Modular Processes in Mind and Brain
Saul Sternberg*

Abstract

One approach to understanding a complex process starts with an attempt to divide it into modules: subprocesses that are independent in some sense, and have different functions. In this paper I discuss an approach to the modular decomposition of neural and mental processes. A dozen examples of process decomposition are presented. Two are of well-established and purely behavioral realizations of the approach (signal detection theory applied to discrimination data; the method of additive factors applied to reaction-time data), and lead to the identification of mental modules. Other examples, leading to the identification of modular neural processes, use brain measures, including the latency and amplitude of electrophysiological events, and fMRI measures. Some measures are pure (reflecting just one process), while others are composite. Two of the examples reveal mental and neural modules that correspond. 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 approach discussed here, in which the criterion for modularity is separate modifiability, is superior for modular decomposition to the more frequently used task comparison procedure (often used in cognitive neuropsychology) and to its associated subtraction method. To demonstrate the limitations of task comparison, I discuss an example involving TMS.

* 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
Acknowledgements

For providing unpublished details of their data I thank Brent Alsop, Stanislas Dehaene, Russell Epstein, Silke Goebel, John Kounios, Lotfi Merabet, Allen Osman, Alvaro Pascual-Leone, Philippe Pinel, Eric Schumacher, and Fren Smulders. For helpful discussions, I thank Geoffrey Aguirre, David Brainard, Russell Epstein, Martha Farah, Joshua Gold, Nancy Kanwisher, John Kounios, David Meyer, Jacob Nachmias, Allen Osman, and Seth Roberts. For helpful comments on the manuscript I thank Stanislas Dehaene, Silke Goebel, Allen Osman, Eric Schumacher, Fren Smulders, Sharon Thompson-Schill, Vincent Walsh, Max Coltheart, and two anonymous reviewers. For computer support I thank Vincent Hurtubise, Christopher Leary, and Roderick Smith.
1. Modules and Modularity

The first step in one approach to understanding a complex process is to attempt to divide it into *modules*: parts that are independent in some sense, and have different functions.\(^1\) 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 perhaps 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 added legitimacy to 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") subprocesses responsible for separate operations such as input, transformation, storage, retrieval, and comparison of internal representations — modules whose arrangement was expressed in systematic flow charts.\(^2\) A useful distinction that is encouraged by comparing the brain to a digital computer is that between processors and the processes they may implement. The existence of functionally specialized processors (either localized or distributed) is a sufficient condition but not a necessary one for functionally distinct processes, the concern of the present paper.

The rise in the 1980s of parallel distributed processing might seem 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" (Rumelhart, McClelland, & the PDP Research Group, 1986, 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 (Jacobs & Jordan, 1992; Jacobs, 1999).

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.

---

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 (1962, 2005) and, in his ‘principle of modular design’, by Marr (1976). Marr (p. 485) 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."
In the present paper I describe an approach to 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. A mental-process module is a part of a 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. The important distinction between processes and processors is sometimes overlooked. Processes occur over time; their arrangement is described by a flow-chart. In contrast, processors (e.g., as described by Erickson, 2001; Goodale, 1996; Op de Beeck, Haushofer, & Kanwisher, 2008) are parts of a physical or biological device (such as the brain); their arrangement can often be described by a circuit diagram.

2. The Process-Decomposition Approach

2.1 Separate modifiability, selective influence, process-specific factors, 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. One purpose of the present paper 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 $F$. Often one starts with hypotheses about what the component processes are, and about corresponding process-specific factors that are likely to influence them selectively. Alternatively, the selectivity of effects may be discovered in experiments with other goals. 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.

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

---

3. In Sternberg, 2001, referred to as "SM" or "SM:[Section]", I discuss and defend the approach in more detail, describe its antecedents, illustrate it with several applications to mental and neural processes, and further discuss its inferential logic.

4. It is perhaps a confusion between time (process) and space (processor) that has led some commentators (Broadbent, 1984, and one of the reviewers of this paper) to believe that a process organized in stages cannot include feedback because it is implemented by a "pipeline": an ordered set of processors through which information passes in a fixed direction.

5. This criterion for modularity seems to be far weaker than the set of module properties suggested by Fodor (1983), according to whom modules are typically innate, informationally encapsulated, domain specific, "hard-wired", autonomous, and fast. However, domain specificity appears to imply separate modifiability.

6. Such double dissociation of subprocesses should be distinguished from the more familiar double dissociation of tasks (Schmidt & Vorberg, 2006), discussed in Section 9. Sternberg (2003) discusses the distinction; see also SM:1.1).
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 $B$.

### Table 1
Inferential logic for pure measures.

<table>
<thead>
<tr>
<th>Joint Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H1$: Processes $A$ and $B$ are modules (separately modifiable).</td>
</tr>
<tr>
<td>$H2$: $M_A$, $M_B$ are pure measures of $A$, $B$.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>We may be able to find factors $F$ and $G$ that influence $M_A(A)$ and $M_B(B)$ selectively:</td>
</tr>
<tr>
<td>$p_1$: $M_A \leftarrow F$, $p_2$: $M_B \leftarrow F$, $p_3$: $M_B \leftarrow G$, $p_4$: $M_A \leftarrow G$.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Alternative Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>We find factors $F$, $G$ that influence $M_A$ and $M_B$ selectively.</td>
</tr>
<tr>
<td>We fail to find such factors.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inferences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support for joint hypothesis $H1 + H2$.</td>
</tr>
<tr>
<td>Refutes one/both of $H1$, $H2$, or we didn’t look enough for $F$, $G$.</td>
</tr>
</tbody>
</table>

**Pure measures.** 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 3.1), and the discriminability and criterion parameters of signal-detection theory (which reflect sensory and decision processes; see Section 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:

$$\text{effect}(F) = M_A(F_2) - M_A(F_1).$$

(1)

For factors with multiple levels the effect can be regarded as a vector of differences associated with successive ordered levels. The logic for inferring separate modules when we hypothesize that we have pure measures is shown in Table 1.8 The $\{p_i\}$ are four properties of the data; $M_A \leftarrow F$ should be read as "$M_A$ is influenced by $F$"; $M_B \leftarrow F$ should be read as $M_B$ is not influenced by $F$. All four of the requirements are critical; the instances of invariance are meaningful only if we also know that both factors are potent and both measures are sensitive. 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.9

7. Table 2 of Sternberg, 2001; reprinted by permission.

8. 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 SM:A.2.3.
Composite measure. Suppose that instead of pure measures, 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 reaction time, $RT$ (which may depend on the durations of more than one process of interest). 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 efficient tests of generality and promote new discoveries at little cost. With a composite measure factorial experiments are essential, to assess how the effects of the factors combine; unfortunately they are rare, despite their efficiency (Section 12.5).

Table 2\textsuperscript{10}
Inferential logic for a composite measure with summation as the combination rule.

<table>
<thead>
<tr>
<th>Joint Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H1$: Processes A and B are modules (separately modifiable).</td>
</tr>
<tr>
<td>$H3$: Contributions $u_A$, $v_B$ of A, B to $M_{AB}(A, B)$ combine by summation.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>We may be able to find factors $F$ and $G$ that influence A and B selectively:</td>
</tr>
<tr>
<td>$p'_1: u_A \leftarrow F$, $p'_2: v_B \leftrightarrow F$, $p'_3: v_B \leftrightarrow G$, $p'_4: u_A \leftrightarrow G$,</td>
</tr>
<tr>
<td>and jointly influence no other process.</td>
</tr>
<tr>
<td>If so, their effects on $M_{AB}$ will be additive.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Alternative Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>We find factors $F$ and $G$ with additive effects on $M_{AB}$.</td>
</tr>
<tr>
<td>We fail to find such factors.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inferences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support for joint hypothesis $H1 + H3$.</td>
</tr>
<tr>
<td>Refutes one/both of $H1$, $H3$, or we didn’t look enough for $F$, $G$.</td>
</tr>
</tbody>
</table>

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.\textsuperscript{11} To understand Table 9.

---

9. A common error of interpretation 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 (Berger & Hsu, 1996; Rogers, Howard, & Vessey, 1993). 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 critical size might be determined by the sizes of effects generated by plausible models. An example of suitable caution about a null effect can be found in Ghorashi, Enns, Klein, and Di Lollo (2010).

10. Table 3 of Sternberg (2001); reprinted by permission.
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

\[
M_{AB}(F_j, G_k) = u(F_j) + v(G_k) \tag{2}
\]

where \( u(F_j) \) 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 \( \bar{u}, \bar{v}, \) and \( \bar{M}_{AB} \) to indicate the means of these random variables, it is convenient that with no further assumptions\(^{12} \), Eq. 2 implies:

\[
\bar{M}_{AB}(F_j, G_k) = \bar{u}(F_j) + \bar{v}(G_k) \tag{3}
\]

In what follows, I treat the levels of factors as ordered, which permits describing changes in level as increases or decreases. From Eq. 2 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 \( G \):

\[
effect(F, G) \equiv M_{AB}(F_2, G_2) - M_{AB}(F_1, G_1) = [u(F_2) + v(G_2)] - [u(F_1) + v(G_1)]
\]

\[
= [u(F_2) - u(F_1)] + [v(G_2) - v(G_1)] = effect(F) + effect(G) \tag{4}
\]

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

\[
effect(F \mid G = G_k) \equiv M_{AB}(F_2, G_k) - M_{AB}(F_1, G_k) = [u(F_2) + v(G_k)] - [u(F_1) + v(G_k)]
\]

\[
= [u(F_2) - u(F_1)], \text{ regardless of } G_k. \tag{5}
\]

A given measure may be pure or composite, depending on the factors being varied and 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 approach.\(^{13} \)

### 2.3 Overview of examples and issues

Much of what follows consists of descriptions of ten examples of the process decomposition approach, the first eight successful. There are also brief summaries of two additional examples, in Section 10, which are included to convey the breadth of the approach. Examples will be referred to by the section numbers in which they are first discussed. Thus, the example discussed in Section 4.3 will be called "Ex. 4.3".

With two exceptions, the successful primary examples involve factorial experiments with two factors. The exceptions are Ex. 3.2, in which the effects of the two

---

11. 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.

12. Matters are not so simple for other combination rules, such as multiplication; see Section 8.2.

13. The reasoning is sometimes erroneously expressed as "If we assume \( H3 \) then additivity confers support on \( H1 \)." This ignores the support that additivity also confers on \( H3 \).
factors are studied in separate experiments, and Ex. 4.3, in which the effects of three factors are considered. 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 (10.1 and 10.2) are intriguing and tantalizing cases where the approach could have been used but was not, because of incompleteness of either the design or the analysis.

In Section 3 I discuss two applications based on electrophysiological measurements at the scalp from which pure measures are derived. In both cases, two neural modules are identified: In Ex. 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. 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. 3.1 are discussed in Section 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 4.2, and applied in Section 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.

As discussed in Section 5, signal detection theory has provided a widely applied method for measuring sensory processes in tasks that also involve decision processes, but has in general failed to isolate those decision processes from sensory factors. Ex. 5 illustrates a variant of the method that succeeds, and thereby demonstrates the modularity of sensory and decision processes, in an experiment with pigeons in which the decision factor is not the traditional payoff matrix.

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 Sections 6.3 and 6.4 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. 6.3 (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. 6.4 (manual choice-reaction with four stimulus-response pairs) the fMRI data support the hypothesis of modular neural processes for stimulus encoding and response selection, but unlike earlier observations of effects on $\overline{RT}$ from several similar paradigms, the effects of the two factors on $\overline{RT}$ interact rather than being additive, raising questions of interpretation. One 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 where both factors were found to have effects.

The examples in Sections 6.5 and 8.1 use the fMRI adaptation method; in both cases the measure, which reflects the dissimilarity between the current percept and previous ones, is composite, raising the issue of the combination rule. Ex. 6.5, concerned with the perception and memory of scenes, shows that different processes are responsible for immediate and delayed adaptation effects in the parahippocampal place area.

Perhaps because factorial experiments are relatively rare in studies of brain activation, we have had little experience in considering the implications of the additivity of effects on composite fMRI measures. It is tempting to infer that modular brain mechanisms are responsible. In Section 7 I consider the conditions under which inferences from additivity to modularity are justified, for both RT and fMRI.

In Section 8 I provide brief summaries of three additional examples of the process decomposition approach, using different composite measures. Ex. 8.1 uses the fMRI adaptation method also used in Ex. 6.5 to test the idea that perceptually separable dimensions are encoded by different neural processes, whereas perceptually integral dimensions are not. One reason for including Ex. 8.2, a classic study that provides evidence for modular spatial-frequency analyzers in visual pattern detection, is that unlike the other examples that use composite measures, the combination rule here is multiplication rather than summation. Ex. 8.3 illustrates the power of using the amplitude of the event-related potential as a composite measure.

In Section 9 I contrast the process-decomposition approach with the task-comparison method. Here, inferences are drawn from the effects on different tasks of factors such as the presence versus absence of sleep deprivation or the presence or absence of a lesion in a particular brain region, together with theories about the processes used to accomplish those tasks. I describe two examples of task comparison. One is based on the effects of repetitive transcranial magnetic stimulation (TMS) on two kinds of tactile perception. The other is 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 (Shallice, 1988, Ch. 11; SM:A.1), the process-decomposition approach incorporates such tests. In Section 9.3 I describe the conflict between the conclusion from Ex. 4.3 that the effect of sleep deprivation is selective and a claim, based on task comparison, that the effect is global.

The final two examples involve the effects of 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 $\overline{RT}$ 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 $\overline{RT}$, the brain region to which TMS is applied probably plays some role in performance of the task. However, the potential of this way of associating brain regions with modular subprocesses 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. 10.1), the absence of variation of a within-task search-specific factor precluded doing so. For number comparison (Ex. 10.2) a comparison-specific factor was varied, but no focused test of the interaction of this factor
with TMS was conducted.

The traditional task of cognitive neuropsychology is to learn about the cognitive functions of the normal brain by studying what happens to them when it is damaged, typically by comparing the effects of damage in different tasks. Can brain damage also be used in the process decomposition approach? This question is considered in Section 11; several reasons lead to a negative answer. Instead, the two approaches can inform each other.

Several additional issues are considered briefly in Section 12.

3. Decomposing Neural Processes with the Lateralized Readiness Potential

3.1 Serial modules (stages) 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.\textsuperscript{14} Let $A_{mc}(t)$ (an index of motor-cortex asymmetry) express this difference as a function of time after 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 is defined to as to become positive when the (correct) response hand is selected; the increase of such asymmetry is called the lateralized-readiness potential (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, Kok, Kenemans, and Bashore (1995). It was a two-choice RT experiment with single-digit 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 — simple — versus a sequence of three keystrokes — complex — made by fingers of the responding hand). 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 the LRP onset times be $T_{sjk}$ (measured with respect to the stimulus) and $T_{rjk}$ (a negative quantity, measured with respect to the response), and let $T_{rjk} = RT_{jk} + T_{rjk}$. $T_{sjk}$ and $T_{rjk}$ are then

\textsuperscript{14} As in some other brain measurements (e.g., PET, fMRI), the poor $S/N$ ratio often means that averaging over trials is required for the measures to be interpretable. 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.
different estimates of the same time point between stimulus and response, measured from the stimulus; let $T_{jk}$ be their mean. Averaging over the four conditions, $RT_\cdot \cdot = 416$ ms and $T_\cdot \cdot \cdot = 264$ 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 permit us to estimate their durations (pure measures of $\alpha$ and $\beta$) averaged over the four conditions: $D_{\alpha \cdot \cdot} = T_\cdot \cdot \cdot = 264$ ms and $D_{\beta \cdot \cdot \cdot} = RT_\cdot \cdot \cdot - T_\cdot \cdot \cdot = 152$ ms.

![Fig. 1](image-url)

**Fig. 1.** Means over 14 subjects of data from Smulders, Kok, Kenemans, and Bashore (1995). Estimated duration $D_{\alpha \cdot \cdot} \cdot \cdot$ of Process/Stage $\alpha$, from stimulus presentation to LRP onset (Panels A1, B1); and duration $D_{\beta \cdot \cdot \cdot}$ 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 33 ± 6 ms (Panel A1) and 1 ± 7 ms (Panel A2); the corresponding main effects of $RC$ are 4 ± 7 ms (Panel B1) and 21 ± 6 ms (Panel B2). The interaction contrasts are 6 ± 13 ms (stage $\alpha$) and $-8 \pm 11$ ms (stage $\beta$). The $RT$ data (discussed in Section 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). (Fig. 15 of Sternberg, 2001; reprinted by permission.)

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

What are stages? They are functionally distinct operations that occur during nonoverlapping epochs such that the response occurs when all stages have been completed. In a process with two stages, the stream of operations between stimulus and response can be cut, separating Stage A (the processes before the cut) from Stage B (the processes after the cut). If A and B are influenced selectively by factors F and G, respectively, this means that F can have an effect only before the cut, and G only after the cut. In most analyses based on behavioral data, the cut is hypothetical, inferred from the additivity of factor effects. However, in the neural process analysis discussed above, the cut corresponds to a particular observable neural event and divides the neural process into two subprocesses whose durations are selectively influenced by the two factors. This more direct observation of processing stages supports the inferences from the RT data.

In contrast to $D_\alpha$ and $D_\beta$, which are hypothesized (and confirmed) to be pure measures, $RT_{jk}$, shown in Panel C, is a composite measure. With two stages between stimulus and response, the RT reflects contributions from both. In Section 4.1 I consider the RT data from this experiment and how the inferred stages A and B relate to the neural processes $\alpha$ and $\beta$ inferred from the LRP data.

### 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, Bashore, Coles, Donchin, and Meyer (1992) devised a clever way to ask 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 Go versus 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.

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, GND₁) or low (similar shapes, GND₂). The other factor (in Exp. 2) is the

---

15. The error variance values reported by Smulders et al. (1995) and the SE estimates provided here are likely to be overestimates (because balanced condition-order effects were treated as error variance); the data required to calculate better values are no longer available (F. T. Y. Smulders, personal communications, September, 1999).
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, $MC_2$).

Fig. 2. Schematic idealized asymmetry functions $A_{mc}(t)$ from Osman, Bashore, Coles, Donchin, and Meyer (1992). 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 (mapping compatibility = $MC = MC_1$) or incompatible ($MC = MC_2$). (Fig. 3 of Sternberg, 2001; reprinted by permission.)

Idealizations of the resulting $A_{mc}(t)$ functions are shown in Fig. 2. Each of the four panels 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 LRP) is the time at which their difference reliably exceeds zero. In Exp. 1, GND influenced $T_2$ (by 43 ms) but not $T_1$ (compare Figs. 2A1 and 2A2). The absence of an effect of GND on $T_1$ shows that the location of the stimulus, but not its category, controlled the start of response preparation. 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.

---

16. The effect of GND on $T_1$ was $2.5 \pm 5.0$ ms; its effect on $T_2$ was $43 \pm 14$ ms; the difference between these effects is $41 \pm 11$ ms ($N = 6; p \approx 0.01$).
In one of the conditions of Exp. 2, the stimulus-response 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 prolong the Go-NoGo discrimination. Osman et al. therefore reduced letter-digit shape discriminability so as to increase $T_2$ from about 240 ms to about 350 ms. In this experiment, $MC$ influenced $T_1$ (by 121 ms) but not $T_2$ (compare Figs. 2B1 and 2B2). 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 1). 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. 2B) 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. 2A) for the effect of $GND$. An 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 $T_1$ (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. 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, a composite behavioral measure ($\overline{RT}$) leads to a similar analysis of corresponding mental processes, as we shall see in Section 4.1. 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, depending on the level of $GND$.

### 4. Mental Processing Stages Inferred From Reaction Times

#### 4.1 Analysis of the reaction-time data in Ex. 3.1

The conclusion in Ex. 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:

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

Suppose we had only those data and not the LRP data. It follows from the partitioning of $\overline{RT}$ shown in Eq. 6 that the combination rule for the contributions of neural processes $\alpha$ and $\beta$ is summation. As discussed in Section 2.2, that and the selective influence of factors $SQ$ and $RC$ on

---

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

18. The effect of $MC$ on $T_1$ was $121 \pm 17$ ms; its effect on $T_2$ was $3.3 \pm 8.8$ ms; the difference between these effects is $118 \pm 21$ ms ($N = 6; p = 0.01$).
**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 the hypothesis that the process contains subprocesses arranged sequentially, in stages, with each of the factors influencing a different subprocess selectively. 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.

---

19. This way of deriving the $\{RT^*_{jk}\}$ forces their means into agreement: $RT^*_{jk} = \overline{RT}$; the question of interest is whether the differences among the two sets of four values agree.
Section 7 for a discussion of the strength of such inferences. The application discussed in the section below exemplifies how, by combining inferences from a pattern of additive and interacting factor effects, it is possible to reveal more complex processing structures.

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, and Van Arkel (1982, 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 of the effect of a factor is a feature of this example that distinguishes it from the others: the inclusion of more than two actors. 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 "noise" in the form of 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). 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. The measure was the RT for trials with correct responses.

Other studies (see Section 6.4) had already suggested that SQ and MF were likely to selectively influence two different stages of processing, stages that might be described as stimulus encoding (S) and response selection (R).

The results in Fig. 3 consist of the RTs from the $2 \times 2 \times 2 = 8$ conditions. Panels A1 and A2 show that at each level of SLP there are additive effects of SQ and MF on RT. 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 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 RT when the stimulus is degraded (98 ms)

---

20. Given an assumption about the durations of different stages that is stronger than zero correlation but weaker than stochastic independence, then assuming stages plus selective influence implies numerous properties of other features of the RT distributions in addition to their means (Sternberg, 1969; Roberts & Sternberg, 1993), such as additive effects on var(RT). However, a stages hypothesis alone doesn’t require this property of var(RT).


22. This conclusion (separately-modifiable sequential processes, or stages) is further strengthened by analyses of complete RT-distributions rather than just RT means, from similar experiments (Sternberg, 1969, Sec. V; Roberts & Sternberg, 1993, Exp. 2).
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.

Fig. 3. Data from Sanders, Wijnen, and Van Arkel (1982, 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 (unfilled squares, broken lines), 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 $\pm SE$ bar was 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. (Fig. 14 of Sternberg, 2001; reprinted by permission.)
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, \( S \), thereby further characterizing \( S \) and \( R \).

5. Sensory and Decision Modules Revealed by Signal-Detection Theory

Probably 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 (Swets, Tanner, & Birdsaill, 1961; Macmillan & Creelman, 2005). 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.

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 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 matrix, 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 $M_S = d'$. 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. Unfortunately, selective influence by PM of D has not been shown; D is also influenced by s-factors. However, a few animal studies, including the one described below, suggest that response bias is selectively influenced 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", $RR$). $RR$ can be defined as $Pr\{R_{NT}|\text{Reward}\}$, the proportion of the total number of rewards (for both kinds of correct responses) that are given for $R_{NT}$.

In a luminance-discrimination experiment with six pigeons, McCarthy and Davison (1984) used a linked concurrent pair of variable-interval (VI) schedules to control $RR$. On each trial in a series, one of two light intensities appeared on the center key of three keys; 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 $Pr\{R_{NT}|\text{Reward}\}$, could be one of three values, 0.2, 0.5, or 0.8.\footnote{If $Pr\{R_{NT}|\text{Reward}\} = 0.2$, for example, for each rewarded $R_{NT}$ response there are four rewarded $R_T$ responses, encouraging a liberal (low) criterion for $R_T$.} 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

\[
\hat{d'} = z(Pr\{R_T|S_T\}) - z(Pr\{R_T|S_{NT}\}) \quad \text{and} \quad \hat{c} = \frac{z(Pr\{R_T|S_T\}) + z(Pr\{R_T|S_{NT}\})}{2}.
\]

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. 4.

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 D), it leaves invariant our measure $d'$ of discriminability (and hence of 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 $LR$.\footnotetext{If $Pr\{R_{NT}|\text{Reward}\} = 0.2$, for example, for each rewarded $R_{NT}$ response there are four rewarded $R_T$ responses, encouraging a liberal (low) criterion for $R_T$.}
Fig. 4. Mean effects of Reinforcement Ratio, $RR = Pr\{R_{NT}|\text{Reward}\}$ (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 (SM:7.2,15.1), which is not statistically significant. The ±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 and Davison (1984); basic data kindly provided by B. Alsop. (Fig. 5 of Sternberg, 2001; reprinted by permission.)

6. Neural Processing Modules Inferred from Brain Activation Maps

6.1 fMRI signals as pure measures

Suppose there is localization of function (see, e.g., Op de Beeck et al. 2008, but also Haxby, 2004), such that two neural processes $\alpha$ and $\beta$ are implemented by different processors $P_\alpha$ and $P_\beta$ in non-overlapping brain regions $R_\alpha$ and $R_\beta$. The search for such processes has sometimes been stimulated by behavioral findings that support the existence of modular mental processes $A$ and $B$ that are influenced selectively by factors $F$ and $G$. 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 $R_\alpha$, a pure measure, should vary with $F$, but not with $G$, and conversely for region $R_\beta$.$^{24}$ 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 such regions were found, it would support the modular decomposition inferred from the behavioral data, and would also support the conclusion that the processors that implement A and B are anatomically localized.

For examples 6.3 and 6.4, it was the modular decomposition into two processing stages A and B 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.

6.2 The fMRI signal as a composite measure

If α and β are implemented by different neural processors, P_α and P_β (or by the same processor P_{αβ}) in one region, R_{αβ}, then the activation level in R_{αβ} is a composite measure that depends on both α and β. To test separate modifiability we must know or show how their contributions to the activation measure are combined. For example, if the combination rule is summation (sometimes assumed without justification) and if factors F and G influence α and β selectively, then the effects of F and G will be additive. Finding such additivity in a factorial experiment would support the combination rule in that brain region, as well as selective influence. (If summation were assumed erroneously, selective influence would be obscured: the effect of each factor would appear to be modulated by the level of the other.) Additivity was found in examples 6.5 and 8.1, both of which used fMRI adaptation and composite measures, as well as in two of the brain regions examined in Ex. 6.4. Some of the evidence that bears on the combination rule for fMRI signals is mentioned in Section 7.2.

6.3 Modular processes in number comparison

In an experiment by Pinel, Dehaene, Rivière, and LeBihan (2001), 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 (Dehaene, 1996) had shown additive effects of N and P on 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 e and γ that are implemented by separately localized processors. When we average absolute effect sizes and SEs over the regions of each type, we find that for the nine N-sensitive regions the N effect was 0.17 ± 0.05% (median p-value = 0.01), while the P effect was 0.06 ± 0.08%; for the seven P-sensitive regions the P effect was 0.32 ± 0.10% (median p-value = 0.01), while the N effect was 0.04 ± 0.04%.

24. Such tests require no assumptions about whether a change in factor level causes an increase or decrease in activation. This contrasts with the assumption, sometimes used to infer modular neural processors (Kanwisher, Downing, Epstein, & Kourtzi, 2001), that stimuli more prototypical of those for which a processor is specialized will produce greater activation.
The fMRI data from three well-behaved regions are shown in Fig. 5, Panels B, C, and D. The concurrently collected RT data (Fig. 5, Panels A) replicated the earlier study, suggesting that we associate the neural modules $e$ 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 $RT$) increased activation in some regions (e.g. Fig. 5, Panels C) and decreased it in others (e.g. Fig. 5, Panels D).\textsuperscript{25}

### 6.4 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 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 mapping compatibility): these studies have shown that stimulus quality and mapping difficulty have additive effects on $RT$, consistent with the idea that there are two processes, arranged in stages, that are selectively influenced by these factors. (These studies include Biederman & Kaplan, 1970, after a session of practice; Frowein & Sanders, 1978; Roberts & Sternberg, 1993, Exp. 2; Sanders, 1977, Exp. II; Sanders, 1980, Exp. 3; Sanders, Wijnen, & Van Arkel, 1982, Exp I; and Schwartz, Pomerantz, & Egeth, 1977, Exp. 2.) 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 choice-reaction task with their versions of $SQ$ and $MC$, Schumacher and D’Esposito (2002) measured fMRI in several brain regions concurrently with RT. In their task, 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 spatial 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. 6.3, 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 $RT$ — also produced the larger fMRI signal. Figures 6.1 and 6.2 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

---

\textsuperscript{25} Without requiring it, this finding invites us to consider that there are two qualitatively different encoding processes $e$, one for each notation, rather than ‘one’ process whose settings depend on $N$. This possibility is 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" (S. Dehaene, personal communication, September 29, 2006). 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 8), but evidence for modularity emerges nonetheless: the proximity effect is invariant across the two task. 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, an RT experiment that could, alone, support such a claim has yet to be devised.
Fig. 5. Reaction-time and selected brain-activation data from Pinel, Dehaene, Rivière, and LeBihan (2001). 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 RT; this effect, from low to high \( P \), is 159 ± 24 ms, while the main effect of \( N \) is 204 ± 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 ± 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 ± 0.09\% \((p = 0.01)\), −0.03 ± 0.03\%, and 0.00 ± 0.04\%. Shown in Panels B2, C2, and D2, the main effects of \( N \) are −0.06 ± 0.06\%, 0.16 ± 0.05\% \((p = 0.01)\), and −0.15 ± 0.05\% \((p = 0.02)\). (Fig. 6.2 of Sternberg, 2004; reprinted by permission.)
the fMRI signal was influenced by one of the factors but not the other) are consistent with S and R being implemented, at least in part, by anatomically distinct populations of neurons. 26

If 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 brain regions (Panels D and E), both factors were found to have effects. One possibility is that these regions each contain two specialized populations of neurons, each of which is influenced 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 S and R are sequential, 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, and again it is plausible that the summed neural activity would be influenced additively by the two factors. As shown in Fig. 7, 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; the supplementary assumptions required for this implication are discussed in Section 7.2. Given these assumptions, 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.

In contrast to the earlier findings with various experimental arrangements mentioned above, the effects of the two factors on RT in the Schumacher-D’Esposito task were unfortunately not additive; as shown in Fig. 7C 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); such an interaction was found in the data for eight of the nine subjects. 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, which suggests that stronger inferences require sampling more brain regions. 27 28 It is possible that whereas the measured neural processes are modular, the processes that underlie the RT in this experiment are not.

26. 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?

27. Schumacher and D’Esposito (2002) suggest that such a process might occur only under the stress of a subject’s being in the scanner, and not under normal conditions. On the other hand, a whole-brain analysis of the fMRI data did not reveal any additional task-sensitive regions (E. Schumacher, personal communication, November 27, 2006), and RT data from the practice session, outside the scanner, showed a non-significant interaction of about the same size and in the same direction.

28. There is an unresolved puzzle about these data that suggests that it would be valuable to replicate this experiment, using a procedure known to produce additive effects on RT. The large SEs associated with the very small mean interaction contrasts for the data shown in panels A and B 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 mean interaction contrasts were significantly \( p < 0.05 \) too small.
Fig. 6.1 & 6.2. Brain activation data relative to a fixation baseline for the six regions measured by Schumacher and D’Esposito (2002). 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.
Fig. 6.1 & 6.2. Brain activation data relative to a fixation baseline for the six regions measured by Schumacher and D’Esposito (2002). 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.
Fig. 7. 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 Schumacher and D'Esposito (2002). 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 RT, and the corresponding interaction contrast. (The mean effect of discriminability when the mapping was easy versus hard is 64 versus 94 ms, respectively, a difference that is 37% of the smaller main effect.)
6.5 Modular short-term and long-term memory processes in scene perception

Consider the process used to recognize a scene and make a judgement about it. This process is influenced by the observer’s previous experiences with the same scene and with other views of the same place. Epstein, Parker, and Feiler (2008) asked whether there are different subprocesses influenced selectively by such previous experiences, depending on whether the experiences are very recent ("short-interval" — within 2 sec) or less recent ("long-interval" — separated by about a half hour). Students at the University of Pennsylvania (Penn) saw two series of visual scenes, Series 1 before their brains were scanned, Series 2 while being scanned. During Series 1, all the scenes were from Penn, and subjects had to judge which half of the campus each scene was from.

Two scenes were presented successively on each trial in Series 2, with onsets separated by 1.2 sec. For two thirds of the trials, both scenes were from Penn; for the remaining third (whose data were discarded), one or both scenes were not from Penn. Subjects had to judge whether both scenes were from Penn, and make one speeded manual response to the second scene if so, the other response if not. Primary interest was in the RT and in the fMRI signal in the parahippocampal place area (PPA) on each trial when both scenes were from Penn, and how these measures were influenced by two orthogonal factors: the short-interval relation: between the second scene and its predecessor, and the long-interval relation between the second scene and scenes that had been displayed during Series 1.

fMRI Adaptation (repetition suppression; Grill-Spector & Malach, 2001) was expected both from Series 1 and from the first scene on each trial in Series 2: To the extent that a stimulus is similar to one shown earlier, the neural response to that stimulus is reduced. (One can also think of recovery from fMRI adaptation as a novelty effect: to the extent that aspects of a stimulus to which the measured neurons are sensitive are novel, their response is greater.) Because the fMRI signal in the PPA as well as the RT vary with both short-interval and long-interval relations, they are both composite measures. Answering the question whether short- and long-interval fMRI adaptation in the PPA is generated by a single process or by two separate processes therefore requires determining how their effects on the neural response combine. Because the RT undoubtedly depends on other brain regions in addition to the PPA, analysis of the two measures might lead to different answers.29

There were three possible short-interval relations between the first and second scenes on a trial: they could be identical (second scene = old place, old view), similar (second scene = old place, new view), or different (second scene = new place). Similarly, there were three possible long-interval relations: the Series 1 scenes could include one that was identical to the second scene, or one that was similar, or none that was either identical or similar. The ingenious design permitted these two kinds of relation to vary orthogonally. For the analysis presented here, the data were partitioned so as to separately examine (1) the view effect: the effect of increasing novelty by showing a different view of a previously seen place rather than the same view, and (2) the place effect: the effect of increasing novelty by showing a new place rather than repeating a previously seen place.30 Thus there are two short-interval factors (sameness or difference of view

29. Suppose that the fMRI analysis leads to the conclusion that the two factors influence separate modules within the PPA. This would not preclude their having interactive effects on the mean RT. This could happen, e.g., if the processes that contribute to the RT include one or more processes other than the one implemented by the PPA that were influenced by both factors. Or it could happen if the processes that contribute to the RT were selectively influenced by the two factors, but were arranged in parallel rather than as stages. On the other hand, suppose that the fMRI analysis leads to the conclusion that there is a single module influenced by the two factors. Then, unless the PPA does not play a role in generating the response, additivity of the RT effects of the two factors would be unexpected.
and place) and two long-interval factors (sameness or difference of view and place), each with two levels. In asking how the effects of short-interval and long-interval factors combine, we can ask how each of the two short-interval factors combines with each of the two long-interval factors, giving us four separate tests.

Means of the PPA fMRI signal strengths over the sixteen subjects are shown in Fig. 8. All four of the main effects are significant, with \( p < 0.01 \). In each case the long- and short-interval effects are remarkably close to being additive. Whereas the long- and short-interval place effects are similar in magnitude (Panel B), with the former slightly greater, the long-interval view effect is substantially smaller than the short-interval view effect (Panel A). Indeed, the difference between the long- and short-interval main effects of the view factor is significantly greater than the difference between the main effects of the place factor, with \( p < 0.01 \). In addition to its significance level attesting to the remarkably high precision of these data, this three-way interaction supports the idea that the long-interval adapting effect is more "viewpoint invariant" than the short-interval effect.

Means of the RT data are shown in Fig. 9. There is one substantial interaction (in Panel D; the contrast has about the same magnitude as the smaller of the two main effects) but it is not statistically significant. The additivity in Panel A should probably not be taken seriously, because the long-interval view effect is small (just as in the fMRI data) and, for the RT data, not statistically significant. As in the fMRI data, lengthening the interval reduces the view effect more than the place effect in the RT data; indeed, it reverses the effect in the RT data. It is interesting to consider how strongly this difference between the view and place effects across long and short intervals would alone argue for the two effects being generated by different processing modules. (See the discussion of differential influence in Section 12.4.)

To the extent that summation is a plausible combination rule for effects on fMRI (see Section 7), the additivity of the effects of short-interval and long-interval factors is consistent with their being generated by separate modular processes. However, the fMRI data alone don’t tell us whether these processes are concurrent (resulting from the activity of different neural populations in the same region) or sequential. The fact that no brain regions were found in which just one process appeared to be operating seems to argue against their being implemented by different neural populations. Both of the neurophysiological accounts suggested by Epstein et al. (2008) are consistent with the processes being sequential. If the additivity of the RT effects were more persuasive, we would have evidence for two processes being arranged as stages. However, as mentioned above, we don’t know the relation between PPA activity and the processes that generate the response. Whatever their temporal arrangement, if two independent (modular) processes mediate the effects of the same experience, depending on how long ago it occurred, this seems to require separate short-term and long-term memory representations.

30. In the present analysis of the effects of place change, the data for same place were collapsed over the view factor. This analysis differs from that of Epstein, et al. (2008).

31. These are means of the signal strengths in the PPA regions of the two hemispheres.

32. In a whole-brain analysis reliable long- and short-interval effects were found in many other brain regions. There was no persuasive evidence that any region had just one of these effects. Furthermore, the number of regions in which the interaction of the two effects was significant (two among 21 tests) can be explained as being the result of type I error. Thus, no regions were found that provided pure measures of either effect, and additional evidence was found of additivity of the two effects on the fMRI signal.
Fig. 8. Combined effects on the fMRI signal in the parahippocampal place area of two kinds of short-interval and long-interval novelty, from Epstein, Parker, and Feller (2008). The interaction contrasts as a proportion of the smaller main effects are: A: 0.19; B: 0.06; C: 0.43; D: 0.07.
Fig. 9. Combined effects on the mean reaction time of two kinds of short-interval and long-interval novelty, from Epstein, Parker, and Feller (2008). The interaction contrasts as a proportion of the smaller main effects are: A: 0.33; B: 0.05; C: 0.54; D: 1.09.

7. Modular Explanations of Additive Effects: How Strong is the Inference?

The inferences from composite measures to modular processes in Sections 4 and 6 based on RT and fMRI data have made use of the reasoning described in Table 2. How persuasive is such reasoning? Suppose that we make an observation — of the additivity of the effects of two factors on a measure — that is consistent with a theory according to which the two factors influence a particular pair of processing modules selectively. Given the observation, how credible is the
theory? This depends on the plausibility of the theory, the plausibility of alternative modular
theories, and the plausibility of single-process (i.e., non-modular) theories that are also consistent
with the observation (Howson & Urbach, 2006).

7.1 Inferences from reaction-time measurements

Let us first consider the case where the measure is $RT$, and where the observation is of the
additivity of the effects of two factors. When the additive-factor method was first proposed
(Sternberg, 1969), there was a highly plausible and historically significant modular explanation:
stages plus selective influence, with the combination rule (summation) inherent in the concept of
stages. And there were no well-known plausible alternatives, either modular or non-modular.
Since then, it has been discovered that under some conditions, other models, quite different in
spirit from stage models, can also generate such additive effects. (Ashby, 1982; McClelland,
1979; Miller, van der Ham, & Sanders, 1995; Roberts & Sternberg, 1993). In all these cases, the
prediction of means additivity derives from modularity plus selective influence; hence, from the
viewpoint of discovering modular processes (but not of how these processes are arranged in time)
the existence of these alternative possibilities does not weaken the argument outlined in Table 2.
Their discovery, however, decidedly weakens the inference from the additivity of factor effects on
$RT$ to the organization in stages of the corresponding processes. To help distinguish among such
modular alternatives one can sometimes use brain measurements, or other aspects of the RT data
(e.g., Roberts & Sternberg, 1993). (Examples of the use of brain measurements are the pure
measures of stage durations in the electrophysiological data of Ex. 3.1, and neurophysiological
data; see Schall, 2003, and references therein, and Woodman, Kang, Thompson, & Schall, 2008.)
As for non-modular theories, I know of no plausible way in which what is truly a single
mechanism can give rise to additive effects of two factors on mean RT.33 This could of course
change, as we learn more about mental and neural mechanisms.

7.2 Inferences from fMRI measurements

For the reasoning that has been applied to RT data also to be valid for the fMRI signal in a
brain region, we have to believe in the truth of two propositions. Proposition 1: Summation is the
combination rule for their contribution to the BOLD signal ($B$) of different processes in that
region, either successive, or concurrent but implemented by different neural populations. (Only
then can additivity of effects support the inference of selective influence.) This requires $B$ to be
linear in the total amount of neural activity in the region, a property discussed below. Proposition
2: Factors that influence the same neural process $\alpha$ do not have additive effects on the amount of
neural activity, $N_{\alpha}$. (If they did, and $B$ satisfied the linearity requirement, then factors that
influence the same neural process would produce additive effects on $B$.) The resulting
requirement of nonlinearity for neural activity is discussed below, together with the concept of
factor-level strength. Consider first two plausible ways in which modular neural processes $\alpha$ and
$\beta$ in a region could give rise to additive effects on fMRI signal strength. Let $F$ and $G$ be factors
that influence them selectively.

Different neural populations. Suppose first that $\alpha$ and $\beta$ are implemented by two different
populations of neurons in the measured region. Let us write $N_{\alpha} = N_{\alpha}(F)_{j}$ and $N_{\beta} = N_{\beta}(G_{k})$ for
the levels of neural activity in those two populations, and assume that the fMRI signal, $B$ (for
BOLD) depends on both $N_{\alpha}$ and $N_{\beta}$, but possibly to different degrees. Assume further that $B$ is
a linear function of $N_{\alpha}$ and $N_{\beta}$: $B(N_{\alpha}, N_{\beta}) = B_{0} + aN_{\alpha} + bN_{\beta}$, where $B_{0}$ is a baseline level, and

33. In the most popular detailed RT model, the diffusion model (e.g., Ratcliff & Smith, 2010), the most natural way
in which a factor has its effect is by changing the rate of evidence accumulation. Additive effects on this rate
produce effects on $RT$ that are overadditive.
of neural activity, discussed below. The effects of factors $F$ and $G$ on $B_{jk}$ must be additive, whatever the form of the functions $N_a(\cdot)$ and $N_b(\cdot)$. The inferential logic for this case is described by Table 2, and, as described in that table, the observation of additivity, as in Ex. 6.4, 6.5, and 8.1, supports the hypothesis that factors $F$ and $G$ influence different modules selectively, together with the hypothesis that the contributions to $B$ of the two populations combine by summation. However, the strength of such a conclusion depends on the validity or plausibility of the assumption that $B$ is linear in the level of neural activity, discussed below.

One neural population, successive processes. Because the fMRI signal reflects neural activity that occurs over a period of seconds, if processes $a$ and $\beta$ are carried out successively and sufficiently rapidly by the same population of neurons in the measured region, they would both contribute to the fMRI signal. With one proviso, even if temporal summation were imperfect (e.g., Dale & Buckner, 1997), an equation like Eq. 8 would apply; the failure of perfect temporal summation would be reflected only in the parameters $a$ and $b$.

Single-process explanation of additive effects. As summarized in Table 2, the observation of additive effects of two factors supports the joint hypothesis consisting of $H1$ and $H3$. An alternative explanation for such a finding is that there is a single process ($a$) influenced by both factors that operates in such a way that their effects on $M_a(F,G)$ are additive. In evaluating the plausibility of this alternative it is important to consider what would be required of a single process for it to have this property, and how the requirement could be tested. For this purpose it is helpful to consider the idea of factor-level strength.

Suppose two factors, $F$ and $G$, each at two levels, in a $2 \times 2$ factorial design, and any composite measure $M = M(F,G)$. If the fMRI signal reflects neural activity that occurs over a period of seconds, if processes $a$ and $\beta$ are carried out successively and sufficiently rapidly by the same population of neurons in the measured region, they would both contribute to the fMRI signal. With one proviso, even if temporal summation were imperfect (e.g., Dale & Buckner, 1997), an equation like Eq. 8 would apply; the failure of perfect temporal summation would be reflected only in the parameters $a$ and $b$.

Single-process explanation of additive effects. As summarized in Table 2, the observation of additive effects of two factors supports the joint hypothesis consisting of $H1$ and $H3$. An alternative explanation for such a finding is that there is a single process ($a$) influenced by both factors that operates in such a way that their effects on $M_a(F,G)$ are additive. In evaluating the plausibility of this alternative it is important to consider what would be required of a single process for it to have this property, and how the requirement could be tested. For this purpose it is helpful to consider the idea of factor-level strength.

Suppose two factors, $F$ and $G$, each at two levels, in a $2 \times 2$ factorial design, and any composite measure $M = M(F,G)$. Associate a factor-level strength, $s_f$ and $s_g$, with each level of each factor; let the strengths for $F_1$, $F_2$, $G_1$, and $G_2$ be $q$, $q+r$, $u$, and $u+v$, respectively. To make it clear that $s_f$ and $s_g$ need to be specified only up to arbitrary multiplicative constants, we can write, for total factor-level strength, $tFLS = bs_f + cs_g$. Then for the four conditions in the experiment, $(F_1, G_1)$, $(F_1, G_2)$, $(F_2, G_1)$, and $(F_2, G_2)$, the total factor-level strengths are $bq + cu$, $bq + cu + bv$, $bq + cu + br$, and $bq + cu + br + cv$, respectively. Subtracting $bq + cu$, they become $0$, $cv$, $br$, and $cv + br$, which makes it clear that the factor-level strengths in the four conditions are additive. (For example, increasing the level of $G$ has the same effect on $tFLS$ when $F = F_1$ as when $F = F_2$.) Because the total strength is additive, a necessary and sufficient condition for the effects of the two factors on $M(F,G)$ to be additive is that $M$ is a linear function of $tFLS$:

$$M(F,G) = M(tFLS) = M(bs_f + cs_g) = a + bs_f + cs_g.$$  

(9)

To the extent that such linearity is plausible or valid, it prevents us from confidently inferring

34. These constants would depend, for example, on the mean proximities of the two populations to the center of the brain region in which the fMRI signal is measured, and on the time relation between the two activations.

35. If the duration of either process is changed sufficiently by the change in factor level so that the temporal distribution of neural activation in the region is altered, additivity could fail. However, the findings in Ex. 6.4 are perhaps reassuring: Despite the fact that the $RT$ data (Fig. 7C) indicated substantial effects on the durations of both processes of interest, the additivity of the fMRI effects in both of the regions influenced by both factors was remarkably good. Apparently these duration changes are small relative to the sluggishness of $B$.

36. For example, $F_1$ and $F_2$ might correspond to two luminance levels of light on half of a small region of the retina, and $G_1$ and $G_2$ might correspond to luminance levels on the other half.
modular processes from additive effects on $M$. Applying this reasoning to fMRI, and assuming that $B_\alpha$ is linear in $N_\alpha$, $B_\beta$ is linear in $tFLS$ if (and only if) $N_\alpha$ is linear in $tFLS$: $N_\alpha(bs_f + cs_g) = a + b's_f + c's_g$. Thus, only to the extent that the linearity of $N$ in $tFLS$ is invalid or implausible does the finding of additivity of factor effects on $B$ support the hypothesis of modular processes.

In short, to justify the inferences in Ex. 6.5 and 8.1, and those based on the additive effects in two of the brain regions in Ex. 6.4, we have to believe that $B$ is (approximately) linear in $N$, and that $N$ is nonlinear in $tFLS$. These requirements appear not to have been recognized by some investigators.

Evidence for linearity of $B(N)$ and nonlinearity of $N$. The neural basis of the BOLD response is controversial, as is its relation to different types of neural activity (e.g., Ekstrom, 2010; Logothetis & Wandell, 2004; Goense & Logothetis, 2008). That $B(N)$ is linear has been concluded from comparing the effects of motion coherence (Rees, Friston, & Koch, 2000) and contrast (Heeger, Huk, Geisler, & Albrecht, 2000) on $B$ in humans with their effects on the single-cell spike rate in monkeys. Predictions from a linear transform model for the effects of certain visual stimulus variations on $B$ were largely confirmed for area V1 in an influential study by Boynton, Engel, Glover, and Heeger (1996); see also Dale and Buckner (1997). The failures of linearity in such studies (such as failures of temporal summation) generally don’t tell us whether it is the neural response, $N$, or $B(N)$, or both that are nonlinear (Boynton & Finney, 2003, Miller, Luh, Liu, Martinez, Obata, et al., 2001; Soltysik, Peck, White, Crosson, & Briggs, 2004). Generalization is problematic, because different brain regions may differ in linearity (Horner & Andrews, 2009, Soltysik, et al. 2004). The authors of a study in which electrophysiological as well as fMRI measures were taken concluded that the source of nonlinearity in V1 is $N$, not $B(N)$ (Wan, Riera, Iwata, Takahashi, Wakabahyashi, & Kawashima, 2006). However, Miller et al.’s (2001) finding of a nonlinear relation between cerebral blood flow and $B$ in primary visual and motor cortex is hard to explain unless $B(N)$ is nonlinear. And Boynton and Finney (2003) conclude that whereas some of the nonlinearity in higher visual areas is neural, some may be in $B(N)$. Taken together, the available evidence seems to permit us to treat the linearity of $B(N)$ only as a tentative working hypothesis.

Also, because pure measures of each factor were found in some brain regions in Ex. 6.4, it seems reasonable that the finding of effects of both factors in two regions should be explained by those regions containing two populations of neurons. These and other instances of additive effects on $B(N)$ provide evidence of its linearity, unless there are compensating nonlinearities in $B(\cdot)$ and $N(\cdot)$, which is unlikely.

Evidence, including that mentioned above, indicates that in response to contrast, duration, number of pulses, and number of trials, neural responses change nonlinearly (Albrecht, Geisler, & Crane, 2003; Carandini & Heeger, 1994; Geisler & Albrecht, 1995; Heeger et al., 2000; Li, Miller, & Desimone, 1993; Meeter, Myers, & Gluck, 2005). This reduces the plausibility of a single neural mechanism producing additive effects by responding linearly to $tFLS$. If we assume this, and therefore associate additive effects of two factors in a brain region, as in Ex. 6.4, 6.5, and 8.1, with their influencing different modules selectively, then such additive effects increase the support for the linearity of $B(N)$, for those regions and under the conditions of those experiments.$^{37}$

$^{37}$ This argument is especially compelling for the instances of additivity found in two of the regions in Ex. 6.4: The finding of pure measures of each process in other regions increases the plausibility of two selectively influenced neuron populations generating the additive effects.
8. Brief Accounts of Three Additional Examples

In this section I provide brief summaries of three additional examples that show the diversity of applications of the process-decomposition approach. The first is an application of fMRI adaptation, a method introduced in Ex. 6.5. The second makes use of a composite measure with a multiplicative combination rule; an outline of the inferential logic for this case is included. The third has considerable power and makes use of a brain measure (ERP amplitude) not used in the other examples.

8.1 Modular processes for perceptually separable dimensions

Another use of fMRI adaptation is exemplified by a study by Drucker, Kerr, and Aguirre (2009). Suppose a set of visual stimuli that vary along two dimensions, and that these dimensions are neurally separable, in the sense that they are evaluated by different modular neural processes (possibly implemented by different populations of neurons). And suppose further that the fMRI signal depends additively on the neural activity associated with these two processes:

\[ B(NA_1, NA_2) = f(NA_1) + g(NA_2) \]

Then the adapting effect of one stimulus, \( S_1 \) on the fMRI response to the next stimulus, \( S_2 \) will be the sum of the adapting effects on the two modules, which in turn will depend separately on the proximities of \( S_1 \) and \( S_2 \) along the two dimensions. In an experiment to test this idea, Drucker et al. used two different sets of outline shapes, each representing several levels of two dimensions. For one set the two dimensions were expected to be neurally separable because they are perceptually separable in the sense of Garner (1974); for the other stimulus set they were expected to be neurally integral, that is, evaluated by a single process, because they are perceptually integral, and were therefore expected to have adapting effects on the two dimensions that would interact rather than being additive.38

Applying a single-df global measure of interaction to each subject’s data for each stimulus set, they found results that supported their hypotheses. Unfortunately, they provided no breakdown of the multiple separate tests of additivity that their data permitted, no comparison of measures of interaction with corresponding main effects, and no way to be sure that the global measure in the additive case was not the resultant of multiple interaction contrasts of opposite sign. Their test is analogous to representing all the information in Fig. 8 as a single number. One reason for the appeal of this study is that the same experimental and analytic methods and the same subjects revealed contrasting fMRI effects, additive and interactive. Another reason is that the results support an attractive neural theory of perceptual integrality and separability.

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

Suppose a task in which a subject says "yes" when either or both of two detection processes respond, and says "no" if neither process responds. (For the present discussion we ignore the complication introduced by "guessed" yes responses that may occur when neither process responds.) If the behavior of the two processes is uncorrelated, then the probability of this joint event is the product of the individual nonresponse probabilities for the two processes. \( Pr[\text{"no"}] \) is thus a composite measure of the two detection processes, with a multiplicative combination rule, sometimes called "probability summation".

To describe the consequences of a multiplicative combination rule for a composite measure, I need to introduce the idea of a proportional effect, or \( p.eff ect \). We saw in Section 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.

38. While viewing each sequence of shapes, subjects reported on the position of a bisecting line.
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 $M_A$ is defined as a ratio (for a factor with two levels):

$$p.\text{effect}(F) = \frac{M_A(F_2)}{M_A(F_1)},$$

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

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

To derive the equivalent of Eq. 3 from Eq. 11 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. 11 that

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

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

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

If the p.effects of the factors are multiplicative, as in Eq. 13, 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 3.

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 (1971) 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 increased $Pr\{\text{non-detect}\}$ by a constant factor, consistent with Eq. 13, thus supporting the joint hypothesis. This model failed badly when the component spatial frequencies in the compound were closer together. These findings provided important early evidence for modular spatial-frequency analyzers and for the multiplicative combination rule.

### 8.3 Evidence from ERP amplitude 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 amplitude 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. Because of the simplicity of the hypothesis and the richness of the possible tests, this approach is especially powerful.

---


40. This follows from the electrical linearity of brain tissue (Nunez & Srinivasan, 2006, Ch. 1.5.)

41. Table 2 of Sternberg, 2001; reprinted by permission.
Kounios (2007) exploited this in a study of the effects of priming on the semantic classification of a sequence of spoken nouns. Most of the words required no response, while 5% were targets (names of body parts) that called for a manual response. The words consisted of primes and probes. The factors (two levels each) were the semantic relatedness (REL) of the probe to the preceding prime, and the semantic satiation (SAT) of that prime (number of immediate repetitions of the prime before the probe). The data were the ERPs elicited by the non-target probes at several locations on the scalp. A composite measure \( M_{\alpha \beta} \) is defined for each location as the mean ERP amplitude at that location during the epoch from 600 to 800 ms after probe onset. Consider the following theory, with three components:

**H1 (Subprocesses):** The complex process of recognizing the probe as a non-target contains (at least) two subprocesses, \( \alpha \) and \( \beta \), carried out by different neural processors, \( P_{\alpha} \) and \( P_{\beta} \).

**H2 (Selective Influence):** \( \alpha \) and \( \beta \) are selectively influenced by SAT and REL, respectively.

**H3 (Combination Rule):** Each process is an ERP source; physics tells us that at any location the combination rule for sources is summation.

This theory implies that the effects of SAT and REL on \( M_{\alpha \beta} \) will be additive at all scalp locations. Kounios found such additivity (mean main effects of REL and SAT were \( 1.3 \pm 0.2 \mu V \) and \( 2.1 \pm 0.4 \mu V \), respectively, while the mean interaction contrast was \( 0.01 \pm 0.3 \mu V, n = 36 \)), supporting the above theory and hence the modularity of \( \alpha \) and \( \beta \) during the 600 to 800 ms epoch.\(^{42}\) Also, the topographies of the two effects (their relative sizes across locations) differ markedly, indicating different locations in the brain for \( P_{\alpha} \) and \( P_{\beta} \), which argues against a single-process explanation.\(^{43}\)

---

\(^{42}\) Support for the theory is support for all of its three components. However, because the combination rule is given by physics in this application, there is no need to test component \( H3 \).
9. Process Decomposition versus Task Comparison

The applications above exemplify a process-decomposition approach whose goal is to divide the complex process by which a particular task is accomplished into modular subprocesses. The 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. The task-comparison method 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 (Schmidt & Vorberg, 2006) 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. See Section 11.) Although it may achieve other goals, task comparison is inferior to process decomposition for discovering the modular subprocesses of a complex process or for investigating their properties: The interpretation usually requires assuming a theory of the complex process in each task, and the method includes no test of such theories. In contrast, process decomposition requires a theory of only one task, and, as illustrated by the examples above, incorporates a test of that theory.\(^{44}\)

9.1 Comparison of two tactile perception tasks

An elegant example of task comparison, but one that is subject to its usual limitations, is provided by Merabet, Thut, Murray, Andrews, Hsiao, and Pascual-Leone (2004) 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. Both tasks involved palpation by the fingers of one hand of a set of tactile dot arrays with varying dot spacings. The judged dimensions were roughness \((r)\) in one task, and distance between dots \((d)\) in the other. 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 \(b = 1.0\); if there is an effect, the value of \(b\) is reduced, so the effect of rTMS can be measured by \(1 - b\). Means over the 11 subjects indicate that performance in the roughness-judgement task is influenced by rTMS\(_r\) of the contralateral somatosensory cortex \((1 - \bar{b}_{rs} = 0.21 \pm 0.07; p = 0.02)\), but negligibly by rTMS\(_o\) of the contralateral occipital cortex \((1 - \bar{b}_{ro} = 0.02 \pm 0.03)\), while performance in the distance-judgement task is influenced by rTMS\(_o\) \((1 - \bar{b}_{do} = 0.16 \pm 0.07; p = 0.04)\), but negligibly by rTMS\(_s\) \((1 - \bar{b}_{ds} = 0.05 \pm 0.04)\), a double dissociation of the two tasks.\(^{45}\)

Plausible theories might include, for each task, processes for control of stimulus palpation \((\alpha)\), for generation of a complex percept \((\beta)\), for extraction of the desired dimension \((\gamma)\), and for conversion of its value into a numerical response \((\delta)\). Any or all of these processes might differ

43. In this application, modularity appears to change over time: During an earlier epoch (400 to 600 ms after probe onset) the two effects had similar topographies, but they interacted substantially. See SM:14 for more details.

44. See SM:A.1.

45. 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 \(\bar{b}_{ro} - \bar{b}_{rs}\) and \(\bar{b}_{ds} - \bar{b}_{do}\) 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, straightforward interpretation of the slope values requires us to assume that effects of practice on those values were negligible.
between tasks. The striking findings indicate that for the two tasks, one or more of these processes depend on different regions of the cortex. In addition to specifying these four processes, a weak pair of task theories might also assert that \( \gamma_d \) and \( \gamma_r \) depend on occipital and somatosensory cortex, respectively. However, 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 \( \alpha_d = \alpha_r = \alpha \) are identical, that \( \beta_d = \beta_r = \beta \) are identical, and that \( \delta_d = \delta_r = \delta \) are identical. This pair of theories would predict the results, which, together with the absence of effects in two of the conditions, would then also imply that none of processes \( \alpha, \beta, \gamma, \) and \( \delta \) are identical. This pair of theories would predict the results, which, together with the absence of inferential limitations such as these characterize the task-comparison method.

9.2 Donders’ subtraction method: Task comparison with a composite measure

Perhaps the most venerable version of the task-comparison method is Donders’ subtraction method (Donders, 1868/1969; Sternberg, 1998b, Appendix 2) for two tasks with measures \( M_1 = RT_1 \) and \( M_2 = RT_2 \). 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 \( A \) (which may consist of more than one modular subprocess).

\( H2 \) (Task Theory 2): Task 2 is accomplished by processes \( A \) 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 \( M_2 = RT_2 \) combine by summation, as in Table 2, and as implied by Donders’ assumption that \( A \) and \( B \) occur sequentially, in stages.

Given these hypotheses, it follows that \( M_1 = RT_1 \) is an estimate \( \hat{T}_A \) of the mean duration of \( A \), \( RT_2 \) is an estimate \( \hat{T}_B \) of the sum of the mean durations of \( A \) and \( B \), and therefore, by subtraction, \( \hat{T}_B = RT_2 - RT_1 \) provides an estimate of the mean duration of \( B \). It is a serious limitation of the subtraction method that it usually embodies a test neither of the combination rule nor of pure insertion. That is, \( H2 \) and \( H3 \) are assumed but not tested.\(^{46,47} \) However, if these hypotheses are correct, then \( \hat{T}_A \) and \( \hat{T}_B \) are pure measures of processes \( A \) and \( B \); the hypotheses can be tested by finding factors that influence the two measures selectively.\(^{48} \) Another way to test the set of hypotheses is to extend it by finding Tasks 3 and 4 that satisfy \( H4 \) (Task 3 is accomplished by \( A \) and \( C \)) and \( H5 \) (Task 4 is accomplished by \( A \), \( B \), and \( C \)), and to extend \( H3 \) by including \( C \). The extended set of hypotheses can then be tested by confirming its prediction that \( \frac{RT_4}{RT_3} = RT_2 - RT_1 \).\(^{49} \)

---

46. Tasks in which the number of iterations of the same process can be controlled, as in some search tasks, provide a special case of the subtraction method in which it is easier to validate the required assumptions. If the numbers of iterations in three variations of the same task are \( n_1, n_2, \) and \( n_3 \), the test is the linearity of \( RT(n_j) \); the slope of the function is an estimate of the duration of the iterated process.

47. Tests of the invariance of the response process across tasks are provided by Ulrich, Matthews, and Miller (1999).

49. This use of two related pairs of tasks is similar to the "cognitive conjunction" method for brain activation experiments introduced by Price and Friston (1997), except that no analogous test is proposed.
An analog of Donders’ method, sometimes called "cognitive subtraction", has frequently been used with brain activation measures (e.g., Petersen, Fox, Posner, Minton, & Raichle, 1988; Cabeza & Nyberg, 1997; Lie, Specht, Marshall, & Fink, 2006). If localization of function were perfect, and if one could design pairs of tasks that were accomplished by processes such as those required by Task 1 and 2 above, differing by just the inclusion of one known subprocess, B, then brain mapping would benefit greatly from the resulting sets of pure measures. Unfortunately, neither of these requirements is typically satisfied. The method has been criticized (Friston, Price, Fletcher, Moore, Frackowiak, & Dolan, 1996; Sartori & Umiltà, 2000), and alternatives proposed (e.g., Friston, et al., 1996; Price, Moore, & Friston, 1997), but the alternatives usually also depend on task comparison with unverified task theories, as has been recognized (Caplan & Moo, 2004; Sartori & Umiltà, 2000). Even some critics of "cognitive subtraction" in brain imaging rely on task comparison with unverified task theories (Jenning, McIntosh, Kapar, Tulving, & Houle, 1997; Sidtis, Strother, Anderson, & Rottenberg, 1999).

9.3 Finding the ‘mechanism of action’ of a manipulation

It was 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 was interpreted in Ex. 4.3. The implication — that sleep deprivation has effects that are process specific — contradicts the conclusion of Dinges and Kribbs (1991, 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, and the finding that performance is impaired by sleep deprivation in a wide range of tasks. If Dinges and Kribbs 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. The problem for the task-comparison method in this application is perhaps the high 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 (Tiplady, Bowness, Stien, & Drummond, 2005).

48. The test would require finding factors F and G such that in a factorial experiment using Task 1 there would be an effect of F but not of G, and in a factorial experiment using Task 2 the effect of F would be equal to its effect in Task 1, and additive with the effect of G. The inclusion of Task 1 in such a study could add to what was learned from Task 2: it would test the pure insertion assumption as well as permitting estimation of $\hat{T}_A$ and $\hat{T}_B$, rather than just of the effects on these quantities. For example, if we define the Smulders et al. (1995) experiment discussed in Section 3.1 and 4.1 as Task 2 (analogous to Donders "b"), with target stimuli for the left and right hand, one could add a Task 1 (analogous to Donders "c") in which the subject would respond with a single keystroke with the right hand if the right-hand target appeared, and make no response otherwise. One test of pure insertion would be to determine whether the effect of SQ on $\overline{RT}$ in Task 1 was equal to its effect on $\overline{RT}$ in Task 2.

50. An alternative approach is exemplified by the use of the same working-memory task with five retention intervals by Haxby, Ungerleider, Horwitz, Rapoport, and Grady (1995), and the use of a different working memory task with four sizes of memory load by Braver, Cohen, Nystrom, Jonides, Smith, and Noll (1997).

51. Electrophysiological evidence that confirms the selectivity of the effect of SLP has been found by Humphrey, Kramer, and Stanny (1994).
The process-decomposition approach might be more helpful. Suppose, for example, that processes $A$ and $B$ in a task are known to be selectively influenced by factors $F$ and $G$, respectively. Now suppose that processes $A$ and $B$ were also selectively influenced by Drugs 1 and 2, respectively. Then, whereas both drugs would affect task performance, the effect of $F$ on that performance would be modulated only by Drug 1 and the effect of $G$ only by Drug 2.

10. The Use of Transcranial Magnetic Stimulation (TMS) to Associate Brain Regions with Mental Processes

Why does sleep deprivation increase RT? An answer to this question was provided by Ex. 4.3, using a task in which the additivity of the effects of two factors ($SQ$ and $MF$) on $RT$ 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 responses by interfering with the stimulus encoding process but not with response selection. Similarly, one can ask why interfering with a particular brain region changes the performance of a task.

In recent years it has been discovered that even 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 (Walsh & Pascual-Leone, 2003). This opens the intriguing possibility of employing TMS within the method of additive factors, just as sleep deprivation was used by Sanders, Wijnen, and Van Arkel (1982). 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 on $RT$ of the other factors are modulated by TMS, it may be 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 $RT$, from which we infer separate stages, $A$ (influenced by $F$) and $B$ (influenced by $G$). Now we add $TMS_R$ (TMS of region $R$) as a third factor, and ask whether it interacts with $F$ or $G$ or neither. 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 $B$, but not of $A$. 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 $B$) and not the others. Unfortunately, there may be as yet no fully satisfactory application of this kind; both of the studies described below have missed

52. An important advantage of TMS over measures of brain activation (Section 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 (Chatterjee, 2005).

53. If the effect of $TMS_R$ is time specific, as is likely with single-pulse TMS, it may not be straightforward to determine with which other factors it interacts. Thus, in the present example, suppose that it is region $R$ that implements process $B$, and that $B$ follows $A$. Because effects on the duration of $A$ influence the starting time of $B$, and hence the time of TMS relative to $B$, a change in the level of $F$ might modulate the effect of $TMS_R$ on $B$, and hence its effect on $RT$. The resulting interaction of $TMS_R$ with $F$ would lead to the erroneous conclusion that region $R$ is involved in the implementation of $A$. This argues for the use of rTMS before the task is performed in such studies, rather than a single TMS pulse during the task.
the opportunity to make such observations, because of inadequate experimental design in Ex. 9.1, and incomplete analysis of data in Ex. 9.2.

10.1 Visual search and TMS

In visual search for a conjunction of features, as in many search tasks, it is usually observed that the $RT$ 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.\(^54\) In search for a single feature, unlike search for a feature conjunction, $RT$ 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 (1997) examined the effect of TMS of the right 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.\(^55\) 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 one or more 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 paper, 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.

In a later experiment where display size was varied (Ellison, Rushworth, & Walsh, 2003), TMS was found to influence the intercept but not the slope in conjunction search, which led some investigators to change their interpretation of the role of the parietal cortex (O’Shea, Muggleton, Cowey, & Walsh, 2006).

10.2 Number comparison and rTMS

In Section 6.3 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

---

\(^{54}\) Which intercept is appropriate depends on details of the search process, and may differ for target-absent and target-present trials.

\(^{55}\) They may have believed that using displays of two or three sizes would require doubling or tripling the size of the experiment.
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 (2001) found that rTMS applied to the left or right angular gyrus influenced the RTs in such a task. 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 6.3 the data were collapsed over the sign of the difference, and the 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 RT to the difference, \(|k - 65|\). To determine whether rTMS 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 its influence on encoding or response processes.

In principle, the data to answer this question were collected in this study. (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 rTMS 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 rTMS modulates the proximity effect, which would indicate an effect on the number comparison process, but the appropriate analysis has not yet been reported.

11. Process Decomposition and the Effects of Brain Damage

Can localized brain damage (i.e. presence vs absence of lesions in particular locations) permit the demonstration of separate modifiability of two or more subprocesses that underlie performance of a task? Consider the hypothesis that a particular task is accomplished by a complex process consisting of modular subprocesses \(A\) and \(B\). Suppose there is localization of function, such that a lesion in region \(R_A\) influences process \(A\), while leaving process \(B\) invariant, and suppose that it does not replace \(A\) with a qualitatively different process. Define a \(R_A\) - lesion factor that has two levels, no lesion and lesion. And suppose there is a factor, \(G^{56}\), that influences process \(B\) but not \(A\). (In a more complex scenario, our hypothesis might include the idea that \(A\) is also selectively influenced by non-lesion factor \(F\). If a lesion in \(R_A\) influences \(A\) but not \(B\), we would expect it to modulate the effect of \(F\) but not \(G\). Given pure measures \(M_A\) and \(M_B\) we would need to show the invariance of \(M_B\) across levels of the \(R_A\) - lesion factor. Given a composite measure \(M_{AB}\) we would need to show that the effect of \(F\) on that measure, but not the effect of \(G\), was modulated by the \(R_A\) - lesion factor.) If the required selective influence could be shown, it would support the hypothesis. However, showing this faces two impediments.

(1) Inadequate precision to demonstrate invariance. Consider first a case where we believe that we have pure measures \(M_A\) and \(M_B\) of \(A\) and \(B\). Suppose the data are sufficiently precise so that the effect of the \(R_A\) - lesion factor on process \(A\) could be persuasively shown, by comparing \(M_A\) in a group of one or more patients with the lesion to a control group without the lesion.

---

56. This factor could, but need not, be the presence or absence of a lesion in a region \(R_B\) within which process \(B\) was carried out.
Unfortunately, it is unlikely that the invariance of process B across levels of the $R_A - lesion$ factor, as indicated by the invariance of $M_B$, could be persuasively demonstrated. The problem is that there is virtually never any pre-lesion measurement,\(^{57}\) which precludes within-subject comparison, and, given the ubiquity of individual differences among "normals", the sample size necessary for persuasive demonstration of equality of $M_B$ between groups of patient and control subjects is likely to be impractically large.\(^ {58}\)

Next consider the case where we have a composite measure $M_{AB}$ of A and B. For example, A and B might be processing stages, and the measure might be mean reaction time, $\bar{RT}$. Assuming that the $R_A - lesion$ factor influenced $\bar{RT}$, we would have to show that this effect was additive with the effect of factor $G$. That is, we would have to show that the effect of $G$ on $\bar{RT}$ was invariant across levels of $R_A - lesion$. As in the case of pure measures, this would almost always require a between-groups comparison.\(^ {59}\)

(2) Likelihood of qualitative change of the process under study. The process-decomposition approach depends on measuring a complex process that is qualitatively the same (i.e., that consists of the same set of subprocesses) under different conditions (i.e., with different levels of factors that influence those subprocesses). The effect of changing the level of a factor cannot be so great as to prevent the process that it influences from working; this would require that if the task is to be accomplished at all, it would require a qualitatively different complex process. Nor can the effect of changing the level of a factor be sufficiently great so as to induce the patient to learn over time to adopt a different "strategy" — a different complex process — from the normal one. But brain damage can sometimes have such effects.\(^ {60}\)

These impediments perhaps justify the use within cognitive neuropsychology of task-comparison (with its more limited goals; see Section 9.1) rather than process-decomposition.

It is possible that TMS may overcome the difficulties described above. First, TMS sometimes appears to be equivalent to a reversible lesion, which enables within-subject comparisons, hence greater precision. Second, the fact that the magnitude of TMS can be

---

57. This problem need not arise when the lesion is produced surgically, as in Farah, Soso, and Dashieff, 1992.

58. There is a sharp distinction between persuasive evidence of invariance and the failure to find a significant effect. In some papers it is concluded that brain damage has no effect merely from the fact that the test of the effect proves not to be statistically significant. Often no confidence intervals or other measures of the effect are provided that would permit one to decide how large an effect would have to be, for significance. One such example can be found in the interesting study of lateral prefrontal damage by Gehring and Knight (2002).

59. If the second factor was a lesion in region $R_B$, and only a composite measure was available, then a test of the hypothesized combination rule would require some patients who had lesions in both regions.

60. Other impediments, well known to the cognitive neuropsychology community are:

(3) The victims of strokes often have widespread cerebrovascular disease. Traumatic head injuries tend to produce widespread minor damage, as well as localized major damage. This may be why damage that appears to be localized seems often to produce at least small effects on many functions.

(4) Even where functionally distinct brain regions are spatially distinct there is no reason to expect that the region of damage due to a stroke (which is determined by the brain’s vascular organization) corresponds, so as to be functionally specific. Indeed, the localized effects of a stroke may be to damage nerve tracts that project to many brain regions.

(5) It may be difficult to find undamaged control subjects with overall levels of performance that are poor enough to be comparable. One approach is to increase the difficulty of the task for these subjects, but such increases may themselves have differential effects on different aspects of performance. Another approach is to select the better performing among the brain damaged subjects, but such selection is also a potential source of bias.

(6) The requirement that despite brain damage the subject can still perform in the task also introduces subject-selection biases with unknown implications.
adjusted may mean that it can be arranged to modulate the process that it influences, rather than preventing it from functioning altogether. And third, because the functional lesion produced by TMS is presumably short-lived (rather than chronic) as well as being reversible, it may be less likely to lead to a change in the "strategy" by which a task is accomplished.

What should be the relation between traditional cognitive neuropsychology (i.e., making inferences about normal cognitive processes from the effects of brain damage) and the process decomposition approach with normals? One goal that they share is the identification of separately modifiable processes of the normal brain. I suggest that each can inform the other. The task theories, modular processes, and processor locations inferred or hypothesized on the basis of effects of brain damage can be tested by using the process-decomposition approach, the modules and task theories tested by using factors that vary the task, the processor locations by using fMRI or TMS. This seems to be close to the view expressed by Gurd and Marshall (2003, p. 194): "Double dissociations help us to make plausible conjectures about the functional architecture of mind, conjectures that can then be further explored with . . . "process-decomposition’ methods . . . ." It also appears to be close to the view expressed by Shallice (2003, p. S148): "Consider cognitive domains in which the functional architecture is poorly understood. Even today this includes nearly all the underpinnings of thought . . . For such poorly understood domains, the use of dissociations was essentially viewed as a discovery procedure, which needs converging evidence other than the dissociation per se in order to support the existence of an isolable system specifically impaired in the relevant patients." And the task theories and modular processes inferred from the process-decomposition approach in normals (along with the inferred processor locations, if process decomposition is augmented with fMRI or TMS) can be used to inform the interpretation of the effects of brain damage.

12. Additional Issues

12.1 Quantitative versus qualitative task changes

As shown by Ex. 6.3, the distinction between process decomposition (with its avoidance of task changes) and task comparison 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, because of the invariance of the effects of proximity on both RT and brain activation, the findings in that example provide evidence for modular processes. In general, qualitative task changes should be avoided because they reduce the likelihood of such invariance, and hence the likelihood of identifying 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 (SM: A.2, A.9).

12.2 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 $T_b$, and such that we are willing to assume that all tasks $T_a$ require 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 \( T_q \). A change in \( \alpha \) would then induce a change in \( \gamma \), so they would not be separately modifiable.

### 12.3 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’ (Teller, 1984) does the answer to this question depend? It would be helpful to have more studies (such as Exs. 3.1, 6.3, 6.4, and 8.1), 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. 6.3 and 6.4, 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. In this respect Ex. 6.4 backfired: although the fMRI data supported the expected modularity of stimulus encoding and response selection, the RT data did not. In general, fMRI measures in some subsets of regions might show modularity (selective effects on pure measures, additive effects on composite measures), but because other regions are involved, the functional processes might not be decomposable.

### 12.4 Differential influence 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 influence more useful than selective influence? It is helpful to be precise about the meaning of differential influence in a way that clarifies its relation to selective influence. Assume that factor levels are assigned so that an increase in level produces an increase in the measure of the corresponding process, consider the case where both measures of interest change in the same direction, and let \( F_j \) and \( G_k \) be numerical factor level strengths of factors with two levels. Let \( M_A(\Delta F) = M_A(F_2) - M_A(F_1) \) represent the effect of a change in factor \( F \) on a measure of process \( A \), etc. Then selective influence can be described as the combination of four properties: \( M_A(\Delta F) > 0, M_B(\Delta G) > 0, M_A(\Delta G) = 0, \) and \( M_B(\Delta F) = 0 \). The first two properties show that factors \( F \) and \( G \) are both potent and that measures \( M_A \) and \( M_B \) are both sensitive. If measures \( M_A \) and \( M_B \) were measures of the same mechanism, the second pair of properties would therefore be impossible. In this terminology, differential influence is satisfied if either \( M_A(\Delta F) > M_A(\Delta G) \) and \( M_B(\Delta F) < M_B(\Delta G) \), or if \( M_A(\Delta F) > M_B(\Delta F) \) and \( M_A(\Delta G) < M_B(\Delta G) \). Suppose that we are willing to assume that the functions relating \( M_A \) and \( M_B \) to factor levels differ by at most additive and multiplicative constants:

\[
M_B(x) = gM_A(x) + h. \tag{12}
\]

Then differences in sensitivity or potency can explain neither of these pairs of properties; \( M_A \) and \( M_B \) must be measures of different mechanisms. However, without that assumption, such pairs of properties could result from differences between measures of the same mechanism. Whether constraint (12) is plausible or valid depends on the nature of \( M_A \) and \( M_B \).

---

61. If one variety of differential influence obtains, one can find factors \( F \) and \( G \) such that both factors influence both processes \( A \) and \( B \), but for \( A \) (or \( B \)) the effect of \( F \) (or \( G \)) is the larger (Kanwisher, et al., 2001). Whether differential or selective influence characterizes modular processors is controversial (Haxby, 2004; Reddy & Kanwisher, 2006).

62. With suitable normalization, these pairs of inequalities are equivalent.
Factorial experiments: Verification and discovery

Despite R. A. Fisher’s (1935) explanation of the advantages of factorial experiments, they are used too seldom with behavioral or brain measurements, and are then usually limited to two factors. Factorial experiments are efficient: there is virtually no loss of precision from adding a second orthogonal factor to a one-factor experiment, or a third factor to a two-factor experiment, for example. With composite measures we have seen that they are essential for determining how effects combine, which is theoretically critical. By including subsidiary factors with those of primary interest (as in "fishing expedition") it is possible to combine discovery with verification. In blocked designs the resulting increase in the number of conditions adds to the difficulty of balancing conditions across levels of practice, but even this issue doesn’t arise in random designs, such as "event-related" fMRI experiments.

When it is possible to define more than two levels of a factor, this should be considered. As well as having other advantages (Section 12.1; SM:A.9), multiple levels permit more powerful tests of interaction, and permit focused tests of monotone interaction, the interaction of most interest. (SM:7,15)

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 (Lennie, 2003). This seems to be consistent with "sparse coding" as well as arguing against the idea of "massive parallelism". Implications of these severe metabolic limitations for the plausibility and possibility of alternative processing architectures, and for the modularity of processors, have to be considered.

13. Conclusion

Considerable evidence has accumulated, some of it documented here, for the existence, within complex mental and neural processes, of modular subprocesses that are separately modifiable and that carry out distinct functions. The identification of such subprocesses is only a first step, of course, in the understanding of the complex processes to which they belong; it needs to be followed by detailed understanding of how each of the modules works, sometimes in the form of quantitative models, and about the relationship of the neural and mental modules. The success thus far across a range of cognitive domains encourages further searching, and raises questions about how best to do so.

Related Discussions

For Hadley’s defense of the existence and plausibility of mental modules against attacks by Fodor (2000), Kosslyn (2001), and Uttal (2001), see Hadley (2003). For other discussion of the properties that Fodor (1983) ascribed to modular processes, see Coltheart (1999) and Jacobs (1997). For discussions of double dissociation of tasks, as in the task-comparison method, a good place to start is with Schmidt and Vorberg (2006). For more extensive discussion of the inferential logic associated with the process decomposition approach, detailed discussion of Ex. 3.2, 4.3, 5, 8.2, and 8.3, and additional examples, both behavioral and neural, see the text and

---

63. For example, suppose that $M_A(x) = x^2$ and $M_B(x) = x$, where $x$ is the level ("strength") of a factor. Let $F_1 = 1.4$, $F_2 = 1.6$, $G_1 = 0.3$, and $G_2 = 0.6$. Then both of the above pairs of properties are satisfied.

64. For example, a $2 \times 2$ design, with four conditions, provides only one df for interaction, whereas a $3 \times 3$ design, with nine conditions, provides four df.
appendices of SM, and references therein. For the method of additive factors and numerous examples of its application, see SM:16, SM:A.16, Roberts and Sternberg (1993), Sternberg (1998a), Sanders (1998), and references therein.

REFERENCES


potentials and the additive-factors method. *Journal of Cognitive Neuroscience, 8,* 47-68.


39, 180-182.


