration  $\bar{D}_\alpha$  is the process that starts with the stimulus and ends with the activation of sensory cells\*. Stage  $\gamma$  with mean duration  $\bar{D}_\gamma$  is the process that starts with the activation of motor cells\* and ends with the overt response. The mean duration of stage  $\beta$ , which starts with the activation of sensory cells\* and ends with the activation of motor cells\*, is estimated by  $\bar{D}_\beta = \bar{RT} - (\bar{D}_\alpha + \bar{D}_\gamma)$ .

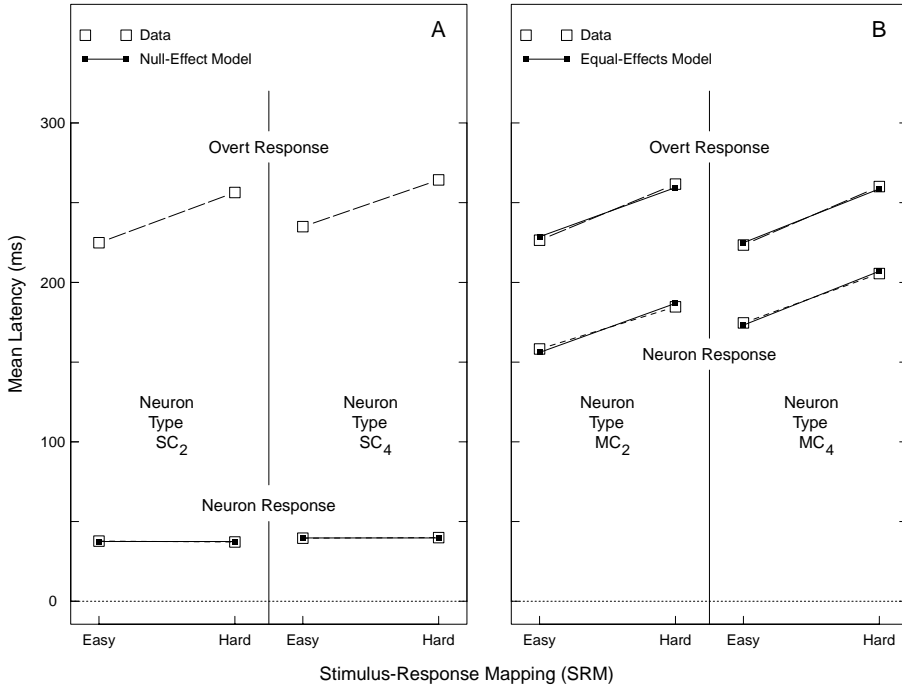


Fig. 8. Mean latencies of neuron activations and overt responses on four sets of trials in the experiment by Mouret and Hasbroucq (2000) with fitted selective-influence models. In each set of trials a group of cells\* of one type was each measured many times. For each type, mean latencies are shown as functions of the level of *SRM*, with levels  $SRM_1 = \text{easy}$  (spatially compatible) and  $SRM_2 = \text{hard}$  (spatially incompatible). Panel A shows the results for trials on which stimulus-related (sensory) cells\* were measured, type SC<sub>2</sub> on the left, type SC<sub>4</sub> on the right. For trials on which SC<sub>2</sub> cells\* were measured the *SRM* effects are  $-0.5 \pm 1.7$  (neuron response) and  $32 \pm 10$  ms (overt response). For trials on which SC<sub>4</sub> cells\* were measured the *SRM* effects are  $0.3 \pm 0.8$  (neuron response) and  $29 \pm 4$  ms (overt response). Also shown for each cell\* type is a null *SRM*-effect model for the neuron activation latency. Panel B shows the results for trials on which response-related (motor) cells\* were measured, type MC<sub>2</sub> on the left, type MC<sub>4</sub> on the right. For trials on which MC<sub>2</sub> cells\* were measured the *SRM* effects are  $27 \pm 8$  (neuron response) and  $35 \pm 8$  ms (overt response); for trials on which MC<sub>4</sub> cells\* were measured the *SRM* effects are  $31 \pm 8$  (neuron response) and  $37 \pm 7$  ms (overt response). Also shown for each cell\* type is a fitted equal *SRM*-effects model (latencies of neuron vs overt response). For MC<sub>2</sub> cells\* the difference between the effects (measuring badness of fit of the model) is  $9 \pm 11$  ms; for MC<sub>4</sub> cells\* the difference is  $6 \pm 7$  ms.

latencies of neuron and overt responses were influenced to about the same extent by *SRM*, which means that  $\overline{D}_\gamma$  is (approximately) invariant with respect to *SRM*. (Note, however, that for both MC<sub>2</sub> and MC<sub>4</sub> cells\*, mean activation latency shows a non-significantly smaller effect of *SRM* than does  $\overline{RT}$ ; the average reduction is 20%. Even this experiment – heroic for both subject and experimenters – provides less precision than would be desirable.) Idealized, these two findings mean that neither  $D_\alpha$  nor  $D_\gamma$  were influenced by *SRM*. It follows that the *SRM* effect was due entirely to a change in  $D_\beta$ , which supports the idea that  $\beta$  can be modified without also modifying  $\alpha$  or  $\gamma$ . However, this support is not as strong as it might be.

### 10.2. Relation of these findings to the demonstration of separate modifiability

To handle three modules – here organized as stages – the argument of Table 2 must be expanded. For example, one might search for three factors,  $F$ ,  $G$ , and  $H$ , each selectively influencing a different stage, which would require testing for nine properties like the four in Table 2. Of those nine properties, these results confirm only three:  $M_\alpha \leftarrow G$ ,  $M_\beta \leftarrow G$ , and  $M_\gamma \leftarrow G$ , where  $G$  is  $SRM$ , and  $M_\alpha$ ,  $M_\beta$ , and  $M_\gamma$  are the three duration estimates  $D_\alpha$ ,  $D_\beta$ , and  $D_\gamma$ , respectively. Hence, while Mouret and Hasbroucq have provided an important and novel demonstration of selective influence in a neural process, the argument for modularity could be considerably strengthened (Section 2.3): we know neither that  $M_\alpha$  and  $M_\gamma$  are sensitive nor that they can be modified separately from  $M_\beta$  or from each other. For example, the present findings leave open the following possibilities for  $\alpha$ :

- (1)  $\alpha$  is influenced by no task factors – it is non-modifiable.
- (2)  $\alpha \leftarrow SRM$ , but this is not reflected in  $M_\alpha$ .
- (3)  $\alpha \leftarrow SRM$ .

If  $M_\alpha$  were now shown to be sensitive (responsive to some other factor), this would falsify (1), reduce the plausibility of (2) (which would now require that  $M_\alpha$  reflect some functional changes in  $\alpha$  but not others), and hence strengthen (3).

### 10.3. Comments

In relation to the goal of identifying modules, the major limitation of the present example is the absence of one or more factors with systematic effects that influence  $\alpha$  and/or  $\gamma$  selectively.<sup>36</sup> Without such information the demonstration of modularity is incomplete (Section 2.3) even for two processes, and certainly for three, where nine properties must be established rather than only the four listed in Table 2.

Idealizing the data, we have learned that of the 244 ms in the  $\overline{RT} = \overline{D}_\alpha + \overline{D}_\beta + \overline{D}_\gamma$ , a substantial part ( $\overline{D}_\alpha + \overline{D}_\gamma = 102$  ms, or 42%) is invariant with respect to  $SRM$ , evidence for the selective influence of  $SRM$  on an epoch within the RT, and hence for a stage model. The strength of this evidence would be greatly enhanced if the sensitivity of  $D_\alpha$  and/or  $D_\gamma$  could be established, even using factors whose influence was non-selective.

The four examples (2, 5, 6, and 8) based on brain measures all require the assumption that the measure (here,  $T_a$ ) corresponds to the neuronal implementation of (part of) the process that controls the behavior. It is in the present example that this assumption is most defensible, because the measurement of individual neurons permits more persuasive tests of task-essentiality (Section 1.7; Appendix A.10.1) than the other brain measures. Applying such criteria to individual neurons seems especially important, given the finding (Appendix A.10.1) that even areas S1 and M1 of the monkey cortex each contain mixtures of sensory and motor cells.

<sup>36</sup> See Appendix A.10.3 for a discussion of effects of the stimulus ( $t$  vs  $f$ ) and the response ( $T$  vs  $F$ ) as factors, in this connection.

Behavior is controlled by populations of neurons (Parker & Newsome, 1998); it is not obvious that the mean is the appropriate statistic of the population latency distribution to compare to the behavioral latency measure  $\overline{RT}$  (Appendix A.10.2). Appendix A.6 contains a brief summary of a study in which the latency of the LRP in humans was used as an indicator of the boundary between two stages, analogous to  $T_a$ .

## 11. Evidence from fMRI for modular neural processes implemented by anatomically delimited processors (Ex. 6)

Among the informal (and controversial) assumptions that appear to underlie the interpretation of localized brain activation in much research using functional brain-imaging (such as PET and fMRI; Frackowiak, Friston, Frith, & Mazziotta, 1997; Posner & Raichle, 1994) are the following four:

- (1) The neural processes ( $\alpha, \beta, \dots$ ) that underlie different functional operations are likely to be implemented by neural processors that are localized in the brain such that they occupy non-overlapping brain regions ( $\mathbf{R}_\alpha, \mathbf{R}_\beta, \dots$ ) and are of a size resolvable by the imaging method.<sup>37</sup>
- (2) If and only if a process (e.g.,  $\alpha$ ) occurs during performance of a task, there is an increase in activation of the region containing the corresponding processor ( $\mathbf{R}_\alpha$ ), relative to a task in which  $\alpha$  does not occur.<sup>38</sup>
- (3) If the same operation (e.g.,  $\alpha$ ) is employed in two different tasks, it is likely to be implemented by the same neural processor (located in region  $\mathbf{R}_\alpha$ ).
- (4) Many laboratory tasks are accomplished by a set of distinct mental-processing operations, some of which are also employed in other tasks.

The “only if” in Assumption 2 (process occurrence *necessary* as well as sufficient) is needed, with the other assumptions, to infer the occurrence of  $\alpha$  from activation of  $\mathbf{R}_\alpha$ . It is equivalent to assuming that the neural processes carried out in regions  $\mathbf{R}_\alpha$  and  $\mathbf{R}_\beta$  are modular. (To see why, replace  $M_A, M_B$ , and processes **A** and **B** in Table 2 by  $M_\alpha, M_\beta$ , and processes  $\alpha$  and  $\beta$ , respectively.)

The best-known paradigm in which functional imaging methods have been used is task comparison (Section 1.6; Appendix A.1), in which we search for dissociations between *tasks*, each accomplished by a qualitatively different complex process, rather than dissociations among parts of the complex process used to carry out a single task. In the simplest case, one selects two tasks, Task 0 and Task 1, and compares the associated patterns of brain activation. Inferences from the differences between these

<sup>37</sup> It has to be recognized that neural processors that are distributed over large regions or intermingled in small regions are less likely to be picked up using fMRI (Geisler & Albrecht, 2000, p. 82n). Also, the possibility that human brain regions contain neuron types with diverse functions is made more plausible by the discovery that primary somatosensory and motor areas in monkeys each contain both “sensory” and “motor neurons” (Appendix A.10.1). See also Phillips, Zeki, and Barlow (1984).

<sup>38</sup> This in turn requires that spatially resolvable brain regions can be found that are functionally homogeneous.

patterns are made in the context of a *task theory* for each task, which specifies the set of mental operations it requires. (While typical inferences require such task theories, they are sometimes only implicit.) Among other things, these inferences depend on the validity of the two task theories. The principal goal of such studies is to use the task theories together with the measured activation differences to discover the locations of the processors that carry out the hypothesized component operations, without necessarily validating the two task theories or assessing the modularity of these operations.<sup>39</sup>

Similar ideas and fewer assumptions lead to the use of fMRI for process decomposition, which can help validate a task theory as well as localize its constituent processes. As there is only one task, we eliminate Assumptions 3 and 4. And to justify the interpretation of brain-image differences due to within-task factor variation we replace Assumption 2 by:

(2') If and only if a process (e.g.,  $\alpha$ ) that occurs during a task is prolonged or otherwise made more difficult, there is an increase in activation of the region containing the corresponding processor ( $\mathbf{R}_\alpha$ ).<sup>40</sup>

Consider a single task for which performance depends on modular processes  $\alpha$  and  $\beta$ . Suppose these processes are accomplished by neural processors that occupy non-overlapping brain regions  $\mathbf{R}_\alpha$  and  $\mathbf{R}_\beta$ , and therefore provide pure measures  $M_\alpha$  and  $M_\beta$  of processor activation. (Given the assumed links between functional and neural processes,  $M_\alpha$  and  $M_\beta$  are also pure measures of functional processes **A** and **B**.) If we can find a factor  $F$  ( $G$ ) that selectively influences **A** (**B**), then if it influences the activation of either region, it should plausibly also selectively influence  $M_\alpha$  ( $M_\beta$ ).<sup>41</sup> One example of a candidate task is a choice–reaction experiment, for which the task theory might assert that the sub-processes include modules for stimulus identification and response selection; the two factors might be *stimulus quality* ( $SQ$ ) and *stimulus-response mapping familiarity* ( $MF$ ). (See Section 16 for relevant behavioral evidence.) Some requirements for a study of this kind are described in Appendix A.11.1; as I have been unable to find a study that meets them, the present example is hypothetical.

Fig. 9 shows mean data from a pair of hypothetical experiments in such a task. In one experiment (panel A), with  $G$  at a fixed level and four levels of  $F$ , changing the level of  $F$  leaves  $M_\beta$  invariant while changing  $M_\alpha$ . In the other experiment (panel B),

<sup>39</sup> A classic example of task comparison in brain imaging is provided by Petersen, Fox, Posner, Mintun, and Raichle (1988). See Appendix A.1 for a more extensive discussion of task comparison, its use with fMRI, and the contrast between task comparison and process decomposition.

<sup>40</sup> Prolonging a process may not increase its instantaneous neural activation level while it is occurring. The fMRI signal is inherently a time average, however; if the period of elevated activation is brief compared to the effective period of the average, then an increase in its duration will increase that average.

<sup>41</sup> With about  $10^5$  neurons/mm<sup>3</sup> (Wandell, 1995), a typical fMRI voxel (volume element) of 70 mm<sup>3</sup> contains almost  $10^7$  neurons. Due to this limit on the spatial resolution of fMRI, meaningful changes in activation patterns within voxels may not be reflected in  $M_\alpha$  or  $M_\beta$ .



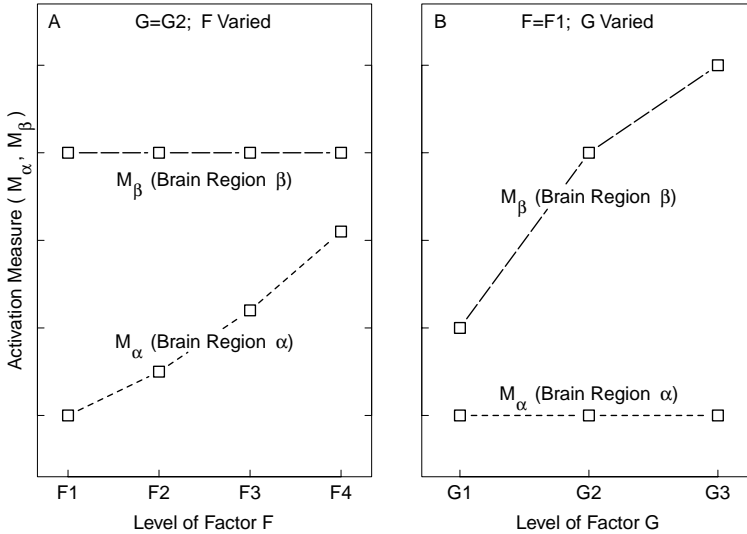


Fig. 9. Idealized hypothetical data from a pair of experiments in which localized measures of brain activation  $M_\alpha$  and  $M_\beta$  are obtained in two regions  $\mathbf{R}_\alpha$  and  $\mathbf{R}_\beta$ . Values shown are the means over  $N$  subjects. Panel A: The two measures as functions of the level of factor  $F$ , with the level of factor  $G$  fixed at  $G_2$ . Panel B: The two measures as functions of the level of factor  $G$ , with the level of factor  $F$  fixed at  $F_1$ . Unlike Fig. 5, the  $x$ -axes have not been rescaled for linearity of the effect of the most potent factor. See Appendix A.11.2 for discussion of statistical issues.

with  $F$  at a fixed level and three levels of  $G$ , changing the level of  $G$  leaves  $M_\alpha$  invariant while changing  $M_\beta$ . (The statistical treatment of such data is considered in Appendix A.11.2.) Such results support the hypotheses that the task involves (at least) the two processes  $\alpha$  and  $\beta$ , that  $M_\alpha$  and  $M_\beta$  are pure measures of them, that  $\alpha$  and  $\beta$  are modular and selectively influenced by  $F$  and  $G$ , respectively, and that they are carried out by neural processors in regions  $\mathbf{R}_\alpha$  and  $\mathbf{R}_\beta$ . Thus, in addition to localizing processes  $\alpha$  and  $\beta$  in the brain, such results provide support for the task theory. However, such data provide us with less confidence about the modularity of **A** and **B** than of  $\alpha$  and  $\beta$  because there may be no support for the assumed  $\alpha$ -**A** and  $\beta$ -**B** linkage. How might such support be provided? One approach would be to search for pure or composite *behavioral* measures of processes **A** and **B**, and seek evidence for **A** (**B**) being influenced selectively by the same factors as  $\alpha$  ( $\beta$ ).<sup>42</sup>

In addition to providing evidence for modularity of processes and processors as well as functional localization, a test of the task theory, and information (from the nature of the factors) that helps to define the processes of interest, the method has four other desirable features: First, a theory for only one task is required. Second, because  $M_\alpha$  and  $M_\beta$  are pure measures, the critical question is whether a factor

<sup>42</sup> Appendix A.6 summarizes such parallel evidence from an actual study, but where the brain measurements were time measures based on ERPs. However, parallel behavioral evidence may not always be available (Section 6.1).

changes an activation measure or leaves it invariant; the thorny issue of the appropriate combination rule for activation levels therefore need not be addressed. Third, because comparisons are of different conditions within the same task, there is no baseline task, hence, again, no need to assume any particular combination rule for baseline and task-dependent activation.<sup>43</sup> And fourth, while the presence vs absence of a process (in the task-comparison method) is a factor with only two levels, this method permits the use of multiple-level (within-task) factors, with its several advantages (Appendices A.2.1, A.9.2).

On the other hand, among the sources of concern associated with this method as well as some other imaging methods are the following: With a time constant on the order of seconds, the hemodynamic response is sluggish in relation to the time scale of interesting mental operations, which could lead to contamination of localized activation measures. Together with trials averaging, this even raises the possibility that events that follow a behavioral response may be interpreted as being required for it to occur, and that the increased activation measure for a region occurs on only a subset of trials. Finally, if the assumption of functional localization were to fail substantially in a particular study, as might be expected from some single-neuron studies (Appendix A.10.1), then we would be unlikely to find evidence for selective modifiability: persuasive invariance of activation in one region concurrent with systematic change in another (such as shown in Figs. 9A and 9B) would be improbable.

*A factorial version of the pair of experiments.* As argued in Appendix A.9, a factorial experiment rather than two one-factor experiments, while not essential, supplements demonstrations of selective influence such as those shown in Fig. 9 with useful tests of generality. Consider panel A, for example. In a  $4 \times 3$  factorial experiment we should find, first, that  $M_\alpha$  behaves identically at each level of  $G$ , and, second, that the invariance of  $M_\beta$  with respect to  $F$  is expressed at each level of  $G$  by a different flat function of  $F$ -level. In addition, such an expanded experiment permits assessing the extent to which any deviations from the expected patterns are systematic (see Exs. 3, 7, and 9). The cost of such considerable enrichment of the results is an increase in the number of conditions from seven ( $3 + 4$ ) to only 12 ( $3 \times 4$ ).

### 11.1. Comments

In this example we consider the assumptions often used in making inference from fMRI measurements in task-comparison studies. I suggest that a smaller and perhaps weaker set of assumptions can be used as a basis for identifying modular processes  $\alpha$  and  $\beta$  used in implementing the complex process for a single task, and thus acquire evidence about the modularity of the processes **A** and **B** that they implement.

<sup>43</sup> The rule assumed in the task-comparison paradigm (Appendix A.1.2.4) is almost always summation; usually its validation is at best indirect, because the conditions that would permit a test are not included.

But how strongly does the modularity of neural processes support the modularity of functional ones? That we observe only *measures* of functional processes and not the processes themselves, and that one measure of one such process might be invariant with respect to a factor while another measure of the same process might not be (Section 2.1), also applies to neural processes. In this example, the conclusion that processes carried out by  $\alpha$  and  $\beta$  are modular is based on the selective influence of  $F$  and  $G$  on spatially and temporally integrated (fMRI) activation measures  $M_\alpha$  and  $M_\beta$ , respectively. One reason for advocating the search for parallel evidence using behavioral process measures  $M_A$  and  $M_B$  is that the outcome of such a search is not obvious: Especially as measures of temporally and spatially integrated neural activity, it is possible that  $M_\alpha$  and  $M_\beta$  are merely crude and partial indicators that might be invariant with respect to a factor that substantially influences the corresponding neural process. (See Section 1.7 for other relevant considerations.) Hence, at least early in such research, it may be desirable to use tasks for which factors with selective influence have already been found in successful functional decomposition.

## 12. Introduction to four examples of inference based on composite measures

In each of the examples of Sections 13–16 a composite measure of the process is used, rather than pure measures of its parts. In the three behavioral examples the composite measures are standard indices of task performance (detectability, response rate, and reaction time); in each case a combination rule is specified and tested (Tables 3, 4, 8). In Ex. 8, where brain measurements are used, the combination rule is known, hence need not be tested.

**Example 7** (*Evidence for modular spatial-frequency analyzers from “probability summation” at threshold*). As discussed in relation to Ex. 4 (where the measures were pure), the selective effects of high-contrast adapting gratings of particular spatial frequencies on detection thresholds over a range of frequencies have provided one argument for modular spatial-frequency analyzers. Another kind of evidence is how the detectabilities of gratings of two different frequencies combine (producing a composite detectability measure) if they are presented together as a compound grating (Sachs et al., 1971). When the frequencies are sufficiently separated, the behavior of the composite measure indicates that they are detected by different modular analyzers (*two-channel model*). The discussion shows why a multiplicative combination rule for the non-detection probabilities of the two analyzers (usually called “probability summation”) is appropriate. In Appendix A.13, we consider Sachs et al.’s *single-channel model*, which fits the data when the two frequencies are close and the two-channel model fails. When the frequencies are separated the single-channel model fails, strengthening the evidence for modularity.

**Example 8** (*Evidence from ERP amplitude for modular neural processes*). The electrical potential at a point on the scalp is a composite measure of the contributions from all the voltage sources in the brain, suitably weighted. In his *additive-amplitude*

method, Kounios (1999) exploits the fact that, unlike our other examples of composite measures, the combination rule (summation) for ERP amplitude is given by physics rather than being one of the hypotheses ( $H3$  in Tables 3, 8) to be tested. Such composite measures can be obtained simultaneously at each of several locations rather than just one, which permits stronger inferences.

**Example 9** (*Response rate as a composite measure with multiplication as the combination rule*). Among the relatively few factorial experiments by students of animal behavior, Roberts (1987) discovered a large number of diverse examples of factors having multiplicative effects on response rate,  $RR$ , one from a study by Clark (1958). Such a pattern can arise if the combination rule for the contributions to  $RR$  of different processes (influenced selectively by those factors) is multiplication; Roberts has considered what sort of process model would create such a combination rule. By transforming Clark's data logarithmically, multiplicative factor effects become additive, and we can apply familiar tests of additivity. Because the factors had more than two levels, we can see the advantages of scaling them numerically (as in Ex. 3), which helps here in assessing the extent to which deviations from additivity are systematic. Together with Ex. 1, this example differs from the others in how responses are controlled: here, individual responses are operants, rather than being elicited by individual stimuli.

**Example 10** (*Selective effects of sleep deprivation*). One application of the separate-modifiability method to a composite measure ( $\overline{RT}$ ) with summation as the combination rule is the AFM, which has been used effectively and in interesting ways by Sanders. Unlike most of the other examples, a study by Sanders et al. (1982) goes beyond merely demonstrating the separate modifiability of two processes, by also showing that two additional factors, one being the amount of sleep deprivation, influences only one of those two processes.<sup>44</sup> While most experiments on sleep deprivation have used the task-comparison approach, leading experts to conclude that deprivation influences all mental processes (Appendix A.1.2.2), this use of the single-task process-decomposition approach indicates otherwise. As well as prolonging  $RT$ , sleep deprivation increases the number of response omissions; in Appendix A.16 we consider the proportion of trials on which responses are produced as a composite measure with multiplication as the combination rule, and also discuss several issues in the interpretation of additive  $\overline{RT}$  effects, and what, exactly, a stage model is.

### 13. Evidence for modular spatial-frequency analyzers from 'probability summation' at threshold (Ex. 7)

In Section 9 we considered evidence for modular spatial-frequency analyzers (or "channels") from the selectivity of the effects of adapting to a sinusoidal grating of

<sup>44</sup> This required Sanders et al. to run a four-factor ( $2^4$ ) experiment.

one spatial frequency on the modulation-contrast thresholds for detecting test gratings over a range of frequencies. If detection thresholds are determined by modular band-limited analyzers, then the absence of an effect far from the adapting frequency (outside the inferred band of the adapted analyzer) favors the hypothesis. In a ground-breaking and once controversial paper, Sachs et al. (1971) adopted a different approach; they considered how gratings of two frequencies in a *compound grating* (consisting of superposed simple vertical gratings at frequencies  $f$  and  $f'$ ) combine in influencing the probability of detection. In their experiments, a brief test stimulus was preceded and followed by a uniform field of the same mean luminance; the observer responded “yes” (“detection”) if he saw any change. Because  $Pr\{Detect\}$  is then a composite measure, reflecting the contributions of both constituent gratings, the test of separate modifiability requires assuming a combination rule, as discussed in Section 3 and exemplified in Tables 3, 4, and 8.

### 13.1. Justification of a multiplicative combination rule for non-detection probability

Suppose spatial frequencies  $f$  and  $f'$  are sufficiently far apart such that the analyzers most sensitive to them,  $\mathbf{A}$  and  $\mathbf{A}'$ , are modules (distinct “channels”). That is, they are separately modifiable, as shown by  $\mathbf{A}$  being influenced by the *contrast of the  $f$  grating* ( $C_f$ ) but not by the *contrast of the  $f'$  grating* ( $C_{f'}$ ), and vice versa: factors  $C_f$  and  $C_{f'}$  influence processes  $\mathbf{A}$  and  $\mathbf{A}'$  selectively. As the stimuli in these experiments are near threshold, the behavior of each analyzer is probabilistic: from trial to trial the same signal may or may not excite it, perhaps depending on the levels of external and internal “noise”. What is selectively influenced by a factor is therefore a probability. For each analyzer on each trial we define a binary indicator random variable ( $\mathbf{x}$  for  $\mathbf{A}$ ,  $\mathbf{x}'$  for  $\mathbf{A}'$ ) that has the value one if its analyzer is excited, and zero otherwise.

As shown below, the appropriate combination rule for  $\mathbf{x}$  and  $\mathbf{x}'$  is multiplication (Section 3.2). Thus, in addition to assuming  $\mathbf{A}$  and  $\mathbf{A}'$  to be independent in the sense of separately modifiable, justification of the combination rule required Sachs et al. also to assume that the behaviors of the two analyzers (expressed by the random variables  $\mathbf{x}$  and  $\mathbf{x}'$ ) are uncorrelated. For example, the fact that  $\mathbf{A}$  is excited ( $\mathbf{x} = 1$ ) on a trial can have no effect on  $Pr\{\mathbf{x}' = 1\}$  on that trial. Also, the occurrence of detection responses on “catch” trials, on which no grating was displayed, led them to assume that, independent of these analyzers, a third “guessing” mechanism,  $\mathbf{GM}$ , may generate a false detection response with some probability regardless of the stimulus; let  $\mathbf{y} = 1$  when such a guess is generated, and  $\mathbf{y} = 0$  otherwise. Random variables  $\mathbf{x}$ ,  $\mathbf{x}'$ , and  $\mathbf{y}$  are then assumed to be mutually independent stochastically.<sup>45</sup> Finally, Sachs et al. (1971) assumed that the observer

<sup>45</sup> As discussed in Appendix A.3.3, whereas the zero-correlation requirement can be violated for processes we would regard as modular, support for stochastic independence strengthens the evidence for modularity.

makes an overt detection response on a trial if and only if at least one of the variables  $\mathbf{x}$ ,  $\mathbf{x}'$ , or  $\mathbf{y}$  is non-zero.<sup>46</sup>

The consequences of these assumptions can be more simply expressed if we consider  $P_{ND} = Pr\{\text{Nondetect}\} = 1 - Pr\{\text{Detect}\}$  rather than  $Pr\{\text{Detect}\}$ . Non-detection occurs if and only if  $\mathbf{x}$ ,  $\mathbf{x}'$ , and  $\mathbf{y}$  are all zero; using the stochastic independence of these three random variables, it follows that  $P_{ND} = Pr\{\mathbf{x} = 0\} \times Pr\{\mathbf{x}' = 0\} \times Pr\{\mathbf{y} = 0\}$ , a multiplicative combination rule similar to Eq. (4). Given contrast levels  $Cf = Cf_j$  and  $Cf' = Cf'_k$ , we let  $\pi(Cf_j) = Pr\{\mathbf{x} = 0\}$  be the probability that  $\mathbf{A}$  is not excited, let  $\pi'(Cf'_k) = Pr\{\mathbf{x}' = 0\}$  be the probability that  $\mathbf{A}'$  is not excited, and let  $p_{ng} = Pr\{\mathbf{y} = 0\}$  be the probability that no guessed “yes” is generated. The assumptions above then imply that

$$P_{ND}(Cf_j, Cf'_k) = \pi(Cf_j) \times \pi'(Cf'_k) \times p_{ng}. \quad (8)$$

That is,  $Cf$  and  $Cf'$  are *multiplicative factors*, because the composite measure is a product and their levels influence different factors in that product (Section 3.2).

### 13.2. Three tests of the joint hypothesis

The joint hypothesis is as in Table 4, with modules  $\mathbf{A}$ ,  $\mathbf{A}'$  and measure  $M_{AA'} = P_{ND}$ , but strengthened by the added hypothesis that  $\mathbf{GM}$  neither affects nor is affected by  $\mathbf{A}$  and  $\mathbf{A}'$ , and by replacement of  $H5$  by the mutual stochastic independence of the contributions of  $\mathbf{A}$ ,  $\mathbf{A}'$ , and  $\mathbf{GM}$ . (See Appendix A.2.3 for an alternative formulation of the inferential logic.) In Sachs et al.’s primary test, they fixed  $f$  at 14 cycles/deg and set  $f'$  at different values in different conditions, some far from  $f$  (the furthest were 2.8 and 28 cycles/deg) and some near  $f$  (the closest was 11.2 cycles/deg). In each condition,  $Cf$  took on levels in a random sequence selected from  $0 = Cf_1 < Cf_2 < \dots < Cf_n$  with  $n \approx 6$ .  $Cf'$  was randomly either  $Cf' = 0$ , generating a *simple* ( $f$ ) grating, or a fixed value  $Cf'_2 > 0$ , generating a *compound* ( $f, f'$ ) grating. Each condition thus consisted of an  $n \times 2$  ( $Cf \times Cf'$ ) factorial design, with the ( $Cf_1, Cf'_1$ ) cell consisting of catch trials (zero contrasts for both  $f$  and  $f'$ ). For the simple grating,

$$P_{ND}(Cf_j, 0) = \pi(Cf_j) \times p_{ng}. \quad (9)$$

Combining this with Eq. (8) for the compound grating, we find

$$P_{ND}(Cf_j, Cf'_2) = \alpha P_{ND}(Cf_j, 0), \quad (0 < \alpha < 1), \quad (10)$$

where  $\alpha$  is the constant  $\pi'(Cf'_2)$ . The result of adding the  $f'$  grating (the result of changing the level of factor  $Cf'$  from zero to a fixed positive value) is thus predicted to be *multiplication* of  $P_{ND}$  by a fixed constant  $0 < \alpha < 1$ , regardless of the level of  $Cf$ .

<sup>46</sup> For discussion related to such models, in which what is combined are binary decisions about each of several sources of information rather than the (possibly continuous) information on which those decisions are based, see, for example, Graham (1989, especially Chapters 4 & 7), Shaw (1982, on evidence favoring “second-order integration”), and Treisman’s (1998) review and elaborations of “probability-summation” models.

As  $\alpha$  is just the p.effect of  $Cf'$ , we see that it is invariant with respect to  $Cf$ ; similarly, the p.effect of changing the level of  $Cf$  from  $Cf_j$  to  $Cf_{j+1}$  (for example) is invariant with respect to  $Cf'$ .

The plots in Fig. 10, of data from three of the conditions shown in Sachs et al.'s Fig. 8, differ from theirs in assuming nothing about the form of the psychometric function other than its monotonicity. To help visualize any systematic pattern in the model-data deviations, I scaled the  $Cf$  levels ( $x$ -axis) in each condition so as to linearize the observed proportions for the simple grating,  $\hat{P}_{ND}(Cf_j, 0)$ , and hence the predicted proportions for the compound grating. From Eq. (10), we then expect the observed proportions for the compound grating,  $\hat{P}_{ND}(Cf_j, Cf'_2)$ , to fall on a linear

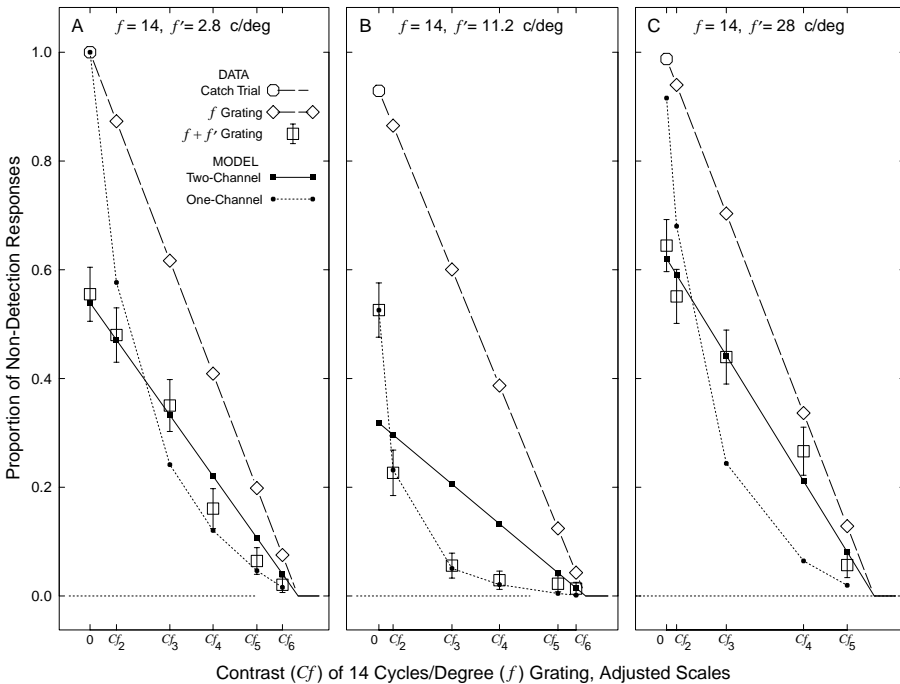


Fig. 10. Data from one subject in three conditions of a grating-detection experiment by Sachs et al. (1971, Fig. 8). For each condition, each open symbol shows the proportion,  $\hat{P}_{ND}$  of  $N = 100$  trials on which the observer reported “no change”, as a function of the contrast ( $Cf$ ) of the grating with spatial frequency  $f = 14$  cycles/deg, for simple gratings (no other frequency present;  $Cf' = Cf'_1 = 0$ ) and for compound gratings (also containing a superposed fixed-contrast grating with frequency  $f'$ ;  $Cf' = Cf'_2 > 0$ ). Panels A and C show conditions in which  $f'$  was 2.8 and 28 cycles/deg, respectively, far from  $f = 14$  cycles/deg. Panel B shows a condition in which  $f'$  was 11.2 cycles/deg, close to  $f$ . For each condition the  $x$ -axis has been scaled so as to linearize the simple-grating data. The two-channel (modular-analyzer) model implies that  $P_{ND}(compound)$  is a fixed proportion ( $\alpha$  in Eq. 10) of  $P_{ND}(simple)$ , as shown by the lower fitted line and filled squares. Non-detection proportions for catch trials (zero contrast at both frequencies) are shown by the highest value of  $\hat{P}_{ND}$  in each panel. Error bars represent  $\pm$ S.E., based on binomial variance. Also shown in each panel are the fitted values for a broadband single-channel model (Appendix A.13). Because the basic data are no longer available, values were determined from the 1971 figure.

function that lies below the  $\hat{P}_{ND}(Cf_j, 0)$  line, and that converges with it where  $P_{ND} = 0$ . The  $\alpha$ -value that best fitted the compound-grating proportions in each condition was determined by least squares; the resulting fitted lines (two-channel model) are shown. Global tests of goodness of fit as well as tests focused on systematic (linear and quadratic) trends in the deviations of the compound-grating data from the fitted model show that for panels A and C (compounds consisting of two frequencies separated by at least a factor of 2) the multiplicative prediction fits well, while it fails dramatically for panel B (compound consisting of two frequencies separated by much less than a factor of 2). Confirming the joint hypothesis, Sachs et al. thus provide us with evidence that detection of compounds of two gratings with sufficiently separated frequencies, but not of those with similar frequencies, is accomplished by modular analyzers.

### 13.3. Comments

In this investigation of separate modifiability, the joint hypothesis is an elaborated version of what is stated in Table 4, including *H1*: Analyzers A and A' are modules; *H4*: Their contributions to  $M_{AA'} = P_{ND}$  combine by multiplication; and *H5*: These contributions are uncorrelated.<sup>47</sup> The measure  $P_{ND}$  is composite because it reflects contributions from both modules; in this experiment subjects did not attempt to report separately on the presence of  $f$  and  $f'$ . That factors  $Cf$  and  $Cf'$  (the contrasts of frequencies  $f$  and  $f'$ , respectively) have multiplicative p.effects on the composite measure  $P_{ND}$  (when  $f$  and  $f'$  are sufficiently separated) supports the joint hypothesis (in that case). Unlike the joint hypotheses for pure measures (Table 2) and for composite measures with summation as the combination rule (Table 3), this one incorporates the hypothesis of stochastic independence. The finding of multiplicative p.effects therefore supports such stochastic independence, providing stronger evidence for modularity than in these other cases (Appendix A.3.3), as we shall also see in Ex. 9.

As mentioned in Section 9.1, one advantage of the present approach to evaluating the modularity of spatial-frequency analyzers is that it clearly involves one task that employs both analyzers, rather than a different task for each. Ignoring this difficulty, the two kinds of evidence in Exs. 4 and 7 complement each other in demonstrating the modularity of such analyzers when their frequencies are sufficiently different: Ex. 4 shows that they can be selectively adapted by high-contrast simple gratings; the present example shows that their detection performance is selectively influenced by the threshold-level contrasts of separated spatial-frequency components in a compound grating.

One aspect of this example is the demonstration (Fig. 10; Appendix A.13) that a competing broadband one-channel model fails when the two-channel model succeeds, and vice versa. This strengthens the argument, first by demonstrating that the

<sup>47</sup> The elaborations of the joint hypothesis result from the need for a third process, **GM**, which is assumed to function independently of A and A', both stochastically and in its mean contribution.



data features considered, together with their precision, are capable of discriminating between plausible models, and second, by showing a dependence of the relative merits of the two models on the  $f-f'$  difference that is plausible, given the idea that spatial-frequency analyzers have measurable bandwidths. Also, because introduction of the competing model can be regarded as simply replacing one candidate composite measure (Eq. (8)) by another (Eq. (19)), this example permits me to emphasize that a particular measure  $M_{AB}$  may or may not have the behavior hypothesized ( $H4$  and  $H5$ , Table 4), just as did Ex. 3 for a pure measure (Appendix A.7.2). Also described in Appendix A.13 are details of fitting the two-channel model.

#### 14. Evidence from amplitude of the event-related potential for modular neural processes (Ex. 8)

The  $ERP(t)$  at an electrode site on the scalp is typically the average over trials of a function that relates the voltage at that site to the time since stimulus presentation. As in some other brain measurements (e.g., PET, fMRI), averaging is often required for the measures to be interpretable, because of the poor  $S/N$  ratio.<sup>48</sup> At any particular time  $t = t_0$ ,  $ERP(t_0)$  is a composite measure of contributions from all the neural processors (“sources” or “generators” of voltage) in the brain that are active at time  $t_0$ . The physics of volume conduction tells us that the potentials from different sources obey linear superposition, and that the combination rule is thus *summation*: the voltage at a site is a weighted sum of the contributions from all the sources, with the weights dependent on variables such as source–site distance and conductivity, and source orientation (Kutas & Dale, 1997; Nunez, 1981).

##### 14.1. The additive-amplitude method

Let the modules  $H1$  in Table 3 be neural processes  $\alpha$  and  $\beta$ . Because the combination-rule part of the joint hypothesis ( $H3$ ) is given by physics, the inferential logic in this case is concerned exclusively with  $H1$ . If we are able to find factors  $F$  and  $G$  that influence  $\alpha$  and  $\beta$  selectively during some epoch (and that jointly influence no other neural process), it follows that the factor effects on  $M_{\alpha\beta} = ERP(t)$  during that epoch will be additive *and* that this will be true at any electrode site. This realization was the basis of Kounios’ (1999) *additive-amplitude method* (see also Kounios, 1996).

How strongly does a finding of additive factor effects support hypothesis  $H1$ ? As mentioned in Section 1.8, the degree to which confirmation of a prediction supports a hypothesis depends (partly) on the existence and plausibility of alternative hypotheses that generate the same prediction. And it can be argued that we know too little about brain circuitry to estimate the likelihood that two factors that influence

<sup>48</sup> Here the “noise” is due partly to brain events probably unrelated to the task being performed, whose contributions are eliminated by subtracting a pre-stimulus baseline level and by an averaging process that reveals only those events that are time-locked to the stimulus.

the *same* neural process will have effects on the scalp potential that are additive.<sup>49</sup> However, because multiple electrodes are typically used, we also have information about the locations of processors  $\mathbf{P}_\alpha$  and  $\mathbf{P}_\beta$ , hence of the sources of the signals that contribute additively to  $M_{\alpha\beta}$ . If  $\alpha$  and  $\beta$  are influenced selectively by  $F$  and  $G$ , and  $\mathbf{P}_\alpha$  and  $\mathbf{P}_\beta$  have different loci (arguably evidence against the two factors influencing the same process), then we expect the effects of these factors to have different *topographies*. That is, we expect that the pattern across the set of electrode sites of the relative magnitudes of the effects that  $F$  produces should differ from the pattern for  $G$ . Thus, if different topographies are found for the main effects of two additive factors, this strengthens the support for *HI* provided by additivity. It is of course possible that two neural processors  $\mathbf{P}_\alpha$  and  $\mathbf{P}_\beta$  that implement modular processes are coextensive. Hence, while sameness of topography of the factor effects weakens the support for *HI* that is provided by additivity, it does not rule *HI* out.

What if two factors interact at some electrode sites while at others they are additive? One possible interpretation is that  $\alpha$  and  $\beta$  are influenced selectively by  $F$  and  $G$ , but that there is a third process  $\gamma$  that is influenced by both factors, but whose processor  $\mathbf{P}_\gamma$  is located sufficiently far from some sites so its contributions there are greatly attenuated. (That is, its weights in the weighted sums for those sites are effectively zero.) A second possibility is that  $\alpha$  and  $\beta$  are both influenced by both factors, but that the resulting interactions are of opposite sign; at sites appropriately located between  $\mathbf{P}_\alpha$  and  $\mathbf{P}_\beta$  the net interaction could then be zero, even though both main effects are non-zero there. Possibilities like these have other implications that can be tested, given the rich data provided by multiple electrodes.

#### 14.2. Application of the additive-amplitude method to word classification

The goal of one of the applications that Kounios (1999) reports was to understand the mechanism of semantic satiation. The subject heard a continuously presented list of spoken nouns, each of which had to be classified by meaning. The list consisted of sublists, each containing 15 presentations of a “prime”, the first and fifteenth followed by different “probes”. Target words (names of body parts), which were randomly interspersed in the list, comprised about 5% of the words and required a manual response; otherwise no overt response was required. The critical data are the ERPs elicited by the probes.

Two factors were manipulated, each at two levels. One factor was *REL*, the *semantic relatedness* of the probe to the prime, that preceded it. For high relatedness (e.g., *dog-cat*),  $REL = REL_1$ ; for low relatedness (e.g. *desk-truck*),  $REL = REL_2$ . The other factor was *SAT*, *semantic satiation* of the prime, which is believed to increase with successive presentations: for probes following the first presentation of a prime  $SAT = SAT_1$ ; for probes following the fifteenth,  $SAT = SAT_2$ . Factor levels are in-

<sup>49</sup> Example of relevant evidence are provided by Carandini and Heeger (1994) and Geisler and Albrecht (1995).

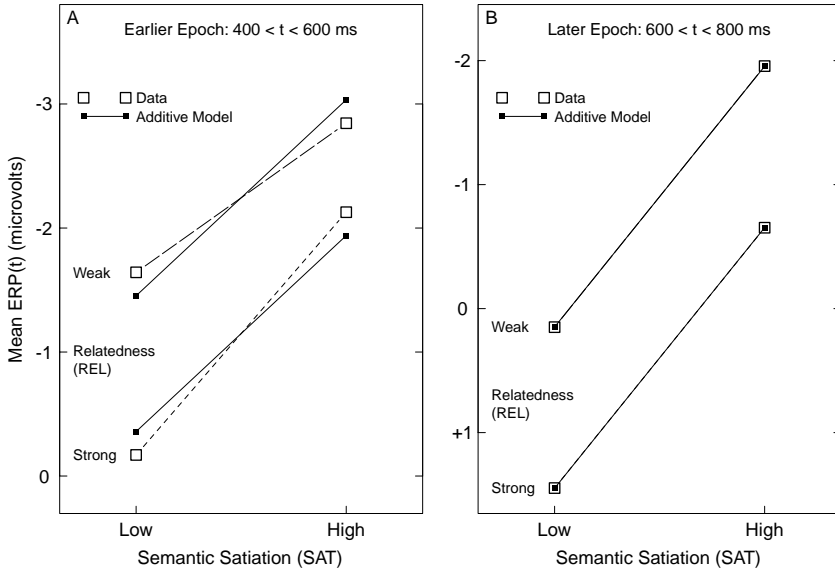


Fig. 11. Means over 36 subjects and five pairs of lateral electrode sites from Kounios (1999, Example 2). In panel A for the earlier ( $i = 1$ ) epoch, unfilled squares show the mean ERP amplitude,  $ERP_{1jk}$  over the five pairs of lateral electrode sites for each of the four conditions. Panel B for the later ( $i = 2$ ) epoch shows the corresponding values of  $ERP_{2jk}$ . Note the y-axis translation between panels. Also shown for each epoch is a fitted additive model. The main effects of  $REL$  ( $ERP_{2\bullet} - ERP_{1\bullet}$ ) for early and late epochs are  $-1.1 \pm 0.2$  and  $-1.3 \pm 0.2$   $\mu\text{V}$ , respectively, a non-significant increase of  $0.20 \pm 0.23$   $\mu\text{V}$  from one epoch to the next. The corresponding main effects of  $SAT$  ( $ERP_{\bullet 2} - ERP_{\bullet 1}$ ) are  $-1.6 \pm 0.4$  and  $-2.1 \pm 0.4$   $\mu\text{V}$ , respectively, a significant increase of  $0.5 \pm 0.3$   $\mu\text{V}$  ( $p = 0.046$ ). For the earlier epoch the additive model fits poorly: using an S.E. based on between-subject variation, the interaction contrast is  $(ERP_{122} - ERP_{121}) - (ERP_{112} - ERP_{111}) = -0.8 \pm 0.3$   $\mu\text{V}$  ( $p = 0.013$ ). For the later epoch the model fits well: the interaction contrast is  $0.01 \pm 0.34$   $\mu\text{V}$ . The change in the interaction contrast between epochs is significant:  $0.8 \pm 0.3$   $\mu\text{V}$  ( $p = 0.013$ ). The inverted y-axis scale is consistent with the ERP response being an increased negativity.

dexed so that a higher index value is associated with a greater (negative) ERP amplitude, averaged over electrode sites.

For each site, the mean  $ERP(t)$  was determined for each of two epochs,  $400 \leq t \leq 600$  ms (early,  $i = 1$ ) and  $600 \leq t \leq 800$  ms (late,  $i = 2$ ), and for each of the four pairs of factor levels, giving  $\overline{ERP}(\text{early}; R_j, S_k) = ERP_{1jk}$  and  $\overline{ERP}(\text{late}; R_j, S_k) = ERP_{2jk}$  ( $j = 1, 2; k = 1, 2$ ). For each epoch and each of the four conditions, the mean amplitude over subjects and over the five pairs of lateral electrode sites is shown in Fig. 11, together with a fitted additive model. There is a clear interaction during the first epoch that has vanished by the second.<sup>50</sup>

<sup>50</sup> Given its variation over subjects, expressed by the S.E., the mean interaction contrast is surprisingly (and significantly) small. This effect can occur if the estimated S.E. is inflated by a factor that was balanced over subjects.

For each epoch, the main effects and interactions of *SAT* and *REL* are shown in Fig. 12, separated by pairs of electrode sites. The figure shows that during both epochs the effects of the two factors had substantially different topographies: the *SAT* effect increases markedly from back to front while the *REL* effect has a gentle peak at the second site and changes relatively little across sites. Furthermore, each main effect has a size and topography that are similar across epochs, which suggests that the loci of  $P_\alpha$  and  $P_\beta$  do not change from one epoch to the next. The striking difference between epochs is in the *REL*  $\times$  *SAT* interaction, which is noticeable and reliable during the first epoch (with satiation reducing the effect of relatedness), but persuasively close to zero at all sites during the second. The three midline electrode sites (not shown) reveal the same patterns.

These results support the following account: During the second epoch, factors *REL* and *SAT* influence separately modifiable neural processes  $\alpha$  and  $\beta$  that are implemented by processors  $P_\alpha$  and  $P_\beta$  at separated loci within the brain, and *REL* and *SAT* do not jointly influence any other process  $\gamma$ . During the first epoch this account is false in one or more ways. What might differ about the first epoch? Possibilities include: (1) Selective influence fails, such that  $\alpha$ ,  $\beta$ , or both are influenced jointly by *REL* and *SAT* then, but not later. If so,  $\alpha$  and  $\beta$  may be modular during

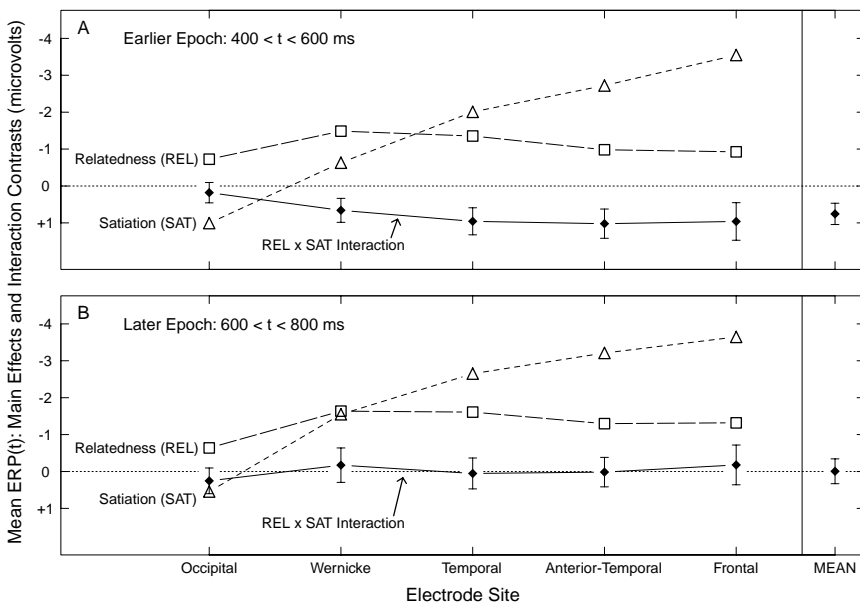


Fig. 12. Means over 36 subjects from Kounios (1999, Example 2), here separated by pairs of lateral electrode sites. Main effects and interactions are defined as in Fig. 11. Panels for earlier ( $i = 1$ , panel A) and later ( $i = 2$ , panel B) epochs each show the main effect of *REL*, the main effect of *SAT*, and their interaction contrast. These values are shown for means of each of five pairs of lateral electrode sites, plotted from posterior to anterior. Also shown for the interaction contrasts are their means over the five sites, and  $\pm$ S.E. bars based on between-subject variation.

both epochs, and an alternative pair of factors might be found that have additive effects during the first epoch, or possibly during both. (2) A third process  $\gamma$  (implemented by processor  $P_\gamma$ ) that is jointly influenced by *REL* and *SAT* is active during the first epoch but not later. Again, this would be consistent with the modularity of  $\alpha$  and  $\beta$  during both epochs. (3)  $\alpha$  and  $\beta$  are modular only during the second epoch; during the first they interact in such a way that they are not separately modifiable. The similarity across epochs of the sizes and topographies of the main effects does not appear to favor one of these possibilities over another.

### 14.3. Comments

As in other cases where the measurements during a task are of neural processes rather than aspects of task performance, we would like to link  $\alpha$  and  $\beta$  to functional processes that are among those required to carry out the task. This would be especially difficult in the present example: while a decision was required as to whether each word was a target, overt responses were made on only the 5% of the trials on which the decision was affirmative, trials not relevant to the questions being asked.<sup>51</sup>

One of the strengths of the ERP method is that measurements can be made at multiple electrode sites, providing a richness of data that behavioral measurements often do not. It is impressive, for example, that during the later epoch, the interaction is shown to be negligible at all five pairs of lateral sites. And we have seen that topography differences between the effects of the two factors can play a useful role in inferences. Another strength of the method is its excellent time resolution compared to some other methods (such as fMRI), without which the change from interactive to additive factor effects between epochs could not have been observed. On the other hand, if the explanation of the difference between epochs is possibility (2), it might be easier to determine whether  $P_\gamma$  is separated from  $P_\alpha$  and  $P_\beta$  with better spatial resolution, which may be provided by advances in ERP source localization, perhaps supplemented with fMRI. The interpretation of provocative data such as these will be aided by additional within-task factorial ERP experiments, along with relevant behavioral measurements, especially ones that evaluate effects of the same factors on both behavioral and brain measures.

## 15. Multiplicative combination rule for response rate (Ex. 9)

In Roberts' (1987) paper on the *multiplicative-factor method*, he described a widespread orderliness of *RR* in the data from 17 factorial experiments with animals,

<sup>51</sup> The possibility has to be considered that by the second epoch the decision had been made. (The strength of the main effects during that epoch might then be puzzling.) However, in similar experiments (Kounios, Kotz, & Holcomb, 2000) the RT for overt responses to targets was 749 ms, measured from the onsets of spoken words with mean duration 544 ms. On many trials, therefore, some or all of the decision process may have been included within the second epoch (from 600 to 800 ms) rather than being completed earlier.

experiments that covered a remarkable range of factors and species. The orderliness takes one of two forms: In one, displayed by several of the data sets, pairs of factors had multiplicative p.effects (Section 3.2) on  $RR$ , supporting the joint hypothesis of Table 4, which includes the multiplicative combination rule expressed by Eq. (4):

$$M_{AB}(F_j, G_k) = u(F_j) \times v(G_k),$$

and leads to the combining of factor effects described by Eqs. (6) and (7) – multiplicative p.effects on  $M_{AB} = RR$ , and additive effects on  $\log(M_{AB}) = \log(RR)$ . The other form, required by many of the data sets, is consistent with a third process **C** that generates responses independently of processes **A** and **B** of Table 4, where **C** may be influenced by  $F$  or  $G$  or neither; this arrangement is associated with a hybrid (multiplicative–additive) combination rule,

$$M_{AB}(F_j, G_k) = [u(F_j) \times v(G_k)] + w, \quad (11)$$

where  $w$  is either a constant with respect to  $j$  and  $k$ , or  $w = w(G_k)$  or  $w(F_j)$ , a function of one of the factors.<sup>52</sup>

One of the simple cases is provided by Clark's (1958) experiment on the effects of *hours of food deprivation (HD)* and *frequency of feeding (FF)* on the rate of bar pressing by rats under VI (variable-interval) reinforcement schedules. After training, each rat was kept at about 85% of its free-feeding weight, and was tested on different days at different delays after the daily feeding, thus varying the level  $HD_j$  of  $HD$ ; the levels were 1, 3, 5, 7, 10, 20, and 23 h. Each of three groups of three rats experienced a different VI schedule, with 3, 2, and 1 min VI schedules generating low, medium, and high levels  $FF_k$  of  $FF$ , respectively.

Fig. 13A displays both the remarkable precision of the data and the excellent fit of the multiplicative model described by Eq. (6): the p.effect of a change in  $HD$  is invariant over levels of  $FF$ , and vice versa. This finding supports the joint hypothesis described in Table 4: There are two modular processes, **A** related to motivation and **B** related to a variety of learning, influenced selectively by  $HD$  and  $FF$ , respectively; their contributions to  $RR$  combine by multiplication and are uncorrelated. In both panels the levels of  $HD$  plotted on the  $x$ -axis have been assigned numerical values so as to linearize the main effect of that factor; the goal is to make any systematic deviations of data from model easier to discern.

Roberts (1987) considered several process models in relation to the multiplicative combination rule, and favored one that is equivalent to the following: Processes **A** and **B** are arranged sequentially. Each operates by receiving a series of discrete signals (“pulses”) and transmitting each pulse with some probability. Let  $\mathbf{x}_A$  and  $\mathbf{x}_B$  be zero-one indicator random variables that correspond to pulse transmission by **A** and **B**, respectively, with  $p_A = Pr\{\mathbf{x}_A = 1\}$  and  $p_B = Pr\{\mathbf{x}_B = 1\}$ . Because **A** and **B** are selectively influenced by  $HD$  and  $FF$ , respectively,  $p_A = p_A(HD)$  and  $p_B = p_B(FF)$ .

<sup>52</sup> See Appendix A.15 for discussion of such hybrid combination rules and their implications.

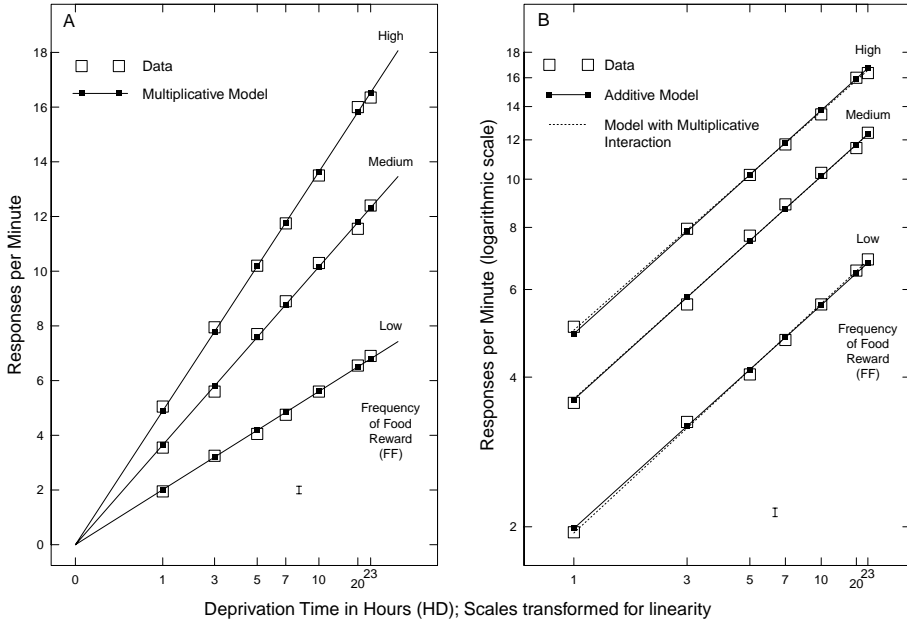


Fig. 13. Means reported by Clark (1958) for three groups of three rats each. Panel A, similar to Roberts (1987, Fig. 13), shows  $RR_{jk}$  vs  $HD$  for three levels of  $FF$ . Levels  $HD_j$  have been assigned numerical values so as to render the  $RR_{jk}$  linear. For each level  $FF_k$ , panel A also shows the best-fitting line constrained to pass through the origin. Panel B shows the same  $RR_{jk}$  data scaled logarithmically:  $RR_{jk}^* = \log(RR_{jk})$ . Here, levels  $HD_j$  have been assigned numerical values so as to linearize the  $RR_{jk}^*$ , the means over  $FF_k$  of  $RR_{jk}^*$ . Panel B also shows an additive model fitted to the data, as well as a model with a multiplicative interaction; see text. Only the mean  $RR_{jk}$ -values are available; individual data points reflect an average of about 1750 responses. Also shown are  $\pm$ S.E. bars, representing the precision of individual data points, based on deviations from the fitted multiplicative (panel A) and additive (panel B) models. Data values were determined from Fig. 1 of Clark (1958).

A series of input pulses at a fixed mean rate  $R_0$  is the input to the first process; each output pulse from the second process is converted into a response. The probability that an input pulse produces a response is then  $Pr\{\mathbf{x}_A\mathbf{x}_B = 1\}$ . Assuming that the events  $\mathbf{x}_A = 1$  and  $\mathbf{x}_B = 1$  are uncorrelated,  $Pr\{\mathbf{x}_A\mathbf{x}_B = 1\} = Pr\{\mathbf{x}_A = 1\} \times Pr\{\mathbf{x}_B = 1\}$ . With such an arrangement, then,

$$RR(HD_j, FF_k) = R_0 \times p_A(HD_j) \times p_B(FF_k), \tag{12}$$

and we have the multiplicative combination rule.<sup>53</sup>

<sup>53</sup> In a variant model, the input signal is continuous, each process functions as a “gain control” or “attenuator”, the response rate is proportional to the output signal level, and any variabilities in the gains of the two processes are uncorrelated.

For the hybrid combination rule (Eq. (11)), Roberts suggested an additional path for generating responses, in parallel with the **A-B** path described above, involving a third process **C** that might or might not be influenced by *FF*.<sup>54</sup> In such a case the three functions in panel A would still be linear, and either would intersect at a common point where  $RR > 0$  (if **C** is influenced by neither *HD* nor *FF*) or would not intersect at a common point (if **C** is influenced by *FF*). Note that to distinguish these alternatives requires at least three levels of factor *FF*.<sup>55</sup>

As shown in panel B, the application of a logarithmic transformation to  $M_{AB}$  to produce  $RR_{jk}^*$  converts a multiplicative pattern of factor effects into an additive one (Eq. (7)); the excellent fit of the additive model is simply another expression of the excellence of the multiplicative-effects model in panel A. For the hybrid combination rule, the use of a logarithmic transformation is less straightforward.

### 15.1. A plausible systematic deviation from additivity measured by multiplicative interaction of scaled factor levels

*Monotone interaction.* The second model fitted to the data in panel B assumes additive effects of *HD* and *FF* together with a multiplicative interaction. It is included to illustrate the use of a focused test for one kind of *systematic* deviation between the data and an additive model, rather than depending solely on a global or “omnibus” test of deviations (interaction) that is less sensitive to a particular pattern. To discuss such a focused test it is convenient to order the levels of each factor in terms of the mean of the composite measure associated with it. In this example, values of  $j$  are assigned such that  $RR_{j\bullet}^*$  increases monotonically with  $j$ , and likewise for  $k$  and  $RR_{\bullet k}^*$ . We can then consider “increasing” or “decreasing” the level of a factor. Especially if the factors are unitary (Appendix A.2.1), the most plausible and interesting systematic deviation is one in which, instead of falling on parallel lines, the data fall on lines that diverge or converge: that is, a deviation from additivity in which each increase in the level of one factor (e.g., *FF*) consistently increases (or decreases) the effect of increasing the level of the other (e.g., *HD*). We thus have a monotone modulation of the effect of one factor by the other – a “monotone interaction”.<sup>56</sup> This kind of interaction is plausible if the mechanism that generates it is qualitatively the “same” at all factor levels.

To think about the quantitative form that such an interaction is likely to take, we need to consider how much each incremental effect of *HD* (for example, from  $HD_1$  to

<sup>54</sup> Where **C** is introduced, an independent source of input pulses with rate  $R'_0$  and a transmission probability  $p_C$  can be added to the process model; to Eq. (12) we would then add the corresponding term  $R'_0 \times p_C$ .

<sup>55</sup> If **C** were influenced by *HD* instead of *FF*, the same statements would apply, but only if the roles of *HD* and *FF* were reversed in the plot; in this case at least three levels of *HD* would be required to distinguish the alternatives.

<sup>56</sup> Monotonicity of a two-way interaction is symmetric: If *FF* modulates the effect of *HD* monotonically, then the converse also obtains.



$HD_2$ , or  $RR_{2k}^* - RR_{1k}^*$ ) is likely to change with a particular increase in the level  $FF_k$  of  $FF$  (from  $k = 3$  to  $k = 4$ , for example). As the choice of factor levels is often arbitrary, there is no basis for assuming that the modulation of the  $HD$  effect induced by increasing the  $FF$  level from  $k = 3$  to  $k = 4$  would be the same as that induced by increasing the  $FF$  level from  $k = 2$  to  $k = 3$ , for example. That is, these two  $FF$ -level increases would not, in general, be equally “potent” for modulating the  $HD$  effect. Similarly, we would not expect two different one-step level increases of  $HD$  to be equally potent for modulating an incremental effect of  $FF$  (for example,  $RR_{j2}^* - RR_{j1}^*$ ).

*Multiplicative interaction.* One solution is to assign numerical values to the levels of each factor that are proportional to their potencies – that reflect the psychological rather than physical magnitudes of the levels – that is, to scale the factor levels psychologically. The amount of modulation expected to be induced by an increase in  $FF$  (from  $FF_k$  to  $FF_{k+1}$ ) on an incremental effect of  $HD$  (e.g., as  $HD$  is increased from  $HD_j$  to  $HD_{j+1}$ ) should be proportional to the potency increases for both factors, and hence to the product of these increases. The amount of modulation by  $FF$  of an incremental effect of  $HD$  should thus increase *linearly* with the potency of  $FF$ ; implemented as suggested below, such an interaction is surprisingly general (footnote 58). This corresponds to a multiplicative interaction (a special case of monotone interaction) between the numerically scaled factor levels, numerical values that I denote “ $\mathcal{H}\mathcal{D}_j$ ” and “ $\mathcal{F}\mathcal{F}_k$ ”. (It is best to define the interaction in terms of the corresponding centered values,  $\mathcal{H}\mathcal{D}_j - \mathcal{H}\mathcal{D}_\bullet$  and  $\mathcal{F}\mathcal{F}_k - \mathcal{F}\mathcal{F}_\bullet$ .) Adding such a multiplicative interaction to an ordinary additive model has at least four advantages over other ways of assessing deviations from additivity: (1) It reflects the pattern of deviations (monotone interaction) that is of most interest in the present context. (2) It adds only one degree of freedom to the model. (3) The estimated size of the interaction (reflected by the coefficient of the interaction term in the model) is expressible as a single number that can be associated with a confidence interval that expresses what we know about it. (4) Because of (2), the multiplicative interaction can be fitted even in a case, like the present one, where we have only one observation per cell.

How should the levels of factors  $HD$  and  $FF$  be scaled to determine the values that define their potencies,  $\mathcal{H}\mathcal{D}_j$  and  $\mathcal{F}\mathcal{F}_k$ ? Given the additivity expected after a logarithmic transformation, one reasonable method is to use the main effect of each factor (the mean  $RR^*$  over the levels that happen to have been chosen of the other factor) to estimate the scale values of its levels. In the present example we would thus define  $\mathcal{H}\mathcal{D}_j \equiv RR_{j\bullet}^*$  and  $\mathcal{F}\mathcal{F}_k \equiv RR_{\bullet k}^*$ . An equivalent way to make this assignment is to fit an additive model, Model 1:

$$RR_{jk}^* = \mu + \alpha_j + \beta_k + \varepsilon_{jk}, \quad (13)$$

where one coefficient  $\alpha_j$  ( $\beta_k$ ) is associated with each level of  $HD_j$  ( $FF_k$ ). These estimated coefficients then determine the scaled values of the factors:  $\hat{\alpha}_j = \mathcal{H}\mathcal{D}_j - \mathcal{H}\mathcal{D}_\bullet$  and  $\hat{\beta}_k = \mathcal{F}\mathcal{F}_k - \mathcal{F}\mathcal{F}_\bullet$ . We then form a new variable, numeric rather than categorical, as the product  $\hat{\alpha}_j \hat{\beta}_k$ , and fit Model 2:

$$RR_{jk}^* = \mu + \alpha_j + \beta_k + \gamma \hat{\alpha}_j \hat{\beta}_k + \varepsilon_{jk}. \quad (14)$$

As the new variable is numeric, Model 2 requires only one additional parameter ( $\gamma$ ) to be estimated.<sup>57</sup> For the data in Fig. 13B,  $\hat{\gamma} = -0.12 \pm 0.07$  (not significantly different from zero). This test for non-additivity was first proposed by Tukey (1949), but motivated differently, and has been widely discussed.<sup>58</sup> One way to get a sense of the size of  $\hat{\gamma}$  is to consider the percentage change in the fitted separation between the top and bottom dotted lines in panel B (for  $FF_3$  and  $FF_1$ ) as  $HD$  changes from  $HD_1$  to  $HD_7$ , which is 6.5%. A second way is to note that if the dotted lines maintained their mean separation but converged fully at  $HD_7$  (23 h deprivation),  $\gamma$  would be 3.62, about 30 times as great. A third way is to note that whereas Model 1 explains 99.89% of the total variance of the  $RR_{jk}^*$  about their mean, Model 2 explains only 0.02% more.

### 15.2. Comments

Because Clark's data exemplify the simple multiplicative combination rule of Eq. (4), it was straightforward to apply a logarithmic transformation to permit also considering the data in relation to an additive model. I suggested that we should be asking about deviations between data and model that are *systematic*, and proposed that for an additive model the deviation pattern of greatest interest is a monotone interaction. While this question is relevant in any assessment of additive effects, it can be asked only if each factor takes on at least three levels. The method proposed for answering it is to scale the factor levels numerically according to their potencies (the scaling that renders their main effects linear), and then to assess the magnitude of a multiplicative interaction expressed by a single number.<sup>59</sup> A monotone interaction is perhaps the kind of deviation most plausibly associated with failure of the joint hypothesis, rather than with an experimental artifact. But it is not the only deviation of interest; it is important to know about any deviation that is consistent across subjects, or across data subsets otherwise defined. Deviations of any kind may lead to rejection of the joint hypothesis, or may indicate non-unitary factors (Appendix A.2.1), or other problems with the experiment.

<sup>57</sup> Such estimation is equivalent to regressing the residuals from Model 1 on just the interaction term of Model 2. In such a procedure the residual degrees of freedom must be adjusted, however, to acknowledge that  $\hat{\alpha}_j$  and  $\hat{\beta}_k$  were estimated from the data.

<sup>58</sup> Scheffé (1959, Section 4.8) shows that if the value of a general interaction term for a cell is a function of the main-effect estimates of the two factors for that cell, and if this function is assumed to be a general second-degree polynomial, then the polynomial must be the simple product we have discussed. Graybill (1976, Section 14.11) shows that with Gaussian-distributed errors, the residuals from Model 2 are independent of the main-effect estimates, which insures the validity of the usual variability estimates and tests even though the data themselves have been used to determine those main-effect estimates and thus to determine the numeric values of the added interaction factor. See also the "differential curvilinear trend test" of Abelson and Prentice (1997), of which the present test appears to be an extension.

<sup>59</sup> A similar approach is used in Section 7.2 in analyzing pure measures.

In Appendix A.15 I discuss a similar numerical scaling of factor levels (but one that linearizes their main effects on the untransformed *RR* data) and explain how Roberts (1987) used this representation to help discriminate among the simple multiplicative model of Eq. (4) and the three hybrid models expressed by Eq. (11).

Because the expectation of multiplicative factor effects depends on the contributions of the two processes **A** and **B** to  $M_{AB}$  being uncorrelated (*H5* in Table 4), the persuasive evidence provided by Clark's data for the effects of *HD* and *FF* being multiplicative supports that hypothesis (as well as *H1* and *H4*), thus strengthening the evidence for modularity in this case, as in Ex. 7 (see Appendix A.3.3).

## 16. Reaction time as a composite measure: selective effects of sleep deprivation (Ex. 10)

Suppose that the complex process between stimulus and response can be partitioned into stages. I mentioned in Section 3.1.1 that RT then becomes an example of a composite measure with summation as the combination rule. If the process is divided into two such stages **A** and **B** with durations  $T_A$  and  $T_B$ , then on any particular trial, these durations are their contributions to the composite measure, and  $RT = T_A + T_B$ . Thus, over a series of trials,  $\overline{RT} = \overline{T}_A + \overline{T}_B$ . It follows that if factors *F* and *G* selectively influence **A** and **B**, respectively, their effects on  $\overline{RT}$  should be additive; the  $\overline{M}_{AB}$  of Eq. (2) is the mean sum of the durations of **A** and **B**, and the equation becomes

$$\overline{RT}(F_j, G_k) = \overline{T}_A(F_j) + \overline{T}_B(G_k). \quad (15)$$

The AFM (see, e.g., Miller, 1988, 1993; Sanders, 1980, 1990, 1998; Sternberg, 1969, 1998b; and references therein) was devised to exploit this prediction from a stage model, given factors that influence the stages selectively; other inferences associated with the AFM concern the interpretation of factor interactions, which depend on the converse of the observation that factors that influence no stage in common must have additive effects on *RT*; it follows that factors found to interact must influence at least one stage in common.<sup>60</sup> By combining such inferences from a pattern of additive and interacting factor effects it is possible to draw conclusions about the underlying processing structure, as partially outlined in Tables 3 and 8.

As one part of his wide-ranging research, Andries Sanders, to whom the present volume is dedicated, has applied and advanced the AFM in important ways. He has provided integrative critical syntheses of large sets of findings, has brought out or suggested requirements and assumptions associated with the method as well as testing them, has analyzed its limitations and possibilities, and with his students and colleagues has contributed a number of interesting applications. One of the most provocative is described by Sanders et al. (1982, Experiment 1; see also Sternberg, 1998b, Section 14.5.1), and leads to the controversial conclusion that the effects of sleep deprivation are process-specific rather than global (Appendix A.1.2.2). Sanders

<sup>60</sup> For one view of the "logic" of the AFM, see Sternberg (1984, 1998b, Sections 14.3 & 14.6.3).

et al. used the AFM to examine the effects of experimentally induced sleep deprivation on different mental operations *within the same task*. What follows is a simplified description of their experiment and findings.

It is a feature of this example that distinguishes it from the others – the inclusion of more than two factors – that permits asking the question of process specificity. The stimuli were the single digits “2”, “3”, “4”, “5”; the responses were their spoken names, “two”, “three”, “four”, and “five”. Four factors were manipulated, each at two levels: The first was *SQ*; the digits, presented as dot patterns, could be *intact* or *degraded* by adding other dots. The second was the *MF*; it could either be *high* (respond to each digit with its name) or *low* (respond to “2”, “3”, “4”, “5” with “three”, “four”, “five”, “two”, respectively). The third was *sleep state (SLP)*, which was either *normal* (data taken during the day after a normal night’s sleep) or *deprived* (data taken during the day after a night awake in the lab). The measure was the RT for trials on which a response was made and was correct. Test sessions occurred in both the morning and afternoon, creating a fourth two-level factor, *time of day (TD)*, discussed below. The  $2^4 = 16$  conditions were run in separate blocks of trials. For simplicity, the data shown in Fig. 14 have been averaged over levels of *TD*.

Other studies had already suggested that *SQ* and *MF* were likely to influence two different processing stages selectively, processes that might be described as stimulus identification (**S**) and response selection (**R**), respectively.<sup>61</sup> If this were confirmed in the present experiment (by the additivity of effects of *SQ* and *MF*) then we could conclude that increasing the level of *SQ* (*MF*) adds to the difficulty of only **S** (**R**). Given the global-effect hypothesis about sleep deprivation, *SLP* should influence both **S** and **R**. Increasing the level of *SLP* should therefore exacerbate both kinds of difficulty: *SLP* should interact with both *SQ* and *MF* by amplifying their effects.

The results in Fig. 14 consist of the  $\overline{RT}$ s from the  $2 \times 2 \times 2 = 8$  conditions. These eight values are plotted in each of the three pairs of panels of Fig. 14 in different arrangements, to show the relations between the two factors in each of the three factor pairs. Panels A1 and A2 show that at each level of *SLP* there are additive effects on  $\overline{RT}$  of *SQ* and *MF*.<sup>62</sup> This evidence supports:

- (1) Performing the task involves at least two modules, arranged as stages.
- (2) Factors *SQ* and *MF* influence no stages in common.

Panels B1 and B2 show that at each level of *SQ* there are additive effects on  $\overline{RT}$  of *MF* and *SLP*. That is, the extra time a subject takes to execute an unfamiliar **S–R**

<sup>61</sup> This conclusion (separately modifiable sequential processes, or stages) has recently been further strengthened by analyses of complete RT-distributions rather than just RT means, from a similar experiment (Roberts & Sternberg, 1993, Experiment 2).

<sup>62</sup> Panels A1 and A2 show that the *SQ* effect decreases slightly with the more difficult mapping (not a statistically significant interaction). Because this study had only two levels per factor (a property shared with many other factorial experiments in psychology, and understandably, given the complexities of experimental design when conditions are blocked) we cannot decide whether this tendency is a systematic one, as was possible, e.g., in Ex. 9.

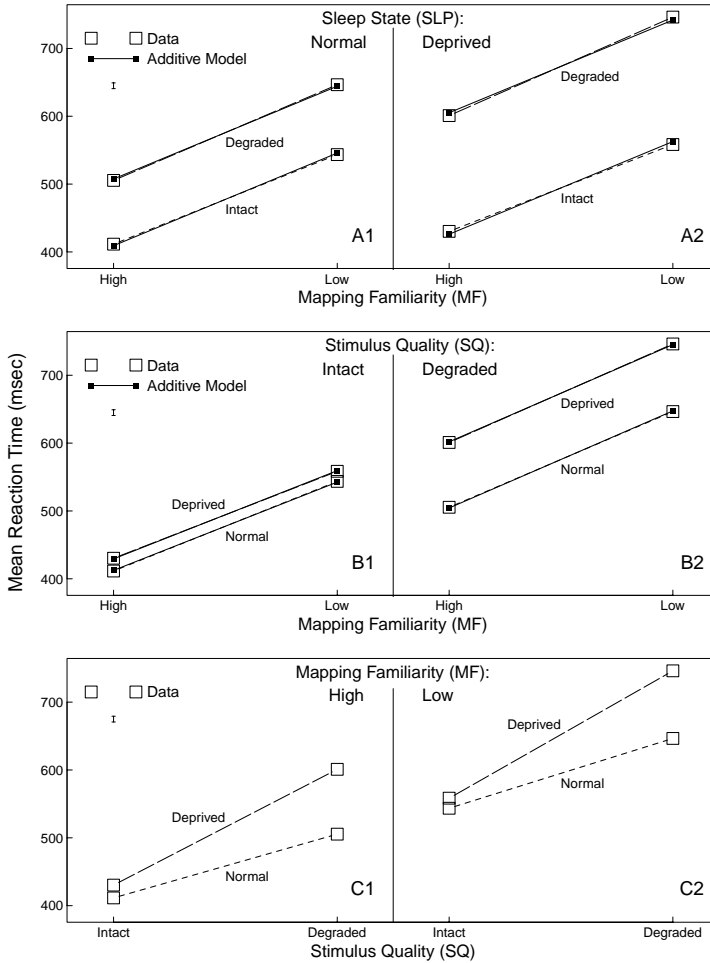


Fig. 14. Data from Sanders et al. (1982, Experiment 1). Means over the two levels of  $TD$ . Each pair of panels shows the same  $2 \times 2 \times 2 = 8$  data points, plotted in different ways. Each point is the mean of about 300 RTs from each of 16 subjects. A fitted additive model is also shown in each of the top four panels. Mean absolute deviations of data from model are 3.3 (panels A1, A2) and 1.0 ms (panels B1, B2). Because basic data are no longer available, values were obtained from Fig. 1 of Sanders et al. (1982). For the same reason, neither within-cell nor between-subject measures of variability are available. The  $\pm$ S.E. bars were therefore determined by separating the data by  $TD$ , fitting a model that assumes the additivity of  $MF$  with  $SQ$ ,  $SLP$ , and  $TD$ , and using the deviations (7 df) to estimate S.E.

association rather than a well-learned one is invariant over sleep states, rather than being increased by sleep deprivation. This evidence lends further support to (1) and also supports:

(3) Factors  $SLP$  and  $MF$  influence no stages in common.

Panels C1 and C2 show that at each level of  $MF$  there are interactive effects of  $SQ$  and  $SLP$ : increasing the level of  $SLP$  has a far greater effect on  $\overline{RT}$  when the stimulus

is degraded (98 ms) than when it is intact (17 ms). That is, sleep deprivation modulates the difficulty of stimulus identification. This evidence supports:

(4) Factors *SLP* and *SQ* influence at least one stage in common.

Taken together, the three pieces of evidence support a theory according to which the process used to perform the task contains at least two modules, **S** and **R**, these modules are arranged as stages (but see Appendix A.16.3), and among the factors *SQ*, *MF*, and *SLP*, *SQ* and *SLP* influence **S**, while *MF* alone influences **R**. Suppose the stimulus is identified during **S**, and the response is selected during **R**. (This is suggested by the nature of the factors *SQ* and *MF* that influence them.) We then arrive at the surprising conclusion that whereas *SLP* influences stimulus identification, it does not influence response selection. The reasoning here is an instance of the AFM: The additivity of the effects of *SQ* and *MF* and of *SLP* and *MF* support the hypothesis that the process by which the task is accomplished can be divided into at least two modules arranged in stages, and that particular factors influence their measures selectively; given this inference the interaction of *SLP* and *SQ* shows that *SLP* influences the same stage as *SQ*.

### 16.1. Comments

The main purpose of this paper has been to explore various ways in which separate modifiability can be used to reveal the modular structure of a process. Partly to simplify the exposition, I have considered primarily the decomposition of a process into just two modules, and have said little about how such analyses can be elaborated by further partitioning one or both of these modules, or by further characterizing them. For pure measures such elaboration is straightforward, as discussed in Section 5.1. For composite measures, however, it requires factorial experiments with three or more factors. Despite Fisher (1926, 1935) demonstration of the efficiency of such experiments, their obvious advantages for assessing generality, and their necessity for determining how the effects of different factors combine, they are often not used when they would be fruitful.

Unlike our other applications of composite measures, the sleep-deprivation findings not only demonstrate separate modifiability and thereby permit us to partition the **S-R** path into two modules (here, stages) **S** and **R** (the former selectively influenced by *SQ*, the latter by *MF*), but also extend the initial analysis, providing an example of localizing the influence of a third factor *SLP* in one of the identified modules, **S**, and thereby further characterizing **S** and **R**.

The pattern of effects of the fourth factor *TD* (not shown in Fig. 14) also permits further elaboration of the analysis. First, to justify averaging over its levels to create the figure, it is important that the pattern of effects of the other three factors is invariant across levels of *TD* (an instance of “stage robustness”, Sanders, 1998). Indeed, at each level of *TD*, *SQ* and *SLP* interact while *MF* is additive with both *SQ* and *SLP*. Second, *TD* itself interacts with *SLP* but is additive with *MF*,

indicating that it influences **R** but not **S** and, consistent with this conclusion, *TD* also interacts with *SQ*. (Such findings about *TD* are not logically required, but can instead be regarded as successful tests of the analysis based on the other three factors.)

There has been much discussion of alternatives to stage models for the interpretation of additive effects on  $\overline{RT}$ , as well as confusion about what a processing stage might be. I discuss these issues in Appendices A.16.2 and A.16.3. In Appendix A.16.1 I briefly describe a multiplicative combination rule for the pattern of response omissions in the present study, which has been found (Schweickert, 1985) to apply to the errors of commission generated by stage models. Another example of process decomposition using  $\overline{RT}$ , but supplemented by pure measures based on the LRP, is briefly described in Appendix A.6.

## 17. General discussion

A priori considerations (Section 1.1) argue for the idea that complex processes and devices may be composed of relatively independent parts, and psychological research has revealed abundant evidence for modular mental processes, composed of separately modifiable and functionally distinct sub-processes. Some of this evidence is included among the examples here discussed. One reasonable starting point for understanding a complex process is therefore to determine whether it is composed of such modules, and if so, try to identify them. And an initial step in doing so is to divide the process into just two parts (Section 1.3), perhaps followed later by dividing those parts in turn. Ultimately this should lead to analysis of the functions of each such module, as well as an understanding how they work together. Also, we would hope to find that different tasks are implemented by different subsets of the same set of basic modules, an idea associated with both the task-comparison method (Appendix A.1) and functional brain localization (Section 11).

In this paper I have explored the idea of separate modifiability as a criterion for modularity, by considering the shared and distinguishing features of a set of diverse examples. The examples fall into three categories, depending on whether the measures of the hypothesized modular processes are believed to be *pure* or *composite* and, among the pure measures, whether they depend on different data (*direct pure*) or different aspects of the same data (*derived pure*). In each case, evidence for separate modifiability consists of finding experimental manipulations (factors) that influence the hypothesized modules selectively. Such evidence consists of pure measures that are selectively influenced by the factors, or of a composite measure for which the effects of the factors combine appropriately. Selective influence is also, in itself, evidence for the modules being functionally distinct.

That the discovery and identification of modules is a central issue in psychology is reflected by the broad range of examples; they include different substantive areas, species, responses (e.g., stimulus-elicited vs operant), measures, combination rules,

experimental designs, and relations between modules; they lead to the decomposition of neural as well as functional processes. By considering such examples together, within a unified structure that highlights their inferential logic, we come to understand them better, and applications in one area benefit from lessons learned in another.

*Inferential logic.* In the course of this exploration I have suggested rational reconstructions of the inferential logic associated with the identification of modules, using measures hypothesized to be pure (Table 2) and composite (Tables 3, 4), with variant reconstructions in Appendix A.2.3. They are “rational reconstructions” because research is often not explicitly guided by such formal considerations, and experimental results may predate the formulation of such hypotheses. Attempts to make the inferential logic explicit should help in thinking about what a set of findings may mean, and may suggest preferences for some kinds of experiment over others.

Several lessons follow from the reconstructions: (1) Along with the researcher’s good fortune in finding suitable experimental manipulations (factors), what is tested in every example is a *joint hypothesis*: two or more distinct individual hypotheses, such that all gain support when evidence for modularity is revealed, while only one need be faulty for us to fail in an attempt to acquire such evidence. More generally, failure of the prediction of a joint hypothesis may be less informative than success. (2) While we are interested in *processes*, all we have is *measures* of them (Section 2.1); whereas change in a measure implies change in the process, the converse need not be true. (3) Nonetheless, to establish separate modifiability, *process invariance* is critically important; it is indicated by invariance of the value of a pure measure across levels of a factor, or invariance of the effect (or p.effect) on a composite measure of one factor across levels of another. Implications of such invariance are sufficiently strong that it should not be asserted simply because the evidence against it is unconvincing. The consequences of this for statistical tests and reliability assessment are mentioned in Section 1.5, Appendix A.11.2, and elsewhere (and illustrated by the use of the S.E.s of relevant contrasts); like much else in the present paper, these consequences need further consideration. (4) While modularity can be demonstrated (assuming sufficiently precise data), non-modularity cannot: as illustrated by Ex. 3, it is always possible that “we didn’t look enough” for appropriate experimental manipulations. Of course, the search for such factors is not blind, but depends on knowledge related to the functions that might be carried out by hypothesized modules; as attempts fail we should become increasingly convinced that these modules reflect an erroneous partition of the process under study, or that the process is simply too integrated to be modular. (5) A poor choice of measure or measures can cause a search for modules to fail (Section 13.3; Appendix A.7.2).

*Experimental design.* Another consequence of the importance of invariance is the desirability of increasing the chance of revealing any systematic violation, by using factorial experiments even with pure measures (illustrated in Section 7 and Appendix A.6, and discussed in Appendix A.9.1 and elsewhere), and by using factors with more than two levels (illustrated by Exs. 3, 6, 7, and 9, and discussed in Appendices A.2.1,



A.9.2, and elsewhere). Though not required with pure measures, factorial experiments permit testing the generality of findings of invariance. And factors with multiple levels permit separating systematic from non-systematic deviations, the former more damaging to claims of invariance. In Exs. 3 and 9, where multiple-level factors were employed, I illustrated the use of one-dimensional indices of invariance violation, based on numerically scaled factor levels, to create focused tests for one interesting and plausible kind of systematic deviation. We need to consider further the meaning of “systematic deviation”, as well as the merit of the meanings proposed in Sections 7.2 and 15.1.

*Brain measures.* The use of brain measures to identify modular neural processes may be informative about the structure of functional processes as well, but brain measures introduce complexities as well as opportunities. Some of the issues are discussed in Section 1.7 and Appendix A.1.2, and in the context of the examples (Sections 6, 10, 11, and 14; Appendices A.10 and A.11). It is advantageous to study effects of the same factors within the same task on both brain and behavioral measures (Sections 6, 10, and 11; Appendix A.6).

*Process decomposition vs task comparison.* Because the separate-modifiability approach to process decomposition has formal similarities to the popular task-comparison method and the associated pattern of double task-dissociation, I consider the latter method in Sections 1.6 and 11, and in Appendix A.1. The two methods have different purposes and strengths; I suggest that task comparison is not especially helpful in identifying the parts of a complex process.

*Limitations.* For simplicity, this paper is limited in at least two ways: First, in illustrating the inferential logic and the treatment of data I discuss the examples largely as if they are isolated cases, rather than considering them in relation to existing knowledge and related studies; conclusions normally depend on more context. Second, the examples are almost entirely restricted to dividing a complex process into only two parts. With a caveat (footnote 3), however, this is a reasonable starting point in decomposing a complex process, and because modules may themselves be modular, further partitioning can follow.

*Open questions.* Among others, these include the following: Are there useful further elaborations of the inferential logic? What is the best statistical method for testing a prediction of invariance or additivity? In which patterns of deviation should we be especially interested? How can we strengthen the inferences from brain measures to the structure of functional processes? What is the relation between the modularity of functional processes and of the brain processes associated with them? What are the relative merits of using pure vs composite measures? For composite measures are there other combination rules that might be of interest? Can we be more specific about what to do when a partitioning attempt fails? How interesting are cases of partial modularity (Section 2.3) or approximate modularity (Section 7.2)? Is separate modifiability too strong or too weak to be a useful criterion for partitioning a process? What are the relative merits of alternative criteria for modularity and alternative approaches to module identification? Does the present approach lead to modules that have other desirable properties? How does it compare to module discovery in other sciences?

## Appendices

The appendices that follow provide extensions, complements, technical details, and comments on the methods, and examples described in the body of this article. Each appendix (e.g., A.1, A.2) refers to a different main section of the paper (e.g., Sections 1, 2). There are no appendices corresponding to Sections 4, 5, 8, 12, or 14.

### A.1. Process decomposition vs task comparison

#### A.1.1. Introduction

The goal of the process-decomposition methods under discussion is to reveal the separately modifiable sub-processes (modules) of a single complex process. For this reason, one tries to find experimental manipulations (factors) that perturb or vary one such complex process, rather than using manipulations that change it into a qualitatively different process – one composed of a different set of modules. In almost all cases, a necessary condition for the complex process to be qualitatively invariant as factor levels change is for the subject's *task* (what the subject is instructed to do) to be kept invariant. However, task invariance in this sense is unfortunately not a sufficient condition: using a fixed set of instructions to define the task does not guarantee the desired qualitative invariance as factor levels are changed.<sup>63</sup> It is for this reason that the choice of factors has to be considered carefully, as discussed in Appendix A.2.

Because separate modifiability is demonstrated by finding factors that selectively influence sub-processes, these process-decomposition methods share some formal similarities with psychology's much-used "task-comparison" method, in which the interest is in finding factors that selectively influence different tasks – i.e., searching for dissociations between *tasks*, each accomplished by a qualitatively different complex process, rather than dissociations between the *sub-processes* of one complex process. Such comparison of distinct but related tasks, such as recall vs recognition has been an important tool in purely behavioral research (especially related to memory; see, e.g., Kahana, 2000; Kelley & Lindsay, 1996, pp. 37–41; Richardson-Klavehn & Bjork, 1988) and also in research on the relations between brain and

---

<sup>63</sup> The "task" of free recall of the words in a studied list, when time into the recall period is the factor, provides an interesting example. The subject responds to one set of instructions throughout the recall period during a trial. The dissociations between the recency and pre-recency part of the serial-position curve (reported as evidence for a short-term/long-term memory dichotomy by Glanzer & Cunitz, 1966) are therefore, nominally, within-task dissociations. However, many believe that there is a qualitative change in the retrieval process during a trial, from a process that involves output from a short-term store and affects the recency part of the curve, to a process that involves search of associative memory and affects the pre-recency part. (See, e.g., Raaijmakers & Shiffrin, 1981 and, for a different account, Howard & Kahana, 1999.) The dissociations found in free recall between early and late performance are therefore dissociations of two distinct complex processes, and exemplify findings from task comparison.

behavior (see Section 11). Because of their formal similarities, it is helpful to consider the task-comparison method and to distinguish it from process decomposition based on separate modifiability.<sup>64</sup>

### A.1.2. Qualitative task comparison

#### A.1.2.1. Behavioral studies of memory

Two tasks are dissociated if a factor  $F$  is varied in each, and is discovered to influence a measure of performance in Task 1 but not Task 2. For example, encoding manipulations during learning, such as variations in “level of processing” ( $F$ ) have been claimed to influence “direct” memory tasks such as recognition that a word is in a previously learned list (Task 1) but not “indirect” tasks such as identification of a degraded word (Task 2), revealing a qualitative *dissociation* between the tasks.<sup>65</sup> In the ideal case, a second factor  $G$  is discovered to influence performance in Task 2 but not Task 1, thus revealing a *double dissociation* between the two tasks.

If  $M_k^T$  for  $k = 1, 2$  is the measure of performance in Task  $k$ , then the four properties called “double dissociation” are formally the same as the four properties that define selective influence, described in the second part of Table 2 and discussed in Section 2.3. As emphasized by the superscript  $T$ , however,  $M_1^T$  and  $M_2^T$  are measures of performance in different tasks, and are assumed to be neither pure measures of particular sub-processes nor different composite measures of the same complex process. Nonetheless, if “process” is replaced by “task” in Table 2, the argument for the importance of demonstrating all four properties (Section 2.3) carries over, and explains the greater power of double dissociation over single.

Such dissociations, indicating the separate modifiability of classes of tasks, have been used as evidence for distinct memory systems (e.g., Tulving, 1983), but, as Richardson-Klavehn and Bjork (1988) point out, the observed patterns of separate and joint factor effects on memory tasks are complex, and the inference from such task dissociations to distinct memory systems is controversial. With Shoben and Ross (1986) they argue that findings of separately modifiable tasks can usually be explained without postulating distinct kinds of memory, and that inferences from such task dissociations can be strengthened by process decomposition of the individual tasks. One way to express the problem is that each task may be accomplished by a complex process composed of a set of several distinct operations, that inferences made from the task-comparison method often require valid theories of the complex processes used in each task, and that process-decomposition methods (such as those discussed in the present paper) can be helpful in developing and testing such task theories.

<sup>64</sup> Roberts (1987, pp. 165–167; 1993, p. 590) has also discussed this distinction.

<sup>65</sup> For a particular example see Jacoby and Dallas (1981); for a review see Richardson-Klavehn and Bjork (1988); for conflicting findings see Thapar and Greene (1994). Another such factor is the presence or absence of the brain damage associated with amnesia.

In their analysis of the logic of inferences based on task dissociations, Dunn and Kirsner (1988) argue that the likelihood of observing a dissociation is high only if two assumptions are satisfied: (1) Factors  $F$  and  $G$  can be found that selectively influence processes  $A$  and  $B$ , respectively (*selective influence of factors on processes*, Section 2.1). (2) Tasks 1 and 2 can be found such that  $A$  and  $B$  contribute selectively to their measures  $M_1^T$  and  $M_2^T$  (i.e.,  $M_1^T$  reflects  $A$  but not  $B$ , and vice versa for  $M_2^T$ ; *selective contribution of processes to task measures*). They also suggest that if either of these conditions is not met but a task dissociation is nonetheless discovered, the usual interpretation – that it excludes a single-process explanation of the two effects – would be incorrect. And they conclude (p. 97): “Because the exact relations between [factors] and processes and between processes and tasks are rarely known, indeed they are the primary object of investigation, [Assumptions 1 and 2] can rarely be confidently met.” The idea that particular pairs of tasks can be created that satisfy the selective contribution property (such as tasks in which performance depends on retrieval from implicit or explicit memory, but not both) has been questioned, e.g., by Jacoby, Toth, and Yonelinas (1993) and Richardson-Klavehn and Bjork (1988), who suggest, for example, that performance on “indirect” tests may reflect both.

#### *A.1.2.2. Effects of sleep deprivation*

Perhaps it is the difficulty of finding tasks that rely exclusively (or not at all) on particular processes that led to the dramatic conflict between inferences from task comparison vs process decomposition in attempts to understand sleep deprivation. The work of Sanders et al. (1982) provides persuasive evidence for process-specific effects: the sleep-state factor  $SLP$  appears to influence stimulus identification, but not response selection (Section 16). Additional support for a purely perceptual locus of sleep-deprivation effects is provided by other within-task process-decomposition data (Humphrey, Kramer, & Stanny, 1994) in which the occurrence time of the P300 peak in the ERP was used to partition the RT into early and late segments. On the other hand, the wide range of tasks in which performance is impaired by sleep deprivation has led others (who approached the question by considering its effects on performance in a range of tasks) to the opposite conclusion: According to Dinges and Kribbs (1991, p. 117), there is “a generalized effect of sleepiness on all cognitive functioning.” In this application, task comparison suffers from the possibility that there may be no task (in which performance can be measured) that is accomplished by a process containing no perceptual operations.

#### *A.1.2.3. Task-specific effects of localized brain lesions*

Another domain in which task comparison has been applied is in the study of the effects of localized brain lesions (e.g., Farah, 1984; Shallice, 1988), which sometimes appear to be remarkably selective, impairing performance (predominantly) on some tasks but not others. Of special interest in such investigations have been instances that approximate the ideal case of double task-dissociation, where (to simplify) a lesion in region  $R_\beta$  impairs performance in Task 1 ( $M_1^T$  reduced), but not in Task 2

( $M_2^T$  invariant), while a lesion in region  $\mathbf{R}_\gamma$  does the reverse. If the presence or absence of a lesion in  $\mathbf{R}_\beta$  ( $\mathbf{R}_\gamma$ ) is thought of as two levels of factor  $F$  ( $G$ ), then these instances fit the framework described in Appendix A.1.2.1. Such findings indicate that  $\mathbf{R}_\beta$  ( $\mathbf{R}_\gamma$ ) is necessary for Task 1 (2), provide evidence for localization of function within the brain,<sup>66</sup> and encourage hypotheses about distinct processes (each included among the component processes used to carry out one of the tasks) that might be implemented by neural processors in  $\mathbf{R}_\beta$  and  $\mathbf{R}_\gamma$ . (See Section 1.1; for some of the limitations of such inferences see Glymour, 1994, and Sarter et al., 1996.) The contrasting process-decomposition approach, with its advantages over task comparison, can also be used to investigate the process-specificity of the effects of localized brain lesions, by using a factor such as  $F$  in conjunction with task factors. A nice example is provided by Rubinstein, Meyer, and Evans (2000) and discussed by (Sternberg, 1998b, Section 14.5.6).

#### A.1.2.4. *Task-specific effects on localized brain activation*

A more recent domain of application is the use of measures of localized brain activation to study the brain functions that underlie the performance of different tasks. One method (complementary to the lesion method mentioned above) uses fMRI, which provides measures of regional blood oxygenation that reflect amounts of regional neural activity during performance of a task. Here, one popular approach is to ask how Task 1 differs from Task 0 (a “baseline” task) in the activation pattern it produces.<sup>67</sup> In a simple ideal case, there is just one region  $\mathbf{R}_\beta$  for which the activation measure  $\mathbf{M}_\beta$  increases as the task is changed from Task 0 to Task 1.<sup>68</sup> In an extension of such a finding to double dissociation, there is another region,  $\mathbf{R}_\gamma$ , not overlapping with  $\mathbf{R}_\beta$ , in which an activation measure  $M_\gamma$  increases as the task is changed from Task 0 to Task 2.

What we can learn from such findings depends on having hypotheses about how the processing operations differ among the three tasks (*functional task theories*), and about how the processing operations are reflected in neural activity (*neural task theories*). Thus, the functional theories for Tasks 0, 1, and 2 might say that the tasks are carried out by processes **A** and **D** (Task 0); **A**, **B**, and **D** (Task 1); and **A**, **C**, and **D** (Task 2). The neural theories would say that processes **B** and **C** are implemented by

<sup>66</sup> While it has been claimed that “lesioning” even a non-modular artificial network can produce double dissociations between tasks (e.g., Plaut, 1995), it has also been argued that this occurs only in relatively small networks (Bullinaria & Chater, 1995).

<sup>67</sup> McDermott, Buckner, Petersen, Kelley, and Sanders (1999) provide an example in which each of four tasks is compared to a baseline.

<sup>68</sup> The interpretation of the *existence* of an increment in  $M_\beta$ , i.e., a difference  $M_\beta(\text{Task 1}) - M_\beta(\text{Task 0}) > 0$ , is probably insensitive to exactly how  $\mathbf{R}_\beta$ -activation associated with Task 0 (the baseline) and  $\mathbf{R}_\beta$ -activation contributed by the task difference are plausibly combined. However, when the *magnitude* of the increment is considered, it is usually obtained by subtraction (resulting in the term “cognitive subtraction” for this method); hence, while seldom tested or even discussed, an additive combination rule is implicitly assumed. Without more secure knowledge of the combination rule, it would thus be difficult to decide whether an increment due to one factor is invariant with respect to another factor that influences activation in the same region.

neural processes  $\beta$  and  $\gamma$ , respectively. These neural task theories also incorporate the hypotheses that component processes are localized at a level that fMRI can resolve spatially, and that the set of processes described in the functional theories map one-to-one onto a set of (localized) processors; see Sarter et al. (1996). The functional theories also include the assumption, usually untested, that processes **A** and **D** remain the same, whether carried out alone or with **B** or **C** inserted: this is the assumption of “pure insertion” on which Donders’ (1868) subtraction method depends.<sup>69</sup> The neural theories also include the analogous assumption that the pattern of brain activation associated with processes  $\alpha$  and  $\delta$  remains invariant as regions  $\mathbf{R}_\beta$  or  $\mathbf{R}_\gamma$  are activated with the insertion of  $\beta$  or  $\gamma$ , respectively.<sup>70</sup> We can then conclude that while some of the neural processes that carry out Tasks 0, 1 and 2 (those that implement processes **A** and **D**) may be the same,  $\beta$  involves activation of region  $\mathbf{R}_\beta$  but not  $\mathbf{R}_\gamma$ , and  $\gamma$  involves activation of  $\mathbf{R}_\gamma$  but not  $\mathbf{R}_\beta$ .

However, such findings give us little evidence that helps us to discover the modular sub-processes of the complex process used to carry out either task. For example, in the idealized case above, whereas the task theories assert that Tasks 1 and 2 are each carried out by three distinct processes, there is nothing in the data that distinguishes between Task 1, for example, being accomplished by  $\alpha$ ,  $\beta$ , and  $\delta$  and its being accomplished by a “single” process,  $\theta_1$  (which differs from the corresponding  $\theta_2$  for Task 2 in one of the brain regions it uses). Other evidence is needed to distinguish between these possibilities, such as evidence from a process-decomposition method (where, for example, in a task carried out by  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$  it can be shown that the level of activation of region  $\mathbf{R}_\beta$  is selectively influenced by a factor  $F$  also shown to selectively influence **B**, and similarly for  $\mathbf{R}_\gamma$ ). Moreover, even if the task theories are correct, our data do not bear on the modularity of the hypothesized operations within either task: we have not learned whether any of them can be

<sup>69</sup> One test of pure insertion is provided if a Task 3 can be found, carried out by all four processes, **A**, **B**, **C**, and **D**. For Donders’ method, Taylor (1966) recognized the prediction  $\overline{RT}_3 - \overline{RT}_2 = \overline{RT}_1 - \overline{RT}_0$ , and tested it. For the analogous fMRI application, let  $A(x)$  be the activation level, as indexed by the strength of the fMRI signal, at any location  $x$  where it is measured. The corresponding prediction is then  $A_3(x) - A_2(x) = A_1(x) - A_0(x)$ , which should be satisfied for each  $x$ . (Unlike the case of  $\overline{RT}$ , where there is only one test for a set of four tasks, here there are as many tests as locations  $x$ .) To carry out such tests it is critical for the presence of processes  $\beta$  and  $\gamma$  to be varied factorially. Suppose instead that we had no Task 2, but just a series of tasks of increasing complexity, Task 0 ( $\alpha, \delta$ ), Task 1 ( $\alpha, \beta, \delta$ ), and Task 3 ( $\alpha, \beta, \gamma, \delta$ ). For such a case, Posner and Raichle (1995, p. 374) suggest testing the prediction  $A_3(x) - A_0(x) = [A_1(x) - A_0(x)] + [A_3(x) - A_1(x)]$ . This test would always be satisfied, however, as it is an algebraic identity. For more on the pure insertion assumption, see Sternberg (1998b, Section 14.4.2), and references therein.

<sup>70</sup> Also relevant to the interpretation of such data are: (a) The fMRI signal is increased by the activity of the numerous inhibitory neurons in the cortex, as well as the excitatory ones (Raichle, 1998), but the relation between these two effects seems not to be known. (It therefore seems possible for the signal to increase when a process is actually deleted.) (b) We do not know the extent to which the fMRI signal reflects neural activity that produces no action potentials (Wandell, 1999, p. 148).

modified at all, or whether they can be modified selectively, nor, if they can be, by which factors.

Thus, whether it is used to study behavior or brain function, qualitative task comparison does not readily permit inferences about how a complex process should be divided into parts.

#### *A.1.3. Quantitative task comparison: subtraction and division methods*

In some cases researchers are willing to create sufficiently detailed theories of two tasks such that, if the theories are correct, quantitative properties of sub-processes (i.e., derived pure measures) can be estimated from the performance measures in the two tasks. The goal here is typically not to test the task theories, but rather to use them to provide such estimates.

##### *A.1.3.1. Donders' subtraction method*

Perhaps the most venerable task-comparison method of this kind is Donders' (1868) *subtraction method*, for two tasks, Tasks 1 and 2, in which only composite measures,  $M_1^T$  and  $M_2^T$  (here, mean RTs) are available. The joint hypothesis consists, first, of a pair of task theories – *HY1* and *HY2* – that specify the constituent processes of each task, and second, *HY3*, a combination rule. For example, we might have the following hypotheses:

*HY1*: Task 1 is accomplished by process **A** (which may consist of more than one sub-process).

*HY2*: Task 2 is accomplished by processes **A** and **B**, where **A** is identical (at least in duration) to the corresponding process in Task 1 (the “pure insertion” assumption).

*HY3*: Contributions of the processes to the composite measures combine by summation (as in Table 3, and implied by Donders' assumption that the processes are arranged in stages).

Given these hypotheses, it follows that  $M_1^T = \overline{RT}_1$  is an estimate  $\hat{t}_a$  of the mean duration of **A**,  $M_2^T = \overline{RT}_2$  is an estimate  $\hat{t}_a + \hat{t}_b$  of the sum of the mean durations of **A** and **B**, and therefore, by subtraction,  $\hat{t}_b = M_2^T - M_1^T$  is an estimate of the mean duration of **B**.<sup>71</sup>

##### *A.1.3.2. Jacoby's division method*

The “process-dissociation” method devised by Jacoby (1991) for separating the contributions of conscious and unconscious memory to performance in memory tasks has aroused considerable interest (see, e.g., Buchner, Erdfelder, & Vaterrodt-

<sup>71</sup> Unlike the methods exemplified in Sections 13–16, in which separate modifiability is tested with composite measures (where no pure-insertion assumption is required, and where a successful test supports the combination rule), application of the subtraction method has often embodied a test of neither the combination rule nor pure insertion. That is, the task theories are assumed but not tested, even though such tests may be possible (footnote 69; Appendix A.1.3.3).

Plünnecke, 1995; Jacoby et al., 1993; McBride & Doshier, 1999) and can be regarded as a modern “division method”, by analogy to the subtraction method, but with a combination rule of multiplication (as in Table 4) rather than summation (as in Table 3). Development of the method was motivated by the idea that qualitative uses of the task-comparison method for “direct” and “indirect” memory tasks (such as old-new recognition and word-fragment completion, respectively) were limited by the incompleteness of our knowledge about the complex processes associated with each such task, and that performance measures in the two tasks were probably not pure measures of “explicit/controlled/conscious” memory on the one hand, and “implicit/automatic/unconscious” memory on the other. The goal was to develop pairs of tasks, each of which is acknowledged to reflect contributions of both kinds of memory, but constructed such that data from the two tasks can be combined to provide separate estimates of the two contributions.

In a simplified illustrative experiment, subjects are first exposed to a list of words, and then given a mixture of two kinds of stem-completion trials, for which the list is a possible source of completions. On Task IN (“inclusion”) trials, subjects are asked to complete the stem to produce any word that comes to mind (including a word from the list). On Task EX (“exclusion”) trials, subjects are asked to complete the stem to produce any word other than a word from the list. The tasks are assumed to depend on retrieval from both conscious memory (process C) and unconscious memory (process U), but in different ways.

As in the subtraction method, the joint hypothesis consists of a pair of task theories and a combination rule. In a simple version of the method, the hypotheses are:

*HY4*: For Task IN, a completion is generated if and only if U and/or C succeed in producing one.

*HY5*: For Task EX, a completion is generated if and only if U succeeds in producing one and C fails, where U and C are identical (at least in success rate) to the corresponding processes in Task IN.

*HY6*: Generation of completions by U and C are uncorrelated events.

Given this joint hypothesis, it will be seen that the measures defined below for the two tasks can each be thought of as estimates of the probabilities of the joint occurrence of two uncorrelated events, thus depending multiplicatively on the individual event probabilities, as in Ex. 7.<sup>72</sup> Let  $P_{IN}$  and  $P_{EX}$  be the probabilities of completion in Task IN and Task EX, respectively. Let  $p_u$  and  $p_c$  be the respective probabilities that U and C succeed in generating completions. And let  $Q_{IN} = 1 - P_{IN}$ ,  $q_u = 1 - p_u$ , and  $q_c = 1 - p_c$  be the probabilities of failure. For Task IN, no completion will be generated if and only if both U and C fail. Hence

<sup>72</sup> The simple task theories described here have had to be complicated by the incorporation of baseline, guessing, and bias effects (Buchner et al., 1995). Moreover, hypothesis *HY6* (on which the multiplicative combination rule depends) is controversial (see, e.g., Curran & Hintzman, 1995).



$$M_{IN}^T \equiv Q_{IN} = q_u \times q_c. \quad (16)$$

For Task EX, a completion will be generated if and only if **U** succeeds and **C** fails. Hence

$$M_{EX}^T \equiv P_{EX} = p_u \times q_c. \quad (17)$$

Dividing the measure for Task IN by the measure for Task EX to form the ratio  $R^T$  (hence “division method”),  $q_c$  is eliminated, and we have a pure measure of  $p_u$ :

$$R^T \equiv \frac{M_{IN}^T}{M_{EX}^T} = \frac{q_u}{p_u} = \frac{1 - p_u}{p_u}, \quad (18)$$

or  $\hat{p}_u = (1 + R^T)^{-1}$ , and it follows from this and Eq. (16) that  $\hat{q}_c = M_{IN}^T + M_{EX}^T$ . We thus have estimates for the underlying processes **U** and **C**, respectively.<sup>73</sup>

#### A.1.3.3. Derived pure measures from subtraction and division methods

In applying Donders’ or Jacoby’s method to obtain estimates of process parameters from measures in two different tasks, one is not testing the assumed task theories. However, assuming the theories to be valid, the estimates can be regarded as derived pure measures of the hypothesized processes to which they correspond. Thus, for the subtraction method,  $M_A = M_1^T$ , and  $M_B = M_2^T - M_1^T$ . And for the division method,  $M_U = R^T$ , and  $M_C = S^T$ , where  $S^T \equiv M_{IN}^T + M_{EX}^T$  is the sum of the two task measures. The response of these pure measures to changes in the levels of suitable factors can be studied, just as described in Table 2, to test the hypothesized processes for separate modifiability. And, insofar as factors are found that influence them selectively, we have support for the task theories that justify the pure measures, as well as for separate modifiability. For example, for the subtraction method, (1) a factor that selectively influences **B** should influence only  $M_2^T$ , and (2) the effects on  $M_1^T$  and  $M_2^T$  of a factor that selectively influences **A** should be equal. (The second property depends jointly on – and hence supports – the pure-insertion assumption together with separate modifiability.)

The important difference between these quantitative task-comparison methods and Exs. 1-6 is that here the data are from two separate tasks, each believed to involve a different complex process. What makes these methods possible is that the theoretical starting points on which the inferences depend are strong, stronger than those in Exs. 1-6: a pair of task theories that specify the exact relationship between the complex processes in two different tasks.

<sup>73</sup> This treatment of Jacoby’s method is equivalent to the one often provided, in which  $M_{in}^T = 1 - M_{IN}^T$ ,  $M_{ex}^T = M_{EX}^T$ , and  $p_c = M_{in}^T - M_{ex}^T$ .

## A.2. Considerations in the choice of factors

### A.2.1. Unitary factors and qualitative process invariance

Our goal is to understand the structure of a complex process – in particular, to identify its modules (if any); our strategy is to vary experimental conditions (factor levels) so as to influence that process. Interpretation of the findings depends on the process being the “same” (in a poorly defined sense) under all the conditions tested. If a factor-level change replaces one operation by (or adds) a qualitatively “different” one, then we are comparing two different processes rather than revealing the structure of one. (Another way to phrase this is that a change in factor level should not change the subject’s “task” into a qualitatively different one. If so, we would be engaged in task comparison rather than process decomposition.) Accordingly, across the ranges of factor levels examined, it should be possible to make a persuasive argument that the process is the same, in the sense of including the same set of processing operations; the strength of our conclusions depends on the persuasiveness of that argument.<sup>74</sup> Such an argument can be based on prior knowledge of the process and the factors; another basis is the pattern of factor effects. The concept of a *unitary factor* can be helpful here: A factor is unitary within the selected range of levels if each change in level influences the same operations and leaves the same other operations invariant. If a factor is unitary, its effects should be qualitatively the same across more than one change in its level, and across several levels of the other factor or factors.<sup>75</sup> This is one of the reasons for selecting factors that can have more than two levels, and for using more than two levels when feasible, despite the complexities this may add to the experimental design. Other reasons are discussed in the context of the examples and in Appendix A.9.2.

### A.2.2. Manipulated vs selective factors

Even if a process is modular, some factors may not reveal this. For example, a factor that influences a module selectively may not be one of those manipulated or controlled in an experiment, but may instead be a function of the levels of two factors that are manipulated, which might then be found to have non-selective effects. As discussed in Section 7, this may have occurred in applications of signal-detection theory to “yes–no” experiments, in which the factor that may influence the decision process selectively is *RR* rather than *PM*. Because *RR* depends on signal discriminability ( $S_T$ , an s-factor) as well as on *PM* (a d-factor), ostensibly pure measures of sensory and decision processes failed to satisfy the requirements of separate modifiability when factors  $S_T$  and *PM* were used, but appear to approximate them well with factors  $S_T$  and *RR*. Thus, in interpreting a test of separate

<sup>74</sup> Factors that determine the number of multiple similar operations (such as the number of non-targets in some search tasks) are unlikely to be regarded as “changing the task” in this sense.

<sup>75</sup> See footnote 63 for a factor (time during free recall) that may be non-unitary. For an example of extending the range of a factor such that it becomes non-unitary, see Sternberg (1998a, p. 389).

modifiability, it has to be kept in mind that the manipulated factors may not be the ones that influence the hypothesized modules selectively, even when such factors exist.

### A.2.3. Inferential logic when hypotheses about modules specify the factors

If the hypotheses about processes **A** and **B** are sufficiently detailed, they may specify particular factors that should influence **A** and **B** selectively. This possibility is not well represented by the formulations in Tables 2–4, which treat *F* and *G* as members of a larger set of plausible candidates to be tested. In alternative formulations of the inferential logic, the specification of factors *F* and *G* can be included in the joint hypothesis, with the remainder of the reasoning adjusted accordingly. For Tables 2 and 3, this is shown by Tables 7 and 8, respectively, in which I have replaced *H1* by *H1\**, which is stronger than *H1* and implies it; a similar reformulation can be created for Table 4.

Given the initial conceptualization of band-limited analyzers, Table 7 is probably more appropriate than Table 2 for Ex. 4. For the same reason, a reformulation of Table 4 is probably more appropriate for Ex. 7. And similarly, Table 8 is probably more appropriate than Table 3 for the part of Ex. 10 involving *SQ* and *MF*, given that the two hypothesized modules are **S** and **R**.

Table 7  
Alternative inferential logic for pure measures<sup>a</sup>

<b>Joint Hypothesis</b>	
<i>H1*</i> : Processes <b>A</b> , <b>B</b> are influenced selectively by factors <i>F</i> , <i>G</i> :	
$p'_1: \mathbf{A} \leftarrow F, p'_2: \mathbf{B} \leftarrow F, p'_3: \mathbf{B} \leftarrow G, p'_4: \mathbf{A} \leftarrow G.$	
<i>H2</i> : $M_A, M_B$ are pure measures of <b>A</b> , <b>B</b> .	
<b>Prediction</b>	
<i>F</i> and <i>G</i> influence $M_A$ and $M_B$ selectively:	
$p_1: M_A \leftarrow F, p_2: M_B \leftarrow F, p_3: M_B \leftarrow G, p_4: M_A \leftarrow G.$	
<b>Alternative Results</b>	
<i>F</i> , <i>G</i> influence $M_A, M_B$ selectively.	They do not.
<b>Inferences</b>	
Support for joint hypothesis <i>H1*</i> + <i>H2</i> .	Refutes one/both of <i>H1*</i> , <i>H2</i> .

<sup>a</sup> See footnote 14 for the distinction between  $p'_k$  and  $p_k$ .

Table 8  
 Alternative inferential logic for a composite measure with summation as the combination rule

<b>Joint Hypothesis</b>	
<p><i>H1*</i>: Processes <b>A</b>, <b>B</b> are influenced selectively by factors <i>F</i>, <i>G</i>:</p> <p><math>p'_1: \mathbf{A} \leftarrow F, p'_2: \mathbf{B} \leftarrow F, p'_3: \mathbf{B} \leftarrow G, p'_4: \mathbf{A} \leftarrow G,</math></p> <p>which jointly influence no other process.</p> <p><i>H3</i>: Contributions <math>u_A, v_B</math> of <b>A</b>, <b>B</b> to <math>M_{AB}</math> combine by <i>summation</i>.</p>	
<b>Prediction</b>	
<p>The effects of <i>F</i> and <i>G</i> on <math>M_{AB}</math> are additive.</p>	
<b>Alternative Results</b>	
<p>Additive effects of <i>F</i>, <i>G</i> on <math>M_{AB}</math>.</p>	<p>Non-additive effects.</p>
<b>Inferences</b>	
<p>Support for joint hypothesis</p> <p><i>H1*</i> + <i>H3</i>.</p>	<p>Refutes one/both of</p> <p><i>H1*</i>, <i>H3</i>.</p>

### A.3. Composite measures, combination rules, and stochastic independence

#### A.3.1. Measures, combination rules, and plausibility

In the illustration of Section 3.1.1, we reasoned from an hypothesized process structure or “mental architecture” (here, stages) to a combination rule (here, summation) that applies to a particular measure  $M_{AB}$  of the complex process (here, duration). It is plausible that the duration of a process would be the sum of durations of its parts, and we can show (Eq. (3)) that, with selective influence of *F* and *G*, we expect their effects on  $\overline{RT}$  to be additive. But we would probably not be interested in considering an application of Table 3 in which  $M_{AB}$  was a non-linear transformation of  $RT$ , such as its square root. Why not? Because it is hard to think of a plausible mechanism in which contributions of two separately modifiable processes to  $\sqrt{RT}$  would combine by summation. Conversely, suppose we discovered that the effects of two factors on  $\sqrt{RT}$  are additive in a particular experiment. This finding can be used as an argument for modular processes (selectively influenced by those factors) only if there is a plausible mechanism that is consistent with summation being the combination rule for that measure; but there appears to be no such mechanism.

For Exs. 7 and 9 there are plausible structures that lead to multiplicative combination rules for certain measures. Section 3.2 shows that such a rule implies

multiplicative factor effects and, also, that if such measures are transformed by taking logarithms, the factor effects become additive. Here, finding additive effects on the logarithmically transformed measure is relevant to the discovery of modules because there is a plausible modular process structure which, when combined with selective influence, would lead to additive effects on that measure.

In Section 13 we saw that a plausible detection mechanism involving two modules,  $\mathbf{A}, \mathbf{A}'$ , leads to a multiplicative combination rule for their contributions to  $M_{AA'} = P_{ND}$ , the probability of not detecting a signal. In contrast, we shall see in Appendix A.13 that under some conditions, the transformed measure  $Odds(P_{ND}) = P_{ND}/(1 - P_{ND})$  reveals multiplicative effects of two factors, and therefore that  $logit(P_{ND}) = \log[Odds(P_{ND})]$  exhibits additive effects. (Actually,  $P_{ND}$  must be “corrected for guessing” for the *Odds* transformation to work in this way.) Here, whereas it is easy to explain such additive effects on this transformation of  $P_{ND}$  as arising if both factors influence a particular plausible single module, there appears to be no plausible structure of two modules selectively influenced by the two factors that could also explain it.

Hence, whereas suitable transformation of a measure may render factor effects additive or multiplicative, and whereas this can be interesting, it may not support a claim of modularity. In general, data may be transformed so as to create measures that behave in orderly, simple ways, but such patterns become important for the inference of modularity only if a plausible modular mechanism can generate them. Of course, what is plausible is subject to change.

#### *A.3.2. Multiplication as the combination rule: implications for data analysis of the zero-correlation requirement*

The requirement expressed in hypothesis *H5* (Table 4) has implications for how data should be treated so as to minimize the likelihood and/or extent to which failures of that hypothesis interfere with testing *H1* and *H4*. Consider the application discussed in Section 15, where the measure is the response rate of a rat in a bar-pressing experiment. For ease of exposition, let our measure instead be the inter-response time, *IRT*, the reciprocal of response rate. The data for a particular pair of factor levels is a summary of the bar-press times on each of the set of trials at those levels. One obvious approach to creating that summary would be to determine the mean *IRT* for each trial, and then average those means over that set of trials. At this point one could apply the logarithmic transformation and test for additivity of factor effects.

In relation to *H5* there are at least two potential sources of correlated variability of  $u_A$  and  $v_B$ : variation across the *IRT*s that comprise a trial, and variation across the trials that comprise the set of trials. If these are substantial and we use the method described above, *H5* will be violated. However, by applying the logarithmic transformation to *individual IRT*s (before any averaging), so that the averaging is of sums rather than products of the contributions of  $\mathbf{A}$  and  $\mathbf{B}$  to the (transformed) measure, we render these two potential sources of correlated variation irrelevant (see, Eqs. (1) and (2)). Whether this eliminates *all* potential sources of correlated

variation depends on details of the mechanism: If single values of  $u_A$  and  $v_B$  determine any particular *IRT*, then there is no other source, Eq. (4) applies to individual *IRT*s, and the need for *H5* in testing independence in mean is eliminated.<sup>76</sup> In general, by applying the logarithmic transformation as early as possible – before any explicit summing or averaging operations – we are likely either to reduce the impact of violations of *H5* in testing *H1* and *H4* or eliminate the need for *H5* altogether.<sup>77</sup>

### A.3.3. Stochastic independence of process contributions as further evidence of modularity

In the discussion of composite measures thus far we have been concerned with whether two processes obey “independence in mean” – that is, whether their *mean* contributions to a composite measure can be separately modified. Whatever the processes and whatever the measure, it is highly likely that their contributions (and the measure itself) have non-zero variability. This raises the question whether such variation is correlated across the two processes. The evidence for the modularity of these processes would be strengthened if their contributions to the measure were found not only to be independent in mean, but also to be *stochastically* independent (or even if they were found to be merely uncorrelated – often a weaker property).<sup>78</sup> As we saw in Section 3.2, the assumption that contributions are uncorrelated (have zero covariance) is *required* when the combination rule is multiplication (*H4* in Table 4), even to get predictions about the pattern of mean p.effects. (Finding multiplicative p.effects therefore supports *H5*; Exs. 7, 9.) When the combination rule is summation, however, the zero covariance or (stronger) stochastic independence property can be tested separately from independence in mean. Thus, if the contributions are uncorrelated, then the combination rule for their contributions to  $var(M_{AB})$  is also summation, and we should find additive factor effects on  $var(M_{AB})$  as well as on its mean; if the contributions are stochastically independent, this implies not only the latter, but also additional powerful constraints on the data from a factorial experiment (Roberts & Sternberg, 1993; Sternberg, 1969). If meaningful pure measures are obtained on individual trials, questions about independence can be asked for such measures also.

If one process directly influenced another (Section 2.1) then they would not be separately modifiable, and we would expect their pure measures (if based on individual trials), or their contributions to composite measures, to be correlated. It has

<sup>76</sup> Alternatively, suppose each *IRT* comprises a sequence of unobserved sub-intervals, each determined by a possibly different  $(u, v)$  pair. Eq. (4) then applies to each sub-interval, but, unless *H5* is true across the sub-intervals of an *IRT*, Eq. (4) does not apply to the *IRT*, the sum of these sub-intervals.

<sup>77</sup> For this approach to be used, all values of the measure to which it is applied must be positive; this excludes cases like Ex. 7.

<sup>78</sup> In research on memory, the relevance of stochastic independence has been recognized for measures of different tasks (Appendix A.1.2.1). Interpretations differ, however; see Kahana (2000) and Richardson-Klavehn and Bjork (1988).

to be recognized, however, that there may be factors that influence both of the two processes in question, are not under the experimenter's control, and whose levels fluctuate from trial to trial. Because this can induce covariation of their contributions, the predictions mentioned above can fail, even when the two processes are modular.

### A.6. Evidence for neural and functional modules that correspond

A remarkable study by Smulders, Kok, Kenemans, and Bashore (1995) provides a supplement to Exs. 2, 5, 6, and 10. In relation to Ex. 2, their findings are among those that give us reason to believe that the onset time of the lateralized readiness potential (LRP) is psychologically meaningful. In relation to Ex. 5, their experiment is an analog in which an aspect of the ERP is used, instead of single-neuron activations, to index a possible boundary between neural processing stages. As in Ex. 10, their RT data can be used as a composite measure to test a hypothesis about modular functional processes arranged as stages. And in relation to Ex. 6, their findings provide evidence for corresponding neural and functional modules from brain and behavioral measurements, respectively, as advocated in Section 11.<sup>79</sup>

In a two-choice RT experiment with single-digit stimuli mapped on left-hand and right-hand responses, Smulders et al. varied two factors,  $SQ_j$  (digit intact vs degraded) and *response complexity*,  $RC_k$  (one keystroke vs a sequence of three keystrokes made by fingers of the responding hand). They measured  $RT_{jk}$ , and also the onset time of the LRP, based on both stimulus-locked ( $LRP_s$ ) and response-locked ( $LRP_r$ ) averaging of the scalp-potential functions. (See Section 6, where we consider only  $LRP_s$ , based on scalp potentials treated as functions of time from stimulus onset.) Let  $T_{sjk}$  and  $T_{rjk}$  be the corresponding LRP onset times measured from the stimulus, and let  $T_{\bullet jk}$  be their mean. Averaging over the four conditions,  $RT_{\bullet\bullet} = 416$  ms, and  $T_{\bullet\bullet\bullet} = 264$  ms. If  $\alpha$  is the process from stimulus to LRP onset, and  $\beta$  is the process from LRP onset to response, then these values give us measures of their mean durations (direct pure measures of  $\alpha$  and  $\beta$ ):  $D_\alpha = T_{\bullet\bullet\bullet} = 264$  ms and  $D_\beta = RT_{\bullet\bullet} - T_{\bullet\bullet\bullet} = 152$  ms.

The LRP data  $\{T_{\bullet jk}\}$  are separated by condition in Fig. 15. Because  $T_{sjk}$  and  $T_{rjk}$  give similar estimates for effects of the two factors on  $D_\alpha$  and  $D_\beta$ , the estimates are based on  $T_{\bullet jk}$ . The evidence indicates that the two factors  $SQ$  and  $RC$  have selective effects on  $D_\alpha$  and  $D_\beta$ , supporting the hypothesis that in this situation the LRP onset

<sup>79</sup> The error variance values reported by Smulders et al. (1995) and the S.E. estimates provided here are likely to be overestimates (because balanced condition-order effects were treated as error variance), and the data required to calculate better values are no longer available (F.T.Y. Smulders, personal communication, 1999).

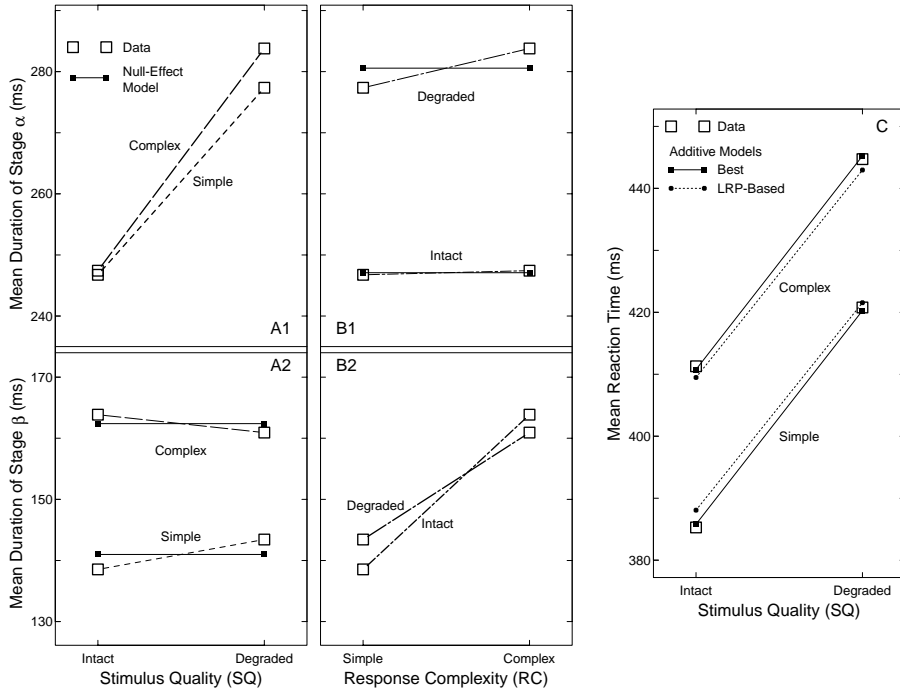


Fig. 15. Means over 14 subjects of data from Smulders et al. (1995). Estimated durations  $D_\alpha$  of Stage  $\alpha$ , from stimulus presentation to LRP onset (panels A1, B1); and  $D_\beta$  of Stage  $\beta$ , from LRP onset to response (panels A2, B2). These are shown as functions of  $SQ$  (panels A1, A2), and of  $RC$  (panels B1, B2). Data in panels A1 and A2 are separated by level of  $RC$ ; those in panels B1 and B2 are separated by level of  $SQ$ . Also shown in panels A2 and B1 are null-effect models. Main effects of  $SQ$  on  $D_\alpha$  and  $D_\beta$  are  $34 \pm 6$  (panel A1) and  $1 \pm 8$  ms (panel A2), respectively; the corresponding main effects of  $RC$  are  $4 \pm 8$  (panel B1) and  $21 \pm 7$  ms (panel B2), respectively. The  $\overline{RT}$  data are shown in panel C, together with two fitted models. One is the best-fitting additive model (mean absolute deviation 0.5 ms); the other is an additive model based on the LRP data (see text; mean absolute deviation 1.8 ms).

indeed defines a boundary between two neural processing modules that are arranged as stages, and consistent with the reasoning in Table 2.

The same study permitted Smulders et al. to use the composite measure  $\overline{RT}$  to ask about the functional structure of the entire process (Section 3.1.1, Table 3). If the RT is generated by two processes **A** and **B** arranged in stages, then it may be possible to find factors that have additive effects on  $\overline{RT}$ . Indeed,  $SQ$  and  $RC$  are such factors, as shown by the excellent fit of the best-fitting additive model displayed in Fig. 15C: the interaction of  $SQ$  and  $RC$  was a negligible  $2 \pm 5$  ms.

Finally, suppose that the (functional) processes **A** and **B** responsible for the additive RT effects are implemented by  $\alpha$  and  $\beta$ , respectively, the (neural) processes demarcated by the LRP. Then, not only is it unsurprising that the same factors influence them selectively, but also, the sizes of their effects should be the same (ignoring some possible problems in estimating  $T_s$  and  $T_r$ ). Agreement among the effect-sizes can be examined by assuming that the two factors indeed have perfectly



selective effects on the  $\alpha$  and  $\beta$  processes demarcated by the LRP, and by using the appropriate subset of the LRP data to “predict” the pattern of the  $\{\overline{RT}_{jk}\}$ . Thus, we should be able to use just the data in panel A1 (averaging over  $RC$  levels) to estimate  $D_\alpha(SQ_1)$  and  $D_\alpha(SQ_2)$ . Similarly, we should be able to use just the data in panel B2 (averaging over  $SQ$  levels) to estimate  $D_\beta(RC_1)$  and  $D_\beta(RC_2)$ . If  $RT = D_\alpha + D_\beta$  we have the “predictions”  $RT_{jk}^* = D_\alpha(SQ_j) + D_\beta(RC_k)$  for the four conditions.<sup>80</sup> Fig. 15C shows that the agreement is good:  $RT_{jk} \approx RT_{jk}^*$ . Numerically, the effects of  $SQ$  and  $RC$  on the composite measure  $\overline{RT}$  are  $35 \pm 3$  and  $25 \pm 7$  ms, respectively, close to their effects (34 and 21 ms) on the pure measures  $D_\alpha$  and  $D_\beta$ .<sup>81</sup>

This study is distinguished from the other examples in the present paper by its independent use of two methods: pure measures to ask about the structure of a neural process, and a composite measure to ask about the structure of a corresponding functional process.

## A.7. Details of the analysis of brightness discrimination by pigeons

### A.7.1. Support for the equal-variance Gaussian detection model

With six birds and five levels of  $LR$ , 30 ROCs are defined, each with three points, one for each level of  $RR$ . The slope of each ROC was estimated by fitting a line to the first and third point. The mean slope of this line is  $0.99 \pm 0.06$ , and there is no significant effect of discriminability on slope. The deviation from linearity was measured by determining the perpendicular distance (in  $z$ - $z$  space) from the second point to the fitted line, and giving it a sign, positive if the second point was above the line, and negative otherwise. The mean *signed distance* is  $-0.03 \pm 0.05$ , and there is no significant effect of discriminability on signed distance. The  $-0.03$  value is only about 1% of the difference between the mean  $d'$ -values for highest and lowest discriminability levels (the distance between the corresponding ROCs) of 3.59. Note, however, that the mean over birds of the *absolute distance* increases significantly with discriminability, with values 0.06, 0.08, 0.26, 0.19, and 0.44, as  $LR$  increases. It is tempting to believe that this reflects the greater variability in  $z(p)$  as  $p$  approaches zero or unity, but an analysis of between-bird variability in both  $z(\text{Pr}\{R_T|S_T\})$  and  $z(\text{Pr}\{R_T|S_{NT}\})$  by condition (after correcting for mean differences among birds) shows it to be larger than the theoretical value (Gourevitch & Galanter, 1967), and approximately constant.

### A.7.2. Evaluation of an alternative measure of the decision process

In most of the examples we have not considered alternatives to the hypothesized measures ( $H2$ , Table 2). One feature of SDT is that more than one appealing and

<sup>80</sup> This way of deriving the  $\{RT_{jk}^*\}$  forces their means into agreement:  $RT_{..}^* = RT_{..}$ ; the question of interest is whether the differences among the four values agree.

<sup>81</sup> Such good agreement is unexpected, given the possibility of artifacts associated with estimating  $T_s$  and  $T_r$  from the ERP voltage wave after averaging it over trials (Section 6.1).

plausible  $M_D$  – reflecting the response bias – has been considered and used. The one considered in Section 7.2 is  $M_D = c$ , a value on the scale on which  $X_T$  and  $X_{NT}$  have Gaussian distributions; the origin for  $c$  is  $(\bar{X}_T + \bar{X}_{NT})/2$ . An alternative criterion is the likelihood ratio associated with  $X = c$ :  $M'_D \equiv \beta \equiv \phi_T(c)/\phi_{NT}(c)$ , where  $\phi_T$  is the density function for  $X_T$ , and  $\phi_{NT}$  for  $X_{NT}$ . One advantage of  $\beta$  is that with a fixed set of costs and benefits (a fixed PM) the criterion that maximizes the expected value of a trial, expressed as the optimal  $\beta$ -value, is invariant over levels of discriminability. For the present data, however, whereas  $c$ -values are approximately invariant across levels of  $LR$  for fixed  $RR$ , the corresponding actual  $\beta$ -values change systematically with  $LR$ , falling from 0.94 to 0.47 for  $RR = 0.2$  ( $F(4, 20) = 14.1$ ,  $p < 0.0001$ ) and rising from 1.05 to 12.82 for  $RR = 0.8$  ( $F(4, 20) = 11.7$ ,  $p < 0.0001$ ).<sup>82</sup> Thus, whereas  $M_D = c$  may be a pure measure of a **D** module,  $M'_D = \beta$  is apparently not.

### A.9. Factorial experiments and multiple-level factors with pure measures

In this section we consider two issues in the design of experiments to assess separate modifiability when pure measures are hypothesized, in the context of Ex. 4.

#### A.9.1. Advantages of a factorial design

For composite measures, levels of the factors  $F$  and  $G$  must be combined factorially to test selective influence, whether or not the combination rule is also to be tested: for each level of  $F$ , measurements must be made at each level of  $G$ . With pure measures a factorial experiment is not required, as shown by Exs. 1, 2, 4, and 6. However, insofar as such a design is feasible it offers advantages, an important one being a test of the generality of selective influence described in Sections 1.4, 5.2, and 11, and illustrated by Ex. 3. (Given the joint hypothesis, selectivity of a factor's effect should normally obtain for all pairs of levels of that factor, and at all levels of the other factor, and not be an accident of particular levels.) Thus, if we find that  $LA$  influences measure  $M_L$  and not  $M_H$  when  $HA = HA_1 = 0\%$ , we would also like to determine whether the effect of  $LA$  is numerically the same (subject to sampling variability) when  $HA$  is changed to  $HA_2 = 80\%$ .

Consider the design that underlies the analysis shown in Fig. 6A and C. The three conditions used, which permit examining the effects of adapting with simple (one-frequency) gratings, are those in bold type in Table 9. A complete factorial design would also include the fourth condition, in which the adapting grating is a compound (superposition of two simple gratings). According to the joint hypothesis, not only should the main effects of  $LA$  and  $HA$  on  $M_L$  and  $M_H$  be appropriately selective, but also the interaction of their effects on each measure should be zero: One factor should not modulate the effect of the other. This property is testable only with the

<sup>82</sup> When data for the highest LR (for which the ROCs were unusually irregular) are removed, the failure of invariance remains: For  $RR = 0.2$  and  $0.8$ ,  $F(3, 15) = 22.2$  ( $p < 0.0001$ ) and  $6.0$  ( $p = 0.007$ ), respectively.

Table 9

A factorial design with each factor at two levels

		Level of LA	
		LA <sub>1</sub>	LA <sub>2</sub>
Level of HA	HA <sub>1</sub>	(0%, 0%)	(80%, 0%)
	HA <sub>2</sub>	(0%, 80%)	(80%, 80%)

complete design. What would we expect if the fourth condition had been included? Several experiments (see Graham, 1989, pp. 118–120) have shown that the adapting effect of a two-frequency ( $f_1, f_2$ ) compound grating on the contrast threshold for  $f_k$ , ( $k = 1$  or  $2$ ) is smaller than the adapting effect of a simple ( $f_k$ ) grating on the contrast threshold for  $f_k$ , ( $k = 1$  or  $2$ ).<sup>83</sup> One explanation is that when analyzers such as **L** and **H** are stimulated at the supra-threshold contrasts of adapting gratings, they inhibit each other rather than being modular; when the compound is used as an adapting stimulus, the amount of adaptation of each analyzer is therefore reduced.<sup>84</sup> Hence, while the separate modifiability of these analyzers can be demonstrated with simple adapting gratings, the generality of selective influence (Section 1.4) that we would normally require does not obtain, and the test of interaction mentioned above would fail. Such failure would inform us about limitations on the selectivity of the influence of LA and HA on **L** and **H** that might have been erroneously inferred from the incomplete design.

#### A.9.2. Advantages of multiple factor levels

In many experiments, factors are selected for which only two levels can be defined, or factors for which multiple levels could be defined are examined at only two. In the present example, a continuum of levels is possible for the factors (such as LA, the contrast of the low-frequency adapting gratings), and it is known from other data that the threshold elevation caused by adaptation increases with adapting contrast (e.g., Graham, 1989, Chapter 3). The use of multiple levels has several advantages. First, it provides additional tests of the generality of selective influence mentioned above. Second, it helps determine whether a slight failure of invariance reflects a systematic trend or random variation, as in Ex. 3. Third, multiple levels are required for testing whether the factor is *unitary* over its range (Appendix A.2.1). And fourth, if the effect of a factor with potentially more than two levels might be non-monotone, then apparent invariance in data taken at just two levels might be an accident. As

<sup>83</sup> For example, consider the thresholds  $M_H$  in Fig. 6A, which were determined when  $HA = HA_1 = 0\%$ . We see an excellent approximation to invariance with respect to LA. The corresponding  $M_H$ -values for  $HA = HA_2 = 80\%$  would be closer to 1% than 0.4% (see upper curve in Fig. 6C). However, if  $HA = HA_2$ , we would expect the  $M_H$  value to decrease from  $LA = LA_1$  (simple-grating adaptation) to  $LA = LA_2$  (compound-grating adaptation), violating the expectation of invariance of  $M_H$  with respect to LA, and indicating that LA modulates the effect of HA.

<sup>84</sup> De Valois and De Valois (1988, pp. 186–187) suggest that such inhibitory interactions are relatively minor.

Table 10

A factorial design with each factor at four levels

		Level of LA			
		LA <sub>1</sub>	LA <sub>2</sub>	LA <sub>3</sub>	LA <sub>4</sub>
Level of HA	HA <sub>1</sub>	<b>(0%, 0%)</b>	<b>(20%, 0%)</b>	<b>(50%, 0%)</b>	<b>(80%, 0%)</b>
	HA <sub>2</sub>	<b>(0%, 30%)</b>	(20%, 30%)	(50%, 30%)	(80%, 30%)
	HA <sub>3</sub>	<b>(0%, 60%)</b>	(20%, 60%)	(50%, 60%)	(80%, 60%)
	HA <sub>4</sub>	<b>(0%, 80%)</b>	(20%, 80%)	(50%, 80%)	(80%, 80%)

discussed in Section 15.1, the use of multiple levels is also helpful with composite measures, where the joint hypothesis is often tested by evaluating an interaction.

Table 10 illustrates a possible multiple-level design that could be used in relation to Ex. 4. A set of conditions that avoids compound adapting gratings (not a factorial experiment) is shown in bold type; a complete factorial design would include the other as well. I was unable to find data from an experiment with both a wide range of adapting frequencies and multiple adapting contrasts.

## A.10. Inferring neural processing stages from single-unit recordings

### A.10.1. Classification of neurons

Analyses included only correct trials. Activations of cells\* classified as sensory-SC<sub>2</sub> had to be stimulus-specific, occurring on all *f-F* and *f-T* trials but on no *t-F* or *t-T* trials, or vice versa. Activations of cells\* classed as sensory-SC<sub>4</sub> had to occur with similar magnitudes on all trials of all four types. Also, SC<sub>4</sub> activations had to be as closely coupled temporally to the stimulus (vs the response) as SC<sub>2</sub> activations. Relative closeness of coupling was measured by  $\log\{[var(T_r - T_a)]/[var(T_a - T_s)]\}$ , large for stimulus-coupled activations, small for response-coupled activations. Activations of cells\* classed as motor-MC<sub>2</sub> had either to be response specific (occurring on all *f-F* and *t-F* trials, but on no *f-T* or *t-T* trials, or vice versa), or to occur on all trials of all four types and show a response preference (such that activation magnitude with one response was at least twice as great as with the other). Activations of cells\* classed as motor-MC<sub>4</sub> had to occur during all trials of all four types with similar magnitudes. Also, MC<sub>4</sub> activations had to be as closely coupled temporally to the response (vs the stimulus) as MC<sub>2</sub> activations.

The number of measured cells\* in classes SC<sub>2</sub>, SC<sub>4</sub>, MC<sub>2</sub>, and MC<sub>4</sub> were 15, 49, 43, and 10, respectively, each measured on many trials.<sup>85</sup> In relation to the functional localization assumption of Section 11, it is noteworthy that of the 64 sensory cells\*, 55% were in area M1 (with about the same  $T_a$  as those from area S1), and of

<sup>85</sup> Among the SC<sub>2</sub> cells\*, 9 responded on *f* trials, 6 on *t* trials. Among the MC<sub>2</sub> cells\*, 37 responded on *F* trials, 6 on *T* trials. Values reported here for SC<sub>2</sub> and MC<sub>2</sub> cells\* are equally weighted means of the values for their two subclasses. The data in Table 2 of Mouret and Hasbroucq (2000) reflect a different way of combining data over these subclasses.

the 53 motor cells\*, 32% were in area S1 (with a  $T_a$  about 50 ms later than those from area M1). These findings of inhomogeneous cell types are similar to those of Requin, Riehle, and Seal (1988, 1993), which demonstrate the “functional heterogeneity of structurally defined neuronal networks” (1988, p. 179).

#### *A.10.2. Combining activation times over neurons*

It is reasonable to believe that large numbers of neurons participate in any mental operation of the kinds that are of interest in the present paper. Yet, in an experiment of this type, only one cell\* of one class is measured on any particular trial. Each measurement provides a single latency; if other cells\* of that class had been measured on the same trial, their latencies would probably differ, and we would have a distribution of latencies. Suppose we had such a distribution. One of the limitations of a study of this kind arises because we know very little about the behavior of populations of neurons as they participate in complex processes; we have little guidance in choosing a statistic of that latency distribution – perhaps a location measure such as the minimum, median (or other quantile), mean, or maximum – that is appropriate for defining a boundary between one stage and the next.<sup>86</sup> That the system would require the activation of only one relevant cell (implying the minimum as the measure) or of all relevant cells (implying the maximum) before proceeding to the next stage both seem implausible. Because computing the mean requires activation times of all participating cells, the mean also seems unlikely to be the appropriate location measure, especially in a speeded task.

What problem is caused by our uncertainty about the appropriate population statistic? To assess separate modifiability in this case we have to quantify the changes in stage durations as we manipulate factor levels (here, levels of S-R compatibility). In the present example we compare the effects of compatibility on the times of neural and overt responses. If factor manipulations cause changes in only the *location* of the population latency distribution, any of the location measures could serve; the mean would be fine. But factor manipulations may also change the spreads or shapes of such distributions. Now, while only one neuron was measured on each trial by Mouret and Hasbroucq (2000), that neuron, and each of many others, was measured on many such trials. It may be possible to use within- and between-neuron variation to provide estimates of population latency distributions, and thus to determine how those distributions change with compatibility level. This could inform us roughly about the extent and direction in which effect-size estimates based on means are likely to differ from estimates based on other location measures. Mouret and Hasbroucq did not obtain such estimates of latency distributions, however, and this issue should be kept in mind in considering the findings. The values reported here are based on first getting the mean of the activation latencies of each cell\* in each condition of the experiment in which it responded, and then determining means of

---

<sup>86</sup> Parker and Newsome (1998) show that evidence from behavioral vs single-neuron perceptual thresholds is divided between favoring the “lower envelope principle” (which suggests the minimum latency) and the “pooling” (physiological summation) model (which suggests a higher quantile).

these values over cells\* within classes and subclasses. As a working hypothesis, I assume, with Mouret and Hasbroucq, that the use of means does not substantially bias the effect-size estimates. This is justified to the extent that S-R compatibility influences only the location of the latency distribution of the MC<sub>2</sub> and MC<sub>4</sub> cells\*, and not the shape of that distribution.

#### A.10.3. Effects of stimulus and response factors on stage durations

Given the Mouret–Hasbroucq experiment, stimulus and response factors can be defined: just as the contrast for activation time  $T_a$  associated with *SRM* is  $(T_{a,tF} + T_{a,tT})/2 - (T_{a,tI} + T_{a,tJ})/2$ , the orthogonal contrast for the *stimulus* factor *STIM* (levels  $f, t$ ) is  $(T_{a,tF} + T_{a,tI})/2 - (T_{a,tJ} + T_{a,tT})/2$ , and for the *response* factor *RESP* (levels  $F, T$ ) is  $(T_{a,tF} + T_{a,tJ})/2 - (T_{a,tI} + T_{a,tT})/2$ . Ideally, *STIM* would influence only  $\alpha$ , and *RESP* would influence only  $\gamma$ . While there was no consistent effect of *RESP* on  $\overline{RT}$ , there was a consistent effect of *STIM*:  $\overline{RT}$  was longer by 19 ms for  $f$  than  $t$ . Unfortunately, insofar as the neuronal data permit localizing the *STIM* effect in one of the three stages, it appears to be in  $\beta$  rather than  $\alpha$ . If these findings were borne out by data with greater precision and, ideally, larger effects, then this would not violate the conclusion that *SRM* selectively influences  $\beta$ , but it would say that if there is a separate stage influenced selectively by *STIM*, the choice of neurons to measure has not isolated it.

### A.11. fMRI and modular processes: requirements and statistical issues

#### A.11.1. Requirements for a process-decomposition study

Some of the requirements for a study of the kind described in Section 11 are as follows:

- (1) The subject should be performing a task as measurements are taken. Even in sensory studies enough evidence has emerged for task effects at early levels of cortical processing (e.g., Martinez et al., 1999) so it is no longer appropriate merely to present stimuli to a passive observer.
- (2) The subject should be performing the same task as factor levels are varied. By “same task” I mean that a persuasive argument can be made that for all combinations of factor levels, the same set of processing operations is involved, varying only “quantitatively”. A change in level should not replace one process with another, or introduce an additional process. Another way to say this is that factors should be unitary (Appendix A.2.1); using the data to provide a partial check of whether this criterion is met requires at least three levels of each factor.
- (3) Because the invariance of a measure (e.g.,  $M_\beta$  in Fig. 9A) across levels of a factor (e.g.,  $F$ ) is at least as important in our inferences as the influence of that factor on another measure (e.g.,  $M_\alpha$ ), it is critical that we have some index of precision for the size of an effect when the effect is claimed to be nil. Insofar as the precision of the data is poor, a test can fail to reach significance even if the “true” effect is large (Section 1.5; Appendix A.11.2).

- (4) Because activation levels of selected voxels are the quantities of interest, it is they that should be reported, rather than only quantities (such as the  $t$ -values of “statistical maps”) that reflect both means and variances of such activation levels. (Compare Figs. 1 and 2 in McDermott et al., 1999, for example.)
- (5) The sets of voxels that define regions such as  $\alpha$  and  $\beta$  should, ideally, be selected before the experiment, based on prior knowledge. If that is not possible, their selection should be based on measurements within the experiment that can be shown to be independent of the measures  $M_\alpha$  and  $M_\beta$  used to test the joint hypothesis. Otherwise the selection process may bias the obtained values of those measures.

Non-essential desiderata include factorial designs and, as mentioned in Section 11, testing for analogous patterns of effects of the same factors in behavioral data.

As researchers gain more experience with fMRI, studies are appearing that increasingly adopt these requirements and desiderata, such as the use of factorial designs (e.g., McDermott et al., 1999) the use of such designs where factors are plausibly unitary (Appendix A.2.1; e.g., Cohen et al., 1997), and the use of factors with multiple levels (e.g., Braver et al., 1997).

#### A.11.2. Statistical issues

Because the data in Fig. 9 are hypothetical and idealized, no tests or measures of precision are reported. With real data, we would of course have to show: (1a) the effects of  $F$  on  $M_\beta$  and of  $G$  on  $M_\alpha$  are convincingly negligible (Section 1.5); as well as (2a) the effects of  $F$  on  $M_\alpha$  and of  $G$  on  $M_\beta$  are convincing. Initial global tests could use four ANOVAs. Insofar as we expect monotonic effects, more focused tests would use one-dimensional measures of effect size such as the mean slopes of lines fitted to the four functions for each subject, each with an estimated S.E. based on between-subject variation. Letting  $M_x^*(Y)$  be the slope of  $M_x(Y)$ , we would then have to show: (1b)  $M_\beta^*(F)$  and  $M_\alpha^*(G)$  are convincingly close to zero; as well as (2b)  $M_\alpha^*(F)$  and  $M_\beta^*(G)$  are convincingly greater than zero.

We could show (2b) if the two mean slopes are substantially greater than their S.E.s. For (1b), consider  $M_\beta^*(F)$ . One minimal requirement is that it does not differ significantly from zero. However, this can occur simply because the experiment is insufficiently precise, the S.E. large. Another minimal requirement, worth checking, is for  $M_\beta^*(F)$  to be significantly less than  $M_\beta^*(G)$ . But it would seem important also to consider the power of the test of  $M_\beta^*(F) \neq 0$  (or of  $M_\beta^*(F) > 0$ ), expressed, for example, by the smallness of its S.E., or by using the S.E. to specify  $\max_{ns}\{M_\beta^*(F)\}$ , the largest value of  $M_\beta^*(F)$ , given the data, that would not differ significantly from zero. However, to determine how small  $\max_{ns}\{M_\beta^*(F)\}$  must be to declare invariance requires a complex judgement, based on relevant information from outside the experiment as well as on (a) the sensitivity of  $M_\beta$  and (b) the potency of  $F$  (Section 2.3). One might think of using  $M_\beta^*(G)$  and  $M_\alpha^*(F)$  as measures of (a) and (b), respectively; unfortunately, however, the former also depends on the potency of  $G$ , and the latter on the sensitivity of  $M_\alpha^*$ .

Instead, it may be worth basing a decision about whether selective influence is satisfied on the relationship between two quantities: One, the mean absolute value of

$M_\alpha^*(G)$  and  $M_\beta^*(F)$ , combines the magnitudes of the evidence *against* the two kinds of invariance that selective influence requires. The other, the mean absolute value of  $M_\alpha^*(F)$  and  $M_\beta^*(G)$ , combines the magnitudes of the evidence *favoring* two kinds of influence that it requires. Both quantities reflect the potency of both factors together with the sensitivity of both measures. To declare invariance, the first quantity could be required to be no greater than a small percentage (e.g., 10%) of the second.

### A.13. Fitting and testing one-channel and two-channel models of detection

In Section 3.2 I noted that the logarithmic transformation can be useful when proportional factor effects are multiplicative. However, in the present example,  $\hat{P}_{ND}$  takes on values that can be close to zero, which introduces considerable heterogeneity in the S.E.s of its log-transformed values. Because this renders plots somewhat misleading, the data displayed in Fig. 10 were not logarithmically transformed. However, more standard fitting and testing of the two-channel model was done by using a generalized linear model with the logarithm as the link function and the variance specified as  $\mu/(1 - \mu)$ .<sup>87</sup> The conclusions are the same as those stated in Section 13.2, in relation to both overall tests and tests focused on linear and quadratic patterns of deviation.

Fig. 10 also shows the results of fitting a broadband single-channel model.<sup>88</sup> For this model, the shape of the psychometric function ( $Pr\{Detect\}$  vs contrast) becomes relevant; Sachs et al. (1971) showed that for most of their sets of simple-grating data it is well-approximated by multiplying  $(1 - p_{ng})$  by the Gaussian cumulative distribution function (cdf), which is similar to the logistic cdf. For the data in panel B, the relation between  $P_{ND}(Cf_j, 0)$  and  $P_{ND}(Cf_j, Cf'_2)$  is better captured by assuming the logistic. Letting  $P_{ND}^* = P_{ND}/p_{ng}$  (the non-detection proportion corrected for guessing) we then have  $P_{ND}^* = (1 + e^{\alpha + \beta x})^{-1}$ , where  $x$  is contrast. If a second grating is added to the first, with  $f'$  sufficiently close to  $f$  so that detectability is determined by a single analyzer, then adding the second grating should be equivalent to an increase in the contrast of the first. The contrast modulation  $x$  can then be regarded as the sum of two parts,  $y = y(Cf_j)$  and  $y' = y'(Cf'_k)$ , due to the  $f$ - and  $f'$ -gratings, respectively. Letting  $Odds_{ND}^* = P_{ND}^*/(1 - P_{ND}^*)$  be the guessing-corrected odds of non-detection, it follows that

$$Odds_{ND}^*(Cf_j, Cf'_k) = e^{-\beta y(Cf_j)} \times e^{-\beta y'(Cf'_k)} \times e^{-\alpha}. \quad (19)$$

Thus, a consequence of the form of the psychometric function is that for the composite measure  $Odds_{ND}^*, Cf$  and  $Cf'$  are multiplicative factors as in Eq. (8). (This

<sup>87</sup> See “quasi-likelihood” in McCullagh and Nelder (1989).

<sup>88</sup> For a broadband analyzer (required for a single-channel model), a composite grating (but not a simple one) varies in amplitude across spatial locations. The single-channel model discussed by Sachs et al. (1971) and described here would today be regarded as primitive, partly because it does not reflect the combining of information across such locations (Graham, 1989, Chapter 5). One consequence is that the behavior of a more modern single-channel model would be closer to that of the multiple-channel model.



emphasizes the situation mentioned in Appendix A.3.1, where transforming a measure causes the expected p.effects of two factors to be multiplicative, but where this property can be plausibly explained if they influence the same module, and not if they selectively influence different modules.) For the present experiment,  $y'$  is either  $y'_1 = 0$  (simple grating) or a fixed positive value  $y' = y'_2$  (compound grating). It follows that

$$Odds_{ND}^*(Cf_j, Cf'_2) = \eta Odds_{ND}^*(Cf_j, 0), \quad (0 < \eta < 1), \quad (20)$$

where  $\eta$  is the constant  $e^{-\beta y'_2}$ . Note the similarity to Eq. (10): here the predicted result of adding the  $f'$ -grating is multiplication of  $Odds_{ND}^*$  (rather than  $P_{ND}$ ) by a constant.

Shown in each panel of Fig. 10 is the fit of Eq. (20) to the compound-grating data (with  $p_{ng}$  and  $\eta$  estimated by least squares), given the simple-grating data. While the one-channel model fails systematically in panels A and C (where the two-channel model fits well), the one-channel model fits very well in panel B, where the two-channel model fails.

### A.15. Numerical scaling of factor levels for multiplicative and hybrid combination rules

In Sections 7.2, 13.2, and 15.1 I pointed out the usefulness of linearized data for discovering and assessing the extent to which deviations from a model are systematic. In addition to this advantage, constraints on the data that are implied by Eq. (5) are rendered simple and easy to test by such linearization. As Roberts (1987) recognized, the same advantage of linearizing also applies to the more elaborate process structure and the associated hybrid combination rules of Eq. (11). For simplicity I disregard sampling error. What follows is a minor reformulation of his findings.

Let us write Eq. (5) as

$$\text{Variant 0: } M_{jk} = u_j v_k, \quad (21)$$

where  $j$  and  $k$  index the levels of factors  $F$  and  $G$ , respectively. Instead of plotting  $M_{jk}$  for each  $k$  as a function of the level index  $j$  of factor  $F$ , plot it as a function of a numeric variate,  $\mathcal{F}_j \equiv M_{j\bullet}$ . The result of such scaling is that for each level  $k$  of factor  $G$ , the values of  $M_{jk}$  vs  $\mathcal{F}_j$  fall on a line,  $M_{jk} = v'_k \mathcal{F}_j$ , where  $v'_k = v_k / v_\bullet$ . Note that all the resulting lines intersect at the same point – the origin (0,0) – and that their mean slope  $v'_\bullet = 1$ . The fitted lines in Fig. 13A provide an example. Next consider three variants that generalize the simple multiplication rule by incorporating an additive term, described in Section 15 as hybrid (multiplicative-additive) combination rules.

In Variant 1 the additive term,  $w$ , is constant with respect to  $j$  and  $k$ :

$$\text{Variant 1: } M_{jk} = u_j v_k + w, \quad (w \neq 0). \quad (22)$$

Again the values of  $M_{jk}$  for each  $k$  fall on a line,  $M_{jk} = v'_k \mathcal{F}_j + (w - wv'_k)$ , and the lines intersect at a common point, but here the point is  $(w, w)$ . Because  $w$  is unknown, the diagnostic feature is the existence of a common point of intersection other than (0,0). Because Variants 0 and 1 are both symmetric with respect to  $j$  and  $k$ , the same properties obtain if  $M_{jk}$  is considered, for each  $j$ , as a function of  $\mathcal{G}_k \equiv M_{\bullet k}$ .

In Variant 2 the additive term  $w_k$  can vary with the level of  $G$ , but not with the level of  $F$ :

$$\text{Variant 2: } M_{jk} = u_j v_k + w_k. \quad (23)$$

Again the values of  $M_{jk}$  for each  $k$  fall on a line,  $M_{jk} = v'_k \mathcal{F}_j + (w_k - w_\bullet v'_k)$ , but here the lines need not intersect at a common point. To distinguish Variants 1 and 2 requires at least two pairs of intersecting lines, hence at least three levels of factor  $G$ .

In Variant 3 the additive term  $w_j$  can vary with the level of  $F$ , but not with the level of  $G$ :

$$\text{Variant 3: } M_{jk} = u_j v_k + w_j. \quad (24)$$

In this case the functions,  $M_{jk} = v'_k \mathcal{F}_j + (w_j - w_j v'_k)$ , need not be linear. However, the plot obtained by reversing the roles of  $F$  and  $G$  and plotting  $M_{jk}$  against  $\mathcal{G}_k$  has the same properties as Variant 2.<sup>89</sup>

In conclusion, for the simple multiplicative combination rule (Variant 0) and all three of the hybrid rules (Variants 1, 2, and 3) the data can be represented, by suitable numeric scaling of factor levels, as points on a set of lines with different slopes. The rules differ in three ways: (a) which of  $\mathcal{F}_j$  or  $\mathcal{G}_k$  produces linearity (both do for Variants 0 and 1, only  $\mathcal{F}_j$  for Variant 2, and only  $\mathcal{G}_k$  for Variant 3); (b) whether there is a common point of intersection for the set of lines (there is for Variants 0 and 1, but not for 2 and 3); (c) the location of a common intersection point if there is one [(0,0) for Variant 0, ( $w, w$ ) for Variant 1]. Which pattern characterizes the data has important implications for the underlying process.

## A.16. Processing stages as modules

### A.16.1. Multiplicative combination rule for the proportion of response omissions

As often found in sleep-deprivation experiments, the rate of response omissions was high; on average, responses were produced on only 89.2% of trials.<sup>90</sup> The way in which factors  $SLP$ ,  $SQ$ , and  $MF$  combine in influencing the probability of responding (like  $\overline{RT}$ , a composite measure) may also provide useful information about the underlying process. For the analysis of errors of commission in processes believed to be arranged in stages, Schweickert (1985) found that in several data sets, factors had multiplicative p.effects (Section 3.2) on the proportion of responses that were correct. He explained this by a plausible multiplicative combination rule similar to Roberts' (1987; Ex. 9): Roughly, for a response to be correct, each required stage must function correctly. If such events are mutually independent stochastically, then  $Pr\{\text{Response Correct}\}$  is a product of the corresponding stage probabilities. (Schweickert's examples are additional instances of analysis of a composite measure

<sup>89</sup> Among the many data patterns considered by Roberts (1987) some are consistent with each of Eqs. (21)–(23); this is also true for the various “linear fan” patterns discussed by Anderson (1996).

<sup>90</sup> Given that a response was produced,  $SLP$  had virtually no influence on whether it was correct.

with multiplication as the combination rule.) For the present experiment we might expect, by analogy, that a response is produced (rather than omitted) on a trial only if each stage functions so as to permit it. Again, if such events are mutually independent we should have a multiplicative combination rule for the contributions of each stage to response occurrence. This implies (Table 4) that factors that influence different stages selectively should have multiplicative p.effects on  $Pr\{Response\}$ .

Fortunately, whether one considers either effects or p.effects, analysis of the proportion of trials on which responses were produced tells the same story. But this story differs slightly from the inferences from  $\overline{RT}$ : As they do for  $\overline{RT}$ , each of the three factors has substantial main effects, and  $SLP$  modulates the effect of  $SQ$  substantially while leaving the effect of  $MF$  virtually invariant. However, for  $Pr\{Response\}$ ,  $MF$  and  $SQ$  do appear to interact slightly.<sup>91</sup> It would of course be of great interest if two different composite measures,  $\overline{RT}$  and  $Pr\{Response\}$  provided convincing evidence for the same process decomposition.

#### A.16.2. What is a 'stage'?

Suppose we are persuaded that a particular process is organized as a sequence of processing stages: what does this mean? A stage is a function carried out during an epoch in time; it is not necessarily associated with a distinct neuroanatomical processor. Too often, an inferred flowchart, which describes the temporal arrangement of a set of operations (and is inherently ordered, because of the nature of time), has been confused with a circuit diagram, which indicates how processors are connected.

When a two-stage model is supported, all we "know" in the simplest case is that the complex process between stimulus and response can be cut at some time point, defining stage  $a$  (before the cut, and the epoch during which process **A** takes place) and stage  $b$  (after the cut, and the epoch during which process **B** takes place).<sup>92</sup> In some cases, such as the present one, it is plausible that the processes are connected, in the sense that **R** (response selection) uses information provided by **S** (stimulus identification). That is, **S** and **R** probably have a relation of *data-dependence*. This does not necessarily mean that the two processes are accomplished by distinct neural processors. However, insofar as distinct functions are carried out by special-purpose neural processors (possibly localized in different regions of the brain), it is also plausible, for the present example, that **S** and **R** are carried out by different pro-

<sup>91</sup> The discussion of omissions in Sanders et al. (1982) suggests that this last interaction, while proportionally greater than in the  $\overline{RT}$  data, is not statistically significant, whereas the main effects and the  $SLP \times SQ$  interaction are. As the basic data are no longer available, and no analysis of p.effects was reported, we do not know about statistical significance of the corresponding p.effect interaction.

<sup>92</sup> The alternation (multiplexing) of two processes,  $A_1 \rightarrow B_1 \rightarrow A_2 \rightarrow B_2, \dots$ , is an example of an arrangement whose behavior is similar in some ways, and, given only  $M_{AB} = \overline{RT}$ , is indistinguishable from  $A \rightarrow B$ . See Appendix A.6 for combining the  $\overline{RT}$  measure with evidence from pure measures of two neural processes, to distinguish the two.

processors, as their functions seem to be so different. In other instances, multiple operations (such as the recognition of each of two unrelated words, see, e.g., Sternberg, 1998b, Section 14.5.11) may be arranged sequentially, not because of data-dependence but perhaps because of capacity limitation. Indeed, in such a two-word recognition task it is plausible that the *same* neural processor is responsible for both operations, as their functions are the same.

The confusion of flow-charts with circuit-diagrams has led to unwarranted over-interpretations of the class of models that the AFM suggests. Broadbent (1984, p. 55) provided one example when he described a stage model as “a pipeline . . . through which information flows from the senses to the effectors” and suggested that stage models are consistent only with “bottom-up” theories.<sup>93</sup> An example of an alternative is a possible model for the recognition of a degraded printed word during reading: There is no obstacle to having one stage **A** during which a first interpretation of a word is developed, based on an initial sample of letter and word-shape features and influenced by context, followed by a second stage **B** during which the same process is repeated, now guided by the first interpretation as well, involving a new sample of features, and producing a revised interpretation. The flow-chart for this part of the process would contain two boxes, with an arrow from **A** to **B** that represents succession in time (not information flow, although some of the information used by **B** is provided by **A**). In contrast, the corresponding circuit diagram might contain just one box, with an arrow (representing the flow of information) from its output to its input. (See Sternberg, 1998b, Section 14.6.1.)

#### *A.16.3. Additive effects of factors on mean reaction time: alternative interpretations*

Additive effects on  $\overline{RT}$  have been of sufficient interest so that alternatives to stage models have been considered as explanations. It has been discovered that under some conditions, other models, quite different in spirit from stage models, can also generate such additive effects. Thus, McClelland’s (1979) cascade model as further developed by Ashby (1982) (see also Roberts & Sternberg, 1993) can produce close approximations to additivity of factor effects on both means and variances. (Additive effects on variance are predicted from a stage model if the model is elaborated by assuming that stage durations are stochastically independent.) Miller, van der Ham, and Sanders (1995) have shown that a range of models with temporally overlapping processes can approximate additive effects on means, and in a few cases on variances as well.<sup>94</sup> And Roberts and Sternberg (1993) described an alternate-pathways model that produces exactly additive effects on means, but fails miserably for variances. In all these cases, the prediction of means additivity derives from modularity plus selective influence; hence, from the viewpoint of discovering modules (but not of how these modules are organized), the existence of these alternative possibilities does not

<sup>93</sup> Others sharing this misunderstanding include, e.g., Neisser (1976, p. 23) and Rabbitt and Maylor (1991, p. 277).

<sup>94</sup> On the other hand, Miller (1993) developed and explored another family of models in which overlapping processes are permitted, and found that in general they do not produce additive factor effects.

weaken the argument outlined in Table 3. Their discovery, however, decidedly weakens the inference from the additivity of factor effects on  $\overline{RT}$  to the organization in stages of the corresponding modules. Other aspects of the RT data can sometimes help distinguish among such alternatives; by elaborating stage models with the assumption that stage durations are stochastically independent, Roberts and Sternberg (1993) have shown how properties of the RT distributions can help to select among competing models; in their analysis of five data sets, the elaborated stage model survived, while its competitors did not. Stage models without the elaborations provided by uncorrelated or (stronger) stochastically independent stage durations are, however, of interest, and at present there appear to be few ways to distinguish them from some of their competitors. Other approaches to selection among such models include techniques such as speed-accuracy decomposition (e.g., Meyer et al., 1988, Section 5) and concomitant measurements that might be electrophysiological (e.g., Appendix A.6; Magliero, Bashore, Coles, & Donchin, 1984; Meyer et al., 1988, Sections 7 & 8) or behavioral (e.g., Sternberg, 1998b, Section 14.5.7).

### **Acknowledgements**

This paper was inspired by an invitation to contribute to the symposium “Looking for Stages” in August 1998, honoring Andries F. Sanders. Parts of it reflect ideas developed by Seth Roberts, and many fruitful discussions with him. For providing unpublished details of their data and/or guidance in interpreting them I thank Brent Alsop, Norma Graham, Thierry Hasbroucq, Victoria Johnstone, John Kounios, Jacob Nachmias, Seth Roberts, and Fren Smulders. I am also especially grateful to Brent Alsop for furnishing basic data from the McCarthy and Davison (1984) experiment. For helpful comments on the manuscript, I thank Mieke Donk, Janice Hamer, Michael Kahana, John Kounios, Caleb McArthur, Jeff Miller, Jacob Nachmias, Allen Osman, Seth Roberts, Eric Schumacher, Jennifer Sternberg, John Swets, Rolf Ulrich, Dirk Vorberg, and especially Teresa Pantzer. For invaluable guidance about data analysis I thank Paul Rosenbaum, and for other helpful discussions I thank Marion D. Cohen, Wim Koster, David H. Krantz, Frank Norman, and Robert Rescorla.

## Glossary

Here I provide a list of the main abbreviations used, with numbers of the sections or appendices in which they are introduced, and brief definitions. It may also be helpful to review Section 1.2.

---

AFM	1	additive-factor method
ERP	1	event-related potential
<i>ERP(t)</i>	6	event-related potential as a measure
<i>FF</i>	15	frequency of feeding
<i>FF</i>	15.1	<i>FF</i> as a numeric variable
fMRI	1.7	functional magnetic resonance imaging
<i>GND</i>	6	Go–NoGo discriminability
<i>H1, H2</i>	2.2	hypotheses
<i>H1*</i>	A.2.3	hypothesis
<i>H3</i>	3.1.1	hypothesis
<i>H4, H5</i>	3.2	hypotheses
<i>HA</i>	9	high-frequency adaptation
<i>HD</i>	15	hours of food deprivation
<i>HD</i>	15.1	<i>HD</i> as a numeric variable
<i>IRT</i>	A.3.2	inter-response time
<i>LA</i>	9	low-frequency adaptation
<i>LR</i>	7.2	luminance ratio
<i>LR</i>	7.2	<i>LR</i> as a numeric variable
LRP	6	lateralized readiness potential
<i>MA</i>	9	medium-frequency adaptation
<i>MF</i>	11	stimulus-response mapping familiarity
<i>PF</i>	5	probability of food
PM	7	payoff matrix
<i>PM</i>	7	payoff matrix as a factor
<i>RC</i>	A.6	response complexity
<i>REL</i>	14.1	semantic relatedness
<i>RESP</i>	A.10.3	response (as a factor)
ROC	7.2	relative (or receiver) operating characteristic
<i>RR</i>	7.1	as a factor (Section 7; Appendix A.2.2): reinforcement ratio
<i>RR</i>	15	as a measure (Section 15): response rate
<i>RR</i>	7.2	<i>RR</i> (reinforcement ratio) as a numeric variable
RT	1	reaction time
<i>RT</i>	1	reaction time as a measure
S.E.	1.5	standard error
<i>SAT</i>	14.1	semantic satiation
SDT	1	signal-detection theory
<i>SLP</i>	16	sleep state
<i>SQ</i>	11	stimulus quality
<i>SRM</i>	6	stimulus-response mapping
<i>STIM</i>	A.10.3	stimulus (as a factor)
<i>TD</i>	16	time of day
<i>TF</i>	5	time of food
VI	7.2	variable-interval reinforcement schedule

---

## References

- Abelson, R. P., & Prentice, D. A. (1997). Contrast tests of interaction hypotheses. *Psychological Methods*, 2, 315–328.
- Alsop, B. (1998). Receiver operating characteristics from nonhuman animals: some implications and directions for research with humans. *Psychonomic Bulletin & Review*, 5, 239–252.
- Alsop, B., & Davison, M. (1991). Effects of varying stimulus disparity and the reinforcer ratio in concurrent-schedule and signal-detection procedures. *Journal of the Experimental Analysis of Behavior*, 56, 67–80.
- Anderson, N. H. (1996). *A functional theory of cognition*. Hillsdale, NJ: Erlbaum.
- Ashby, F. G. (1982). Deriving exact predictions from the cascade model. *Psychological Review*, 89, 599–607.
- Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity: decomposition and localization as strategies in scientific research*. Princeton, NJ: Princeton.
- Berger, R. L., & Hsu, J. C. (1996). Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Statistical Science*, 11, 283–319.
- Blakemore, C., & Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology*, 203, 237–260.
- Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *Neuroimage*, 5, 49–62.
- Broadbent, D. E. (1984). The Maltese cross: a new simplistic model for memory. *Behavioral and Brain Sciences*, 7, 55–94.
- Buchner, A., Erdfelder, E., & Vaterrodt-Plünnecke, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *Journal of Experimental Psychology: General*, 124, 137–160.
- Bullinaria, J. A., & Chater, N. (1995). Connectionist modeling: implications for cognitive neuropsychology. *Language and Cognitive Processes*, 10, 227–264.
- Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology*, 197, 551–566.
- Carandini, M., & Heeger, D. J. (1994). Summation and division by neurons in primate visual cortex. *Science*, 264, 1333–1336.
- Catania, A. C. (1970). Reinforcement schedules and psychophysical judgements: a study of some temporal properties of behavior. In W. N. Schoenfeld, *The theory of reinforcement schedules* (pp. 1–42). New York: Appleton-Century-Crofts.
- Christie, L. S., & Luce, R. D. (1956). Decision structure and time relations in simple choice behavior. *Bulletin of Mathematical Biophysics*, 18, 89–112.
- Clark, F. C. (1958). The effect of deprivation and frequency of reinforcement on variable-interval responding. *Journal of the Experimental Analysis of Behavior*, 1, 221–228.
- Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature*, 386, 604–608.
- Cowey, A. (1985). Aspects of cortical organization related to selective attention and selective impairments of visual perception: a tutorial review. In M. I. Posner, & O. S. M. Marin, *Attention and performance XI* (pp. 41–62). Hillsdale, NJ: Erlbaum.
- Curran, T., & Hintzman, D. L. (1995). Violations of the independence assumption in process dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 531–547.
- De Valois, R. L., & De Valois, K. K. (1988). *Spatial vision*. New York: Oxford.
- Dinges, D. F., & Kribbs, N. B. (1991). Performing while sleepy: effects of experimentally-induced sleepiness. In T. H. Monk, *Sleep, sleepiness and performance* (pp. 97–128). London: Wiley.
- Donders, F. C. (1868). Over de snelheid van psychische processen (On the speed of mental processes). *Onderzoekingen gedaan in het Physiologisch Laboratorium der Utrechtsche Hoogeschool, 1868–1869, Tweede reeks, II*, 92–120. (Transl. by W. G. Koster (1969). In W. G. Koster, *Attention and performance II. Acta Psychologica*, 30, 412–431).
- Dunn, J. C., & Kirsner, K. (1988). Discovering functionally independent mental processes: the principle of reversed association. *Psychological Review*, 95, 91–101.

- Duaso, A. E. (1975). Treatments of bias in detection and recognition models: a review. *Perception & Psychophysics*, *17*, 167–178.
- Duaso, T. (1983). Isobias curves in some detection tasks. *Perception & Psychophysics*, *33*, 403–412.
- Eimer, M. (1998). The lateralized readiness potential as an on-line measure of central response activation processes. *Behavior Research Methods, Instruments, & Computers*, *30*, 146–156.
- Farah, M. J. (1984). The neurological basis of mental imagery: a componential analysis. *Cognition*, *18*, 245–272.
- Fisher, R. A. (1926). The arrangement of field experiments. *Journal of the Ministry of Agriculture of Great Britain*, *33*, 503–513.
- Fisher, R. A. (1935). *The design of experiments*. Edinburgh: Oliver & Boyd.
- Frackowiak, R. S. J., Friston, K. J., Frith, C. D., & Mazziotta, J. (1997). *Human brain function*. San Diego, CA: Academic Press.
- Geisler, W. S., & Albrecht, D. G. (1995). Bayesian analysis of identification performance in monkey visual cortex: Nonlinear mechanisms and stimulus certainty. *Vision Research*, *35*, 2723–2730.
- Geisler, W. S., & Albrecht, D. G. (2000). Spatial vision: single neurons and perception. In K. K. De Valois, *Seeing* (pp. 79–128). San Diego, CA: Academic Press.
- Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, *5*, 351–360.
- Glymour, C. (1994). On the methods of cognitive neuropsychology. *British Journal for the Philosophy of Science*, *45*, 815–835.
- Gourevitch, V., & Galanter, E. (1967). A significance test for one-parameter isosensitivity functions. *Psychometrika*, *32*, 25–33.
- Graham, N. V. S. (1989). *Visual pattern analyzers*. New York: Oxford.
- Graham, N. (1970). *Spatial frequency channels in the human visual system: effects of luminance and pattern drift rate*. Doctoral dissertation. University of Pennsylvania, Philadelphia.
- Graham, N. (1972). Spatial frequency channels in the human visual system: effects of luminance and pattern drift rate. *Vision Research*, *12*, 53–68.
- Graham, N., Kramer, P., & Haber, N. (1985). Attending to the spatial frequency and spatial position of near-threshold visual patterns. In M. I. Posner, & O. S. M. Marin, *Attention and performance XI* (pp. 269–284). Hillsdale, NJ: Erlbaum.
- Graybill, F. A. (1976). *Theory and application of the linear model*. North Scituate, MA: Duxbury.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Happel, B. L. M., & Murre, J. M. J. (1994). Design and evolution of modular neural network architectures. *Neural Networks*, *7*, 985–1004.
- Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 923–941.
- Humphrey, D. G., Kramer, A. F., & Stanny, R. R. (1994). Influence of extended wakefulness on automatic and nonautomatic processing. *Human Factors*, *36*, 652–669.
- Jacobs, R. A. (1997). Nature, nurture, and the development of functional specializations: a computational approach. *Psychonomic Bulletin & Review*, *4*, 299–309.
- Jacobs, R. A., & Jordan, M. I. (1992). Computational consequences of a bias toward short connections. *Journal of Cognitive Neuroscience*, *4*, 323–336.
- Jacobs, R. A., Jordan, M. I., & Barto, A. G. (1991). Task decomposition through competition in a modular connectionist architecture: the what and where vision tasks. *Cognitive Science*, *15*, 219–250.
- Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Computation*, *3*, 79–87.
- Jacoby, L. L. (1991). A process dissociation framework: separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541.
- Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, *110*, 306–340.
- Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: measuring recollection. *Journal of Experimental Psychology: General*, *122*, 139–154.
- Johnstone, V., & Alsop, B. (2000). Reinforcer control and human signal-detection performance. *Journal of the Experimental Analysis of Behavior*, *73*, 275–290.



- Kahana, M. J. (2000). Contingency analysis of memory. In E. Tulving, & F. I. M. Craik, *Oxford handbook of memory* (pp. 58–72). New York: Oxford.
- Kelley, C. M., & Lindsay, D. S. (1996). Conscious and unconscious forms of memory. In E. L. Bjork, & R. A. Bjork, *Memory* (pp. 31–63) San Diego, CA: Academic Press.
- Kounios, J. (1996). On the continuity of thought and the representation of knowledge: electrophysiological and behavioral time-course measures reveal levels of structure in semantic memory. *Psychonomic Bulletin & Review*, 3, 265–286.
- Kounios, J. (1999). *Neurocognitive modules revealed by event-related brain potentials* (submitted).
- Kounios, J., Kotz, S. I., & Holcomb, P. J. (2000). On the locus of the semantic satiation effect: evidence from event-related brain potentials. *Memory & Cognition* (in press).
- Kutas, M., & Dale, A. (1997). Electrical and magnetic readings of mental functions. In M. D. Rugg, *Cognitive neuroscience* (pp. 197–242). Cambridge, MA: MIT Press.
- Luce, R. D. (1986). *Response times: their role in inferring elementary mental organization*. New York: Oxford.
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: a user's guide*. New York: Cambridge.
- Magliero, A., Bashore, T. R., Coles, M. G. H., & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. *Psychophysiology*, 21, 171–186.
- Markowitz, J., & Swets, J. A. (1967). Factors affecting the slope of empirical ROC curves: comparison of binary and rating responses. *Perception & Psychophysics*, 2, 91–100.
- Marr, D. (1976). Early processing of visual information. *Philosophical Transactions of the Royal Society of London B*, 275, 483–524.
- Martinez, A., Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., Wong, E. C., Hinrichs, H., Heinze, H. J., & Hillyard, S. A. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nature Neuroscience*, 2, 364–369.
- McBride, D. M., & Doshier, B. A. (1999). Forgetting rates are comparable in conscious and automatic memory: a process-dissociation study. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 25, 583–607.
- McCarthy, D., & Davison, M. (1984). Isobias and alloibias functions in animal psychophysics. *Journal of Experimental Psychology: Animal Behavior Processes*, 10, 390–409.
- McClelland, J. L. (1979). On the time relations of mental processes: an examination of systems of processes in cascade. *Psychological Review*, 86, 287–330.
- McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.
- McDermott, K. B., Buckner, R. L., Petersen, S. E., Kelley, W. M., & Sanders, A. L. (1999). Set- and code-specific activation in the frontal cortex: an fMRI study of encoding and retrieval of faces and words. *Journal of Cognitive Neuroscience*, 11, 631–640.
- Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental chronometry. *Biological Psychology*, 26, 3–67.
- Miller, J. O. (1993). A queue-series model for reaction time, with discrete-stage and continuous-flow models as special cases. *Psychological Review*, 100, 702–715.
- Miller, J. (1988). Discrete and continuous models of human information processing: theoretical distinctions and empirical results. *Acta Psychologica*, 67, 191–257.
- Miller, J., van der Ham, F., & Sanders, A. F. (1995). Overlapping stage models and the additive factor method. *Acta Psychologica*, 90, 11–28.
- Mouret, I., & Hasbroucq, T. (2000). The chronometry of single neuron activity: testing discrete and continuous models of information processing. *Journal of Experimental Psychology Human Perception and Performance*, 26, 1622–1638.
- Neisser, U. (1976). *Cognition and reality*. San Francisco, CA: Freeman.
- Nunez, P. (1981). *Electric fields of the brain*. New York: Oxford.
- Osman, A. (1998). Brainwaves and mental processes: electrical evidence of attention, perception, and intention. In D. Scarborough, & S. Sternberg, *An invitation to cognitive science, vol. 4: methods, models, and conceptual issues* (pp. 865–915). Cambridge, MA: MIT Press.
- Osman, A., Bashore, T. R., Coles, M. G. H., Donchin, E., & Meyer, D. E. (1992). On the transmission of partial information: inferences from movement-related brain potentials. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 217–232.

- Parker, A. J., & Newsome, W. T. (1998). Sense and the single neuron: probing the physiology of perception. *Annual Review of Neuroscience*, *21*, 227–277.
- Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, *331*, 585–589.
- Phillips, C. G., Zeki, S., & Barlow, H. B. (1984). Localization of function in the cerebral cortex: Past, present and future. *Brain*, *107*, 328–361.
- Plaut, D. C. (1995). Double dissociation without modularity: evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropsychology*, *17*, 291–321.
- Posner, M.I., & Raichle, M.E. (1994). *Images of mind*. New York: Freeman.
- Posner, M. I., & Raichle, M. E. (1995). Précis of images of mind. *Behavioral and Brain Sciences*, *18*, 327–383.
- Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*, 93–134.
- Rabbitt, P. M., & Maylor, E. A. (1991). Investigating models of human performance. *British Journal of Psychology*, *82*, 259–290.
- Raichle, M. E. (1998). Behind the scenes of functional brain imaging: a historical and physiological perspective. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 765–772.
- Requin, J., Riehle, A., & Seal, J. (1988). Neuronal activity and information processing in motor control: from stages to continuous flow. *Biological Psychology*, *26*, 179–198.
- Requin, J., Riehle, A., & Seal, J. (1993). Neuronal networks for movement preparation. In D. E. Meyer, & S. Kornblum, *Attention and performance XIV: synergies in experimental psychology, artificial intelligence, and cognitive neuroscience – a silver jubilee* (pp. 745–769). Cambridge, MA: MIT Press.
- Richardson-Klavehn, A., & Bjork, R. A. (1988). Measures of memory. *Annual Review of Psychology*, *39*, 475–543.
- Roberts, S. (1981). Isolation of an internal clock. *Journal of Experimental Psychology: Animal Behavior Processes*, *7*, 242–268.
- Roberts, S. (1982). Cross-modal use of an internal clock. *Journal of Experimental Psychology: Animal Behavior Processes*, *8*, 2–22.
- Roberts, S. (1983). Properties and function of an internal clock. In R. L. Mellgren, *Animal cognition and behavior* (pp. 345–397). Amsterdam: North-Holland.
- Roberts, S. (1987). Evidence for distinct serial processes in animals: the multiplicative-factors method. *Animal Learning and Behavior*, *15*, 135–173.
- Roberts, S. (1993). Use of independent and correlated measures to divide a time discrimination mechanism into parts. In D. E. Meyer, & S. Kornblum, *Attention and performance XIV: synergies in experimental psychology, artificial intelligence, and cognitive neuroscience – a silver jubilee* (pp. 589–610). Cambridge, MA: MIT Press.
- Roberts, S. (1998). The mental representation of time: uncovering a biological clock. In D. Scarborough, & S. Sternberg, *An invitation to cognitive science, vol. 4: methods, models, and conceptual issues* (pp. 53–105). Cambridge, MA: MIT Press.
- Roberts, S., & Sternberg, S. (1993). The meaning of additive reaction-time effects: tests of three alternatives. In D. E. Meyer, & S. Kornblum, *Attention and performance XIV: synergies in experimental psychology, artificial intelligence, and cognitive neuroscience – a silver jubilee* (pp. 611–653). Cambridge, MA: MIT Press.
- Rubinstein, J., Meyer, D. E., & Evans, J. E. (2000). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance* (accepted).
- Rugg, M. D., & Coles, M. G. H. (1995). The ERP and cognitive psychology: conceptual issues. In M. D. Rugg, & M. G. H. Coles, *Electrophysiology of mind: event-related brain potentials and cognition* (pp. 27–39). Oxford: Oxford.
- Sachs, M. B., Nachmias, J., & Robson, J. G. (1971). Spatial-frequency channels in human vision. *Journal of the Optical Society of America*, *61*, 1176–1186.
- Sanders, A. F. (1980). Stage analysis of reaction processes. In G. E. Stelmach, & J. Requin, *Tutorials in motor behavior* (pp. 331–354). Amsterdam: North-Holland.
- Sanders, A. F. (1990). Some issues and trends in the debate on discrete vs continuous processing of information. In E. J. Stoffels, M. W. van der Molen, & P. J. Keuss, *Stage analysis of the reaction process: models, methods, and applications. Acta Psychologica*, *74*, 123–169.

- Sanders, A. F. (1998). *Elements of human performance: reaction processes and attention in human skill*. Mahway, NJ: Erlbaum.
- Sanders, A. F., Wijnen, J. L. C., & Van Arkel, A. E. (1982). An additive factor analysis of the effects of sleep loss on reaction processes. *Acta Psychologica*, *51*, 41–59.
- Sarter, M., Berntson, G. G., & Cacioppo, J. T. (1996). Brain imaging and cognitive neuroscience: toward strong inference in attributing function to structure. *American Psychologist*, *51*, 13–21.
- Scheffé, H. (1959). *The analysis of variance*. New York: Wiley.
- Schweickert, R. (1985). Separable effects of factors on speed and accuracy: memory scanning, lexical decision, and choice tasks. *Psychological Bulletin*, *97*, 530–546.
- Shallice, T. (1981). Neurological impairment of cognitive processes. *British Medical Bulletin*, *37*, 187–192.
- Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge, UK: Cambridge.
- Shaw, M. L. (1982). Attending to multiple sources of information: I. The integration of information in decision making. *Cognitive Psychology*, *14*, 353–409.
- Shoben, E. J., & Ross, B. H. (1986). The crucial role of dissociations. *Behavioral and Brain Sciences*, *9*, 568–571.
- Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, *106*, 467–482.
- Smulders, F. T. Y., Kok, A., Kenemans, J. L., & Bashore, T. R. (1995). The temporal selectivity of additive factor effects on the reaction process revealed in ERP component latencies. *Acta Psychologica*, *90*, 97–109.
- Sternberg, S. (1969). The discovery of processing stages: extensions of Donders' method. In W. J. Koster, *Attention and performance II*. *Acta Psychologica*, *30*, 276–315.
- Sternberg, S. (1984). Stage models of mental processing and the additive-factor method. *Behavioral and Brain Sciences*, *7*, 82–84.
- Sternberg, S. (1998a). Inferring mental operations from reaction-time data: how we compare objects. In D. Scarborough, & S. Sternberg, *An invitation to cognitive science, vol. 4: methods, models, and conceptual issues* (pp. 365–454). Cambridge, MA: MIT Press.
- Sternberg, S. (1998b). Discovering mental processing stages: the method of additive factors. In D. Scarborough, & S. Sternberg, *An invitation to cognitive science, vol. 4: methods, models, and conceptual issues* (pp. 703–863). Cambridge, MA: MIT Press.
- Swets, J. A. (1998). Separating discrimination and decision in detection, recognition, and matters of life and death. In D. Scarborough, & S. Sternberg, *An invitation to cognitive science, vol. 4: methods, models, and conceptual issues* (pp. 635–702). Cambridge, MA: MIT Press.
- Swets, J. A., Tanner, Jr., W. P., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review*, *68*, 301–340.
- Taylor, D. H. (1966). Latency components in two-choice responding. *Journal of Experimental Psychology*, *72*, 481–487.
- Teller, D. Y. (1984). Linking propositions. *Vision Research*, *24*, 1233–1246.
- Thapar, A., & Greene, R. L. (1994). Effects of level of processing on implicit and explicit tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 671–679.
- Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge, UK: Cambridge.
- Treisman, M. (1998). Combining information: probability summation and probability averaging in detection and discrimination. *Psychological Bulletin*, *3*, 252–265.
- Tukey, J. W. (1949). One degree of freedom for nonadditivity. *Biometrics*, *5*, 232–242.
- Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford.
- Vallé-Inclan, F., & Redondo, M. (1998). On the automaticity of ipsilateral response activation in the Simon effect. *Psychophysiology*, *35*, 366–371.
- Van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information processing in the primate visual system: an integrated systems perspective. *Science*, *255*, 419–423.
- Wandell, B.A. (1995). *Foundations of vision*. Sunderland, MA: Sinauer.
- Wandell, B. A. (1999). Computational neuroimaging of human visual cortex. *Annual Review of Neuroscience*, *22*, 145–173.