

## Modular Processes in Mind and Brain

Saul Sternberg\*

### Abstract

One approach to understanding a complex process or system starts with an attempt to divide it into *modules*: parts that are independent in some sense, and having different functions. In this chapter I discuss a method for the modular decomposition of neural and mental processes. Several applications of process decomposition are presented. Two are well-established and purely behavioral (discriminability and signal detection theory; reaction time and the method of additive factors), and lead to the identification of *mental modules*. One involves electrophysiological brain measurements, and leads to the identification of *neural modules*. Two use both neural (ERP or fMRI) and behavioral (RT) measurements, and lead to the identification of neural and mental modules that correspond. Three of the examples raise questions without answering them: One involves both neural (fMRI) and behavioral (RT) measurements, but leads to the identification of only neural modules. Two promising but incomplete applications are attempts to relate brain regions to behaviorally defined processing modules, using a brain manipulation (transcranial magnetic stimulation, TMS), but suffer from flaws in design or analysis. I show why the process-decomposition method discussed here, in which the criterion for modularity is *separate modifiability*, is superior for modular decomposition to the more frequently used task comparison procedure and to its associated subtraction method. As a contrast, an example of task comparison involving TMS is discussed. Finally, five additional applications are briefly described, to indicate the wide range of organisms, measures, and modules to which the process-decomposition method can be applied.

---

\* E-mail: saul@psych.upenn.edu  
Tel: 215-898-7162  
Fax: 215-898-7210  
Address:  
Department of Psychology  
University of Pennsylvania  
3401 Walnut Street, C-Wing  
Philadelphia, PA 19104-6228, USA

## Modular Processes in Mind and Brain

Saul Sternberg

### 1.1 Modules and Modularity

The first step in one approach to understanding a complex process or system is to attempt to divide it into *modules*: parts that are independent in some sense, and having different functions.<sup>1</sup> Early in the last century, scientific psychology, dominated by behaviorism, emphasized the directly observable relations between stimuli and responses, and devoted little effort to describing the perception, memory, and thought processes that intervene. During the second half of the century there was a change in the kinds of questions that psychologists asked and in the acceptable answers. This change was influenced by the growth of computer science, which persuaded psychologists that programming concepts might be acceptable as precise descriptions of information processing by people as well as by machines. And the software-hardware distinction supported the legitimacy of theories couched in terms of abstract information-processing operations in the mind rather than only neurophysiological processes in the brain. In the "human information processing" approach, complex activities of perception, decision, and thought, whether conscious or unconscious, came to be conceptualized in terms of functionally distinct and relatively independent ("modular") sub-processes responsible for separate operations such as input, transformation, storage, retrieval, and comparison of internal representations — modules whose arrangement was expressed in systematic flow charts.<sup>2</sup> A useful distinction that is encouraged by comparing the mind to a digital computer is that between processors and the processes they may implement. The existence of functionally specialized processors is a sufficient condition but not a necessary one for functionally distinct processes, the concern of the present chapter.

The rise in the 1980s of parallel distributed processing might seem on the surface to conflict with the idea of modular organization of processes, but it need not: such models "do not deny that there is a macrostructure," and are intended to "describe the internal structure of the larger (processing) units" ([4], p. 12). Furthermore, even starting with a relatively unstructured neural network, there is reason to believe that over time and with experience it will develop functionally specialized processing modules, and hence, functionally specialized processes [5].

Toward the end of the century, it became possible (using fMRI, for example) to measure the human brain in action with previously unattainable spatial resolution. Because functions of the brain are often implemented by specialized neural processors that are anatomically localized, these new measurement methods encouraged scientists to attempt the modular decomposition of complex neural processes, just as they had been doing for complex mental processes.

In the present chapter I shall be describing a method for the modular decomposition of two kinds of complex process, mental and neural, along with several examples of its application from the psychology and cognitive neuroscience literature.<sup>3</sup> A mental-process module is a part of a

- 
1. A module may itself be composed of modules.
  2. Heuristic arguments for the modular organization of complex biological computations have been advanced by Simon [1, 2] and, in his 'principle of modular design', by Marr ([3], p. 485). Marr argued that "Any large computation should be split up and implemented as a collection of small sub-parts that are as nearly independent of one another as the overall task allows. If a process is not designed in this way, a small change in one place will have consequences in many other places. This means that the process as a whole becomes extremely difficult to debug or to improve, whether by a human designer or in the course of natural evolution, because a small change to improve one part has to be accompanied by many simultaneous compensating changes elsewhere."

process, functionally distinct from other parts, and investigated with behavioral measures. Such modules will be denoted **A**, **B**, etc. A neural-process module is a part of a neural process, functionally distinct from other parts, and investigated with brain measures. Such modules will be denoted  $\alpha$ ,  $\beta$ , etc. (When either kind of module is meant, I shall use **A**, **B**, etc.) The distinction between processes and processors is sometimes overlooked. Processes occur over time; their arrangement is described by a flow-chart. In contrast, processors are parts of a physical or biological device (such as the brain); their arrangement can often be described by a circuit diagram.

## 1.2 The Process-Decomposition Method

### 1.2.1 Separate modifiability, process-specific factors, selective influence, and functional distinctness

Much thinking by psychologists and brain scientists about the decomposition of complex processes appeals either implicitly or explicitly to *separate modifiability* as a criterion for modularity: Two (sub)processes **A** and **B** of a complex process (mental or neural) are modules if and only if each can be changed independently of the other.<sup>4</sup> One purpose of the present chapter is to explicate by example the notion of separate modifiability and the conditions under which one can assert it. To demonstrate separate modifiability of **A** and **B**, we must find an instance of *selective influence*. That is, we must find experimental manipulations (factors) *F* and *G* that influence **A** and **B** selectively, i.e. such that **A** is influenced by *F* but is invariant with respect to *G*, whereas **B** is influenced by *G* but is invariant with respect to *F*. Usually one starts with hypotheses about what the component processes are, and about corresponding *process-specific factors* that are likely to influence them selectively. Separate modifiability of **A** and **B** is also evidence for their *functional distinctness*; information about what a process does is provided by the sets of factors that do and don't influence it; if two processes had the same function they would be influenced by the same factors.<sup>5</sup>

### 1.2.2 Processes and their measures, pure and composite, and combination rules

How do we demonstrate that a process is influenced by a factor, or invariant with respect to it? We know only about one or more hypothesized *measures*  $M_A$  of process **A**, not about the process as such. Depending on the available measures, there are two ways to assess separate modifiability of **A** and **B**. Suppose we have *pure measures*  $M_A$  and  $M_B$  of the hypothesized modules: A pure measure of a process is one that reflects changes in that process only. Examples include the durations of two different neural processes (see Section 1.3.1), and the discriminability and criterion parameters of signal-detection theory (which, in suitable experiments, reflect sensory and decision processes; see Section 1.5). To show that *F* and *G* influence **A** and **B** selectively, we must demonstrate their selective influence on  $M_A$  and  $M_B$ . That is, we must show that  $M_A$  is influenced by *F* and invariant with respect to *G*, and vice versa for  $M_B$ . If  $F_j$  has two levels,  $j = 1, 2$ , the effect of *F* on  $M_A$  is a difference:

$$\text{effect}(F) = \bar{M}_A(F_2) - \bar{M}_A(F_1) . \quad (1)$$

For factors with multiple levels the effect can be regarded as a vector of differences associated

- 
3. In [6] I discuss and defend the method, describe its antecedents, illustrate it with a dozen applications to mental and neural processes, and further discuss its inferential logic.
  4. This criterion for modularity seems to be far weaker than the set of module properties suggested by Fodor [7], according to whom modules are typically innate, informationally encapsulated, domain specific, "hard-wired", autonomous, and fast. Domain specificity appears to imply separate modifiability.
  5. Such double dissociation of subprocesses should be distinguished from the more familiar double dissociation of tasks, discussed in Section 1.7. For comments on the distinction see [8].

with successive levels. The logic for inferring separate modules when we hypothesize that we have pure measures is shown in Table 1.<sup>6</sup>

Table 1  
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 $M_A(\mathbf{A})$ and $M_B(\mathbf{B})$ selectively: $p_1: M_A \leftarrow F$ , $p_2: M_B \not\leftarrow F$ , $p_3: M_B \leftarrow G$ , $p_4: M_A \not\leftarrow G$ .	
<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 $F$ , $G$ .

The influence and invariance requirements are both critical. Unfortunately, it is seldom appreciated that persuasive evidence for invariance cannot depend solely on failure of a significance test of an effect: such a failure could merely reflect variability and low statistical power.<sup>7</sup>

Instead of pure measures, suppose we have a *composite measure*  $M_{AB}$  of the hypothesized modules — a measure to which they both contribute. Examples of composite measures are the event-related potential (ERP) at a particular point on the scalp (which may reflect several ERP sources in the brain), and mean reaction time,  $\overline{RT}$  (which may depend on the durations of several processes). To support a hypothesis of selective influence in this case we must also know or have evidence for a *combination rule* — a specification of how the contributions of the modules to the measure combine. With pure measures factorial experiments (rather than separate experiments for different factors) are desirable, because they provide tests of generality. However, with a composite measure factorial experiments are essential, to assess how the effects of the factors combine; unfortunately they are rare.

- 
6. When the hypotheses about **A** and **B** are sufficiently detailed to specify particular process-specific factors that should influence them selectively, this leads to an alternative formulation of the inferential logic, in which the specification of  $F$  and  $G$  is included in the joint hypothesis, with the remainder of the reasoning adjusted accordingly. For a discussion of such alternatives, see [4], Section A.2.3.
  7. Perhaps the most common error of interpretation in psychology is to assert the nonexistence of an effect or interaction merely because it fails to reach statistical significance. In evaluating a claim that an effect is null, it is crucial to have at least an index of precision (such as a confidence interval) for the size of the effect. An alternative is to apply an *equivalence test* that reverses the asymmetry of the standard significance test [9, 10]. In either case we need to specify a critical effect size (depending on what we know and the particular circumstances) such that it is reasonable to treat the observed effect as null if, with high probability, it is less than that critical size.

The logic for inferring separate modules using a composite measure when we either know or hypothesize that the combination rule is *summation* is shown in Table 2.<sup>8</sup>

Table 2  
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 modifi able).	
<i>H3</i> : Contributions $u_A, v_B$ of <b>A, B</b> to $M_{AB}(\mathbf{A}, \mathbf{B})$ combine by <i>summation</i> .	
<b>Prediction</b>	
We may be able to fi nd 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 fi nd factors $F$ and $G$ with <i>additive effects</i> on $M_{AB}$ .	We fail to fi nd such factors.
<b>Inferences</b>	
Support for joint hypothesis <i>H1 + H3</i> .	Refutes one/both of <i>H1, H3</i> , or we didn't look enough for $F, G$ .

To understand Table 2 it is important to keep in mind what the effect of a factor is, and what it means for effects of different factors to be *additive*. To simplify the discussion, let us assume that there are two factors, each with just two levels. Let  $u$  and  $v$  be the contributions of processes **A** and **B** to  $M_{AB}$ . If summation is the combination rule, we have  $M_{AB} = u + v$ . If **A** and **B** are selectively influenced by factors  $F$  and  $G$ , we have

$$M_{AB}(F_j, G_k) = u(F_j) + v(G_k) , \quad (2)$$

where, for example,  $u(F_j)$  is a function that describes the relation between the 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  $u, v$ , and  $M_{AB}$  as random rather than deterministic variables. Using  $\bar{M}, \bar{u}$ , and  $\bar{v}$  to indicate the means of these random variables, it is convenient that with no further assumptions<sup>9</sup>, Eq. (2) implies:

$$\bar{M}_{AB}(F_j, G_k) = \bar{u}(F_j) + \bar{v}(G_k) . \quad (3)$$

In what follows, I treat the levels of factors as ordered, which permits describing changes in level as increases or decreases. From Eq. (3) it is easy to show that  $F$  and  $G$  are *additive factors*: the combined effect on  $M_{AB}$  of increasing the levels of both  $F$  and  $G$  is the sum of the effect of increasing only  $F$  and the effect of increasing only  $G$ :

8. Whereas properties  $\{p_k\}$  (Table 1) apply to observable quantities, the analogous properties  $\{p'_k\}$  (Table 2) apply to contributions to a composite measure that are not directly observable.

9. Matters are not so simple for other combination rules, such as multiplication; see Section 1.9.3.

$$\begin{aligned} effect(F, G) &\equiv \bar{M}_{AB}(F_2, G_2) - \bar{M}_{AB}(F_1, G_1) = [\bar{u}(F_2) + \bar{v}(G_2)] - [\bar{u}(F_1) + \bar{v}(G_1)] \\ &= [\bar{u}(F_2) - \bar{u}(F_1)] + [\bar{v}(G_2) - \bar{v}(G_1)] \equiv effect(F) + effect(G) . \end{aligned} \quad (4)$$

Eq. (2) also implies that the effect of each factor will be invariant over levels of the other. Thus,

$$\begin{aligned} effect(F | G = G_k) &\equiv \bar{M}_{AB}(F_2, G_k) - \bar{M}_{AB}(F_1, G_k) = [\bar{u}(F_2) + \bar{v}(G_k)] - [\bar{u}(F_1) + \bar{v}(G_k)] \\ &= [\bar{u}(F_2) - \bar{u}(F_1)], \text{ regardless of } G_k . \end{aligned} \quad (5)$$

A given measure may be pure or composite, depending on the hypothesized modules of interest. This attribute of a measure is part of the joint hypothesis that is tested as part of the process-decomposition method.

### 1.2.3 Overview of examples and issues

Much of what follows consists of descriptions of eight primary examples of the process decomposition method, the first six successful. There are also brief summaries of five secondary examples, in Section 1.9, which are included to convey the breadth of the method.) Examples will be referred to by the section numbers in which they are first discussed. Thus, the example discussed in Section 1.4.3 will be called "Ex. 1.4.3". Some features of the examples are provided in Table 3, including references to papers in which they are described more fully.

With two exceptions, the successful primary examples involve factorial experiments with two factors. The exceptions are Ex. 1.3.2, in which the effects of the two factors are studied in separate experiments, and Ex. 1.4.3, in which we consider the effects of three factors. In all cases the factors have been selected because it is hoped that they will be "process specific": that they will selectively influence only one of the two or more processes that are hypothesized to underlie performance of the task. The final two examples (1.8.1 and 1.8.2) are intriguing and tantalizing cases where the method could have been used but was not, because of incompleteness of either the design or the analysis.

In Section 1.3 I discuss two applications based on electrophysiological measurements at the scalp, treated as pure measures, and shown to be such. In both cases, two neural modules are identified: In Ex. 1.3.1, in which the modules are associated with preparation of two different aspects of the response, they are found to operate successively, as *stages*. In Ex. 1.3.2, in which the modules are associated with encoding two different aspects of the stimulus, they are found to operate in parallel. The RT data from Ex. 1.3.1 are discussed in Section 1.4.1, where the RT is treated as a composite measure and shown to lead to the identification of two mental modules that correspond to the neural modules inferred from the electrophysiological data discussed earlier. (By "correspond to" I mean that their durations are influenced selectively by the same factors and to the same extent.) The inference from the RT data exemplifies the method of additive factors (AFM), discussed more generally in Section 1.4.2, and applied in Section 1.4.3 to the problem of locating the effect of a manipulation of interest (sleep deprivation) within a pair of already established mental modules, one for encoding the stimulus, the other for selecting the response. In Section 1.4.4 I comment on the conflict between the conclusion from the AFM that the sleep deprivation effect is selective and a claim, based on the more popular task-comparison method, that its effect is global.

Table 3  
Some Features of the Examples

Ex#	Example	Measure(s)	Factors	Module Type(s)	Measure Type(s)	References
1.3.1	Serial modules for preparing two response features	LRP RT	SQ Complexity	N M	P C(sum)	[11] a226
1.3.2	Parallel modules for encoding two stimulus features	LRP Go-NoGo divergence	GND MC	N	P P	[12] a167
1.4.3	Selective effect of sleep deprivation	RT	SQ MF SLP	M	C(sum)	[13] a206, a237
1.5.2	Sensory and decision modules (pigeon)	d-prime Criterion	Luminance RR	M	P P	[14] a172, a228
1.6.1	Number encoding and comparison	fMRI RT	Numeral vs. name Proximity	N M	P C(sum)	[15] b132
1.6.2	Modules for stimulus encoding and response selection	fMRI RT	SQ MC	N	P,C(sum)	[16] —
1.7.1	Tactile perception (Task Comparison, not Process Decomposition)	Subjective magnitude	Dot spacing rTMS Task	?	?	[17] b129
1.8.1	Visual conjunction search (Incomplete design)	RT	TMS	?	?	[18] —
1.8.2	Number comparison (Incomplete analysis)	RT	Proximity Hand rTMS	?	?	[19] —
1.9.1	Isolation of a timing module (rat)	Peak rate Peak time	Food time Food prob.	M	P P	[20] a164
1.9.2	Evidence for modular spatial-frequency analyzers from selective adaptation	High-freq threshold Low-Freq threshold	High-freq adaptation Low-freq adaptation	M	P P	[21] a179, a229
1.9.3	Evidence for modular spatial-frequency analyzers from 'probability summation' in detection	$1 - Pr\{\text{detect}\}$	High-freq contrast Low-freq contrast	M	C(mult)	[22] a191, a235
1.9.4	Evidence from ERP amplitude for modules	ERP (by location)	Relatedness Satiation	N	C(sum)	[23] A196
1.9.5	Modular processes for learning and motivation (rat)	Response rate	Deprivation hrs. Food freq.	M	C(mult)	[24], a200, [25] a236

Module Types: M = Mental, N = Neural.

Measures: P = Pure, C = Composite, with combination rule (summation, multiplication) indicated.

References: aXXX are starting page numbers in [6]; bXX is a starting page number in [26].

As discussed in Section 1.5.1, signal detection theory has provided a widely applied method for isolating and measuring sensory processes in tasks that also involve decision processes, but has in general failed to isolate those decision processes. A variant of the method, in which the decision factor that is controlled differs from the traditional payoff matrix, has been used to successfully demonstrate the modularity of decision as well as sensory processes in Ex.

1.5.2, an experiment with pigeons.

To the extent that there is localization of function in the brain, so that each of several modular processes are implemented in disjoint regions, and to the extent that the level of activation in a region varies with changes in the process it implements, the level of such activation can function as a pure measure of the process. In Section 1.6 I discuss two examples in which fMRI signals in different brain regions were measured for this purpose; in both cases, RTs were measured as well. In Ex. 1.6.1 (number comparison), in which RT measurements in a similar experiment had already indicated separate stages for encoding the test number and comparing it to the target, both the new RT data and the fMRI data support this analysis, suggesting mental and neural modules that correspond. However, the fact that the direction of the effect of the encoding factor on the fMRI response differs in different brain regions raises interesting questions of interpretation. In Ex. 1.6.2 (a novel choice-reaction paradigm with four stimulus-response pairs) the fMRI data support the hypothesis of modular processes for stimulus encoding and response selection, but unlike earlier observations of effects on  $\overline{RT}$  from several similar paradigms, effects of relevant factors on  $\overline{RT}$  interact rather than being additive, raising questions about interpretation. One unexpected and potentially important finding in this example is the additivity of effects of the encoding and response-selection factors on the fMRI measure in the two brain regions found where both factors have effects.

In Section 1.7 I contrast process decomposition with the more familiar *task-comparison* method, already mentioned, in Section 1.4.4, in connection with the sleep deprivation example. I describe two examples of task comparison, one based on the effects of repetitive TMS, and the other Donders' subtraction method, originally developed for RT experiments, but, in recent years, applied to brain-activation data. Unlike task comparison, which is often used in a way that requires various assumptions (including modularity) to be made without test ([27], Ch. 11; [4], Appendix A.1.), the process-decomposition method incorporates such a test.

The final two primary examples both investigate the effects of transcranial magnetic stimulation (TMS) on RT measurements. The goal is to relate brain regions to mental modules. In these experiments, TMS in certain brain regions is found to increase the RT without otherwise disrupting performance. The presence and absence of TMS in such a region can be regarded as two levels of a factor. If this factor influences RT, the brain region to which TMS is applied probably plays some role in performance of the task. However, the potential of this method to associate brain regions with component processes is realized only when the TMS factor is used together with other, process-specific factors, to determine which of their effects, if any, are modulated by TMS. For visual search (Ex. 1.8.1), the absence of variation of a within-task search-specific factor precluded doing so. For number comparison (Ex. 1.8.2) a comparison-specific factor was varied, but no focused test of the interaction of this factor with TMS was conducted.

In Section 1.9 I provide brief summaries of five additional examples of application of the process decomposition method, examples that illustrate its wide range of organisms, measures, combination rules, and modules. Included with Ex. 1.9.3 is an outline of the inferential logic associated with a multiplicative combination rule.

### **1.3 Decomposing Neural Processes with the Lateralized Readiness Potential**

#### **1.3.1 Serial modules for preparing two response features**

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 detected as an increase in the averages over trials of the difference between electrical potentials (ERPs) at the two corresponding scalp locations.<sup>10</sup> Let  $A_{mc}(t)$  (an index of *motor-cortex asymmetry*) express this difference as a function of time from stimulus onset (if "stimulus locked"), or as a function of time before the overt response (if "response locked").  $A_{mc}(t)$  is normally zero, but becomes positive when the response hand is selected; the increase of such asymmetry is called the *lateralized-readiness potential*, or *LRP*. The onset time of the *LRP* is thus an estimate of the time at which the side of the response (left or right) has been selected. Having an estimate of this time makes it possible to ask whether the neural process responsible for selecting the side of the response is separate from the neural process responsible for preparing other aspects of the response.

This possibility was exploited in an experiment by Smulders et al. [11]. It was a two-choice RT experiment with single-numeral stimuli mapped on left-hand and right-hand responses. Two factors were varied orthogonally across blocks of trials: stimulus quality ( $SQ_j$ : digit intact versus degraded) and response complexity ( $RC_k$ : one keystroke —s imple —versus a sequence of three keystrokes —complex —made by fi ngers of the responding hand).<sup>11</sup> In addition to the composite measure  $RT_{jk}$ , Smulders et al. measured the onset time of the LRP, based on both stimulus-locked ( $LRP_s$ ) and response-locked ( $LRP_r$ ) averaging of the  $A_{mc}(t)$  functions. Let  $T_{sjk}$  and  $T_{rjk}$  be the corresponding LRP onset times measured with respect to stimulus and response, and let  $T_{\cdot jk}$  be their mean. Averaging over the four conditions,  $RT_{\cdot\cdot} = 416$  ms and  $T_{\cdot\cdot} = 264$  ms. If  $\alpha$  is the neural process from stimulus to LRP onset, and  $\beta$  is the neural process from LRP onset to response, these values give us measures of their durations (pure measures of  $\alpha$  and  $\beta$ ) averaged over the four conditions:  $D_{\alpha\cdot\cdot} = T_{\cdot\cdot} = 264$  ms and  $D_{\beta\cdot\cdot} = RT_{\cdot\cdot} - T_{\cdot\cdot} = 152$  ms.

-----  
 Insert Fig. 1 about here  
 -----

Fig. 1. Means over 14 subjects of data from Smulders et al. [11]. Estimated durations  $D_{\alpha jk}$  of Process/Stage  $\alpha$ , from stimulus presentation to LRP onset (Panels A1, B1); and  $D_{\beta jk}$  of Process/Stage  $\beta$ , from LRP onset to response (Panels A2, B2). These are shown as functions of *Stimulus Quality* ( $SQ_j$ , Panels A1, A2); and of *Response Complexity* ( $RC_k$ , 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  $34 \pm 6$  ms (Panel A1) and  $1 \pm 8$  ms (Panel A2); the corresponding main effects of  $RC$  are  $4 \pm 8$  ms (Panel B1) and  $21 \pm 7$  ms (Panel B2). The  $\overline{RT}$  data (discussed in Section 1.4.3) are shown in Panel C, together with two fitted models. One (unbroken lines) is the best-fitting additive model (mean absolute deviation 0.5 ms); the other (dotted lines) is an additive model based on estimates of process durations from the LRP data (mean absolute deviation 1.8 ms).

Shown in the A and B panels of Fig. 1 are the estimated durations of processes  $\alpha$  and  $\beta$  separated by condition:  $D_{\alpha jk} = T_{\cdot jk}$ , and  $D_{\beta jk} = RT_{jk} - T_{\cdot jk}$ . 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  $T_{\cdot jk}$ . The results show that the two factors  $SQ$  and  $RC$  have selective effects on  $D_{\alpha}$  and  $D_{\beta}$ . This supports

- 
10. As in some other brain measurements (e.g., PET, fMRI), averaging over trials is often required for the measures to be interpretable, because of the poor  $S/N$  ratio. Here, the "noise" is due partly to neural events unrelated to the task being performed, whose contributions are eliminated by combining subtraction of the pre-stimulus baseline level with an averaging process that reveals only those events that are consistently time-locked to the stimulus or the response.
11. Because there were two levels of each of two factors, and all combinations of levels of the two factors were used (a "complete factorial" experiment), the experiment contained four conditions.

the hypothesis that in this situation the LRP onset indeed defines a boundary between two neural modules which function sequentially, as stages, consistent with the reasoning in Table 2. In contrast to  $D_\alpha$  and  $D_\beta$ , which are hypothesized (and confirmed) to be pure measures, the  $\{RT_{jk}\}$ , shown in Panel C, are composite measures. With two stages between stimulus and response, the RT reflects contributions from both. In Section 1.4.1 I consider the RT data from this experiment and how they relate to the LRP data.

### 1.3.2 Parallel modules for discriminating two stimulus features

Consider a situation in which two different features of the same stimulus must be discriminated to determine how to respond. Are modular neural processes involved in doing so? And, if so, how are they organized temporally? Osman et al. [12] devised a clever way of asking these questions, using the LRP. On each trial, the visual stimulus had two features. Its *position* (left versus right, which was rapidly discriminated) indicated which response to make should a response be required. Its *category* (letter versus 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 no response should be made). Under these conditions, the LRP occurs even on trials with no overt response, and with an onset that is indistinguishable from the LRP on "Go" trials.

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 stimulus location has been discriminated and the response selected. Let us denote this *location discrimination* process by  $\alpha$ . Normally (on "Go" trials)  $A_{mc}$  continues to rise until the overt response is initiated. If a "NoGo" signal tells the subject not to respond, however,  $A_{mc}$  starts falling. 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 category (the NoGo signal) is discriminated and response preparation ceases. Let us denote this *category discrimination* process by  $\beta$ .  $T_1$  and  $T_2$  are hypothesized to be pure measures of their respective processes.

-----  
 Insert Fig. 2 about here  
 -----

Fig. 2. Schematic idealized asymmetry functions  $A_{mc}(t)$  from Osman et al. [12]. Event 1 is the LRP onset; Event 2 is the onset of the divergence of  $A_{mc}(t; Go)$  from  $A_{mc}(t; NoGo)$ . Panels A1 and A2: Asymmetry functions from Exp. 1, in which *Go-NoGo Discriminability* could be high ( $GND = GND_1$ ) or low ( $GND = GND_2$ ). Panels B1 and B2: Asymmetry functions from Exp. 2, in which the *Stimulus-Response Mapping* could be compatible ( $SRM = SRM_1$ ) or incompatible ( $SRM = SRM_2$ ).

-----  
 Insert Fig. 3 about here  
 -----

Fig. 3. Mean effects of two factors on the latencies of Events 1 and 2 in Osman et al. [12]. Panel A: Effects of *Go-NoGo Discriminability*,  $GND$  (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 *Stimulus-Response Mapping*,  $SRM$  (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.

Events 1 and 2 indicate the completion of processes  $\alpha$  and  $\beta$ . Can response preparation start when the location 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 *high* (letter and digit with dissimilar shapes,  $GND_1$ ) or *low* (similar shapes,  $GND_2$ ). The other factor (in Exp. 2) is the

spatial compatibility of the stimulus-response mapping,  $MC$ , which should influence  $\alpha$ ; it could be *compatible* (respond with the hand on the same side as the stimulus,  $MC_1$ ) or *incompatible* (respond with the hand on the opposite side,  $MC_2$ ).

Idealizations of the resulting  $A_{mc}(t)$  functions are shown in Fig. 2, and the observed values of  $\bar{T}_1$  and  $\bar{T}_2$  in Fig. 3. Each of the four panels of Fig. 2 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 Exp. 1, *GND* influenced  $\bar{T}_2$  (by 43 ms) but not  $\bar{T}_1$  (compare Figs. 2A1 and 2A2, and see Fig. 3A). The absence of any effect of *GND* on  $\bar{T}_1$  ( $T_1 \approx 170$  ms on both Go and NoGo trials) shows that the location 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 stimulus-response (S-R) mapping was incompatible, which was expected to delay selection of the response. To ensure that stimulus location had an opportunity to influence response preparation on NoGo trials in both conditions, it was important to delay the Go-NoGo discrimination.<sup>12</sup> Osman et al. therefore reduced letter-digit shape discriminability so as to increase  $\bar{T}_2$  to about 350 ms. In this experiment,  $MC$  influenced  $\bar{T}_1$  (by 121 ms) but not  $\bar{T}_2$  (compare Figs. 2B1 and 2B2, and see Fig. 3B). Increasing the level of mapping difficulty from  $MC_1$  to  $MC_2$  therefore *reduced* the interval between Event 1 and Event 2.

Taken together, the two experiments show that  $MC$  and *GND* influenced the two measures  $T_1$  and  $T_2$  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, as 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 between 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; Fig. 3B) that the effect of  $MC$  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; Fig. 3A) for the effect of *GND*. The alternative to a stages arrangement is that  $\alpha$  and  $\beta$  operate in parallel, such that the  $\overline{RT}$  (on Go trials) is 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  $MC$  on  $\overline{RT}$  (16 ms) is dramatically smaller than its effect on  $\bar{T}_1$  (121 ms). This can happen because regardless of how much the duration of  $\alpha$  is shortened by decreasing the level of  $MC$ , response initiation on a Go trial must await the appropriate completion of  $\beta$  as well.

These findings about  $\overline{RT}$  contrast to those of Ex. 1.3.1, in which pure measures based on the *LRP* provide evidence for a serial arrangement of two neural processes. In that case, unlike this one (as we shall see in Section 1.4.1) a composite behavioral measure ( $\overline{RT}$ ) leads to a similar analysis of corresponding mental processes. In the present example, however,  $T_1$  is a measure of a process (response selection, separated from response execution) for which there may be no pure behavioral measure, and whose contribution to the composite  $\overline{RT}$  measure may be large or small,

12. It is this requirement that would have made it difficult to implement a suitable factorial experiment.

depending on the level of  $GND$ .

## 1.4 Mental Processing Stages Inferred From Reaction Times

### 1.4.1 Analysis of the reaction-time data in Ex. 1.3.1

The conclusion in Ex. 1.3.1 is that the time between the stimulus digit and the response (the RT) can be partitioned into two intervals, from stimulus to LRP (duration  $D_\alpha$ , influenced by  $SQ$  but not  $RC$ ), and from LRP to response (duration  $D_\beta$ , influenced by  $RC$  but not  $SQ$ ). Consider just the mean reaction time, measured under the four factor-level combinations:

$$\overline{RT}_{jk} = D_{\alpha j} + D_{\beta k} = D_\alpha(SQ_j) + D_\beta(RC_k). \quad (6)$$

Suppose we had only those data and not the LRP data. It follows from partitioning of the  $\overline{RT}$  shown in Eq. (6) that the combination rule for the contributions of  $\alpha$  and  $\beta$  is summation. As discussed in Section 1.2.2, that and the selective influence of factors  $SQ$  and  $RC$  on  $\alpha$  and  $\beta$  imply that the effects of  $SQ$  and  $RC$  on  $\overline{RT}$  are additive. In general, factors that selectively influence the durations of sequential components of the RT must have additive effects on the composite measure. Thus, if a hypothesis asserts that the RT in a particular task is the duration of two modular mental processes **A** and **B** arranged sequentially (as stages) with durations  $T_A$  and  $T_B$ , and selectively influenced by factors  $F$  and  $G$ , we should expect:

$$\overline{RT}(F_j, G_k) = T_A(F_j) + T_B(G_k). \quad (7)$$

The goodness of fit of the parallel unbroken lines in Fig. 1C confirms the expectation of additivity for the RTs in the experiment by Smulders et al. and supports the joint hypothesis of Table 2. (The interaction contrast of  $SQ$  and  $RC$  was a negligible  $2 \pm 5$  ms.) However, because in this case we also have measures of the durations of neural modules (from the analysis of the LRP data) we can go further: we can ask whether the mental modules derived from the purely behavioral analysis of the composite measure correspond to the neural modules inferred from the LRP-based pure measures.

Thus, suppose that the mental modules **A** and **B** responsible for the additive RT effects are implemented by the neural modules  $\alpha$  and  $\beta$ , demarcated by the LRP. Then, not only should the same factors influence them selectively, but also, the sizes of their effects should be the same. Agreement among the effect sizes can be examined by assuming that the two factors indeed have perfectly selective effects on  $\alpha$  and  $\beta$ , 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 Fig. 1A1 (averaging over  $RC$  levels) to obtain the estimates  $\hat{D}_\alpha(SQ_1)$  and  $\hat{D}_\alpha(SQ_2)$ . Similarly, we should be able to use just the data in Fig. 1B2 (averaging over  $SQ$  levels) to obtain the estimates  $\hat{D}_\beta(RC_1)$  and  $\hat{D}_\beta(RC_2)$ . If  $RT = D_\alpha + D_\beta$  we have the "predictions"  $RT_{jk}^* = \hat{D}_\alpha(SQ_j) + \hat{D}_\beta(RC_k)$  for the four conditions.<sup>13</sup> The dotted lines in Fig. 1C show that the agreement is good:  $RT_{jk} \approx RT_{jk}^*$ . Numerically, the effects of  $SQ$  and  $RC$  on the composite measure  $\overline{RT}$  are  $35 \pm 3$  ms and  $25 \pm 7$  ms, close to their mean estimated effects (34 and 21 ms) on the pure measures  $D_\alpha$  and  $D_\beta$ .

Another illustration of the independent use of two methods is provided by Ex. 1.6.1: pure measures to ask about the structure of a neural process, and a composite measure to ask about the structure of a corresponding mental process.

13. This way of deriving the  $\{RT_{jk}^*\}$  forces their means into agreement:  $RT_{..}^* = RT_{..}$ ; the question of interest is whether the differences among the two sets of four values agree.

### 1.4.2 The method of additive factors

The approach to decomposing complex mental processes into subprocesses that is exemplified by the RT analysis described above depends on the observation that if a process can be partitioned into subprocesses arranged in stages, then the RT becomes an example of a composite measure with summation as the combination rule; in this case if two factors  $F$  and  $G$  change  $\overline{RT}$  but influence no stages in common ("selective influence"), their effects on mean reaction time should be additive, as described in Table 2. Conversely, if factors  $F$  and  $G$  interact, so that  $G$  modulates the effect of  $F$  rather than leaving it invariant, then  $F$  and  $G$  must influence at least one stage in common. Suppose, then, that we have a process in which RT measurements have revealed two or more factors with additive effects. This supports<sup>14</sup> the hypothesis that the process contains subprocesses arranged sequentially, in stages, with each of the factors influencing a different subprocess selectively.<sup>15</sup> Thus one approach to searching for the modular decomposition of a complex process is the method of additive factors (AFM), which involves determining whether two or more factors have additive or interacting effects on mean RT.<sup>16</sup> The application discussed in the section that follows exemplifies how, by combining inferences from a pattern of additive and interacting factor effects, it is possible to reveal more complex processing structures.

### 1.4.3 Selectivity of the effect of sleep deprivation

One of the most provocative applications of the additive factor method is described by Sanders, Wijnen, & Van Arkel ([13], Exp. 1), and leads to the controversial conclusion that the effect of sleep deprivation is selective (process-specific) rather than global. What follows is a simplified description of their experiment and findings.

What permits asking the question of process specificity is a feature of this example that distinguishes it from the others: the inclusion of more than two factors. The stimuli were the single digits "2", "3", "4", and "5"; the responses were their spoken names, "two", "three", "four", and "five". Four factors were manipulated, each at two levels: The first was *Stimulus Quality (SQ)*; the digits, presented as dot patterns, could be *intact*, or *degraded* by adding other dots. The second was the *Mapping Familiarity (MF)* from digits to names; it could either be *high* (respond to each digit with its name) or *low* (respond to "2", "3", "4", and "5" with "three", "four", "five", and "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 with correct responses. Test sessions occurred in both the morning and afternoon, creating a fourth two-level factor, *Time of Day (TD)*. The  $2^4 = 16$  conditions were run in separate blocks of trials. For simplicity the data shown in Fig. 4 have been averaged over levels of *TD*.

Other studies (See Section 1.6.2) had already suggested that *SQ* and *MF* were likely to influence two different processing stages selectively, processes that might be described as

---

14. As in testing any set of hypotheses, the degree to which a hypothesis is supported by confirmation of a prediction is diminished to the extent that alternative hypotheses that generate the same prediction are plausible [28]. In this case the alternative contenders (some discussed in [29]) do not involve processors arranged in stages, but they do involve different processes that are selectively influenced by the factors, i.e., they involve modules. To discriminate among such alternatives, other aspects of the data may have to be considered.

15. Given an assumption about the durations of different stages that is stronger than zero correlation, but weaker than stochastic independence, stages plus selective influence implies numerous properties of other features of the RT distributions in addition to their means [29].

16. See [30], [31], and Section 1.11 below.

stimulus encoding (**S**) and response selection (**R**).<sup>17</sup>

-----  
 Insert Fig. 4 about here  
 -----

Fig. 4. Data from Sanders, Wijnen, and Van Arkel [13], Experiment 1. Means over the two levels of *Time of Day* (*TD*). Each pair of panels show the same  $2 \times 2 \times 2 = 8$  data points, plotted in different ways. Each point is the mean of about 300 *RT*'s 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 ms (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  $\pm SE$  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 *SE*.

The results in Fig. 4 consist of the  $\overline{RT}$ 's from the  $2 \times 2 \times 2 = 8$  conditions. Panels A1 and A2 show that at each level of *SLP* there are additive effects on  $\overline{RT}$  of *SQ* and *MF*. This evidence supports:

- (1) Performing the task involves at least two modules, arranged as stages, and
- (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 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 effect of the difficulty of stimulus encoding. 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, and among the factors *SQ*, *MF*, and *SLP*, *SQ* and *SLP* influence **S**, while *MF* alone influences **R**. It is reasonable to suppose that the stimulus is identified during **S**, and the response selected during **R**. (This is suggested by the nature of the factors *SQ* and *MF* that influence them.) The AFM has thus led us to the surprising conclusion that whereas *SLP* influences stimulus encoding, it does not influence response selection.

Unlike some other applications of composite measures, the findings from this experiment 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 that analysis, providing an example of localizing the influence of a third factor *SLP* in one of the identified modules, **S**, hereby further characterizing **S** and **R**.

#### 1.4.4 Finding the 'mechanism of action' of a manipulation: Module identification

It is in the context of the evidence for stages selectively influenced by *SQ* and *MF* that the interaction of *SLP* with *SQ* and its additivity with *MF* can be interpreted. The implication—that

---

17. This conclusion (separately-modifiable sequential processes, or stages) is further strengthened by analyses of complete *RT*-distributions rather than just *RT* means, from a similar experiment ([30], Sec. V; [29], Exp. 2).

sleep deprivation has effects that are process specific —contradicts the conclusion of Dinges and Kribbs ([32], p. 117), according to which there is "a generalized effect of sleepiness on all cognitive functioning". They based their conclusion not on the process-decomposition approach but on results of the more common "task-comparison method" (see Section 1.7 below), and the finding that performance is impaired by sleep deprivation in a wide range of tasks. If Dinges and Kribbs [32] were correct, *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 the effects of both, contrary to what Sanders et al. found.<sup>18</sup> The problem for the task-comparison method in this application is perhaps the likelihood that all tasks in which performance can be measured involve some perceptual operations, i.e., processes akin to **S**. Process-decomposition might be more fruitful than task-comparison also in investigating the mechanism of action of different drugs, for similar reasons: Even if a drug influences processes in class **A** and not in class **B**, it may be difficult to find any task that does not involve processes in both classes. This may be why, for example, in a study that used task comparison to determine which processes are affected by clonidine versus temazepam (known to have different pharmacological mechanisms) in which 11 tasks were used, each drug produced statistically significant effects on all except one of the tasks [34]. Using the process-decomposition approach, if factor *F* (*G*) is known to selectively influence process **A** (**B**) in just one task, and process **A** (**B**) is selectively influenced by Drug 1 (Drug 2), then both drugs would affect task performance, but the effect of *F* (*G*) on that performance would be modulated only by Drug 1 (Drug 2).

## 1.5 Signal-Detection Theory: Sensory and Decision Modules

### 1.5.1 The failure to find full modularity in traditional sensory experiments

The most influential approach to deriving pure measures of two processes underlying the performance of a task is the one associated with signal detection theory (SDT, [35, 36]). At the heart of this approach is the recognition that even simple psychophysical tasks involve decision processes as well as sensory mechanisms. Consider a psychophysical experiment in which two types of trials 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's call these representations  $X_T$  and  $X_{NT}$ , for  $S_T$  and  $S_{NT}$ . 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.

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

The data from such an experiment can be described by four numbers arranged in a  $2 \times 2$  matrix, 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

18. Electrophysiological evidence that confirms the selectivity of the effect of *SLP* can be found in [33].

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 to  $\bar{X}_T - \bar{X}_{NT}$ , which increases with discriminability. The other is an estimate of the 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-specific or *s-factors*), and some expected to influence just a decision process (decision-specific 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, 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**. In analogous studies with animals, partial reinforcement is often used: a food reward is provided for only a fraction of the correct responses of each kind; let these fractions be  $\pi_T$  and  $\pi_{NT}$  for  $R_T$  and  $R_{NT}$ .

In terms of the inferential logic of Table 1, 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 [37, 38], or animals [39]. The SDT approach thus seems to provide only "partial modularity".

Why has SDT failed in this respect? The possibility considered by Alsop [39] and by others whose work he reviews, is that the problem may be a failure of the assumption that the payoff matrix *PM* influences **D** selectively and controls response bias. A few animal studies suggest that response bias is controlled, not by the conditional probabilities described by the payoff matrix, but by the distribution of rewards over the two alternative responses (sometimes described as the "reinforcement ratio", *RR*).

One way to characterize *RR* is as  $Pr\{R_{NT}|Reward\}$ , the proportion of the total number of rewards (for both kinds of correct responses) that are given for  $R_{NT}$ . Assuming equal frequencies of  $S_T$  and  $S_{NT}$ ,  $Pr\{R_{NT}|Reward\} = r/(1+r)$ , where  $r$ , the ratio of the expected numbers of rewarded  $R_{NT}$  responses to rewarded  $R_T$  responses, is given by  $[\pi_{NT}Pr\{R_{NT}|S_{NT}\}] / [\pi_TPr\{R_T|S_T\}]$ . As well as 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 criterion. 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. Fortunately it is possible to control *RR* itself instead of the payoff matrix.

### 1.5.2 Modularity of sensation and decision when reinforcement ratio is controlled

In a luminance-discrimination experiment with six pigeons, McCarthy & Davison [14] 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; these 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 most difficult  $LR$  level, a value equal to the brighter luminance. The *reinforcement ratio* ( $RR$ ), described by  $Pr\{R_{NT}|Reward\}$ , could be one of three values, 0.2, 0.5, or 0.8.<sup>19</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 summarized by two proportions,  $Pr\{R_T|S_T\}$  and  $Pr\{R_T|S_{NT}\}$ . If the distributions of  $X_T$  and  $X_{NT}$  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 thirty such curves (6 birds  $\times$  5 levels of  $LR$ ) supports this expectation, and hence the equal-variance Gaussian model.

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  $\hat{c} = [z(Pr\{R_T|S_T\}) + z(Pr\{R_T|S_{NT}\})]/2$ . The origin for the criterion measure is the midpoint between  $\bar{X}_{NT}$  and  $\bar{X}_T$ ; the sign of the criterion thus expresses the direction of the bias. Means over birds of these two measures are shown in Fig. 5.

-----  
 Insert Fig. 5 about here  
 -----

Fig. 5. Mean effects of *Reinforcement Ratio*,  $RR = Pr\{R_{NT}|Reward\}$  (Panels A1, A2) and *Luminance Ratio*,  $LR$  (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). Filled points and heavy solid lines in Panels A2 and B1 represent fitted models in which  $c$  and  $d'$  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, which is not statistically significant. The  $\pm SE$  error bars reflect estimates of the variability of each plotted point after removing mean differences between 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 & Davison [14]; basic data kindly provided by B. Alsop.

The left side of the figure 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 invariance model fits well. Thus we 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  $d'$  of discriminability (and hence of  $\mathbf{S}$ ).

Effects of  $LR$  are shown on the right side of the figure. Panel B2 shows the orderly effect of  $LR$  on  $d'$ ; discriminability ranges widely, from  $d' \approx 0$  to  $d' \approx 3.6$ . Panel B1 shows that to a good approximation the criterion is uninfluenced by  $LR$ .

Overall, this study along with others (see [39]) 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  $\mathbf{S}$  and  $\mathbf{D}$  can be demonstrated. While these data are from pigeons,

19. If  $Pr\{R_{NT}|Reward\} = 0.2$ , 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  $R_T$ .

Johnstone and Alsop [40] have shown with humans that when  $RR$  is controlled, the criterion  $c$  is approximately invariant over changes in discriminability.

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  $LR$  (luminance ratio) and the d-factor  $PM$  (payoff matrix), the modularity of  $\mathbf{S}$  and  $\mathbf{D}$  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 approximation they influence  $M_S$  and  $M_D$  selectively, thus demonstrating the separate modifiability of  $\mathbf{S}$  and  $\mathbf{D}$ . One way to think of this is that  $M_D$  depends on  $LR$  as well as  $PM$  because they both affect  $RR$  if it is permitted to vary; 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.

## 1.6 Neural Processing Modules Inferred from Brain Activation Maps

Modular neural processes can be discovered by applying process decomposition to the kinds of activation measures provided by PET and fMRI. Suppose there is localization of function, such that two such processes  $\alpha$  and  $\beta$  are implemented by different processors  $P_\alpha$  and  $P_\beta$  in non-overlapping brain regions  $R_\alpha$  and  $R_\beta$ . Then activation levels in  $R_\alpha$  and  $R_\beta$  are pure measures of  $\alpha$  and  $\beta$ , and, with sufficiently precise data and factors that influence the processes selectively, separate modifiability is easy to demonstrate. However, if  $\alpha$  and  $\beta$  are implemented by different neural processors,  $P_\alpha$  and  $P_\beta$  (or by the *same* processor  $P_{\alpha\beta}$ ) in *one* region,  $R_{\alpha\beta}$ , then the activation level in  $R_{\alpha\beta}$  is a composite measure that depends on both  $\alpha$  and  $\beta$ , and to test separate modifiability we must know or show how their contributions to the activation measure are combined.<sup>20</sup>

Both of the examples below involve tasks in which modular mental processes were expected, based on earlier evidence from RT measurements. Suppose two such processes,  $\mathbf{A}$  and  $\mathbf{B}$ , influenced selectively by  $F$  and  $G$ . In such cases it is tempting to ask whether the operations carried out by these inferred mental processes are implemented by modular neural processes  $\alpha$  and  $\beta$  in anatomically separate brain regions. Because of their effectiveness in assessing the level of activity in localized brain regions, PET and fMRI are good techniques for this purpose. Then, because process  $\alpha$  should be influenced selectively by  $F$ , the activation level of region  $R_\alpha$  should vary with  $F$ , but not with  $G$ , and conversely for region  $R_\beta$ . Thus, the existence of regions whose fMRI signals are influenced selectively by  $F$  and  $G$  provides evidence for modular *neural* processes that correspond to the modular *mental* processes inferred from the behavioral data. If this were found, it would support the modular decomposition inferred from the behavioral data. Such evidence would also support the conclusion that the processors that implement  $\mathbf{A}$  and  $\mathbf{B}$  are anatomically localized.

For both of the examples below, it was a modular decomposition into two processing stages inferred from RT data that suggested a new experiment that would incorporate fMRI measurements to search for corresponding neural processes. And in both cases, concurrent RT data were taken, along with the fMRI data. While the RT data in the first example confirmed earlier findings, results in the second example did not, probably because of paradigm differences, which raises interesting questions about the fMRI findings.

---

20. For example, if the combination rule is *summation* (sometimes assumed without justification) and if factors  $F$  and  $G$  influence  $\alpha$  and  $\beta$  selectively, then the effects of  $F$  and  $G$  will be additive. Finding such additivity in a factorial experiment (as in Ex. 1.6.2) would support the combination rule in that brain region, as well as selective influence. If summation is assumed erroneously, selective influence would be obscured: the effect of each factor would appear to be modulated by the level of the other.

### 1.6.1 Modular processes in number comparison

In an experiment by Pinel *et al.* [15], subjects had to classify a sequence of visually displayed numbers,  $k$ , as being greater or less than 65. One factor was *notation* ( $N$ ): the numbers  $k$  could be presented as Arabic numerals (e.g. '68') or number names (e.g. 'SOIXANTE-HUIT'). The other was *numerical proximity* ( $P$ ), defined as the absolute difference  $|k - 65|$ , and grouped into three levels. The interesting phenomenon here is the "symbolic distance effect": the smaller the value of  $P$  (the closer the proximity), the slower the response. A similar experiment [41] had shown additive effects of  $N$  and  $P$  on  $\overline{RT}$ ; this was interpreted to indicate two modular subprocesses arranged as stages: encoding (**E**), influenced by  $N$ , which determines the identity of the stimulus and is slower for number names than numbers, and comparison (**C**), influenced by  $P$ , which uses the stimulus identity in performing the comparison and is slower for closer proximities. In the new experiment, in which fMRI as well as RT measurements were taken, most of the sixteen brain regions examined whose activation was influenced by  $N$  or by  $P$  were influenced significantly by only one of them, consistent with two separately modifiable neural processes  $\epsilon$  and  $\gamma$  that are implemented by separately localized processors. Averaging absolute effect sizes and SEs over the regions of each type, for the nine  $N$ -sensitive regions the  $N$  effect was  $0.17 \pm 0.05\%$  (median  $p$ -value = 0.01), while the  $P$  effect was  $0.06 \pm 0.08\%$ ; for the seven  $P$ -sensitive regions the  $P$  effect was  $0.32 \pm 0.10\%$  (median  $p$ -value = 0.01), while the  $N$  effect was  $0.04 \pm 0.04\%$ .

-----  
 Insert Fig. 6 about here  
 -----

Fig. 6. Reaction-time and selected brain-activation data from Pinel *et al.* [15]. The same data are plotted on the left as functions of  $P$  (proximity), with  $N$  (notation) the parameter, and on the right as functions of  $N$ , with  $P$  the parameter. Means over nine subjects of median RTs for correct responses are shown in Panels A, with a fitted additive model. The three levels of  $P$  have been scaled to linearize the main effect of  $P$  on  $\overline{RT}$ ; this effect, from low to high  $P$ , is  $159 \pm 24$  ms, while the main effect of  $N$  is  $204 \pm 34$  ms. SEs are based on variability over the nine subjects. The difference across levels of  $N$  between the simple effects of  $P$  from low to high (a measure of interaction) is a negligible  $4 \pm 20$  ms. (The SE may be inflated by unanalyzed condition-order effects.) Mean activation measures from three sample brain regions, relative to an intertrial baseline, are shown in Panels B, C, and D, accompanied by fitted null-effect models in Panels B2, C1, and D1. Shown in Panels B1, C1, and D1, the main effects of  $P$  (from low to high, using fitted linear functions) are  $0.29 \pm 0.09\%$  ( $p \approx 0.01$ ),  $-0.03 \pm 0.03\%$ , and  $0.00 \pm 0.04\%$ . Shown in Panels B2, C2, and D2, the main effects of  $N$  are  $-0.06 \pm 0.06\%$ ,  $0.16 \pm 0.05\%$  ( $p \approx 0.01$ ), and  $-0.15 \pm 0.05\%$  ( $p \approx 0.02$ ).

The fMRI data from three well-behaved regions are shown in Fig. 6, Panels B, C, and D. The concurrently collected RT data (Fig. 6, Panels A) replicated the earlier study, suggesting that we associate the neural modules  $\epsilon$  and  $\gamma$  with the mental modules **E** and **C**; it is important that the mental and neural modules were selectively influenced by the same factors. However, while the direction of the effect of  $P$  was the same in all the brain regions it influenced, the direction of the effect of  $N$  was not: the change from numeric to verbal notation (which increased  $\overline{RT}$ ) increased activation in some regions (e.g. Fig. 6, Panels C) and decreased it in others (e.g. Fig. 6, Panels D).<sup>21</sup>

### 1.6.2 Modular processes for stimulus encoding and response selection

A common finding has emerged from several studies of choice-reaction time (one of them discussed in Section 1.4.3), using various experimental arrangements and various realizations of the factors  $SQ$  (stimulus quality) and stimulus-response mapping difficulty (either  $MF$ , mapping familiarity, or  $MC$ , spatial compatibility): these studies have shown that stimulus quality and

mapping difficulty have additive effects on  $\overline{RT}$ , consistent with the idea that there are two processes, arranged in stages, that are selectively influenced by these factors. (These studies include [29], Exp. 2; [13], Exp I; [43], after a session of practice; [44]; [45], Exp. II; and [46], Exp. 3. The notion is that in initiating a response to a stimulus, the stimulus must first be identified (one stage, **S**) and then, given the identity, the response must be determined (a second stage, **R**).

Using a new choice-reaction task and their versions of *SQ* and *MC*, Schumacher and D'Esposito [16] measured fMRI in several brain regions concurrently with RT. In the task they used, the stimulus was a row of four circular patches, one patch brighter than the others. The response was to press one of four keys, depending on which of the patches was the brighter one. The two factors, each at two levels, were the discriminability of the brighter patch from the others (*SQ*), and the S-R compatibility of the patch-to-key mapping (*MC*). Each subject was tested under all four combinations of factor levels, with measurements of fMRI signals in six different brain regions, and of RT. Unlike Ex. 1.6.1, in every region where a factor had an effect on the fMRI signal, the "more difficult" level of that factor —the level that produced the longer  $\overline{RT}$ — also produced the larger fMRI signal. Figures 7.1 and 7.2 show that in one of the regions, only *SQ* had a reliable effect (Panels A), in two of the regions only *MC* had reliable effects (Panels B and C), in two regions both factors had reliable effects (Panels D and E), and in one region neither factor had a reliable effect (Panels F). The selective effects found in three of the regions (where the fMRI signal was influenced by one of the factors but not the other) are consistent with the implementation of **S** and **R** being accomplished, at least in part, by anatomically distinct populations of neurons.<sup>22</sup>

-----  
 Insert Figs. 7.1 and 7.2 about here  
 -----

Fig. 7. Brain activation data relative to a fixation baseline for the six regions measured by Schumacher & D'Esposito [16]. Mean main effects over nine subjects (eight for anterior cingulate) of Stimulus Discriminability and S-R Mapping in each region, with null effect models shown for each non-significant effect.

If the two processes **S** and **R** are the only processes involved in the task, as suggested by the earlier RT experiments, then there should be no process influenced by both factors. But in two regions (Panels D and E), both factors were found to have effects. However, it is possible that these regions each contain two specialized populations of neurons, each of which is influenced

- 
21. It is unexpected, hence noteworthy, for brain activation to decline when RT increases, if only because lengthening a process that produces even an unchanging level of activation in a region should increase the integrated fMRI signal in that region. Without requiring it, this finding invites us to consider that there are two qualitatively different encoding processes  $\epsilon$ , one for each notation, rather than 'one' process whose settings depend on  $N$ . (This possibility is now supported by the observation that "the notation factor affects the circuit where information is processed, not just the intensity of the activity within a fixed circuit" [42].) If so, we have a case where a change in the level of a factor (here,  $N$ ) induces a task change (one operation replaced by another; see Section 1.7), but evidence for modularity emerges nonetheless. Whereas the (multidimensional) activation data from such a simple (two-factor) experiment can support a claim of operations replacement, based on the idea that the processes implemented by different processors are probably different, if any (unidimensional) RT data could support such a claim it would require a more complicated experiment, such as one that showed a suitable modulation, by  $N$ , of the effect of a third factor.
22. It is noteworthy and requires explanation that in each of the five cases where an effect is not statistically significant, it is nonetheless in the same direction as in those cases where the effect is significant. Is this because the neural populations that implement the **S** and **R** processes are incompletely localized, or because the measured regions don't correspond to the populations, or for some other reason?

selectively by a different one of the two factors. If so, the total amount of neural activity in each of these regions would be influenced additively by the two factors. Alternatively, if processes **S** and **R** are sequential (arranged in stages), as suggested by the RT data from the earlier experiments, the same neural processor in the same region could contribute to the implementation of both processes, but, again, the summed neural activity would be influenced additively by the two factors. As shown in Fig. 8, Panels A and B, the effects of *SQ* and *MC* on the fMRI signal were remarkably close to being perfectly additive. Additive effects on the amount of neural activity by itself does not imply additive effects on the fMRI signal, but the supplementary assumptions required for this implication are relatively weak.<sup>23</sup> Thus the additivity of the effects of *SQ* and *MC* on the fMRI signal supports additivity of their effects on the amount of neural activity, which in turn supports the idea that separate processes within the regions shown in Panels D and E contribute to the implementation of both **S** and **R**.

-----  
 Insert Fig. 8 about here  
 -----

Fig. 8. Panels A, B show mean simple effects of the two factors on the brain activation measure, in the two regions where both main effects were significant in [16]. Also shown are fitted additive models, and the interaction contrasts that measure the badness of fit of these models. Panel C shows simple effects of the two factors on  $\overline{RT}$ , and the corresponding interaction contrast.

In contrast to the earlier findings with various experimental arrangements, the effects of the two factors on  $\overline{RT}$  in the Schumacher-D'Esposito task were unfortunately not additive; as shown in Figure 8C, there was a reliable interaction: the effect of raising the level of each factor was greater when the level of the other factor was higher (an "overadditive" interaction). This finding seems inconsistent with the fMRI data, all of which support the idea that no neural process is influenced by both factors. One possibility is that there is such a process, but it happens not to be localized in any of the six regions that were examined.<sup>24</sup> This possibility suggests that stronger inferences require sampling of more brain regions.

## 1.7 Process Decomposition versus Task Comparison

The examples above exemplify a process-decomposition method whose goal is to divide the complex process by which a particular task is accomplished into modular subprocesses. The

- 
23. One condition under which additivity of amount of neural activity would be reflected by additivity of the fMRI (blood-oxygenation level dependent, BOLD) signal, would be if the fMRI signal were linear in amount of neural activity. There is debate about whether such linearity obtains, and about whether the fMRI signal displays linear temporal summation ([47, 48]). With respect to linearity of the relationship, however, it is important that differences in the fMRI signal strength between experimental and control conditions are very small fractions of that signal strength. For example, the mean value of the largest difference observed in the Schumacher-D'Esposito experiment is 0.43 percent. Even if the BOLD signal ( $B$ ) grew nonlinearly with amount of neural activity,  $N$  (e.g., as a function of the form  $B = 1 - e^{-N}$ ), the approximation to linearity for such small fractions would be good. Because of the possibility that **S** and **R** occur sequentially, the goodness of temporal summation may seem important. However, to a first approximation this does not matter: For example, suppose a failure of temporal summation in which the  $B/N$  ratio is smaller for the later process **R** than for the earlier process **S**: Additivity of the effects of *SQ* and *MC* on  $N$  would still be reflected by the additivity of their effects on  $B$ .
24. There is an unresolved puzzle about these data that suggests that it would be valuable to replicate this experiment, using a procedure that is known to produce additive effects on  $\overline{RT}$ . The large SEs associated with the very small mean interaction contrasts for the data shown in panels D and E reflect the fact that the variability of the interaction contrast over subjects is quite large —so large relative to the mean that the reported F-statistics in both cases were 0.00. In other words, relative to the variability, the reported interaction contrasts were significantly ( $p < 0.05$ ) too small.

factor manipulations are not intended to produce ‘qualitative’ changes in the complex process (such as adding new operations, or replacing one operation by another), which may be associated with a change in the task, just ‘quantitative’ ones that leave it invariant.<sup>25</sup> The task-comparison method, mentioned in Section 1.4.4, is a more popular approach to understanding the structure of complex processes. Here one determines the influence of factors on performance in different tasks, rather than on different parts of the complex process used to carry out one task. The data pattern of interest is the selective influence of factors on tasks, i.e. the single and double dissociation of tasks. (A classical factor used in brain studies is the amount, usually presence versus absence, of damage in a particular brain region, which may affect performance in some tasks and not others.) Although it may achieve other goals, task comparison is inferior to process decomposition for discovering the modular subprocesses of a complex process or investigating their properties: The interpretation of task comparison often requires assuming a theory of the complex process in each task; the method includes no test of such assumptions. In contrast, process decomposition requires a theory of only one task, and, as illustrated by the examples above, incorporates a test of that theory.

### 1.7.1 Comparison of two tactile perception tasks

An elegant example of task comparison is provided by Merabet et al. [17] in their experiment on the effects of repetitive transcranial magnetic stimulation (rTMS) of different brain regions on subjective numerical scaling of two tactile perceptual dimensions, based on palpation of a set of tactile dot arrays with varying dot spacings by the fingers of one hand. The two judged dimensions were roughness ( $r$ ) and distance between dots ( $d$ ). Where rTMS had an effect, it reduced the sensitivity of the obtained scale values to the differences among dot arrays. One measure of *relative sensitivity* is the slope,  $b$ , of the linear regression of post-rTMS scale values on non-rTMS scale values. If there were no effect we would have  $b = 1.0$ ; the magnitude of the effect of rTMS is measured by  $1 - b$ . The data (Fig. 9) indicate that performance in the roughness-judgement task is influenced by rTMS of the contralateral somatosensory cortex (rTMS<sub>s</sub>, Panel A1), but negligibly by rTMS of the contralateral occipital cortex (rTMS<sub>o</sub>, Panel A2), while performance in the distance-judgement task is influenced by rTMS<sub>o</sub> (Panel B2), but negligibly by rTMS<sub>s</sub> (Panel B1), a double dissociation of the two tasks.<sup>26</sup>

-----  
 Insert Fig. 9 about here  
 -----

Fig. 9. Selective effects on two subjective scaling tasks of repetitive transcranial magnetic stimulation of two brain regions [17]. Mean sensitivity of scale values from 11 subjects relative to their non-rTMS scales are shown for the scaling of roughness (Panels A1, A2) and distance (Panels B1, B2), and for rTMS of somatosensory (rTMS<sub>s</sub>, Panels A1, B1) and occipital (rTMS<sub>o</sub>, Panels A2, B2) cortex. Also shown are null-effect models in Panels A2 and B1. Effects on roughness scaling, measured by  $1 - b$ , are  $1 - \overline{b_{rs}} = 0.21 \pm 0.07$  (Panel A1,  $p = 0.02$ ) and  $1 - \overline{b_{ro}} = 0.02 \pm 0.03$  (Panel A2). Effects on distance scaling are  $1 - \overline{b_{do}} = 0.16 \pm 0.07$  (Panel B2,  $p = 0.04$ ) and  $1 - \overline{b_{ds}} = 0.05 \pm 0.04$  (Panel B1).

Plausible theories might include, for each task, processes for control of stimulus palpation

25. See Section 1.10.2, and [6], Sections A.2.1 and A.9.2.

26. Subscripts  $d$  and  $r$  refer to the two tasks; subscripts  $s$  and  $o$  refer to the two stimulated brain regions. SEs are based on between-subject variability. Also supporting the claim of double dissociation, the differences,  $\overline{b_{ro}} - \overline{b_{rs}}$  and  $\overline{b_{ds}} - \overline{b_{do}}$  are significant, with  $p = 0.01$  and  $p = 0.04$ , respectively. However, because non-rTMS measurements were made only before rTMS, rather than being balanced over practice (a fault in the experimental design), straightforward interpretation of the slope values requires us to assume that effects of practice on those values were negligible.

( $\alpha_d, \alpha_r$ ), for generation of a complex percept ( $\beta_d, \beta_r$ ), for extraction of the desired dimension ( $\gamma_d, \gamma_r$ ), and for conversion of its value into a numerical response ( $\delta_d, \delta_r$ ). Any or all of these processes might differ between tasks. The striking findings indicate that the members of one or more of these pairs of processes depend on different regions of the cortex. A weak pair of task theories might also assert that  $\gamma_d$  and  $\gamma_r$  depend on occipital and somatosensory cortex, respectively, but if nothing is said about the other processes, this would be insufficient to predict the results. A stronger pair of task theories might add the assumptions that  $\alpha_d$  and  $\alpha_r$  are identical, that  $\beta_d$  and  $\beta_r$  are identical, and that  $\delta_d$  and  $\delta_r$  are identical. This pair of theories would predict the results, which would then also suggest that none of processes  $\alpha$ ,  $\beta$ , or  $\delta$  is sensitive to either rTMS<sub>s</sub> or rTMS<sub>o</sub>, indicating that they are implemented by processors in neither of the stimulated regions. But unfortunately the findings do not bear on the validity of such hypothesized task theories, weak or strong, or even on the question whether the operations in either task can be decomposed into modular subprocesses such as  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$ .

### 1.7.2 Donders' subtraction method: Task comparison with a composite measure

Perhaps the most venerable version of the task-comparison method is Donders' *subtraction method* [49], for two tasks with measures  $M_1 = \overline{RT}_1$  and  $M_2 = \overline{RT}_2$ . The joint hypothesis consists, first, of a pair of task theories that specify the constituent processes of each task, and second, a combination rule:

*H1* (Task Theory 1): Task 1 is accomplished by process **A**.

*H2* (Task Theory 2): Task 2 is accomplished by **A** and **B**, where **A** is identical, at least in duration, to the corresponding process in Task 1. (That is, addition of **B** satisfies a "pure insertion" assumption.)

*H3* (Combination Rule): Contributions  $u = T_A$  of **A** and  $v = T_B$  of **B** to  $M_2 = \overline{RT}_2$  combine by *summation*, as in Table 3, and as implied by Donders' assumption that **A** and **B** occur sequentially, in stages.

Given these hypotheses, it follows that  $M_1 = \overline{RT}_1$  is an estimate  $\hat{T}_A$  of the mean duration of **A**,  $\overline{RT}_2$  is an estimate  $\hat{T}_A + \hat{T}_B$  of the sum of the mean durations of **A** and **B**, and therefore, by subtraction,  $\hat{T}_B = \overline{RT}_2 - \overline{RT}_1$  provides an estimate of the mean duration of **B**. It is a serious limitation of the subtraction method that it usually embodies a test of neither the combination rule nor pure insertion. That is, *H2*, and *H3* are assumed but not tested.

Despite this problem, an analog of Donders' method has also often been used with brain activation measures (e.g., [50, 51]). Suppose we are interested in studying a subprocess  $\beta$  of a complex neural process. If  $\beta$  were implemented by a localized processor  $P_\beta$  in brain region  $R$ , then the level of activation of  $R$  might be a pure measure of the subprocess. However, suppose instead that  $P_\beta$  is not localized [52], and the activation measure  $M_{\alpha\beta}$  is a composite measure that reflects contributions from  $\alpha$  as well as  $\beta$ . (Evidence for contributions from two processes was found for two of the brain regions measured in Ex. 1.6.2.) In this case, possible justifications of summation as the combination rule include an assumption that  $\alpha$  and  $\beta$  are implemented by different populations of neurons in  $R$  that contribute independently to the brain activation measure, plus linearity of the function relating the fMRI signal to the total amount of neural activity. (Observations of fMRI additivity, as in Ex. 1.6.2, support summation as the combination rule, but unfortunately, brain-imaging experiments have seldom employed the factorial design required for making such observations.)

The hypotheses imply that  $M_1$  and  $M_2 - M_1$  are estimates of  $u$  and  $v$ , and can thus play the roles of pure measures of  $\alpha$  and  $\beta$ . But again, having these measures provides no test of the hypotheses. If summation proves to be incorrect as the combination rule, other analytic strategies

may be available. For example, suppose measured activation were shown to be a decelerating function of the amount of brain activity. If it was a logarithmic function we would have  $M_{\alpha\beta} = \log(u + v)$ , and the subtraction method could be applied to the mean of the transformed activation measure  $M'_{\alpha\beta} = \exp(M_{\alpha\beta}) = u + v$ . Alternatively, the function might be  $M_{\alpha\beta} = 1 - \exp[-(u + v)]$ ; here it can be shown that  $M'_{\alpha\beta} = \log(1 - M_{\alpha\beta}) = f(u) + f(v)$ , and again the subtraction method could be applied to the mean of the transformed activation measure.

## 1.8 The Use of TMS to Associate Mental Processes with Brain Regions

Why does sleep deprivation increase RT? An answer to this question was provided by Ex. 1.4.3, using a task in which the additivity of the effects of two factors ( $SQ$  and  $MF$ ) on  $\overline{RT}$  had furnished evidence for two modular processes, arranged in stages. A factorial experiment that incorporated two levels of sleep state as well as variation in  $SQ$  and  $MF$ , to determine which of these factor effects is modulated by sleep state, led to the conclusion that sleep deprivation slows RT by interfering with the stimulus encoding process but not with response selection.

In recent years it has been discovered that a single pulse of TMS at certain times and in some brain regions can prolong the RT in some tasks without reducing accuracy very much [53]. This opens the intriguing possibility of employing TMS within the method of additive factors, just as sleep deprivation was used by Sanders et al. [13].<sup>27</sup> By using the TMS manipulation of region  $R$  in a factorial experiment, while varying the levels of other factors that are believed to influence different processing stages selectively—and determining which effects of the other factors on  $\overline{RT}$  are modulated by TMS, it is possible to learn whether region  $R$  is involved in the implementation of any of those processing stages. For example, suppose there is a task in which we find that factors  $F$  and  $G$  have additive effects on  $\overline{RT}$ , from which we infer separate stages,  $\mathbf{A}$  (influenced by  $F$ ) and  $\mathbf{B}$  (influenced by  $G$ ). Now we add  $TMS_R$  (TMS of region  $R$ ) as a third factor, and ask whether it interacts with  $F$  or  $G$ . In the ideal result of such an experiment  $TMS_R$  (1) would have an effect on RT, (2) would interact with (modulate the effect of) one of the other factors, say,  $G$ , and (3) would not interact with (would have an effect that was additive with) the effect of the other factor, say,  $F$ . We would then have evidence that region  $R$  is involved in the implementation of  $\mathbf{B}$ , but not  $\mathbf{A}$ . Such findings would strengthen the inferences made without TMS about the organization of the underlying processes, and would identify brain region  $R$  as being associated with one process (such as  $\mathbf{B}$ ) and not the others.<sup>28</sup> Unfortunately, while there are several interesting studies employing TMS in this new approach, there is as yet no fully satisfactory one; both of the applications described below have missed the opportunity to make such observations, because of inadequate experimental design in Ex. 1.8.1, and incomplete analyses of data in Ex. 1.8.2.

27. An important advantage of TMS over measures of brain activation (Section 1.6) in determining which brain regions are involved in implementing a process is that whereas activation of a region in conjunction with process occurrence does not mean that such activation is *necessary* for that process, interference with a process by stimulation of a region does indicate the necessity of that region for the process to occur normally, just as does interference by a lesion in that region [54].

28. Note, however, that if the effect of  $TMS_R$  is time specific, as is likely with single-pulse TMS, it may not be straightforward to determine which other factors it interacts with. Thus, in the present example, suppose that it is region  $R$  that implements process  $\mathbf{B}$ , and that  $\mathbf{B}$  follows  $\mathbf{A}$ . Because effects on the duration of  $\mathbf{A}$  influence the starting time of  $\mathbf{B}$  (and hence the time of TMS relative to  $\mathbf{B}$ ), a change in the level of  $F$  might modulate the effect of  $TMS_R$  on  $\mathbf{B}$ , and hence on  $\overline{RT}$ . This argues for the use of repetitive TMS (rTMS) before the task is performed in such studies, rather than a single TMS pulse during the task.

### 1.8.1 Visual search and TMS

In visual search for a conjunction of features, as in many search tasks, it is usually observed that the  $\overline{RT}$  for both present and absent responses increases approximately linearly with the number of elements to be searched (the "display size"), suggesting a process of serial comparison of the search target with the displayed elements. If a linear function is fitted to such data, the slope of the function is often interpreted to reflect the time per comparison, while the intercept reflects the summed durations of residual processes whose durations are not influenced by display size.<sup>29</sup> In search for a single feature, unlike search for a feature conjunction,  $\overline{RT}$  may increase very little or not at all with display size, and the intercepts may also differ from those for conjunction search. Thus, feature and conjunction search appear to differ in residual processes as well as in the search process itself.

Ashbridge, Walsh, and Cowey [18] examined the effect of TMS of the parietal cortex during feature and conjunction search. In a preliminary experiment in which TMS was not used but display size was varied, along with search type (feature versus conjunction), their observations conformed to the description above. However, in the experiment in which they applied TMS, they studied only one level of the display-size factor. They found TMS to interact with search type, having a substantially greater effect on conjunction than feature search. However, they missed an opportunity: If they had varied display size, thus obtaining a measure of its effect (the slope) with and without TMS, their findings could tell us whether TMS produced its effect by influencing the serial comparison process and/or residual processes. Without the display-size factor, we do not know which subprocess is responsible for the effect of parietal TMS on conjunction search.

There may be several differences between the complex processes that underlie conjunction search and feature search. So, while finding that TMS influences one type of search but not the other is probably telling us something important, it is not clear which difference between processes is responsible. In the language of this chapter, the search-type factor may determine which of two different tasks is being carried out, rather than varying something about the same task. And because theories for the two tasks are not detailed enough to specify exactly how the associated complex processes differ, interpretation of the differential effects of parietal TMS on performance of the two tasks (and the inference, from this, of the role of parietal cortex in the serial-comparison process in conjunction search) requires speculation.

### 1.8.2 Number comparison and rTMS

In Section 1.6.1 I discussed an experiment in which subjects had to classify visually displayed numbers as being greater or less than 65. RT measurements, supported by fMRI data, indicated that performance in this task depends on (at least) two processing stages, one to determine the identity of the comparison stimulus, and the other to perform the comparison. The processing stage of particular interest—number comparison—is that which produces the effect of proximity—slower responding for stimulus numbers that are closer to the criterion. Goebel, Walsh, and Rushworth [19] found that repetitive TMS (rTMS) applied to the left or right angular gyrus influenced the RTs in such an experiment. Do these effects mean that the angular gyrus is involved in implementing the number comparison process? Alternatively, it might be involved in another process that contributes to the RT. They varied two other factors in addition to the presence of rTMS (and the brain region to which it was applied): (a) the magnitude and (b) the sign of the difference between the stimulus number and the criterion. (In the experiment considered in Section 1.6.1 the data were collapsed over the sign of the difference, and the

---

29. Which intercept is appropriate depends on details of the search process, and may differ for target-absent and target-present trials, where, e.g., it might be the zero- and one-intercept, respectively.

magnitude—the absolute value—of the difference was termed "proximity".) The shape of the function on each side of the criterion is sufficiently close to linear so it would be reasonable to use the slope of a fitted linear function as the measure of the proximity effect. The measures needed to assess the proximity effect are the slopes of the functions for positive and negative differences ( $k > 65$  and  $k < 65$ ) that relate  $\overline{RT}$  to the difference,  $|k - 65|$ . To determine whether TMS influences the comparison process, we need to know whether it modulates the proximity effect—i.e., changes either or both of these slopes. Effects of rTMS on the *heights* of these functions might reflect influences of TMS on encoding or response processes, as well as on the comparison process.

In principle, the data to answer this question were collected in these experiments. (I say "in principle" because the data could be sufficiently imprecise to preclude drawing either of the two interesting conclusions—that the angular gyrus does or does not play a role in implementing the comparison process.) But, unfortunately, a focused test of the effect of TMS on the relevant slopes was not conducted. We thus have a case where the design of the experiment seems ideal for examining whether and how TMS modulates the proximity effect, which would indicate an effect on the number comparison process, but the appropriate analysis has not yet been reported.

## 1.9 Brief Summaries of Five Other Examples

In this section I provide brief summaries of five additional examples that show the diversity of applications of the process-decomposition approach. Two of them make use of composite measures with multiplicative combination rules; included is a discussion of the inferential logic for this case.

### 1.9.1 Isolation of a timing module in the rat

On some trials ("food trials") a rat is rewarded for pressing a bar after the passage of a fixed time interval,  $TF$ , from the onset of a signal. On nonfood trials after performance stabilizes, the response rate starts low with signal onset, grows to a peak at about time  $TF$ , and declines thereafter. Two measures of performance are the time at which the peak is reached (*peak time*) and the response rate at the peak (*peak rate*). Two factors studied by Roberts [20] are  $TF$  (the time at which food becomes available) and  $PF$  (the proportion of food trials). Roberts discovered that peak time and peak rate are selectively influenced by  $TF$  and  $PF$ , respectively. His findings support the idea that the process that controls response rate contains two modules, a clock process and a response process, of which peak time and peak rate are pure measures.

### 1.9.2 Evidence for modular spatial-frequency analyzers from selective adaptation

The thresholds for sinusoidal gratings of sufficiently separated high and low spatial frequencies respond selectively to adaptation by high- and low-frequency gratings. This supports the hypothesis that there exist modular analyzers that detect the high and low frequencies, and of which the thresholds for high and low frequencies are pure measures ([21]; [4], Section 1.9).

### 1.9.3 Evidence for modular spatial-frequency analyzers from the detectability of compound gratings

Suppose a task in which a subject says "yes" when either or both of two detectors respond, and says "no" if neither detector responds. For the present discussion we ignore the complication introduced by "guessed" yes responses that may occur when neither detector responds. The subject says "no" if and only if both detectors fail to respond. If the behavior of the two detectors is uncorrelated, then the probability of this joint event is the product of the individual nonresponse probabilities for the two detectors.  $Pr\{\text{"no"}\}$  is thus a composite measure of the two detectors, with a multiplicative combination rule.

To describe the consequences of a multiplicative combination rule for a composite measure, I need to introduce the idea of a proportional effect, or *p.effect*. We saw in Section 1.2.2 that the *effect* of a factor on a measure is defined as a *difference* (for a factor with two levels), as in Eq. 1, or can be defined as a vector of differences (for a factor with multiple levels). Similarly, the *p.effect* of a factor on a measure is defined as a *ratio* (for a factor with two levels):

$$p.\text{effect}(F) = \frac{\bar{M}_A(F_2)}{\bar{M}_A(F_1)} , \quad (8)$$

or a vector of ratios (for a factor with multiple levels). Suppose we have a composite measure with a multiplicative combination rule:

$$M_{AB}(F_j, G_k) = u(F_j) \times v(G_k) . \quad (9)$$

To derive the equivalent of Eq. 3 from Eq. 9, however, requires us to assume that the contributions  $u$  and  $v$  from processes **A** and **B** to  $M_{AB}$  are uncorrelated. In that case, it follows from Eq. 9 that

$$\bar{M}_{AB}(F_j, G_k) = \bar{u}(F_j) \times \bar{v}(G_k) . \quad (10)$$

By analogy to Eq. 4, it can then be shown that

$$p.\text{effect}(F, G) = p.\text{effect}(F) \times p.\text{effect}(G) . \quad (11)$$

If the *p.effects* of the factors are multiplicative, as in Eq. 11, this supports the joint hypothesis that processes **A** and **B** are separately modifiable, that their contributions to  $M_{AB}$  combine by multiplication, and that their contributions are uncorrelated. The inferential logic in this case is outlined in Table 4.

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> to $M_{AB}(\mathbf{A}, \mathbf{B})$ 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 $F$ 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>proportional effects</i> on $M_{AB}$ will be <i>multiplicative</i> .	
<b>Alternative Results</b>	
We find factors $F$ and $G$ 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 we didn't look enough for F, G.</i>

This observation was exploited in a famous experiment in which the detectors were hypothesized spatial-frequency analyzers sensitive to different frequency bands. Sachs,

Nachmias, and Robson [22] independently varied the contrasts ( $F$ ,  $G$ ) of widely separated frequencies in a compound grating whose presence subjects had to detect. They found that reducing the contrast of one of the frequencies in the compound reduced  $Pr\{\text{non-detect}\}$  by a constant factor, consistent with Eq. 11, thus supporting the joint hypothesis. Their findings provided important early evidence for modular spatial-frequency analyzers and for the multiplicative combination rule.

#### 1.9.4 Evidence from the amplitude of the ERP for modular processes in semantic classification

At any particular time, the ERP at any point on the scalp is a composite measure of all the neural processors ("sources") in the brain that are active at that time. Furthermore, the physics of volume conduction tells us that the combination rule is summation. Hence, unlike most other cases, the combination rule is not a part of the hypothesis that must be tested. Suppose two modular neural processes  $\alpha$  and  $\beta$ , implemented by processors  $P_\alpha$  and  $P_\beta$ , and factors  $F$  and  $G$  that influence them selectively. It follows that the effects of  $F$  and  $G$  on the ERP will be additive at all scalp locations. Furthermore, if  $P_\alpha$  and  $P_\beta$  are at different locations in the brain, the topographies of the effects of  $F$  and  $G$  (the way their sizes vary with location on the scalp) will differ.

Kounios [23] exploited this in a study of the effects of semantic priming on the classification of spoken nouns. The words consisted of *primes* and *probes*. The two factors were the *semantic relatedness* of the probe to the preceding prime, and the number of immediate repetitions of the prime before the probe, which determined the *semantic satiation* of the prime. The measures of interest were the ERPs elicited by the nontarget (non-response) probes at several locations on the scalp. The hypothesis was that the two factors influence two modular neural processes,  $\alpha$  and  $\beta$ , selectively. For the epoch from 600 to 800 ms after the probe the predictions were confirmed: at each location the factor effects were additive, supporting the modularity of processes  $\alpha$  and  $\beta$ . Furthermore, the topographies of the two effects differed, consistent with  $P_\alpha$  and  $P_\beta$  being localized differently within the brain.

#### 1.9.5 Modular processes for learning and motivation

Roberts [24] has drawn attention to the orderliness of response rate in factorial experiments across a remarkable range of factors and animal species. Several of the data sets he examined showed multiplicative effects of two factors on response rate. One of these cases is provided by Clark's [25] experiment on the effects of *hours of food deprivation* (a motivational factor with seven levels) and *frequency of feeding* (a learning factor with three levels) on the rate of bar pressing by rats under variable-interval reinforcement schedules. The effect of each factor was found to be invariant over levels of the other, supporting the existence of uncorrelated modular processes influenced selectively by the two factors, and a multiplicative combination rule.

### 1.10 Comments and Questions

#### 1.10.1 Task-general processing modules

One plausible expectation is that different tasks are accomplished by different subsets of a small set of 'basic' modular processes. To test this expectation we need a reasonable number of tasks for which persuasively successful decompositions have been achieved.<sup>30</sup> On the other hand, to get adequate data we require subjects to learn a task to a point of stable performance. With such intensive practice, it seems possible that the brain is sufficiently flexible that special-purpose routines would be developed that are specific to that task. Thus, an alternative plausible

---

30. For speeded tasks in which processes are arranged as stages, Sanders ([55], Chapter 3) has amassed some suggestive evidence for a small set of mental modules.

expectation is that at least some modular subprocesses are task-specific rather than task-general. In that sense, perhaps there is no 'fundamental architecture of the mind', but rather a flexible architect, who has some stylistic tendencies worth studying.

### 1.10.2 Quantitative versus qualitative task changes

As shown by Ex. 1.6.1, the distinction between process decomposition (with its 'quantitative' task changes) and task comparison (in which the task changes are 'qualitative') can be subtle. In that example, the fMRI data suggest that the effect of the notation factor is probably better thought of as qualitative rather than quantitative —as replacing one encoding process by another, rather than influencing the settings or parameters of the "same" encoding process. Nonetheless, that example provides evidence for modular processes. In general, qualitative task changes should be avoided because they reduce the likelihood of discovering modules. However, evidence is required to assert qualitative task invariance. One kind of evidence is the pattern of factor effects: for each factor, each change in level should influence the same operations and leave the same other operations invariant. The usefulness of such evidence is one of several reasons for using factors with more than two levels. Unfortunately, few studies have done so.

### 1.10.3 Specialized processors and modular processes

Does the existence of a localized neural processor that implements a particular process imply the modularity of that process? To address this question, consider one kind of evidence used to establish the existence of the processor:  $T_a$  and  $T_b$  are two classes of tasks, such that brain region  $R_\alpha$  is activated during  $T_a$ , but not during  $T_b$  (or is activated *more* during one than the other), and such that we are willing to assume that all tasks  $T_a$  require a particular process  $\alpha$  to be carried out, whereas none of tasks  $T_b$  do. While it may seem plausible, such task-specificity of  $R_\alpha$  does not imply that the process  $\alpha$  that it implements in a given task is a modular subprocess in the sense of being modifiable separately from other subprocesses in that task. Suppose, for example, that  $\alpha$  provides a motivational or attentional resource that is required by one or more other processes  $\gamma$  that differ across tasks  $T_a$ . A change in  $\alpha$  would then induce a change in  $\gamma$ , so they would not be separately modifiable.

### 1.10.4 Relation between mental and neural modules

Consider modular mental processes in a task, supported by behavioral evidence, and modular neural processes in that task, supported by brain measurements. Does either of these imply the other? On which psychophysical-physiological 'linking propositions' [56] does the answer to this question depend? It would be helpful to have more studies (such as Exs. 1.3.1, 1.6.1, and 1.6.2), in which both brain and behavioral measures are taken, both directed at process decomposition. One starting point would be to take cases where behavioral data already exist that persuasively favor a modular decomposition, as was done in Exs. 1.6.1 and 1.6.2, and ask whether there is a corresponding decomposition based on brain data into modular neural processes that are influenced by the same factors, and invariant with respect to the same other factors.

### 1.10.5 Separate modifiability as a criterion for modularity

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? Is the weaker differential modifiability<sup>31</sup> more useful than separate modifiability?

---

31. If differential modifiability obtains, one can find factors  $F$  and  $G$  such that both factors influence both processes  $A$  and  $B$ , but for  $A$  ( $B$ ) the effect of  $F$  ( $G$ ) is the larger.

### **1.10.6 Implications of brain metabolism constraints**

The metabolic requirements of brain activity are large relative to the available energy supply, with the implication that, given the spike rates of active neurons, no more than about 1% of the neurons in the brain can be concurrently active [57]. This seems to be consistent with "sparse coding" and to argue against the idea of "massive parallelism". Other implications of these severe metabolic limitations for the plausibility and possibility of alternative processing architectures, and for the modularity of processors, have still to be worked out.

### **1.11 Further Reading**

For more extensive discussion of the inferential logic associated with the process decomposition method, and detailed discussion of the examples in Sections 1.3, 1.4, 1.5, and 1.9, see the text and appendices of [4], and references therein. For further discussion of the material in Sections 1.6.1 and 1.7.1, see [26]. For the additive factor method and numerous examples of its application, see [4 (Sections 16, A.16.2-3), 29, 31, 55] and references therein. For Hadley's defense of the existence and plausibility of mental modules against attacks by Fodor [58], Kosslyn [59], and Uttal [60], see [61]. For other discussion of the properties that Fodor [5] ascribed to modular processes, see [62] and [63].

### **Acknowledgements**

For providing unpublished details of their data I thank Brent Alsop, Stanislaus Dehaene, Norma Graham, John Kounios, Lotfi Merabet, Jacob Nachmias, Alvaro Pascual-Leone, Philippe Pinel, Seth Roberts, Eric Schumacher, and Fren Smulders.

## REFERENCES

- [1] Simon, H.A., The architecture of complexity, *Proc. Am. Phil. Soc.*, 106, 467, 1962.
- [2] Simon, H.A., The structure of complexity in an evolving world: The role of near decomposability, in *Modularity: Understanding the Development and Evolution of Natural Complex Systems*. Callebaut, W. and Rasskin-Gutman, D., Eds., MIT Press, Cambridge, MA, 2005, ix.
- [3] Marr, D., Early processing of visual information, *Phil. Trans. Roy. Soc. Lond. B*, 275, 483, 1976.
- [4] Rumelhart, D.E., McClelland, J.L., and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations*, MIT Press, Cambridge, MA, 1986.
- [5] Jacobs, R. A., Computational studies of the development of functionally specialized neural modules, *Trends Cognit. Sci.*, 3, 31, 1999.
- [6] Sternberg, S., Separate modifi ability, mental modules, and the use of pure and composite measures to reveal them, *Acta Psychol.*, 106, 147, 2001.
- [7] Fodor, J., *The Modularity of Mind: An Essay on Faculty Psychology*. MIT Press, Cambridge, MA, 1983.
- [8] Sternberg, S., Process decomposition from double dissociation of subprocesses, *Cortex*, 39, 180, 2003.
- [9] Berger, R.L. and Hsu, J.C., Bioequivalence trials, intersection-union tests and equivalence confidence sets, *Stat. Sci.*, 11, 283, 1996.
- [10] Rogers, J.L., Howard, K.I., and Vessey, J.T., Using significance tests to evaluate equivalence between two experimental groups, *Psychol. Bull.*, 113, 553, 1993.
- [11] Smulders, F.T.Y., et al., The temporal selectivity of additive factor effects on the reaction process revealed in ERP component latencies, *Acta Psychol.*, 90, 97, 1995.
- [12] Osman, A., et al., On the transmission of partial information: Inferences from movement-related brain potentials, *J. Exp. Psychol. Hum. Percept. Perform.*, 18, 217, 1992.
- [13] Sanders, A. F., Wijnen, J. L. C., and Van Arkel, A. E., An additive factor analysis of the effects of sleep loss on reaction processes, *Acta Psychol.*, 51, 41, 1982.
- [14] McCarthy, D., and Davison, M., Isobias and alloibias functions in animal psychophysics, *J. Exp. Psychol. Anim. Behav. Process.*, 10, 390, 1984.
- [15] Pinel, P., et al., Modulation of parietal activation by semantic distance in a number comparison task, *NeuroImage* 14, 1013, 2001.
- [16] Schumacher, E.H. and D'Esposito, M., Neural implementation of response selection in humans as revealed by localized effects of stimulus-response compatibility on brain activation, *Hum. Brain Mapp.*, 17, 103, 2002.
- [17] Merabet, L., et al., Feeling by sight or seeing by touch?, *Neuron*, 42, 173, 2004.
- [18] Ashbridge, E., Walsh, V., and Cowey, A., Temporal aspects of visual search studied by transcranial magnetic stimulation, *Neuropsychologia*, 35, 1121, 1997.

- [19] Goebel, S., Walsh, V., and Rushworth, M.F.S., The mental number line and the human angular gyrus, *NeuroImage*, 14, 1278, 2001.
- [20] Roberts, S., Isolation of an internal clock, *J. Exp. Psychol. Anim. Behav. Process.* 7, 242, 1981.
- [21] Graham, N., Spatial frequency channels in the human visual system: effects of luminance and pattern drift rate, *Vis. Res.*, 12, 53, 1970.
- [22] Sachs, M.B., Nachmias, J., and Robson, J.G., Spatial-frequency channels in human vision, *J. Opt. Soc. Am.*, 61, 1176, 1971.
- [23] Kounios, J., Functional modularity of semantic memory revealed by event-related brain potentials, in *Neural Basis of Semantic Memory*, Hart Jr., J. and Kraut, M.A., Eds., Cambridge University Press, Cambridge, In Press.
- [24] Roberts, S., Evidence for distinct serial processes in animals: The multiplicative-factors method, *Anim. Learn. Behav.* 15, 135, 1987.
- [25] Clark, F.C., The effect of deprivation and frequency of reinforcement on variable-interval responding, *J. Exp. Anal. Behav.*, 1, 221, 1958.
- [26] Sternberg, S., Separate modifi ability and the search for processing modules, in *Attention and Performance XX: Functional Neuroimaging of Visual Cognition*, Kanwisher, N. and Duncan, J., Eds., Oxford University Press, Oxford, 2004, 125.
- [27] Shallice, T., *From Neuropsychology to Mental Structure*, Cambridge University Press, Cambridge, 1988.
- [28] Howson, C. and Urbach, P., *Scientifi c Reasoning: The Bayesian Approach*. 3rd ed., Open Court, Chicago, 2006.
- [29] Roberts, S. and Sternberg, S., The meaning of additive reaction-time effects: Tests of three alternatives, in *Attention and Performance XIV: Synergies in Experimental Psychology, Arti ficial Intelligence, and Cognitive Neuroscience*. Meyer, D.E. and Kornblum, S., Eds., MIT Press, Cambridge, MA, 1993, 611.
- [30] Sternberg, S. The discovery of processing stages: Extensions of Donders' method, in *Attention and Performance II*, Koster, W.G., Ed., *Acta Psychol.*, 30, 276, 1969.
- [31] Sternberg, S., Discovering mental processing stages: The method of additive factors, in *An Invitation to Cognitive Science, Volume 4: Methods, Models, and Conceptual Issues*, Scarborough, D. and Sternberg, S., Eds., MIT Press, Cambridge, MA, 1998, 703.
- [32] Dinges, D.F. and Kribbs, N.B., Performing while sleepy: Effects of experimentally-induced sleepiness, in *Sleep, Sleepiness and Performance*, Monk, T.H., Ed., Wiley, London, 1991, 97.
- [33] Humphrey, D.G., Kramer, A.F., and Stanny, R.R., Influence of extended wakefulness on automatic and nonautomatic processing, *Hum. Factors*, 36, 652, 1994.
- [34] Tiplady, B., et al., Selective effects of clonidine and temazepam on attention and memory, *J. Psychopharmacol.*, 19, 259, 2005.
- [35] Swets, J. A., Tanner, Jr., W. P., and Birdsall, T. G., Decision processes in perception, *Psychol. Rev.*, 68, 301, 1961.
- [36] Macmillan, N. A. and Creelman, C. D., *Detection Theory: A User's Guide*, 2nd ed., Erlbaum, Mahway, N.J., 2005.

- [37] Dusoior, A. E., Treatments of bias in detection and recognition models: A review. *Percept Psychophys*, 17, 167, 1975.
- [38] Dusoior, A. E., Isobias curves in some detection tasks. *Percept Psychophys*, 33, 403, 1983.
- [39] Alsop, B., Receiver operating characteristics from nonhuman animals: Some implications and directions for research with humans, *Psychonomic Bull. & Rev.* 5, 239, 1998.
- [40] Johnstone, V. and Alsop, B., Reinforcer control and human signal-detection performance, *J Exp Anal Behav*, 73, 275, 2000.
- [41] Dehaene, S., The organization of brain activations in number comparison: Event-related potentials and the additive-factors method. *Journal of Cognitive Neuroscience* 8, 47, 1996
- [42] Dehaene, S. Personal communication, 2006.
- [43] Biederman, I., and Kaplan, R., Stimulus discriminability and stimulus response compatibility: Evidence for independent effects on choice reaction time, *J. Exp. Psychol.*, 86, 434, 1970.
- [44] Frowein, H.W. and Sanders, A.F., Effects of visual stimulus degradation, S-R compatibility and foreperiod duration on choice reaction time and movement time, *Bull. Psychonomic Soc.*, 12, 106, 1978.
- [45] Sanders, A.F., Structural and functional aspects of the reaction process, in *Attention and Performance VI*. Dornic, S., Ed., Erlbaum, Mahway, N.J., 1977, 3.
- [46] Sanders, A.F., Some effects of instructed muscle tension on choice reaction time and movement time, in *Attention and Performance VIII*, Nickerson, R.S., Ed., Erlbaum, Mahway, N.J., 1980, 59.
- [47] Heeger, D.J., et al., Spikes versus BOLD: What does neuroimaging tell us about neuronal activity?, *Nat. Neurosci.*, 3, 631, 2000.
- [48] Heeger, D.J. and Ress, D., What does fMRI tell us about neuronal activity?, *Nat. Rev. Neurosci.*, 3, 142, 2002.
- [49] Donders, F.C., 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, 1868. Transl. by Koster, W.G., in *Attention and Performance II*, Koster, W.G., Ed., *Acta Psychol.*, 30, 412, 1969.
- [50] Petersen, S. E., et al., Positron emission tomographic studies of the cortical anatomy of single-word processing, *Nature* 331, 585, 1988.
- [51] Lie, C.-H., Using fMRI to decompose the neural processes underlying the Wisconsin Card Sorting Test, *NeuroImage* 30, 1038, 2006.
- [52] Haxby, J.V., Analysis of topographically organized patterns of response in fMRI data: Distributed representations of objects in ventral temporal cortex, in *Attention and Performance XX: Functional Neuroimaging of Visual Cognition*, Kanwisher, N. and Duncan, J., Eds., Oxford University Press, Oxford, 2004, 83.
- [53] Walsh, V. and Pascual-Leone, A., *Transcranial Magnetic Stimulation: A Neurochronometrics of Mind*. MIT Press, Cambridge, MA, 2003.
- [54] Chatterjee, A., A madness to the methods in cognitive neuroscience?, *J. Cognit. Neurosci.*, 27, 847, 2005.

- [55] Sanders, A.F., *Elements of Human Performance: Reaction Processes and Attention in Human Skill*, Erlbaum, Mahway, NJ, 1998.
- [56] Teller, D., Linking propositions, *Vis. Res.*, 24, 1233, 1984.
- [57] Lennie, P., The cost of cortical computation, *Curr. Biol.*, 13, 493, 2003.
- [58] Fodor, J., *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*, MIT Press, Cambridge, MA, 2000.
- [59] Kosslyn, S.M., The strategic eye: another look, *Minds and Machines*, 11, 287, 2001.
- [60] Uttal, W.R., *The New Phrenology: The Limits of Localizing Cognitive Processes in the Brain*, MIT Press, Cambridge, MA, 2001.
- [61] Hadley, R.F., A defense of functional modularity, *Connection Science*, 15, 95, 2003.
- [62] Coltheart, M., Modularity and cognition, *Trends Cognit. Sci.*, 3, 115, 1999.
- [63] Jacobs, R.A., Nature, nurture, and the development of functional specializations: A computational approach, *Psychonomic Bull. & Rev.* 4, 299, 1997.

## CONTENTS

1.1	Modules and Modularity .....	2
1.2	The Process-Decomposition Method .....	3
1.2.1	Separate modifi ability, process-specifi c factors, selective influence, and functional distinctness	3
1.2.2	Processes and their measures, pure and composite, and combination rules	3
1.2.3	Overview of examples and issues	6
1.3	Decomposing Neural Processes with the Lateralized Readiness Potential .....	8
1.3.1	Serial modules for preparing two response features	8
1.3.2	Parallel modules for discriminating two stimulus features	10
1.4	Mental Processing Stages Inferred From Reaction Times .....	12
1.4.1	Analysis of the reaction-time data in Ex. 1.3.1	12
1.4.2	The method of additive factors	13
1.4.3	Selectivity of the effect of sleep deprivation	13
1.4.4	Finding the ‘mechanism of action’ of a manipulation: Module identi fication	14
1.5	Signal-Detection Theory: Sensory and Decision Modules .....	15
1.5.1	The failure to fi nd full modularity in traditional sensory experiments	15
1.5.2	Modularity of sensation and decision when reinforcement ratio is controlled	16
1.6	Neural Processing Modules Inferred from Brain Activation Maps .....	18
1.6.1	Modular processes in number comparison	19
1.6.2	Modular processes for stimulus encoding and response selection	19
1.7	Process Decomposition versus Task Comparison .....	21
1.7.1	Comparison of two tactile perception tasks	22
1.7.2	Donders’ subtraction method: Task comparison with a composite measure	23
1.8	The Use of TMS to Associate Mental Processes with Brain Regions .....	24
1.8.1	Visual search and TMS	25
1.8.2	Number comparison and rTMS	25
1.9	Brief Summaries of Five Other Examples .....	26
1.9.1	Isolation of a timing module in the rat	26
1.9.2	Evidence for modular spatial-frequency analyzers from selective adaptation	26
1.9.3	Evidence for modular spatial-frequency analyzers from the detectability of compound gratings	26
1.9.4	Evidence from the amplitude of the ERP for modular processes in semantic classifi cation	28
1.9.5	Modular processes for learning and motivation	28
1.10	Comments and Questions .....	28
1.10.1	Task-general processing modules	28
1.10.2	Quantitative versus qualitative task changes	29
1.10.3	Specialized processors and modular processes	29
1.10.4	Relation between mental and neural modules	29
1.10.5	Separate modifi ability as a criterion for modularity	29
1.10.6	Implications of brain metabolism constraints	30
1.11	Further Reading .....	30
	Acknowledgements .....	30