Ψ Psychology Press
Taylor & Francis Group

Modular processes in mind and brain

Saul Sternberg

University of Pennsylvania, Philadelphia, PA, USA

One approach to understanding a complex process starts with an attempt to divide it into *modules*: sub-processes that are independent in some sense, and have distinct functions. In this paper, I discuss an approach to the modular decomposition of neural and mental processes. Several examples of process decomposition are presented, together with discussion of inferential requirements. Two examples are of well-established and purely behavioural 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 fMRI signal, the latencies of electrophysiological events, and their amplitudes. Some measures are *pure* (reflecting just one process), while others are *composite*. Two of the examples reveal mental and neural modules that correspond. Attempts to associate brain regions with be-haviourally defined processing modules that use a brain manipulation (transcranial magnetic stimulation, TMS) are promising but incomplete. 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 describe the erroneous conclusion to which it has led about sleep deprivation, and the interpretive difficulties in a TMS study.

TABLE OF CONTENTS

Downloaded by [173.59.66.157] at 17:52 17 April 2012

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 distinct functions.[1] Early in the last century, scientific psychology, dominated by behaviourism, 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 the input, transformation, storage, retrieval, and comparison of internal representations—modules whose arrangement was expressed in systematic flow charts.[2] Comparing the brain to a digital computer encourages the distinction between processors and the processes they 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: PDP 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.

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 behavioural measures. Such modules will be denoted **A**, **B**, etc. A neural-process module is a part of a neural process, functionally

---

[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), who argued (p. 485) 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."

distinct from other parts, and investigated with brain measures. Such modules will be denoted α, β, etc.

The important distinction between processes and processors is sometimes overlooked. *Functional* decomposition leads to process*es* that occur over time; their arrangement is described by a flow-chart. In contrast, *structural* decomposition leads to process*ors* (e.g., as described by Erickson, 2001; Goodale, 1996; Op de Beeck, Haushofer, & Kanwisher, 2008) that are parts of a physical or biological device (such as the brain); their arrangement can sometimes be described by a circuit diagram.[3]

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.[4] 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 have the same function they are likely to be influenced by the same factors.[5] With separate modifiability as the criterion, it should be clear that modularity is a relation between or among processes, not an absolute property of one process.

2.2. Processes and their measures, pure and composite, and combination rules

How do we demonstrate that a process is influenced by a factor, or invariant with respect to it? We know only about one or more hypothesized *measures* M_A of process **A**, not about the process as such. Depending on the available measures, there are two ways to assess separate modifiability of **A** and **B**.

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), 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.

---

[3] Machamer, Darden, and Craver (2000) distinguish "activities" and "entities".

[4] 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.

[5] Such double dissociation of subprocesses should be distinguished (Sternberg, 2003) from the more familiar double dissociation of tasks (Schmidt & Vorberg, 2006), discussed in Section 9.

Table 1. *Inferential logic for pure measures*

Joint Hypothesis
H1: Processes **A** and **B** are modules (separately modifiable). *H2*: M_A, M_B are pure measures of **A, B**.

Prediction
We may be able to find factors *F* and *G* that influence M_A (**A**) and M_B(**B**) selectively: $p_1{:}M_A \leftarrow F, \; p_2{:}M_B \nleftarrow F, \; p_3{:}M_B \leftarrow G, \; p_4{:}M_A \nleftarrow G.$

Alternative Results	
We find factors *F*, *G* that influence M_A and M_B selectively.	We fail to find such factors.

Corresponding Inferences	
Support for joint hypothesis *H1 + H2*.	Refutes one/both of *H1, H2, or* we didn't look enough for *F, G*.

If F_j has two levels, $j = 1, 2$, the *effect* of *F* on M_A is a difference:

$$effect(F) = M_A(F_2) - M_A(F_1). \qquad (1)$$

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.[6,7] The $\{p_k\}$ are four properties of the data; $M_A \leftarrow F$ should be read as "M_A is influenced by *F*"; $M_B \nleftarrow F$ 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.[8]

Composite measures. 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 possible

---

[6] Adapted from Table 2 of Sternberg (2001) by permission.

[7] 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 Sternberg (2001, Section A.2.3).

[8] 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. One 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. Bayesian methods (e.g., Gallistel, 2009; Rouder, Speckman, Sun, Morey, & Iverson, 2009) provide another alternative, especially if the null is framed as an appropriate interval hypothesis rather than a point hypothesis. An example of suitable caution about inferring a null effect can be found in Ghorashi, Enns, Klein, and Di Lollo (2010).

Table 2. *Inferential logic for a composite measure with summation as the combination rule*

Joint Hypothesis
H1: Processes **A** and **B** are modules (separately modifiable). *H3:* Contributions u_A, v_B of **A**, **B** to M_{AB} (**A**, **B**) combine by *summation*.

Prediction
We may be able to find factors *F* and *G* that influence **A** and **B** selectively: $p_1': u_A \leftarrow F, \ p_2': v_B \nleftarrow F, \ p_3': v_B \leftarrow G, \ p_4': u_A \nleftarrow G,$ and jointly influence no other process. If so, their effects on M_{AB} will be *additive*.

Alternative Results	
We find factors *F* and *G* with *additive effects* on M_{AB}.	We fail to find such factors.

Corresponding Inferences	
Support for joint hypothesis *H1 + H3*.	Refutes one/both of *H1*, *H3*, *or* we didn't look enough for *F*, *G*.

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 such experiments are rare, despite their efficiency (Section 12.5).

The logic for inferring separate modules using a hypothesized composite measure when we either know or hypothesize that the combination rule is *summation* is shown in Table 2.[9,10] To understand Table 2, it is important to keep in mind what the effect of a factor is, and what it means for effects of different factors to be *additive*. To simplify the discussion, let us assume that there are two factors, each with just two levels. Let *u* and *v* be the contributions of processes **A** and **B** to M_{AB}. If summation is the combination rule, $M_{AB} = u + v$. If **A** and **B** are selectively influenced by factors *F* and *G*,

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

where $u(F_j)$ is a function that describes the relation between the level of *F* and the contribution of **A** to M_{AB}.

---

[9] Adapted from Table 3 of Sternberg (2001) by permission.

[10] 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.

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 $\overline{M}$, $\overline{u}$, and $\overline{v}$ to indicate the means of these random variables, it is convenient that with no further assumptions,[11] Eq. 2 implies:

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

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:

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

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

$$\begin{aligned}
effect&(F | G = G_k) \\
&\equiv M_{AB}(F_2, G_k) - M_{AB}(F_1, G_k) \\
&= [u(F_2) + v(G_k)] - [u(F_1) + v(G_k)] \\
&= [u(F_2) - u(F_1)], \qquad (5)
\end{aligned}$$

regardless of G_k. 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.[12,13]

2.3. Overview of examples and issues

Much of what follows consists of descriptions of successful examples of the process-decomposition approach. Examples will be referred to by the section numbers in which they are first discussed. With two exceptions, the successful examples involve factorial experiments with two factors. The exceptions are Ex. 3.1, in which the effects of the two 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. Three examples (10.2, 10.3, and 10.4) are intriguing and tantalizing cases where the approach could have been used but was not, because of incompleteness of design or 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 encoding two different aspects of the stimulus, they are found to operate in parallel. In Ex. 3.2, in which the modules are associated with preparation of two different aspects of the response, they are found to operate successively, as processing "stages". The RT data from Ex. 3.2 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 of Ex. 3.2 exemplifies the method of additive factors (AFM), discussed more generally in Section 4.2, and applied

---

[11] Matters may not be so simple for other combination rules, such as multiplication; see Section 8.2.

[12] The reasoning described in Table 2 is sometimes erroneously expressed as "If we assume $H3$ then additivity confers support on $H1$." This ignores the support that additivity also confers on $H3$.

[13] It is important to note that whereas factors that selectively influence serially arranged processes will have additive effects on mean RT, this is not the only possible basis for additive effects on an interesting measure, despite beliefs to the contrary (e.g., Poldrack, 2010, p. 148; Jennings, McIntosh, Kapur, Tulving, & Houle, 1997, p. 237). The critical requirement for additivity is combination by summation, whatever its basis.

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 the 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 two or more 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 (BOLD) 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 is important. 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. In these regions, the fMRI signal appears to be a composite measure.

Examples 6.5 and 8.1 use the fMRI adaptation method; in both cases the measure, whose magnitude increases with the dissimilarity of the current stimulus to 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 modular processes are responsible for immediate and delayed adaptation effects in the parahippocampal place area, which seems to require different short-term and long-term memory representations.

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 fMRI measures. It is tempting to infer that modular brain processes are responsible. In Section 7 I consider for both RT (7.1) and fMRI (7.2) the conditions under which inferences from additivity to modularity are justified, and describe inferential requirements and limitations that are sometimes overlooked.

In Section 8 I provide brief summaries of four additional examples of the process-decomposition approach, using different composite measures, to show the diversity of applications. Ex. 8.1 uses the fMRI adaptation method 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, and the measure is the proportion of errors. An outline of the inferential logic is included. In Ex. 8.3, the composite measure with a multiplicative combination rule is the proportion of correct responses; the application is to speeded lexical decisions. Ex. 8.4 uses the amplitude of the event-related potential at multiple scalp locations as a set of related composite measures, powerful because of its fine temporal resolution and multiple locations.

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 various factors, together with

theories about the processes used to accomplish those tasks. Examples of factors that have been used are the presence or absence of sleep deprivation and the presence or absence of a lesion in a particular brain region. To illustrate limitations of the method, I describe a comparison of the effects of repetitive transcranial magnetic stimulation (rTMS) on two tasks of tactile perception (Section 9.1). One variety of task comparison is Donders' subtraction method, originally developed for RT experiments, but, in recent years, applied to brain-activation measures (Section 9.2). 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), 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 claims, based on task comparison, that the effect is global.

In Section 10, I consider experiments in which the effects of transcranial magnetic stimulation (TMS) on RTs in visual search and number-comparison judgements were measured. In these experiments, TMS of a brain region R 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, TMS_R. If this factor influences $\overline{RT}$, region R is inferred to play a role in performance of the task. However, the potential of TMS to associate brain regions with modular subprocesses and to provide evidence about separate modifiability is realized only when TMS_R is used together with other, process-specific factors, to determine which of their effects are modulated by TMS, and which are not. While promising with respect to these goals, the studies discussed are incomplete.

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 brain damage in different tasks. Can brain damage also be used in the process-decomposition approach? This question is considered in Section 11; several considerations lead to a negative answer. Instead, the two approaches can inform each other.

Several additional issues are considered briefly in Section 12.

In Sternberg (2001), referred to as "SM:<Section>", I discuss and defend the process-decomposition approach in more detail, further discuss its inferential logic, describe its antecedents, provide more detail about Exs. 3.1, 3.2, 4.3, 5, 8.2, and 8.4, describe other examples using different species and different behavioural and neural measures, discuss the treatment of data, and consider issues of experimental design.

3. DECOMPOSING NEURAL PROCESSES WITH THE LATERALIZED READINESS POTENTIAL

3.1. Parallel modules for discriminating two stimulus features

Consider a trial in a choice-reaction experiment where two alternative responses are made by the two hands. 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.[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

---

[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 reduced 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.

zero, but is defined so 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.

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 *response selection* process by $\boldsymbol{\alpha}$. 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 stimulus category (the Go versus NoGo signal) is discriminated and response preparation ceases. Let us denote this *category discrimination* process by $\boldsymbol{\beta}$.

Events 1 and 2 indicate the completion of processes $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. 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. (1992) examined the effects of two factors: One (in Exp. 1) is *Go-NoGo Discriminability*, *GND*, which should influence $\boldsymbol{\beta}$; it could be *easy* (letter and digit with dissimilar shapes, GND_1) or *hard* (similar shapes, GND_2). The other factor (in Exp. 2) is the spatial compatibility of the stimulus-response mapping, *MC* (*Mapping Compatibility*), which should influence $\boldsymbol{\alpha}$; it could be *compatible* (respond with the hand on the same side as the stimulus, MC_1) or *incompatible* (respond with the hand on the opposite side, MC_2).

Idealizations of the resulting $A_{mc}(t)$ functions are shown in Figure 1. 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 LRPs) is the time at which their *difference* reliably exceeds zero. In Exp. 1, *GND* influenced $\overline{T}_2$ (by 43 ms) but not $\overline{T}_1$ (compare Figures 1A1 and 1A2), showing that *GND* influenced $\boldsymbol{\beta}$ but not $\boldsymbol{\alpha}$. ($\overline{T}_1 \approx 170$ ms on both Go and NoGo trials.)[15] Because the stimulus influenced response preparation before both of its features were discriminated, these findings from Exp. 1 demonstrate the transmission of "partial information" from the perceptual process to the response process.

In one of the conditions of Exp. 2, the stimulus-response 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.[16] Osman et al. therefore reduced letter-digit shape discriminability so as to increase $\overline{T}_2$ from about 240 ms to about 350 ms. In this experiment, *MC*

---

[15] The effect of *GND* on $\overline{T}_1$ was 2.5 ± 5.0 ms; its effect on $\overline{T}_2$ was 43 ± 14 ms; the difference between these effects is 41 ± 11 ms ($N = 6$; $p \approx .01$).

[16] It is this requirement that would have made it difficult to implement a suitable factorial experiment.
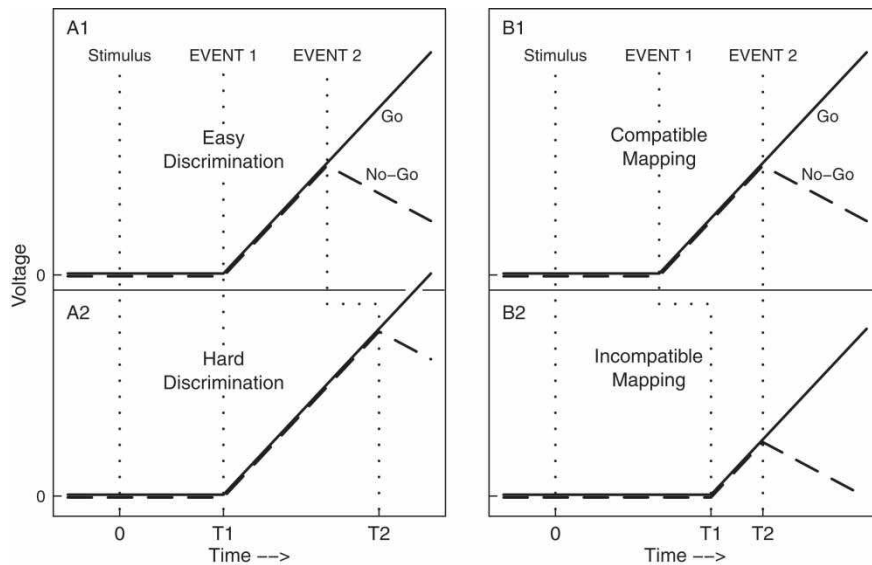
Figure 1. *Schematic idealized asymmetry functions* $A_{mc}(t)$ *from Osman, Bashore, Coles, Donchin, and Meyer (1992). Event 1 is the LRP onset; Event 2 is the onset of the divergence of* $A_{mc}(t; Go)$ *from* $A_{mc}(t; NoGo)$. *Panels A1 and A2: Asymmetry functions from Exp. 1, in which* Go-NoGo Discriminability *could be easy* (GND = GND$_1$) *or hard* (GND = GND$_2$). *Panels B1 and B2: Asymmetry functions from Exp. 2, in which* Mapping Compatibility *could be compatible* (MC = MC$_1$) *or incompatible* (MC = MC$_2$). *(Fig. 3 of Sternberg, 2001; adapted by permission.)*

influenced $\overline{T}_1$ (by 121 ms) but not $\overline{T}_2$ (compare Figures 1B1 and 1B2), showing that *MC* influenced $\boldsymbol{\alpha}$ but not $\boldsymbol{\beta}$. ($\overline{T}_2 \approx 350$ ms on both Go and NoGo trials.)[17] 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 $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are arranged in time. Suppose they were arranged sequentially, as stages. Prolonging the first of two stages by Δt ms should delay completion of the second by the same amount: the prolongation Δt 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.[18] The finding (Exp. 2; Figure 1B) 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; Figure 1A) for the effect of *GND*. An alternative to a stages arrangement is that $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ 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 $\overline{T}_1$ (121 ms). This can happen because regardless of how much the duration of $\boldsymbol{\alpha}$ is shortened by changing the level of *MC*, response initiation on a Go trial must await the completion of $\boldsymbol{\beta}$ as well.

---

[17] The effect of *MC* on $\overline{T}_1$ was 121 ± 17 ms; its effect on $\overline{T}_2$ was 3.3 ± 8.8 ms; the difference between these effects is 118 ± 21 ms ($N = 6$; $p \approx .01$).

[18] T_2 would then be a composite measure, influenced by both factors, with summation as the combination rule.

These findings about $\overline{RT}$ contrast to those of Ex. 3.2, below, 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 behavioural 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 behavioural measure, and whose contribution to $\overline{RT}$ may be large or small, depending on the level of *GND*.

3.2. Serial modules (stages) for preparing two response features

The LRP provides an estimate of the time when the side of the response is selected; this 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 from the stimulus, and using LRP$_s$), and T_{rjk}^* (a negative quantity, measured from the response, and using LRP$_r$), and let $T_{rjk} = RT_{jk} + T_{rjk}^*$. T_{sjk} and T_{rjk} are then alternative estimates of the same time point between stimulus and response, measured from the stimulus; let $T_{\cdot jk}$ be their mean.[19] Averaging over the four conditions, $RT_{..} = 416$ ms and $T_{...} = 264$ ms. If $\boldsymbol{\alpha}$ is the neural process from stimulus to LRP onset, and $\boldsymbol{\beta}$ is the neural process from LRP onset to response, these values permit us to estimate their durations (pure measures of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$) averaged over the four conditions: $D_{\alpha..} = T_{...} = 264$ ms and $D_{\beta..} = RT_{..} - T_{...} = 152$ ms.

Shown in the A and B panels of Figure 2 are the estimated durations of processes $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ separated by condition: $D_{\alpha jk} = T_{\cdot jk}$, and $D_{\beta jk} = RT_{jk} - T_{\cdot jk}$. Because T_{sjk} and T_{rjk} give similar estimates for effects of the two factors on D_α and D_β, the estimates are based on $T_{\cdot jk}$. The results show that the two factors SQ and RC have selective effects on D_α and D_β. This supports the hypothesis that in this situation the LRP onset indeed defines a temporal boundary between two neural modules that function sequentially, as stages, consistent with the reasoning in Table 2.[20]

What are stages? They are functionally distinct operations that occur during nonoverlapping epochs.[21] In a process with two stages, the stream of operations between stimulus and

---

[19] When a mean is taken over values of a subscript, that subscript is replaced by a dot.

[20] 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 communication, September, 1999).

[21] It is perhaps a confusion between time (process) and space (processor) that has led some commentators (e.g., Broadbent, 1984) to believe that a process whose modules are organized in stages cannot include feedback because it must be implemented by a "pipeline": an ordered set of processors through which information passes in a fixed direction from input to output. Broadbent's "pipelines" constrain the relation between process and representation: later processes must operate on representations that have been processed more highly—that are "further from the input". Stage models need not be constrained in this way; they merely partition processing operations into temporally successive components. There is no reason why a later stage cannot make use of new sensory information (such as feedback) in (re)processing earlier sensory information. For further discussion of Broadbent's (1984) critique of stage models, and the distinction among three kinds of stage (completion-controlled, outcome-contingent, and data-dependent), see Sternberg (1984).

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 behavioural data, the cut is hypothetical, inferred from the additivity of factor effects. 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. Note that by examining the effects of other factors, **A** (or **B**) might be further decomposed into stages or non-stage modules, only one of which needs to be influenced by F (or G).

In contrast to D_α and D_β, which are hypothesized (and confirmed) to be pure measures, RT_{jk}, shown in Panel C, is a composite measure: With two stages contributing to the time from stimulus to 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 α and β inferred from the LRP data.

4. MENTAL PROCESSING STAGES INFERRED FROM REACTION TIMES

4.1. Analysis of the reaction-time data in Ex. 3.2

The conclusion in Ex. 3.2 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_α, influenced by SQ but not RC), and from LRP to response (duration D_β, influenced by RC but not SQ). Consider just the mean reaction time, measured under the four factor-level combinations:

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

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 α and β is summation. As discussed in Section 2.2, that and the selective influence of factors SQ and RC on α and β imply that the effects of SQ and RC on $\overline{RT}$ are additive. In general, factors that selectively influence the durations of distinct sequential components of the RT must have additive effects on the composite measure. Thus, if a hypothesis asserts that the RT in a particular task is the summed duration of two modular mental processes **A** and **B** arranged sequentially (as stages) with durations T_A and T_B, and selectively influenced by factors F and G, we should expect:

$$\overline{RT}(F_j, G_k) = \overline{T}_A(F_j) + \overline{T}_B(G_k). \quad (7)$$

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

Thus, suppose that the mental modules **A** and **B** responsible for the additive RT effects are implemented by the neural modules α and β, demarcated by the LRP. Then, not only should the same factors influence them selectively, but also the sizes of their effects should be the same. Agreement among the effect sizes can be examined by assuming that the two factors indeed have perfectly selective effects on α and β, and by using the appropriate subsets of the LRP data to "predict" the pattern of the $\{\overline{RT}_{jk}\}$. Thus, we should be able to use just the data in Figure 2A1 (averaging over RC levels) to obtain
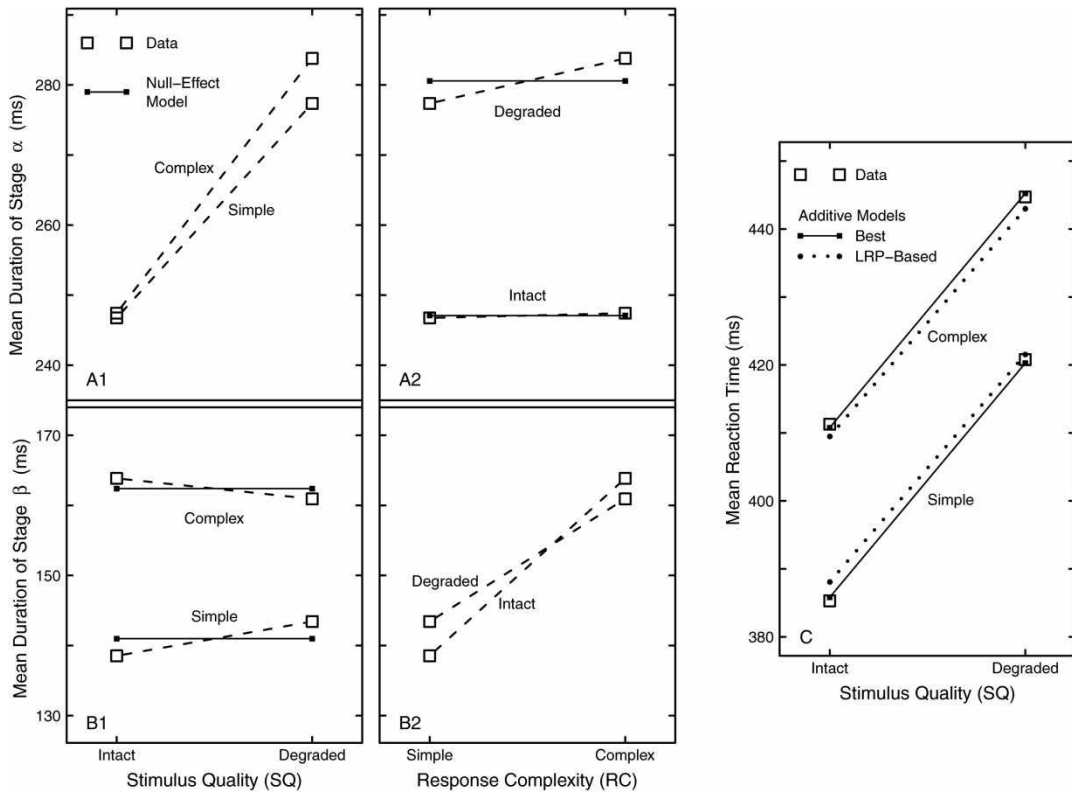
Figure 2. *Means over 14 subjects of data from Smulders, Kok, Kenemans, and Bashore (1995). The panels in each pair, A1–A2 and B1–B2, show the same values, plotted differently. Estimated duration $D_{\alpha jk}$ of Stage α, from stimulus presentation to LRP onset (Panels A1, A2); and duration $D_{\beta jk}$ of Stage β, from LRP onset to response (Panels B1, B2). These are shown as functions of Stimulus Quality (SQ_j, Panels A1, B1); and of Response Complexity (RC_k, Panels A2, B2). Data in Panels A1 and B1 are separated by level of RC; those in Panels A2 and B2 are separated by level of SQ. Also shown in Panels B1 and A2 are null-effect models. Main effects of SQ on D_α and D_β (with $\pm$ SE) are 33 ± 6 ms (Panel A1) and 1 ± 7 ms (Panel B1); main effects of RC on D_α and D_β are 4 ± 7 ms (Panel A2) and 21 ± 6 ms (Panel B2). The interaction contrasts are 6 ± 13 ms (stage α) and -8 ± 11 ms (stage β). The $\overline{RT}$ data (discussed in Section 4.1) 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). For the $\overline{RT}$s, the normalized interaction contrast is 7.0% [For a dimensionless normalized measure of the interaction in 2 × 2 data, I express the interaction contrast as a percentage of the geometric mean of the absolute values of the main effects. This facilitates comparison of the interaction magnitudes associated with different measures. See Roberts & Sternberg (1993), Table 26.1]. (Fig. 15 of Sternberg, 2001; adapted by permission.)*

the estimates $\hat{D}_\alpha(SQ_1)$ and $\hat{D}_\alpha(SQ_2)$. Similarly, we should be able to use just the data in Figure 2B2 (averaging over SQ levels) to obtain the estimates $\hat{D}_\beta(RC_1)$ and $\hat{D}_\beta(RC_2)$. If $\overline{RT} = D_\alpha + D_\beta$ we have the "predictions" $RT^*_{jk} = \hat{D}_\alpha(SQ_j) + \hat{D}_\beta(RC_k)$ for the four conditions.[22] The dotted lines in Figure 1C show that the agreement is good: $\overline{RT}_{jk} \approx RT^*_{jk}$. Numerically, the effects of SQ and RC on the composite measure $\overline{RT}$ are 35 ± 3 ms and 25 ± 7 ms, close to their mean estimated effects (34 and 21 ms) on the pure measures D_α and D_β.

Another illustration of the independent use of two measures is provided by Ex. 6.3: again,

---

[22] This way of deriving the $\{RT^*_{jk}\}$ forces their means into agreement: $RT^*_{..} = \overline{RT}_{..}$; the question of interest is whether the differences among the four values agree.

pure measures to ask about the structure of a neural process, and a composite measure to ask about the structure of a corresponding mental process.

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 the effects of factors F and G on $\overline{RT}$ in a process organized in stages interact, so that F modulates the effect of G rather than leaving it invariant, then F and G must influence at least one stage in common. Suppose 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 selectively influencing a different subprocess.[23] 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. See Section 7.1 for a discussion of the strength of such inferences. Example 4.3 illustrates how it is possible to reveal more complex processing structures by combining inferences from a pattern of additive and interacting factor effects.

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.

This example is especially powerful because it includes more than two factors. The stimuli were the single digits "2", "3", "4", and "5"; the responses were their spoken names, "two", "three", "four", and "five". Four factors were manipulated, each at two levels: The first was *Stimulus Quality* (*SQ*); the digits, presented as dot patterns, could be *intact*, or *degraded* by adding "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 Figure 3 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 influence two different stages of processing selectively, stages that might be described as stimulus encoding (**S**) and response selection (**R**).[24]

---

[23] Given a constraint on the durations of different stages that is stronger than zero correlation but weaker than stochastic independence, the assumption of stages plus selective influence implies numerous properties of aspects of the RT distributions in addition to their means (Sternberg, 1969; Roberts & Sternberg, 1993), such as additive effects on var(RT). However, without this constraint, stages plus selective influence don't require effects on var(RT) to be additive.

[24] The conclusion that *SQ* and *MF* influence 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).
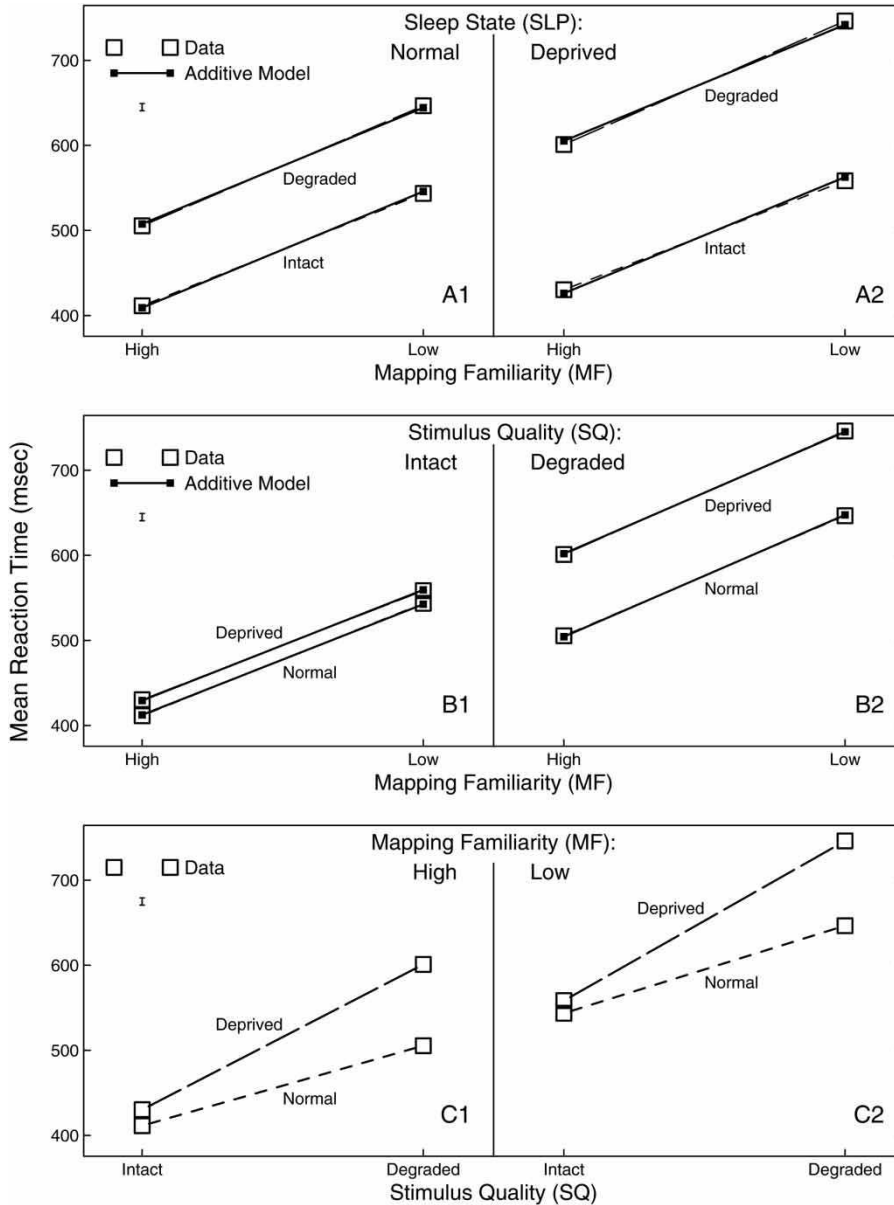
Figure 3. *Data from Sanders, Wijnen, and Van Arkel (1982, Experiment 1.) Means over the two levels of* Time of Day *(TD). The three pairs of panels show the same* $2 \times 2 \times 2 = 8$ *data points (unfilled squares, broken lines), plotted differently. Each point is the mean of about 300 RTs from each of 16 subjects. A fitted additive model is also shown in each of the top four panels. Mean absolute deviations of data from model are 3.3 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. The normalized interaction contrasts (see Figure 2 caption) for the six panels are A1: 8%, A2: 11%, B1: 9%; B2: 3%; C1: 88%; C2: 92%. (Fig. 14 of Sternberg, 2001; reprinted by permission.)*

The results in Figure 3 consist of the $\overline{RT}$s from the $2 \times 2 \times 2 = 8$ conditions. Panels A1 and A2 show that at each level of SLP there are additive effects of SQ and MF on $\overline{RT}$. 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 $\overline{RT}$ of MF and SLP. That is, the extra time a subject takes to execute an unfamiliar S-R association rather than a well-learned one is invariant over sleep states, rather than being increased by sleep deprivation. This evidence lends further support to (1) and also supports:

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

Panels C1 and C2 show that at each level of MF there are interactive effects of SQ and SLP: Increasing the level of SLP has a far greater effect on $\overline{RT}$ when the stimulus is degraded (98 ms) than when it is intact (17 ms). That is, sleep deprivation modulates the effect of the difficulty of stimulus encoding. This evidence supports

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

Taken together, the three pieces of evidence support a theory according to which the process used to perform the task contains at least two modules, **S** and **R**, these modules are arranged as stages, and among the factors SQ, MF, and SLP, SQ and SLP influence **S**, while MF alone influences **R**. It is reasonable to suppose that the stimulus is identified during **S**, and the response selected during **R**. (This is suggested by the nature of the factors SQ and MF that influence them.) The AFM has thus led us to the surprising conclusion that whereas SLP influences stimulus encoding, it does not influence response selection.

Unlike some other applications of composite measures, the findings from this experiment not only demonstrate separate modifiability and thereby permit us to divide the processes between stimulus and response into two modules

(here, stages) **S** and **R** (the former selectively influenced by SQ, the latter by MF); they 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**. And the additivity of the effects of SLP and SQ is of course further evidence for the separate modifiability of **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, & Birdsall, 1961; Macmillan & Creelman, 2004). At the heart of this approach is the recognition that even simple psychophysical tasks involve decision processes as well as sensory processes. 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 response proportions arranged in a 2×2 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 $\mathbf{S}$, and proportional to $\overline{X}_T - \overline{X}_{NT}$, which increases with discriminability. The other is an estimate of the criterion, presumed to be a pure measure of the decision process $\mathbf{D}$.

Many factors have been used in attempts to influence $\mathbf{S}$ and $\mathbf{D}$, 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, 2×2 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 $\mathbf{D}$. Unfortunately, selective influence by PM of $\mathbf{D}$ has not been shown; $\mathbf{D}$ also appears to be 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 $Pr\{R_{NT}|Reward\}$, 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 centre key of three keys; these two trial types were equally frequent. The correct response was to peck the left key (R_T) if the centre key was "bright," and to peck the right key (R_{NT}) if it was "dim". Correct responses were reinforced with food, with a mean probability of about 0.37, controlled by the VI schedules. Two factors were varied orthogonally: The *luminance ratio* (LR) of the two lights was varied by letting the dimmer luminance be one of five values, including, for the most difficult LR level, a value equal to the brighter luminance. The *reinforcement ratio* (RR), described by $Pr\{R_{NT}|Reward\}$, could be one of three values, 0.2, 0.5, or 0.8.[25] 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 0.8 to 0.5 to 0.2, is expected to be linear with unit slope. 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

---

[25] If $Pr\{R_{NT}|Reward\} = 0.2$, for example, for each rewarded R_{NT} response there are four rewarded R_T responses, encouraging a liberal (low) criterion for R_T.
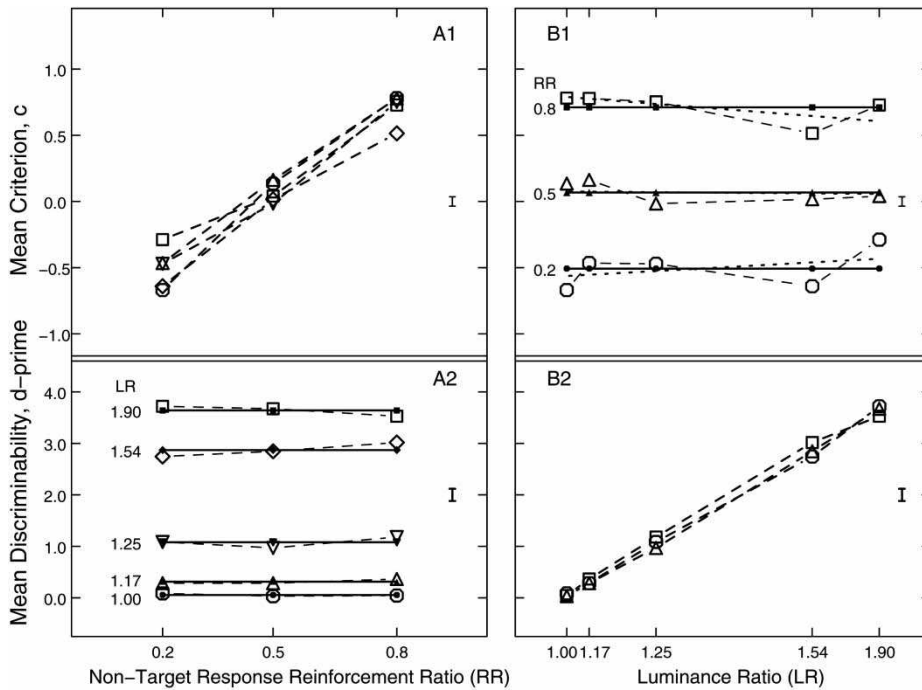
Figure 4. *Mean effects of* Reinforcement Ratio, RR = Pr{R_NT|Reward} *(Panels A1, A2) and* Luminance Ratio, LR *(Panels B1, B2) on criterion* ĉ *(Panels A1, B1) and on discriminability* d̂′ *(Panels A2, B2) are shown by unfilled points and broken lines. RR and LR levels have been scaled so as to linearize their mean effects on ĉ-value and d̂′-value (Panels A1 and B2). Filled points and unbroken 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 (see SM:7.2, SM: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.)*

and criterion measures for each condition are

$$\hat{d}' = z(Pr\{R_T|S_T\}) - z(Pr\{R_T|S_{NT}\}), \text{ and}$$

$$\hat{c} = -[z(Pr\{R_T|S_T\}) + z(Pr\{R_T|S_{NT}\})]/2.$$

The origin for the criterion measure is the midpoint between $\overline{X}_{NT}$ and $\overline{X}_T$; the sign of the criterion thus expresses the direction of the bias. Means over birds of these two measures are shown in Figure 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**). The complementary 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*.

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 α and β are implemented by different processors P_α and P_β in non-overlapping brain regions R_α and R_β. The search for such processes has sometimes been stimulated by behavioural 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 such searches. Then, because process α should be influenced selectively by F, the activation level of region R_α, a pure measure, should vary with F, but not with G, and conversely for region R_β.[26] 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 behavioural data. If such regions were found, it would support the modular decomposition inferred from the behavioural 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 inferred from RT data that suggested a new experiment that would incorporate fMRI measurements to search for corresponding neural processes. And in both experiments, concurrent RT data were taken, along with the fMRI data. While the RT data in Ex. 6.3 confirmed earlier findings, results in Ex. 6.4 did not, probably because of paradigm differences, which raises 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 $P_{\alpha\beta}$) in *one* region, $R_{\alpha\beta}$, then the activation level in $R_{\alpha\beta}$ 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 might 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, as well as in the two brain regions in which it could be tested in Ex. 6.4. Some of the evidence that bears on the combination rule for fMRI signals is discussed 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

---

[26] 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.
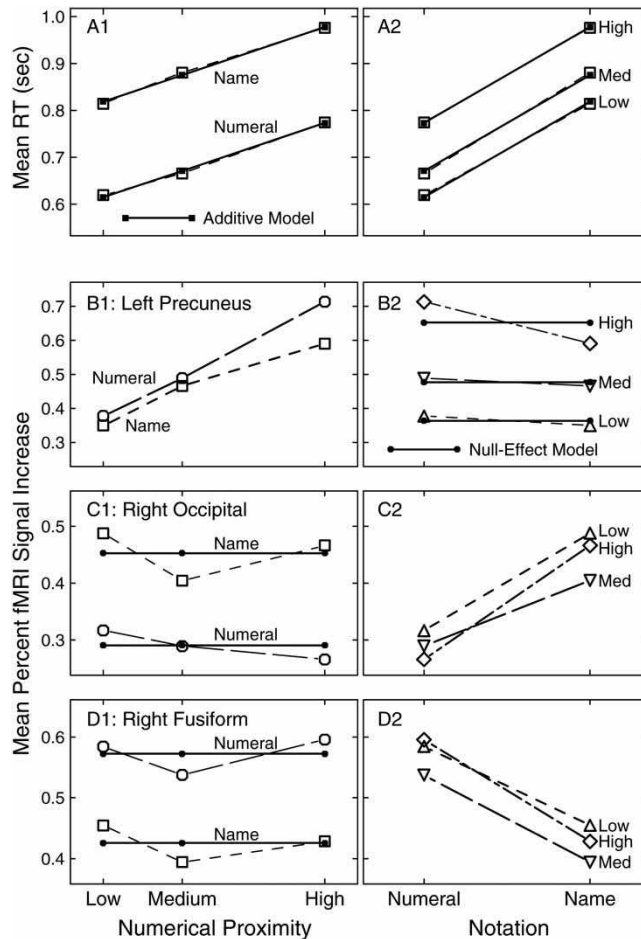
Figure 5. *Reaction-time and selected brain-activation data from Pinel, Dehaene, Rivière, and LeBihan (2001). The same data are plotted on the left as functions of* P *(Proximity, where "high" means closer), with* N *(Notation) the parameter, and on the right as functions of* N, *with* P *the parameter. Means over nine subjects of median RTs for correct responses are shown in Panels A, with a fitted additive model. The three levels of* P *have been scaled to linearize the main effect of* P *on* $\overline{RT}$; *this effect, from low to high* P, *is* 159 ± 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.) The normalized interaction contrast (see Figure 2 caption), based on the four corners of the* 3×2 *design, is 4%. Mean activation measures from three sample brain regions, relative to an intertrial baseline, are shown in Panels B, C, and D, accompanied by fitted null-effect models in Panels B2, C1, and D1. Shown in Panels B1, C1, and D1, the main effects of* P *(from low to high, using fitted linear functions) are* $0.29 \pm 0.09\%$ *(*$p \approx .01$*),* $-0.03 \pm 0.03\%$, *and* $0.00 \pm 0.04\%$. *Shown in Panels B2, C2, and D2, the main effects of* N *are* $-0.06 \pm 0.06\%$, $-0.16 \pm 0.05\%$ *(*$p \approx .01$*), and* $-0.15 \pm 0.05\%$ *(*$p \approx .02$*). (Fig. 6.2 of Sternberg, 2004; reprinted by permission.)*

difference $|k - 65|$, and grouped into three levels. The interesting phenomenon here is the "symbolic distance effect" (Moyer & Landauer, 1967): 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 $\overline{RT}$; this was interpreted to indicate two modular subprocesses arranged as stages: encoding (**E**), influenced by N, which determines the identity of the stimulus and is

slower for number names than numbers, and comparison (**C**), influenced by P, which uses the stimulus identity in performing the comparison and is slower for closer proximities. In the new experiment, in which fMRI as well as RT measurements were taken, most of the sixteen brain regions examined whose activation was influenced by N or by P were influenced significantly by only one of them, consistent with two separately modifiable neural processes ϵ 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 $0.17 \pm 0.05\%$ (median p-value $= .01$), while the P effect was $0.06 \pm 0.08\%$; for the seven P-sensitive regions the P effect was $0.32 \pm 0.10\%$ (median p-value $= .01$), while the N effect was $0.04 \pm 0.04\%$.

The fMRI data from three well-behaved regions are shown in Figure 5, Panels B, C, and D. The concurrently collected RT data (Figure 5, Panels A) replicated the earlier study, suggesting that we associate the neural modules ϵ and γ with the mental modules **E** and **C**; it is important that the mental and neural modules were selectively influenced by the same factors. However, while the direction of the effect of P was the same in all the brain regions it influenced, the direction of the effect of N was not: the change from numeric to verbal notation (which increased $\overline{RT}$) increased activation in some regions (e.g., Figure 5, Panels C) and decreased it in others (e.g., Figure 5, Panels D).[27,28]

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 $\overline{RT}$, 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 Shwartz, 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, starting with 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 RT and, concurrently, measured fMRI in six brain regions.[29] 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

---

[27] Without requiring it, this finding invites us to consider that there are two qualitatively different encoding processes ε, 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 12.1), but evidence for modularity emerges nonetheless: the proximity effect is invariant across the two tasks. Based on the idea that the processes implemented by different processors are probably different, the (multidimensional) activation data from such a simple (two-factor) experiment can support a claim of operations replacement. In contrast, an RT experiment that alone could support such a claim has yet to be devised.

[28] Using a dual-task experiment, Sigman and Dehaene (2005) have added to the evidence that distinguishes **E** from **C** : **E** could occur concurrently with all stages of the initial task, whereas **C** had to await completion of the "central" stage of the initial task.

[29] Extrastriate cortex was expected to respond to SQ; previous studies had implicated the remaining five regions (see Figure 6) in response selection.
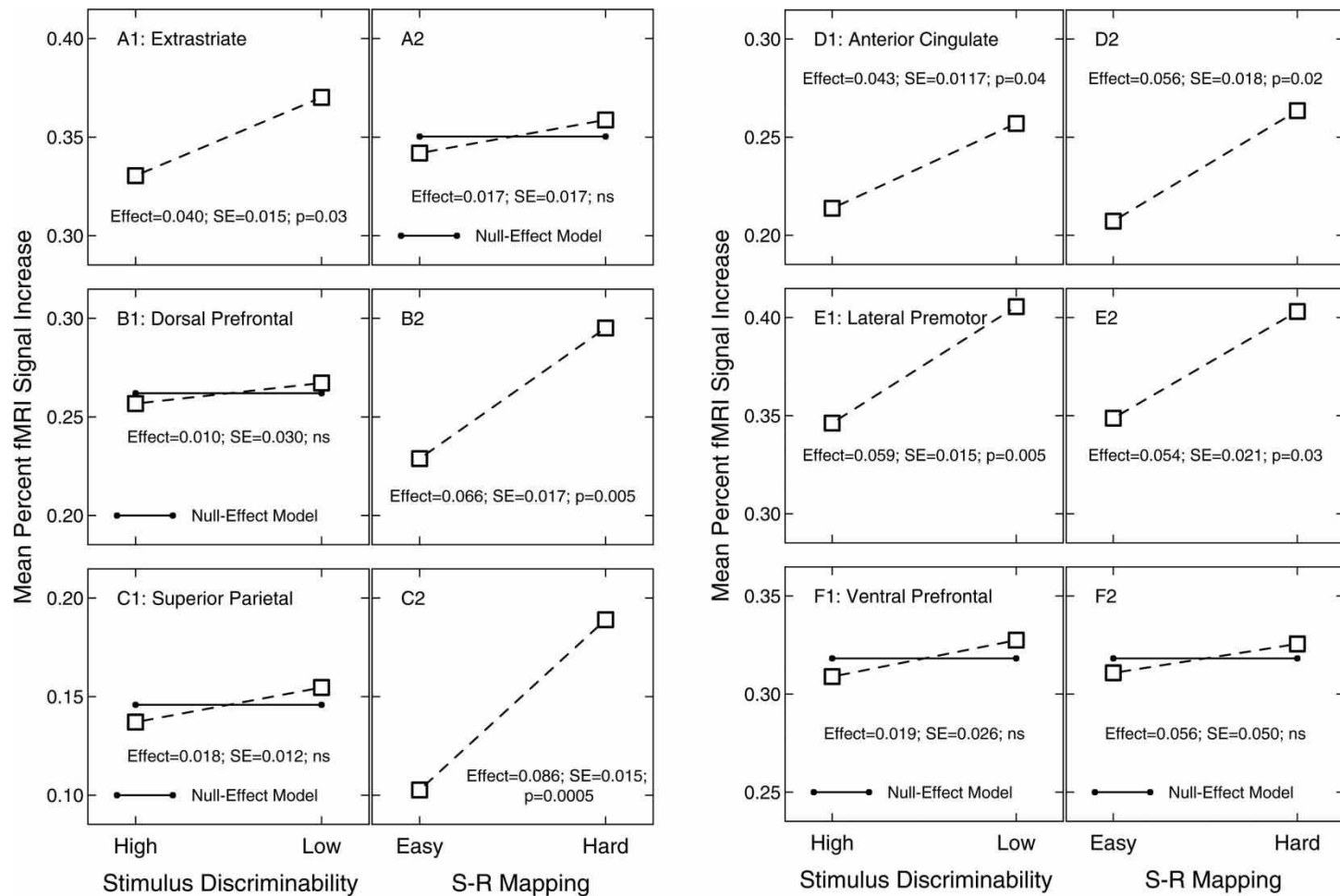
Figure 6. *The six pairs of panels of Figure 6 provide brain activation data relative to a fixation baseline for the six regions measured by Schumacher and D'Esposito (2002). Mean main effects over nine subjects (eight for anterior cingulate) of* Stimulus Discriminability *and* S-R Mapping *in each region, with null effect models shown for each non-significant effect.*

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. 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 $\overline{RT}$—also produced the larger fMRI signal. Figure 6 shows that in one of the regions, only *SQ* had a reliable effect (Panels A), in two of the regions only *MC* had reliable effects (Panels B and C), in two regions both factors had reliable effects (Panels D and E), and in one region neither factor had a reliable effect (Panels F). The selective effects found in three of the regions (where the fMRI signal was influenced by one of the factors but not the other) are consistent with **S** and **R** being implemented, at least in part, by anatomically distinct populations of neurons.[30]

The additivity found in earlier RT experiments suggests that there is 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 Figure 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; 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 **S** and **R**.

In contrast to the earlier findings with various experimental arrangements mentioned above, the effects of the two factors on $\overline{RT}$ in the Schumacher-D'Esposito task were unfortunately not additive; as shown in Figure 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.[31,32]

---

[30] 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?

[31] 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. However, RT data from the practice session, outside the scanner, showed a non-significant interaction of about the same size and in the same direction. It is also of interest that a whole-brain analysis of the fMRI data did not reveal any additional task-sensitive regions (E. Schumacher, personal communication, November 27, 2006).

[32] 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 $\overline{RT}$. The large SEs associated with the very small mean interaction contrasts for the data shown in panels 7A and 7B 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. Indeed, relative to the variability, the reported mean interaction contrasts were significantly ($p < .05$) too small.
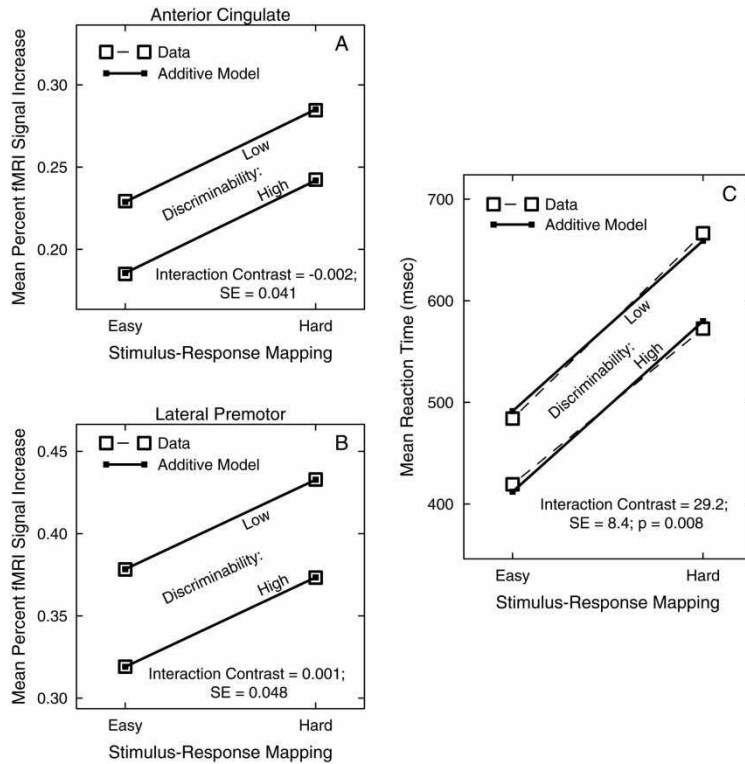
Figure 7. *Panels A, B show mean simple effects of the two factors on the brain activation measure, in the two regions where both main effects were significant in Schumacher and D'Esposito (2002). Also shown are fitted additive models, and the interaction contrasts that measure the badness of fit of these models. Panel C shows simple effects of the two factors on* $\overline{RT}$*, and the corresponding interaction contrast. The normalized interaction contrasts (see Figure 2 caption) are A: 4%; B: 1%; C: 25%. (The mean effect of* Discriminability *on* $\overline{RT}$ *when the* Mapping *was easy versus hard is 64 versus 94 ms, respectively.)*

6.5. Modular short-term and long-term memory processes in scene perception

Consider the process used to 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.

In Series 2, two scenes were presented successively on each trial, 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 the first on trials in Series 2, and the *long-interval relation* between the second scene and scenes that had been displayed during Series 1.

fMRI Adaptation (also called "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.[33]) Consistent with expectation, the fMRI signal in the PPA as well as the RT varied with both short-interval and long-interval relations. Answering the question whether short- and long-interval fMRI adaptation in the PPA are generated by a single process or by two separate processes therefore requires determining how their effects on the fMRI signal 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.[34]

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.[35] Thus there are two short-interval factors (*Sameness or Difference of View* 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 Figure 8.[36] All four of the main effects are significant, with $p < .01$. 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 $p < .01$. In addition

---

[33] The relation between fMRI adaptation and neuronal activity is controversial; see, e.g., Sawamura, Orban, and Vogels (2006), and Grill-Spector (2006).

[34] 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(s) implemented by the PPA, that are influenced by both factors. Or it could happen if the processes that contribute to the RT are selectively influenced by the two factors, but are 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.

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

[36] These are means of the BOLD signal strengths over the PPA regions in the two hemispheres.
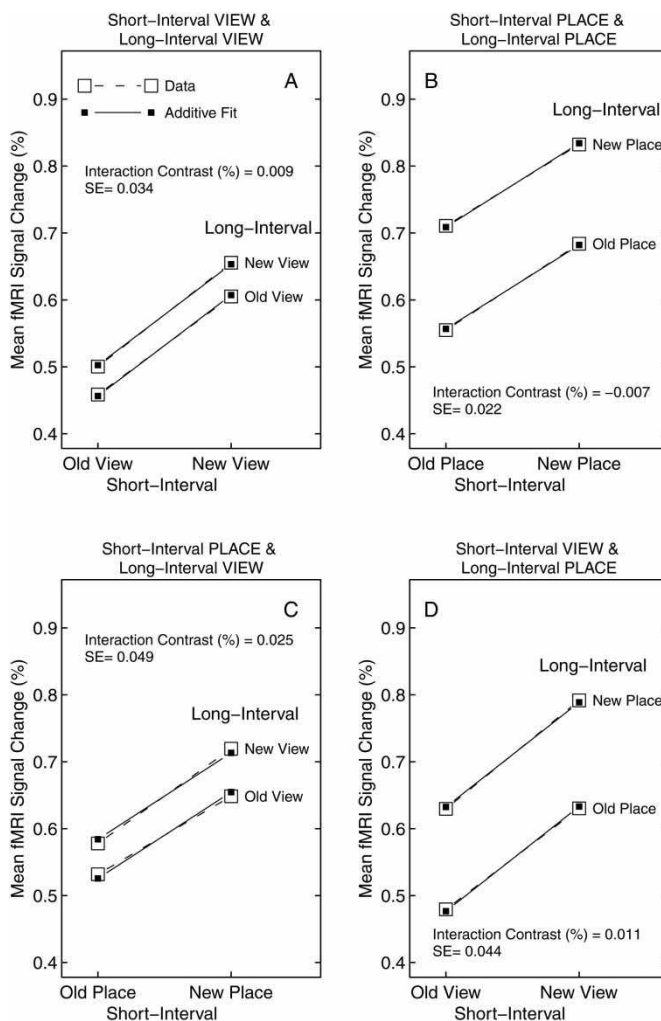
Figure 8. *Combined effects on the fMRI signal in the parahippocampal place area of two kinds of short–interval and long–interval novelty, from Epstein, Parker, and Feiler (2008). The normalized interaction contrasts (see Figure 2 caption) are A: 11%; B: 5%; C: 29%; D: 7%.*

to its significance level attesting to the remarkably high precision of these data, this three-way inter-action supports the idea that the long-interval adapt-ing effect is more "viewpoint invariant" than the short-interval effect.[37]

Means of the RT data are shown in Figure 9. There is one substantial interaction (in Panel D), 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

---

[37] In a whole-brain analysis, reliable long- and short-interval effects were found in many other brain regions. There was no per-suasive 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 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.

as in the fMRI data) and 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 whether the differences between short- and long-interval effects of view and place changes would alone argue that the short- and long-interval effects are generated by different processes. (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 7.2), 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 appears 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- and long-term memory representations.

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 $\overline{RT}$, 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 $\overline{RT}$ to the organization in *stages* of the corresponding processes. To help distinguish among such modular alternatives one can sometimes use other aspects of the RT data (e.g., Roberts & Sternberg, 1993) or brain measurements. Examples of the use of brain measurements for this purpose include neurophysiological data (see Schall, 2003, and references therein, and Woodman, Kang, Thompson, & Schall, 2008) and the pure stage duration measures in the electrophysiological data of Ex. 3.2. As for non-modular theories, I know of no plausible way
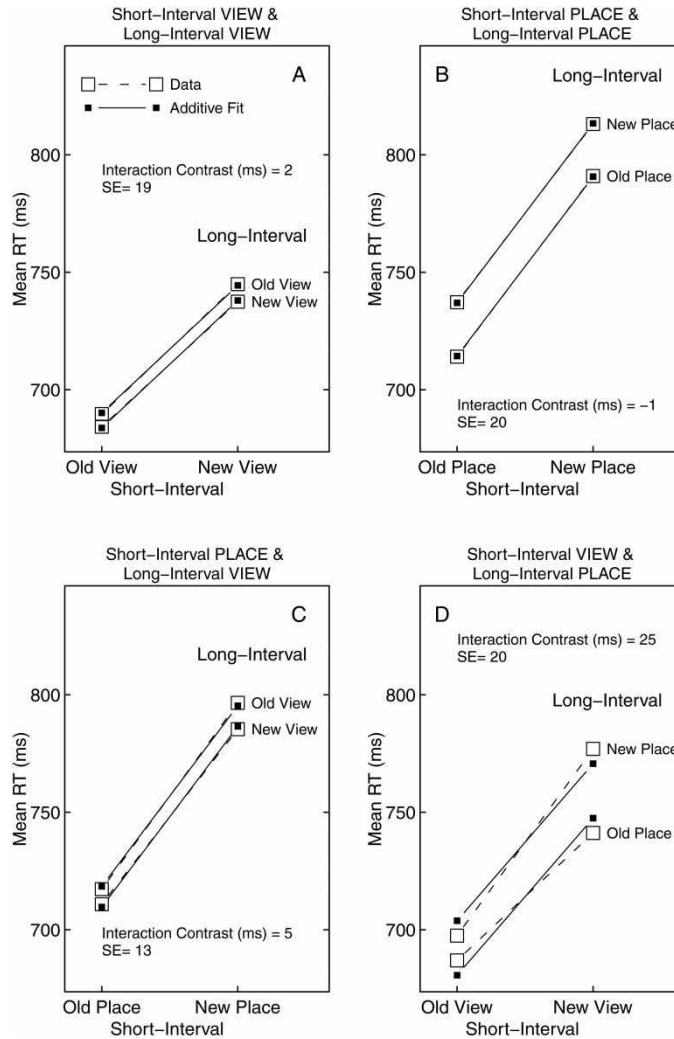
Figure 9. *Combined effects on $\overline{\text{RT}}$ of two kinds of short-interval and long-interval novelty, from Epstein, Parker, and Feiler (2008). The normalized interaction contrasts (see Figure 2 caption) are A: 11%; B: 3%; C: 18%; D: 64%.*

in which what is truly a single process can give rise to additive effects of two factors on mean RT.[38] This could of course change, as we learn more about mental and neural processes.

7.2. Inferences from fMRI measurements

Suppose that we observe that the effects of two factors on the fMRI signal in a brain region are additive. It is tempting to conclude that they

---

[38] In the popular 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 $\overline{RT}$ that are overadditive.

influence different neural processes in that region. To do so, two propositions should be true:

Proposition 1: Summation is the combination rule for their contributions to the BOLD signal (*B*) of different processes in that region, either concurrent and implemented by different neural populations, or successive. (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.

Proposition 2: Factors that influence the *same* neural process do *not* have additive effects on the amount of neural activity. (If they did, and *B* satisfied the linearity requirement, then additive effects could be produced by factors that influence the same neural process.) This leads to a requirement of *non*linearity for neural activity.

To see why these propositions are critical, consider what is required to correctly conclude in favour of modular processes in a region, and how we could be misled.

Different neural populations implement modular processes. Suppose first that α and β, modular processes selectively influenced by *F* and *G*, 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* depends on both N_α and N_β, but possibly to different degrees. Assume further that *B* is a *linear* function of N_α and N_β: $B(N_\alpha, N_\beta) = B_0 + aN_\alpha + bN_\beta$, where B_0 is a baseline level, and *a* and *b* are nonnegative constants.[39] We can then write:

$$B_{jk} = B(F_j, G_k) = B_0 + aN_\alpha(F_j) + bN_\beta(G_k). \quad (8)$$

Because changes in *F* influence only the second term, and changes in *G* influence only the third term, the effects of F_j and G_k on B_{jk} must be additive, whatever the form of the functions $N_\alpha()$ and $N_\beta()$. The inferential logic for this case is described by Table 2, and, as described in that table, the

observation of additivity, as in Exs. 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, which in turn depends on *B* being linear in the level of neural activity.

One neural population implements successive modular processes. Because the fMRI signal reflects neural activity that occurs over a period of seconds, if processes α and β, selectively influenced by *F* and *G*, are implemented 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,[40] even if temporal summation were imperfect (e.g., Dale & Buckner, 1997), Eq. 8 would apply; the failure of perfect temporal summation would be reflected only in the parameters *a* and *b*.

Additive effects from a single process. As summarized in Table 2, the observation of additive effects of two factors *F* and *G* supports the joint hypothesis consisting of *H*1 and *H*3. An alternative explanation for such a finding is that there is a single process (α) influenced by both factors that operates in such a way that their effects on $M_\alpha(F, G)$ are additive. In evaluating the plausibility of this alternative we need to consider what would be required of the 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 2×2 factorial design, and a 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,

---

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

[40] 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 $\overline{RT}$ data (Figure 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, even these duration changes are small relative to the sluggishness of *B*.

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 $bq + cu$, $bq + cu + cv$, $bq + br + cu$, and $bq + br + cu + cv$, 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. 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*:

$$M(F, G) = M(tFLS) = M(bs_f + cs_g)$$
$$= a + b's_f + c's_g. \qquad (9)$$

To the extent that such linearity is plausible or valid, it prevents us from confidently inferring modular processes from additive effects on M. Applying this reasoning to fMRI, and assuming that B_α is linear in N_α, B_α is linear in *tFLS* if (and only if) N_α 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 implausible or invalid does the finding of additivity of factor effects on B support the hypothesis of modular processes.

In short, to justify the inferences in Exs. 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) (Proposition 1) and nonlinearity of N (Proposition 2). The neural basis of the BOLD response is controversial (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 $B(N)$, or the neural response 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) concluded that whereas some of the nonlinearity in higher visual areas is neural, some may be in $B(N)$.

Evidence, including that mentioned above, indicates that in response to contrast, duration, number of stimulus repetitions, and number of trials, the magnitude of the neural response N changes nonlinearly (Albrecht, Geisler, & Crane, 2003; Carandini & Heeger, 1994; Geisler & Albrecht, 1995; Heeger et al., 2000; Li, Miller, & Desimone, 1993; Meeter, Myers, & Gluck, 2005; Sawamura, Organ, & Vogels, 2006). This reduces the plausibility of a single neural process producing additive effects by responding linearly to *tFLS*. If we therefore assume that $N(tFLS)$ is nonlinear, we can conclude from additive effects of two factors on B in a brain region (as in Exs. 6.4, 6.5, and 8.1) that they influence different modules selectively.[41] Such instances of additive

---

[41] This argument is stronger for Ex. 6.4, where pure measures of each factor were found in three regions: This makes it more likely that the finding of effects of both factors in two other regions should be explained by those regions containing two populations of neurons, each responsive to one factor.

effects on B then provide evidence of the linearity of $B(N)$. However, taken together, the available evidence seems to permit us to treat the linearity of $B(N)$ only as a tentative working hypothesis.

8.   BRIEF ACCOUNTS OF FOUR ADDITIONAL EXAMPLES WITH COMPOSITE MEASURES

8.1.   Modular neural processes for perceptually separable dimensions

We considered an application of fMRI adaptation in Section 6.5; another example is a study by Drucker, Kerr, and Aguirre (2009). Suppose a set of visual stimuli that vary along two dimensions that are *neurally separable*, in the sense that they are evaluated by different modular neural processes (possibly implemented by different populations of neurons). Let S_1 and S_2 be successive stimuli, and let F_j and G_k be their proximities along the two dimensions. Then, given Eq. 8, the adapting effect of S_1 on the fMRI response to S_2 will be the sum of the adapting effects on the two processes, which in turn will depend additively on F_j and G_k. In an experiment to test this idea, Drucker et al. used two different sets of outline shapes, each varying over several levels of each of two dimensions. For one stimulus set the two dimensions were expected to be neurally separable because they are perceptually separable in the sense of Garner (1974);[42] 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.[43]

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 permit, 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 about the combination of effects in Figure 8 as a single number.) One appeal of this study is that the same experimental and analytic methods and the same subjects revealed contrasting fMRI effects, additive and interactive, for the two sets of stimuli. Another appeal is that the results support an attractive neural theory of perceptual integrality and separability.

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

Consider 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 behaviour of the two processes is uncorrelated, then the probability of neither process responding 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; this relationship is sometimes called "probability summation".

To describe the consequences of a multiplicative combination rule for a composite measure, it is helpful to introduce the idea of a proportional effect, or *p.effect*. 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, 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*

---

[42] For two perceptually separable (integral) dimensions, variation in one does not (does) interfere with making decisions based on the other, and perceptual distances obey a city-block (Euclidean) metric. See also, e.g., Ashby and Maddox (1994).

[43] While viewing each sequence of shapes, subjects reported on the position of a bisecting line.

Table 3. *Inferential logic for a composite measure with multiplication as the combination rule*

Joint Hypothesis
H1: Processes **A** and **B** are modules (separately modifiable).
H4: Contributions u_A, v_B of **A**, **B** to $M_{AB}(\mathbf{A}, \mathbf{B})$ combine by *multiplication*.
H5: Contributions of **A** and **B** to M_{AB} are uncorrelated.

Prediction
We may be able to find factors F and G that influence **A** and **B** selectively:
$p'_1{:}u_A \leftarrow F,\ \ p'_2{:}v_B \nleftarrow F,\ \ p'_3{:}v_B \leftarrow G,\ \ p'_4{:}u_A \nleftarrow G,$
and jointly influence no other process.
If so, their *proportional effects* on M_{AB} will be *multiplicative*.

Alternative Results	
We find factors F and G with *multiplicative p.effects* on M_{AB}.	We fail to find such factors.

Corresponding Inferences	
Support for joint hypothesis $H1 + H4 + H5$.	Refutes one/more of *H1, H4, H5,* *or* we didn't look enough for *F, G*.

(for a factor with two levels):

$$p.effect(F) = \frac{M_A(F_2)}{M_A(F_1)}, \qquad (10)$$

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

$$M_{AB}(F_j, G_k) = u(F_j) \times v(G_k). \qquad (11)$$

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.[44] In that case, it follows from Eq. 11 that

$$\overline{M}_{AB}(F_j, G_k) = \bar{u}(F_j) \times \bar{v}(G_k), \qquad (12)$$

and, by analogy to Eq. 4, that

$$p.effect(F, G) = p.effect(F) \times p.effect(G). \qquad (13)$$

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.[45]

---

[44] This is a weaker requirement than stochastic independence, but may nonetheless be important.
[45] Adapted from Table 4 of Sternberg (2001) by permission.

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 two widely separated frequencies that comprised a compound grating whose presence subjects had to detect. They found that reducing the contrast of one of the frequencies in the compound caused $Pr\{$non-detect$\}$ to increase 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 analyzers for separated spatial frequencies and for the multiplicative combination rule.[46] One reason why this experiment was especially persuasive is that whereas one of the two gratings was either present or absent, the contrast of the other grating was varied over more than two levels.

More formally, let S_1 and S_2 represent the behaviour of the two modular subprocesses, here detection processes, with $S_j = 1$ indicating success (detection) and $S_j = 0$ otherwise, and assume that S_1 and S_2 are independent. In the present case, the combined process is successful (a "yes" response) if *either* of its subprocesses is successful. Under these conditions, it is the probability of *nonsuccess* ("nondetect") that is related to the subprocesses by a multiplicative combination rule:

$$Pr\{nonsuccess\} = Pr\{S_1 = 0 \; and \; S_2 = 0\}$$
$$= Pr\{S_1 = 0\} \times Pr\{S_2 = 0\}. \quad (14)$$

Shaw (1980) showed that Eq. 14 was a consequence of the attention-sharing version of an independent decisions model, and Mulligan and Shaw (1980) found that Eq. 13 (and neither of two alternative models) described the behaviour of three of their four observers detecting simultaneous auditory and visual signals in a 2×2

experiment in which each signal was either present or absent.

8.3. Modules inferred from multiplicative effects on the accuracy of lexical decisions

In contrast, suppose a situation in which the combined process is successful if and only if *both* of its independent subprocesses are successful. (This could occur if the output of the first is the input of the second.) It is then the probability of *success* that is related to the subprocesses by a multiplicative combination rule:

$$Pr\{success\} = Pr\{S_1 = 1 \; and \; S_2 = 1\}$$
$$= Pr\{S_1 = 1\} \times Pr\{S_2 = 1\}. \quad (15)$$

Equations 14 and 15 generalize, in the obvious way, to complex processes in which more than two (mutually independent) subprocesses are manipulated. One note of caution: If effects are *additive* within each of several subsets of the data (such as data from different subjects), they will also be additive in the mean data. (Additivity is preserved by the arithmetic mean.) However, if effects are *multiplicative* within data subsets, and are averaged using ordinary means (rather than geometric means, or arithmetic means after a logarithmic transformation), effects in the mean data may deviate from being multiplicative. (Multiplicative relationship may not be preserved by the arithmetic mean.) Pooling data from different subjects or from different levels of practice is also likely to induce correlation. Hence, when effects might be multiplicative, the units of analysis should be homogeneous subsets of the data, and experiments should be run so as to provide stability.[47]

In an important paper, Schweickert (1985) showed that effects of factors on $Pr\{correct\}$ in three diverse experiments satisfy a multiplicative combination rule. For example, one of the data sets (Schuberth, Spoehr, & Lane, 1981, Exp. 2) are the accuracies of the lexical decision process in a

---

[46] See SM:13 and SM:A.13 for more details.

[47] Numerical experiments show that under some plausible conditions, Eq. 12 is well approximated even when the contributions u and v to M_{AB} are highly correlated.

speeded task with orthogonal variation of three two-level factors: legibility of the test word, frequency of the word in English, and semantic congruity of the word with a sentence context. Schweickert found that a multiplicative combination rule is consistent with the effects of these factors on $Pr\{correct\}$, supporting theories in which the three factors influence three different mutually independent modular processes, all of which must succeed for a correct response. Because the error rate in these data never exceeded 8%, a multiplicative model for $Pr\{correct\}$ is virtually indistinguishable from an additive model for $Pr\{correct\}$, as Schweickert noted (which is equivalent to an additive model for $Pr\{error\}$); the mean absolute difference between fitted values of percent correct for multiplicative and additive models is a negligible 0.02%. This is a problem for inference only if a plausible additive model is available, which at this writing it appears not to be. For precise tests of models for $Pr\{correct\}$, more data are needed than in the Schuberth et al. experiment. For these data, with 672 observations per condition, the SEs of the mean $Pr\{correct\}$ for the eight conditions are large (0.3% to 1.0%), given the small main effects of 1.6%, 2.3%, and 2.9%.

8.4.   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.[48] Hence, unlike most other cases, the combination rule is not a part of the hypothesis that must be tested. Suppose there are two modular neural processes $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, implemented by processors P_α and P_β, and influenced selectively by factors F and G. It follows that the effects of F and G on the ERP amplitude will be additive at all scalp locations. Furthermore, if P_α and P_β are at different locations in the brain, the topographies of the effects of F

and G (the way the effect sizes vary with location on the scalp) will differ. Because of the simplicity of the hypothesis and the richness of the possible tests, as well as the fine temporal resolution of the ERP, this approach is especially powerful.

Kounios (2007) exploited these properties 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:

*H*1 (Subprocesses): The complex process of recognizing the probe as a non-target contains (at least) two subprocesses, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, carried out by different neural processors, P_α and P_β.

*H*2 (Selective Influence): $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are selectively influenced by *SAT* and *REL*, respectively.

*H*3 (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: The mean main effects of *REL* and *SAT* were $1.3 \pm 0.2\,\mu\text{V}$ and $2.1 \pm 0.4\,\mu\text{V}$, respectively, while the mean interaction contrast was $0.01 \pm 0.3\,\mu\text{V}$, making the normalized interaction contrast (see Figure 2 caption) 0.6% ($n = 36$). This finding supports the above theory and hence the modularity of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ during the 600 to 800 ms epoch.[49] Also, the topographies of the two effects (their relative sizes across locations) differ markedly, indicating different locations in the brain for P_α and P_β,

---

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

[49] 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 *H*3.

which adds to the evidence against a single-process explanation.[50]

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. 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 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.[51]

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$_s$ of the contralateral somatosensory cortex ($1 - \bar{b}_{rs} = 0.21 \pm 0.07$; $p = .02$), 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$_o$ ($1 - \bar{b}_{do} = 0.16 \pm 0.07$; $p = .04$), but negligibly by rTMS$_s$ ($1 - \bar{b}_{ds} = 0.05 \pm 0.04$), a double dissociation of the two tasks.[52]

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

---

[50] In this application, modularity appears to change over time: During an earlier epoch (400 to 600 ms after probe onset) the two effects interacted substantially, while their topographies changed little from one epoch to the next. See SM:14 for more details.

[51] See SM:A.1.

[52] 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 significantly greater than zero, with $p = .01$ and $p = .04$, 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 negligible effects of practice on those values.

that one or more of these processes in the two tasks depend on different regions of the cortex. In addition to specifying these four processes, a weak pair of task theories might also assert that γ_d and γ_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 α, β, or δ is sensitive to either rTMS$_s$ or rTMS$_o$; this in turn would suggest that they are implemented by processors in neither of the stimulated regions. Unfortunately, the findings do not bear on the validity of such hypothesized task theories, weak or strong, or even on the question whether the operations in either task can be decomposed into modular subprocesses such as α, β, γ, and δ; 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 $M_1 = \overline{RT}_1$ and $M_2 = \overline{RT}_2$. The joint hypothesis consists, first, of a pair of task theories that specify the constituent processes of each task, and second, a combination rule:

H1 (Task Theory 1): Task 1 is accomplished by process **A** (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 $M_2 = \overline{RT}_2$ combine by *summation*, as in Table 2, and as implied by Donders' assumption that **A** and **B** occur sequentially, as stages.

Given these hypotheses, it follows that $M_1 = \overline{RT}_1$ is an estimate $\hat{T}_A$ of the mean duration of **A**, $M_2 = \overline{RT}_2$ is an estimate $\hat{T}_A + \hat{T}_B$ of the sum of the mean durations of **A** and **B**, and therefore, by subtraction, $\hat{T}_B = \overline{RT}_2 - \overline{RT}_1$ provides an estimate of the mean duration of **B**. It is a serious limitation of the subtraction method that it usually embodies a test neither of the combination rule nor of pure insertion. That is, *H2* and *H3* are assumed but not tested.[53,54] However, if these hypotheses are correct, then $\hat{T}_A$ and $\hat{T}_B$ are pure measures of processes **A** and **B**. One way to test the hypotheses is to find factors that influence the two measures selectively.[55] Another way to test them is to extend them by finding Tasks 3 and 4 that satisfy *H4* (Task 3 is accomplished by **A** and **C**)

---

[53] 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 $\overline{RT}(n_j)$; the slope of the function is an estimate of the duration of the iterated process.

[54] Tests of the invariance of the response process across tasks are provided by Ulrich, Mattes, and Miller (1999).

[55] 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 be 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 3.2 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.

and *H*5 (Task 4 is accomplished by **A**, **B**, and **C**), and to include **C** in *H*3. The extended set of hypotheses can then be tested by confirming its prediction that $\overline{RT}_4 - \overline{RT}_3 = \overline{RT}_2 - \overline{RT}_1$.[56]

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; Poldrack, 2010). 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 & Friston, 1997; 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 (Jennings, McIntosh, Kapur, Tulving, & Houle, 1997; Sidtis, Strother, Anderson, & Rottenberg, 1999).[57]

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 conclusions of others. For example, according to Dinges and Kribbs (1991, p. 117), there is "a generalized effect of sleepiness on all cognitive functioning", and according to Balkin, Rupp, Picchioni, and Wesensten (2008, p. 654), sleep loss "impairs some essential capacity that is basic to cognitive performance in general". (See also Lim and Dinges, 2010.) They based their conclusion not on the process-decomposition approach (which would lead to asking how sleep deprivation modulates the effects of process-specific factors), but on the finding, from dozens of studies using task comparison, that sleep deprivation impairs performance in a wide range of tasks. If they 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.[58] 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** for many classes **A** and **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

---

[56] 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 they appear to have proposed no analogous test.

[57] An alternative approach (sometimes called "parametric design") is exemplified by variation of attentional load over six levels by Culham, Cavanagh, and Kanwisher (2001), 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).

[58] Electrophysiological evidence that confirms the selectivity of the effect of *SLP* has been found by Humphrey, Kramer, and Stanny (1994).

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). 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. TRANSCRANIAL MAGNETIC STIMULATION (TMS) AND THE SEARCH FOR MODULES

10.1. An ideal experiment

It was recently 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). Repetitive TMS (rTMS) either before or during performance of a task has produced similar results. This opens the intriguing possibility of employing TMS within the method of additive factors, just as sleep deprivation was used in Ex. 4.3.[59] By using TMS of a region R (TMS_R) in a factorial experiment, while varying

the levels of other factors that are believed to influence different processing stages selectively, and determining which effects of the other factors on $\overline{RT}$ are modulated by TMS_R, it may be possible to learn whether R is involved in the implementation of any of those processing stages. For example, suppose there is a task in which we find that factors F and G have additive effects on $\overline{RT}$, from which we infer separate stages, **A** (influenced by F) and **B** (influenced by G). Now we add TMS_R as a third factor, and ask whether it interacts with F or G or neither. In the ideal results of such an experiment, TMS_R would (1) have an effect on RT, (2) interact with (modulate the effect of) one of the other factors, say, G, and (3) not interact with (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**. The selectivity of the effect would strengthen the inferences made without TMS about the existence of modular processes **A** and **B**, and would also associate region R with one process (**B**) and not the other.[60]

At this writing there appear to have been no satisfactory experiments of this kind. The primary outcome of many TMS experiments has been to associate a brain region with a task rather than with a subprocess; the goals have not included using TMS as a tool for process decomposition, even when doing so would not require a larger experiment. Although the examples below include ones where TMS is

---

[59] 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 is better evidence for that region being necessary for the process to occur normally, just as does interference by a lesion in that region (Chatterjee, 2005). However, it needs to be kept in mind that the mechanism of TMS action is controversial (Harris, Clifford, & Miniussi, 2008; Johnson, Hamidi, & Postle, 2010; Miniussi, Ruzzoli, & Walsh, 2010; Siebner, Hartwigsen, Kassuba, & Rothwell, 2009).

[60] If the effect of TMS_R is time specific, as is likely with single-pulse TMS or a burst of rTMS after the trial starts ("on-line" TMS), the interpretation of its interactions with other factors may not be straightforward. 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 $\overline{RT}$. 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 using rTMS before the task is performed ("off-line" TMS) in such studies, rather than using one or more TMS pulses during the task. When doing so, note that the cognitive aftereffects of TMS are short-lived and plausibly decline over time, which suggests that tests should be balanced in small blocks and that, to reduce error variance, trends over trials should be estimated and corrected.

shown to modulate effects generated by interesting processes, and hence to influence those processes, its influence has not been shown to be selective.

10.2. Visual search and parietal TMS

In visual search for a conjunction of features, it is usually observed that the $\overline{RT}$ for both present and absent responses increases approximately linearly with the number of elements to be searched (the *Display Size*, *DS*), suggesting a process of serial (or capacity-limited parallel) comparison of the search target with the displayed elements. If a linear function is fitted to such data, the slope of the function is sometimes interpreted to reflect the time per comparison, while the intercept reflects the summed durations of residual processes whose durations are not influenced by *DS*.[61] In search for a single feature, unlike search for a feature conjunction, $\overline{RT}$ may increase very little or not at all with *DS*, 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 comparison process. The subprocess of most interest is the comparison process, whose measure is the slope of the function $\overline{RT}(DS)$ that relates $\overline{RT}$ to *DS*. To determine whether TMS influences the comparison process in conjunction search we need to discover whether it modulates the effect of *DS*—i.e., changes the slope of $\overline{RT}(DS)$. If TMS interferes with the comparison process, what effect would we expect? If the process is one of serial comparison, the time per comparison and hence the slope of $\overline{RT}(DS)$ function should increase. If the process is limited-capacity parallel, the capacity should decrease, which would also increase the slope.

In one of the first studies using TMS to investigate visual search, Ashbridge, Walsh, and Cowey (1997) examined the effect of single-pulse TMS of the right parietal cortex during feature and conjunction search. In a preliminary experiment in which TMS was not used but *DS* 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 *DS*.[62] They found TMS to interact with search type, having an effect on conjunction search but not on feature search. However, they missed an opportunity: If they had varied *DS*, thus obtaining a measure of its effect (the slope of $\overline{RT}(DS)$) with and without TMS, their findings could tell us whether TMS produces its effect on conjunction search by influencing the comparison process, the residual processes, or both. Without varying *DS*, 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. Thus, 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. 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 comparison process in conjunction search) requires speculation.

Unfortunately, among 15 later attempts to determine the effects of TMS on RTs in visual search published during the past dozen years, *DS* was varied in only two, those by Ellison, Rushworth, and Walsh

---

[61] Which intercept is appropriate depends on details of the search process, and may differ for target-absent and target-present trials.

[62] O'Shea, Muggleton, Cowey, and Walsh (2006, p. 948) say that "a single set size was used because adding a set size doubles the number of trials ...". This reason is valid if a specified level of precision is desired for the effect of TMS_R for each level of *DS*, but not if the goal is to achieve a specified level of precision for the main effect of TMS_R, unless variability accelerates with *DS*. (But in their data, both *SD* and SD^2 decelerate with *DS*.) Thus, if one runs 50 TMS trials each with $DS = 4$ and $DS = 12$, instead of 100 trials with $DS = 8$ (and likewise for the non-TMS control condition), the precision of the estimated main effect of TMS_R (which depends on the means over *DS* levels) would be about the same, and, at minimal added cost one would also have an estimate of the effect of *DS*, and therefore an estimate of the extent to which TMS modulates that effect.

(2003, Exp. 3) and Rosenthal, Walsh, Mannan, Anderson, Hawken, and Kennard (2006). In both studies, rTMS of the right posterior parietal cortex starting with presentation of the search array was found to increase the slope of $\overline{RT}(DS)$.[63] More studies in this promising direction are needed.

10.3. Number comparison and rTMS of the angular gyrus

In Ex. 6.3 subjects had to classify 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 (encoding, **E**, influenced by *Notation, N*) to derive the identity of the comparison stimulus, and the other (comparison, **C**, influenced by numerical *Proximity, P*) to compare the derived identity to the criterion. The processing stage of particular interest, **C**, is that which produces the effect of *P*: slower responding for stimulus numbers that are closer to the criterion. In a pioneering study, Goebel, Walsh, and Rushworth (2001) found that brief rTMS applied to the left or right angular gyrus influences the RTs in such a task. Do these effects mean that the angular gyrus is involved in implementing **C**? Alternatively, it might be involved in some other process that contributes to the RT, such as **E**. They varied two other factors in addition to TMS_R: the magnitude and the sign of the difference between the stimulus number and the criterion. Measures to assess the effect of *P* could be the slopes of the (approximately linear) functions for positive and negative differences ($k > 65$ and $k < 65$) that relate $\overline{RT}$ to the difference, $|k - 65|$. To determine whether TMS_R influences **C**, we need to know whether it changes either or both of these slopes. Effects of TMS_R on the *heights* of these functions might reflect its influence on **E** or on other processes.

In principle, the data to answer this question were collected in this study. But focused tests of the effects of TMS_R on the relevant slopes

were 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 might indicate an effect on **C**, but the appropriate analysis was not reported.

10.4. Number comparison and rTMS of the intraparietal sulcus

Cappelletti, Barth, Fregni, Spelke, and Pascual-Leone (2007, Exp. 1) recognized the importance, for evaluating the relevance of a brain region to **C**, of determining whether TMS_R modulates the effect of numerical proximity. Using the same task as in Ex. 6.3, but with Arabic numerals only, they found that relative to sham TMS, after 1 *Hz* rTMS of the left intraparietal sulcus (IPS), but not of either angular gyrus, the effect of *P* increased, showing that this region is involved in the implementation of **C**. Figure 10 shows some of their results, based on an analysis different from theirs.

In this study, a different opportunity was missed. For findings such as these to contribute to process decomposition, we need to know not only which subprocesses a brain region implements, but also which subprocesses it does not implement. And the issue of functional specificity is also important in interpreting the association of brain regions with subprocesses, one of the primary goals of such studies. In addition to **C**, the left IPS might be important for other processes used to perform the task. And for this task there is a wealth of evidence for **E**, separate from **C**, and a factor (*N*) known to influence it selectively (Ex. 6.3). It would be helpful to use TMS to confirm that **E** and **C** are implemented in different brain regions, as suggested by fMRI. Without enlarging the experiment, the stimuli on half of the trials could have been number names ("SIXTY-EIGHT") rather than Arabic numerals, thus

---

[63] The first study used neither sham TMS nor TMS of a different brain region as the control condition. The authors, who reported only data from the target-present trials, in which the mean slope increased from 18.5 to 22.4 ms/item, claimed that they had found no effect of TMS on the slope, but they report neither a test of the slope difference nor a confidence interval. In the second study, in which the control condition was TMS of a different brain region, $\overline{RT}$s were shorter, slopes were smaller, and the effect of TMS on the slope was greater.
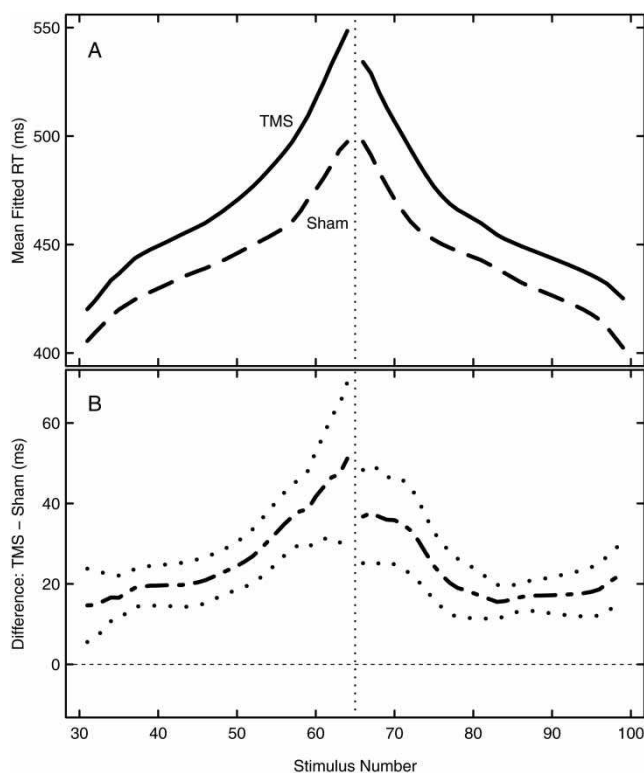
Figure 10. *Four functions fitted [64] to the RT data for correct responses in the left-IPS TMS and Sham TMS conditions from Cappelletti, Barth, Fregni, Spelke, and Pascual-Leone (2007, Exp. 1), in which subjects decided whether stimulus numbers were greater or smaller than 65. Panel A gives means of the four functions averaged over 10 subjects. Panel B gives means of the differences between the TMS and Sham functions (the TMS_R effect) together with their standard errors. Note the difference in scale. If the effects of P and TMS_R were additive, the two difference functions would be flat. Their increase with proximity shows the augmentation by rTMS of the P-effect, i.e., the interaction of the effects of the P and TMS_R factors.* [65]

including two levels of N. Given suitable balancing, data from the two subsets of trials could have been averaged to estimate the effect of P under TMS and control conditions. But in addition, by telling us whether the effect of N is modulated by rTMS of the left IPS, such an experiment would also permit asking whether that region is also important for **E**. Ideally, it would answer negatively, demonstrating the functional specificity of the left IPS by

showing invariance of the effect of N across levels of TMS_R.

11. PROCESS DECOMPOSITION AND THE EFFECTS OF BRAIN DAMAGE

Can localized brain damage (i.e., presence vs absence of lesions in particular locations) permit

---

[64] Data were retrieved for 11 of the 12 subjects; one of these, a clear outlier, was omitted. For each of the two conditions there were only about 50 RTs per subject for trials with each of the two responses. Starting with robust locally fitted polynomial regression (loess), a monotonic function was fitted to each of the four sets of observations for each subject. No adjustment was made for any decline over time in the aftereffect of rTMS.

[65] Future studies should concentrate observations on the closer proximities, where the P-effect is greater. Also, because there is considerable variation across individuals in the magnitude of the P-effect, subjects should perhaps be selected for large P-effects.

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 $LesionR_A$ factor that has two levels, no lesion and lesion. And suppose there is a factor, G, that influences process **B** but not **A**.[66] Given pure measures M_A and M_B we would need to show that $LesionR_A$ influences M_A while leaving M_B (and hence the effect of G) invariant. Given a composite measure M_{AB} with summation as the combination rule, we would need to show that the effects of $LesionR_A$ and G on that measure are additive. In a more complex scenario, our hypothesis might include the idea that **A** is also selectively influenced by non-lesion factor F. We would then also expect $LesionR_A$ to modulate the effect of F on M_A or M_{AB}.

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.* Suppose first that we have pure measures and that the data are sufficiently precise so that the effect of $LesionR_A$ 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. Unfortunately, it is unlikely that the invariance of M_B across levels of $LesionR_A$ could be persuasively demonstrated. The difficulty in showing

the absence of an effect is that there is virtually never any pre-lesion measurement,[67,68] which precludes within-subject comparison; 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.[69] Next consider the case where we have a composite measure M_{AB}. For example, **A** and **B** might be processing stages, and the measure might be mean reaction time, $\overline{RT}$. Assuming that $LesionR_A$ influences $\overline{RT}$, additivity requires that the effect of G on $\overline{RT}$ is invariant across levels of $LesionR_A$. Again, the required precision would almost always require a comparison of large groups.[70]

(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 how much those subprocesses must accomplish). 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.

---

[66] 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.

[67] This problem need not arise when the lesion is produced surgically, as in Farah, Soso, and Dashieff (1992). However, because such lesions are produced to ameliorate some other pathology, such as epilepsy, they may be associated with more than one effect.

[68] Crawford and Garthwaite's (2006) methods are concerned with showing the presence of an effect in a single patient, not its absence.

[69] 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 the required 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).

[70] Furthermore, 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.

These impediments perhaps suggest the use within cognitive neuropsychology of task-comparison with its more limited goals, rather than process decomposition.[71]

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 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 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

---

[71] Other impediments to using brain damage as a factor in process decomposition 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.

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 (see SM:A.2, SM:A.9).

12.2. Specialized processors and modular processes

Does the existence of localized neural processors that implement functionally distinct processes imply the modularity of those process? To address this question, consider one kind of evidence used to establish the existence of two processors: Suppose that T_1, T_2, and T_3 are three tasks, such that distinct brain regions R_α and R_β are activated during T_1, R_α but not R_β is activated during T_2, and neither region is activated during T_3. Many would conclude that processes $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ (carried out in R_α and R_β respectively) are functionally distinct, and that tasks T_1, and T_2 use $\boldsymbol{\alpha}$, task T_1 uses $\boldsymbol{\beta}$, and task T_3 uses neither. While it may seem plausible, such task-specificity of R_α and R_β does not require the processes they implement in T_1 to be modular, in the sense of being separately modifiable. For example, suppose, that $\boldsymbol{\alpha}$ provides a motivational or attentional resource required by $\boldsymbol{\beta}$, or controls which of several variants of $\boldsymbol{\beta}$ is being used. A change in $\boldsymbol{\alpha}$ would then induce a change in $\boldsymbol{\beta}$, and the effect of a change in $\boldsymbol{\beta}$ would depend on the status of $\boldsymbol{\alpha}$.

12.3. Relation between mental and neural processing modules

Consider modular mental processes in a task, supported by behavioural 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. 3.2, 6.3, 6.4, and 6.5), in which both brain and behavioural measures are taken, both directed at process decomposition. One starting point would be to take cases where behavioural data already exist that persuasively favour 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 mental 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?[72] 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,

---

[72] 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** (**B**) the effect of F (G) is the larger (Kanwisher et al., 2001). Whether differential or selective influence characterizes process*ors* is controversial (Haxby, 2004; Reddy & Kanwisher, 2006).

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* process, 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 $M_A(\Delta G) < M_B(\Delta G)$.[73] Suppose 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:

$$M_B(x) = gM_A(x) + h. \qquad (16)$$

Then differences in sensitivity or potency can explain neither of these pairs of properties; M_A and M_B must be measures of different processes. However, without that assumption, such pairs of properties could result from differences between measures of the same process.[74] Whether constraint (16) is plausible or valid depends on the nature of M_A and M_B.

12.5. Factorial experiments: Verification plus discovery

Despite R. A. Fisher's (1935) explanation of the advantages of factorial experiments, they are used too seldom with behavioural 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 orthogonal factor to a two-factor experiment, for example, without adding trials. In the case of composite measures we have seen that factorial experiments 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, as well as testing the generality of the effect of primary interest. 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 and many TMS experiments. For examples of missed opportunities to learn more by including additional factors in TMS experiments without increasing their size, see Sections 10.2 and 10.4.

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,[75] and permit focused tests of monotone interaction, the interaction of most interest. (SM:7, SM:15)

12.6. Implications of brain metabolism constraints

The metabolic requirements of brain activity are large relative to the available energy supply, with the implication that, given the spike rates of active neurons, no more than about 1% of the neurons in the brain can be concurrently active (Lennie, 2003). 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

---

[73] With suitable normalization, these pairs of inequalities are equivalent.

[74] 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.

[75] For example, a 2×2 design, with four conditions, provides only one df for interaction, whereas a 3×3 design, with nine conditions, provides four df.

subprocesses that are separately modifiable and that carry out distinct functions. Of course, the identification of such subprocesses is only a first step 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 on the process-decomposition approach see Sternberg (2001) (described in Section 2.3) 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.

AUTHOR NOTE

Supplementary data (a table of features of the nineteen examples in Sternberg, 2001, and this article) is published online alongside this article at: www.psypress.com/cogneuropsychology.

ABBREVIATIONS GLOSSARY

Listed here are the main abbreviations used, the numbers of the sections where they are introduced, and brief definitions.

A,B	1	processes: mental
$A_{mc}(t)$	3.1	motor cortex asymmetry as a function of time
B	7.2	level of the BOLD fMRI signal
$B(N)$	7.2	function relating B to amount of neural activity, N
BOLD	2.3	blood-oxygen-level dependence (fMRI signal)
C	6.3	process: comparison
c	5	criterion
D	**5**	process: decision
d	9.1	distance dimension to be judged
d', d-prime	5	discriminability measure
d-factors	5	factors that are likely to influence the decision process
DS	10.2	factor: display size
D_α, D_β	3.2	durations of processes α, β
$D_{\alpha jk}$	3.2	duration of α when $SQ = SQ_j$ and $RC = RC_k$
$D_{\alpha \cdot \cdot}$	3.2	mean duration of α, over levels of SQ and RC
E	6.3	process: encoding
ERP	2.2	event-related potential
F,G	2.1	factors
fMRI	1	functional magnetic resonance imaging
F_j	2.2	Factor F at level j
GND	3.1	factor: go-nogo discriminability
$H1$	2.2	hypothesis 1
$LesionR_A$	11	factor: presence or absence of a lesion in region R_A
LRP	3.1	lateralized-readiness potential
LRP_s, LRP_r	3.2	stimulus-locked LRP, response-locked LRP
MC	3.1	factor: mapping compatibility
MF	4.3	factor: mapping familiarity
M_A	2.2	(pure) measure of process **A**
M_{AB}	2.2	composite measure to which **A** and **B** contribute
N	6.3	factor: notation
N_α, N_β	7.2	amount of neural activity in populations that implement α and β
P	6.3	factor: numerical proximity |k - 65|
$p.effect$	8.2	proportional effect
PET	6.1	positron emission tomography
PM	5	factor: payoff matrix
PPA	6.5	parahippocampal place area
$Pr\{nonsuccess\}$	8.2	probability of nonsuccess
P_α, P_β	6.1	neural processors that implement α, β
$P_{\alpha\beta}$	6.2	neural processor that implements both α and β
R	4.3	process: response selection
r	9.1	roughness dimension to be judged
RC	3.2	factor: response complexity
REL	8.4	factor: semantic relatedness
RR	5	factor: reinforcement ratio: $Pr\{R_{NT}\|Reward\}$
R_α, R_β	6.1	brain regions that contain P_α, P_β
$R_{\alpha\beta}$	6.2	brain region that contains $P_{\alpha\beta}$
R_T, R_{NT}	5	target and nontarget responses
rTMS	2.3	repetitive transcranial magnetic stimulation
RT, $\overline{RT}$	2.2	reaction time, mean reaction time
S	4.3	process: stimulus encoding, or process: sensory
SAT	8.4	factor: semantic satiation
s-factors	5	factors that are likely to influence the sensory process
SLP	4.3	factor: sleep state
SM	2.3	reference to Sternberg (2001)
SQ	3.2	factor: stimulus quality
S_T, S_{NT}	5	target and nontarget stimuli
TD	4.3	factor: time of day
TMS	2.3	transcranial magnetic stimulation
TMS_R	2.3	factor: presence or absence of TMS of brain region R
u, v	2.2	contributions of **A** and **B** to M_{AB}
X_T, X_{NT}	5	target and nontarget representations
α, β	1	processes: neural
ε	6.3	process: neural that implements **E**
γ	6.3	process: neural that implements **C**

REFERENCES

Albrecht, D. G., Geisler, W. S., & Crane, A. M. (2003). Nonlinear properties of visual cortex neurons: Temporal dynamics, stimulus selectivity, neural performance. In L. Chalupa & J. Werner (Eds.), *The visual neurosciences* (pp. 747–764). Cambridge, MA: MIT Press.

Ashbridge, E., Walsh, V., & Cowey, A. (1997). Temporal aspects of visual search studied by transcranial magnetic stimulation. *Neuropsychologia, 35,* 1121–1131.

Ashby, F. G. (1982). Deriving exact predictions from the cascade model. *Psychological Review, 89,* 599–607.

Ashby, F. G., & Maddox, W. T. (1994). A response time theory of separability and integrality in speeded classification. *Journal of Mathematical Psychology, 38,* 423–466.

Balkin, T. J., Rupp, T., Picchioni, D., & Wesensten, N. J. (2008). Sleep loss and sleepiness: Current issues. *Chest, 134,* 653–660.

Berger, R. L., & Hsu, J. C. (1996). Bioequivalence trials, intersection–union tests and equivalence confidence sets. *Statistical Science, 11,* 283–319.

Biederman, I., & Kaplan, R. (1970). Stimulus discriminability and stimulus response compatibility: Evidence for independent effects on choice reaction time. *Journal of Experimental Psychology, 86,* 434–439.

Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *Journal of Neuroscience, 16,* 4207–21.

Boynton, G. M., & Finney, E. M. (2003). Orientation–specific adaptation in human visual cortex. *Journal of Neuroscience, 23,* 8781–7.

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *Neuroimage, 5,* 49–62.

Broadbent, D. E. (1984). The Maltese cross: A new simplistic model for memory. *Behavioral and Brain Sciences, 7,* 55–68.

Cabeza, R., & Nyberg, L. (1997). Imaging cognition: An empirical review of PET studies with normal subjects. *Journal of Cognitive Neuroscience, 9,* 1–26.

Caplan, D., & Moo, L. (2004). Cognitive conjunction and cognitive functions. *Neuroimage, 21,* 751–756.

Cappelletti, M., Barth, H., Fregni, F., Spelke, E. S., & Pascual-Leone, A. (2007). rTMS over the intraparietal sulcus disrupts numerosity processing. *Experimental Brain Research, 179,* 631–642.

Carandini, M., & Heeger, D. J. (1994). Summation and division by neurons in primate visual cortex. *Science, 264,* 1333–1336.

Chatterjee, A. (2005). A madness to the methods in cognitive neuroscience? *Journal of Cognitive Neuroscience, 27,* 847–849.

Coltheart, M. (1999). Modularity and cognition. *Trends in Cognitive Sciences, 3,* 115–120.

Crawford, J. R., & Garthwaite, P. H. (2006). Methods of testing for a deficit in single-case studies: Evaluation of statistical power by Monte Carlo simulation. *Cognitive Neuropsychology, 23,* 877–904.

Culham, J. C., Cavanagh, P., & Kanwisher, N. G. (2001). Attention response functions: Characterizing brain areas using fMRI activation during parametric variations of attentional load. *Neuron, 32,* 737–745.

Dale, A. M., & Buckner, R. L. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping, 5,* 329–340.

Dehaene, S. (1996). The organization of brain activations in number comparison: Event-related potentials and the additive-factors method. *Journal of Cognitive Neuroscience, 8,* 47–68.

Dinges, D. F., & Kribbs, N. B. (1991). Performing while sleepy: Effects of experimentally-induced sleepiness. In T. H. Monk (Ed.), *Sleep, sleepiness and performance* (pp. 97–128). London, UK: Wiley.

Donders, F. C. (1868/1969). Over de snelheid van psychische processen [On the speed of mental processes]. *Onderzoekingen gedaan in het Physiologisch Laboratorium der Utrechtsche Hoogeschool, 1868-1869, Tweede reeks, II,* 92-120. Transl. by Koster, W. G. (1969). In *Attention and performance II*, Koster, W. G., Ed., *Acta Psychologica, 30,* 412–431.

Drucker, D. M., Kerr, W. T., & Aguirre, G. K. (2009). Distinguishing conjoint and independent neural tuning for stimulus features with fMRI adaptation. *Journal of Neurophysiology, 101,* 3310–3324.

Ekstrom, A. (2010). How and when the fMRI BOLD signal relates to underlying neural activity: The danger in dissociation. *Brain Research Reviews, 62,* 233–244.

Ellison, A., Rushworth, M., & Walsh, V. (2003). The parietal cortex in visual search: A visuomotor hypothesis. *Clinical Neurophysiology Supplement, 56,* 321–330.

Epstein, R. A., Parker, W. E., & Feiler, A. M. (2008). Two kinds of fMRI repetition suppression? Evidence for dissociable neural mechanisms. *Journal of Neurophysiology, 99,* 2877–2886.

Erickson, R. P. (2001). The evolution and implications of population and modular neural coding ideas. *Progress in Brain Research, 130,* 9–29.

Farah, M. J., Soso, M. J., & Dashieff, R. M. (1992). Visual angle of the mind's eye before and after unilateral occipital lobectomy. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 241–246.

Fisher, R. A. (1935). "VI. The factorial design in experimentation". In *The design of experiments*. Edinburgh, UK: Oliver & Boyd.

Fodor, J. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Fodor, J. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. Cambridge, MA: MIT Press.

Friston, K. J., Price, C. J., Fletcher, P., Moore, C., Frackowiak, R. S. J., & Dolan, R. J. (1996). The trouble with cognitive subtraction. *Neuroimage*, *4*, 97–104.

Frowein, H. W., & Sanders, A. F. (1978). Effects of visual stimulus degradation, S-R compatibility and foreperiod duration on choice reaction time and movement time. *Bulletin of the Psychonomic Society*, *12*, 106–108.

Gallistel, C. R. (2009). The importance of proving the null. *Psychological Review*, *116*, 439–453.

Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Lawrence Erlbaum Associates.

Gehring, W. J., & Knight, R. T. (2002). Lateral prefrontal damage affects processing selection but not attention switching. *Cognitive Brain Research*, *13*, 267–279.

Geisler, W. S., & Albrecht, D. G. (1995). Bayesian analysis of identification performance in monkey visual cortex: Nonlinear mechanisms and stimulus certainty. *Vision Research*, *35*, 2723–2730.

Ghorashi, S., Enns, J. T., Klein, R. M., & Di Lollo, V. (2010). Spatial selection and target identification are separable processes in visual search. *Journal of Vision*, *10*, 1–12.

Goebel, S., Walsh, V., & Rushworth, M. F. S. (2001). The mental number line and the human angular gyrus. *NeuroImage*, *14*, 1278–1289.

Goense, J. B. M., & Logothetis, N. K. (2008). Neurophysiology of the BOLD fMRI signal in awake monkeys. *Current Biology*, *18*, 631–640.

Goodale, M. A. (1996). Visuomotor modules in the vertebrate brain. *Canadian Journal of Physiology and Pharmacology*, *74*, 390–400.

Grill-Spector, K. (2006). Selectivity of adaptation in single units: Implications for fMRI experiments. *Neuron*, *11*, 170–171.

Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: A tool for studying the functional properties of human cortical neurons. *Acta Psychologica*, *107*, 293–321.

Gurd, J. M., & Marshall, J. C. (2003). Dissociations: Double or quits? *Cortex*, *39*, 192–195.

Hadley, R. F. (2003). A defense of functional modularity. *Connection Science*, *15*, 95–116.

Harris, J. A., Clifford, C.W. G., & Miniussi, C. (2008). The functional effect of transcranial magnetic stimulation: Signal suppression or neural noise generation? *Journal of Cognitive Neuroscience*, *20*, 734–740.

Haxby, J. V. (2004). Analysis of topographically organized patterns of response in fMRI data: Distributed representations of objects in ventral temporal cortex. In N. Kanwisher & J. Duncan (Eds.), *Attention and performance XX: Functional neuroimaging of visual cognition* (pp. 83–97). Oxford, UK: Oxford University Press.

Haxby, J. V., Ungerleider, L. G., Horwitz, B., Rapoport, S. I., & Grady, C. L. (1995). Hemispheric differences in neural systems for face working memory: A PET-rCBF study. *Human Brain Mapping*, *3*, 68–82.

Heeger, D. J., Huk, A. C., Geisler, W. S., & Albrecht, D. G. (2000). Spikes versus BOLD: What does neuroimaging tell us about neuronal activity? *Nature Neuroscience*, *3*, 631–633.

Horner, A. J., & Andrews, T. J. (2009). Linearity of the fMRI response in category-selective regions of human visual cortex. *Human Brain Mapping*, *30*, 2628–2640.

Howson, C., & Urbach, P. (2006). *Scientific reasoning: The Bayesian approach* (3rd ed.). Chicago, IL: Open Court.

Humphrey, D. G., Kramer, A. F., & Stanny, R. R. (1994). Influence of extended wakefulness on automatic and nonautomatic processing. *Human Factors*, *36*, 652–669.

Jacobs, R. A. (1997). Nature, nurture, and the development of functional specializations: A computational approach. *Psychonomic Bulletin & Review*, *4*, 299–309.

Jacobs, R. A. (1999). Computational studies of the development of functionally specialized neural modules. *Trends in Cognitive Sciences*, *3*, 31–38.

Jacobs, R. A., & Jordan, M. I. (1992). Computational consequences of a bias toward short connections. *Journal of Cognitive Neuroscience*, *4*, 323–336.

Jennings, J. M., McIntosh, A. R., Kapur, S., Tulving, E., & Houle, S. (1997). Cognitive subtractions may not

add up: The interaction between semantic processing and response mode. *Neuroimage*, *5*, 229–239.

Johnson, J. S., Hamidi, M., & Postle, B. R. (2010). Using EEG to explore how rTMS produces its effects on behavior. *Brain Topography*, *22*, 281–293.

Kanwisher, N., Downing, P., Epstein, R., & Kourtzi, Z. (2001). Functional neuroimaging of visual recognition. In R. Cabeza & A. Kingstone (Eds.), *Handbook of functional neuroimaging of cognition* (pp. 109–151). Cambridge MA: MIT Press.

Kosslyn, S. M. (2001). The strategic eye: Another look. *Minds and Machines*, *11*, 287–291.

Kounios, J. (2007). Functional modularity of semantic memory revealed by event-related brain potentials. In J. Hart, Jr. & M. A. Kraut (Eds.), *Neural basis of semantic memory* (pp. 65–104). Cambridge, UK: Cambridge University Press.

Lennie, P. (2003). The cost of cortical computation. *Current Biology*, *13*, 493–497.

Li, L., Miller, E. K., & Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. *Journal of Neurophysiology*, *69*, 1918–1929.

Lie, C.-H., Specht, K., Marshall, J. C., & Fink, G. R. (2006). Using fMRI to decompose the neural processes underlying the Wisconsin Card Sorting Test. *NeuroImage*, *30*, 1038–1049.

Lim, J., & Dinges, D. F. (2010). A meta-analysis of the impact of short-term sleep deprivation on cognitive variables. *Psychological Bulletin*, *136*, 375–389.

Logothetis, N. K., & Wandell, B. A. (2004). Interpreting the BOLD signal. *Annual Review of Physiology*, *66*, 735–769.

Machamer, P., Darden, L., & Craver, E. (2000). Thinking about mechanisms. *Philosophy of Science*, *67*, 1–25.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.

Marr, D. (1976). Early processing of visual information. *Philosophical Transactions of the Royal Society, London B*, *275*, 483–524.

McCarthy, D., & Davison, M. (1984). Isobias and alloiobias functions in animal psychophysics. *Journal of Experimental Psychology: Animal Behavior Processes*, *10*, 390–409.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*, 287–330.

Meeter, M., Myers, C. E., & Gluck, M. A. (2005). Integrating incremental learning and episodic memory models of the hippocampal region. *Psychological Review*, *112*, 560–585.

Merabet, L., Thut, G., Murray, B., Andrews, J., Hsiao, S., & Pascual-Leone, A. (2004). Feeling by sight or seeing by touch? *Neuron*, *42*, 173–179.

Miller, J., van der Ham, F., & Sanders, A. F. (1995). Overlapping stage models and reaction time additivity: Effects of the activation equation. *Acta Psychologica*, *90*, 11–28.

Miller, K. L., Luh, W-M., Liu, T. T., Martinez, A., Obata, T., Wong, E. C., Frank, L. R., & Buxton, R. B. (2001). Nonlinear temporal dynamics of the cerebral blood flow response. *Human Brain Mapping*, *13*, 1–12.

Miniussi, C., Ruzzoli, M., & Walsh, V. (2010). The mechanism of transcranial magentic stimulation in cognition. *Cortex*, *46*, 128–130.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*, 1519–1520.

Mulligan, R. M., & Shaw, M. L. (1980). Multimodal signal detection: Independent decisions vs. integration. *Perception & Psychophysics*, *28*, 471–478.

Nunez, P. L., & Srinivasan, R. (2006). *Electric fields of the brain: The neurophysics of EEG*. New York: Oxford University Press.

Op de Beeck, H. P., Haushofer, J., & Kanwisher, N. G. (2008). Interpreting fMRI data: Maps, modules and dimensions. *Nature Reviews Neuroscience*, *9*, 123–135.

O'Shea, J., Muggleton, N. G., Cowey, A., & Walsh, V. (2006). On the roles of the human frontal eye fields and parietal cortex in visual search. *Visual Cognition*, *14*, 934–957.

Osman, A., Bashore, T. R., Coles, M. G. H., Donchin, E., & Meyer, D. E. (1992). On the transmission of partial information: Inferences from movement-related brain potentials. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 217–232.

Petersen, S. E., Fox, P. T., Posner, M. I., Minton, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, *331*, 585–589.

Pinel, P., Dehaene, S., Rivière, D., & LeBihan, D. (2001). Modulation of parietal activation by semantic distance in a number comparison task. *NeuroImage*, *14*, 1013–1026.

Poldrack, R. A. (2010). Subtraction and beyond: The logic of experimental designs for neuroimaging. In S. J. Hanson & M. Bunzl (Eds.), *Foundational*

issues in human brain mapping (pp. 147–159). Cambridge, MA: MIT Press.

Price, C. J., & Friston, K. J. (1997). Cognitive conjunction: A new approach to brain activation experiments. *Neuroimage, 5,* 261–270.

Price, C. J., Moore, C. J., & Friston, K. J. (1997). Subtractions, conjunctions, and interactions in experimental design of activation studies. *Human Brain Mapping, 5,* 264–272.

Ratcliff, R., & Smith, P. L. (2010). Perceptual discrimination in static and dynamic noise: The temporal relation between perceptual encoding and decision making. *Journal of Experimental Psychology: General, 139,* 70–94.

Reddy, L., & Kanwisher, N. (2006). Coding of visual objects in the ventral stream. *Current Opinion in Neurobiology, 16,* 408–414.

Rees, G., Friston, K., & Koch, C. (2000). A direct quantitative relationship between the functional properties of human and macaque V5. *Nature Neuroscience, 3,* 716–723.

Roberts, S., & Sternberg, S. (1993). The meaning of additive reaction-time effects: Tests of three alternatives. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience – a silver jubilee* (pp. 611–653). Cambridge, MA: MIT Press.

Rogers, J. L., Howard, K. I., & Vessey, J. T. (1993). Using significance tests to evaluate equivalence between two experimental groups. *Psychological Bulletin, 113,* 553–565.

Rosenthal, C. R., Walsh, V., Mannan, S. K., Anderson, E. J., Hawken, M. B., & Kennard, C. (2006). Temporal dynamics of parietal cortex involvement in visual search. *Neuropsychologia, 44,* 731–743.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16,* 225–237.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (1986). *Parallel Distributed Processing: Explorations in the microstructure of cognition, Volume 1: Foundations.* Cambridge, MA: MIT Press.

Sachs, M. B., Nachmias, J., & Robson, J. G. (1971). Spatial-frequency channels in human vision. *Journal of the Optical Society of America, 61,* 1176–1186.

Sanders, A. F. (1977). Structural and functional aspects of the reaction process. In S. Dornic (Ed.), *Attention and performance VI* (pp. 3–25). Hillsdale, NJ: Erlbaum.

Sanders, A. F. (1980). Some effects of instructed muscle tension on choice reaction time and movement time. In R. S. Nickerson (Ed.), *Attention and performance VIII* (pp. 59–74). Hillsdale, NJ: Erlbaum.

Sanders, A. F. (1998). *Elements of human performance: Reaction processes and attention in human skill.* Mahwah, NJ: Erlbaum.

Sanders, A. F., Wijnen, J. L. C., & Van Arkel, A. E. (1982). An additive factor analysis of the effects of sleep loss on reaction processes. *Acta Psychologica, 51,* 41–59.

Sartori, G., & Umiltà, C. (2000). How to avoid the fallacies of cognitive subtraction in brain imaging. *Brain and Language, 74,* 191–212.

Sawamura, H., Orban, G. A., & Vogels, R. (2006), Selectivity of neuronal adaptation does not match response selectivity: A single-cell study of the fMRI adaptation paradigm. *Neuron, 11,* 307–318.

Schall, J. D. (2003). Neural correlates of decision processes: Neural and mental chronometry. *Current Opinion in Neurobiology, 13,* 182–186.

Schmidt, T., & Vorberg, D. (2006). Criteria for unconscious cognition: Three types of dissociation. *Perception & Psychophysics, 68,* 489–504.

Schuberth, R. E., Spoehr, K. T., & Lane, D. M. (1981). Effects of stimulus and contextual information on the lexical decision process. *Memory & Cognition, 9,* 68–77.

Schumacher, E. H., & D'Esposito, M. (2002). Neural implementation of response selection in humans as revealed by localized effects of stimulus-response compatibility on brain activation. *Human Brain Mapping, 17,* 193–201.

Schweickert, R. (1985). Separable effects of factors on speed and accuracy: Memory scanning, lexical decision, and choice tasks. *Psychological Bulletin, 97,* 530–546.

Shallice, T. (1988). *From neuropsychology to mental structure.* Cambridge, UK: Cambridge University Press.

Shallice, T. (2003). Functional imaging and neuropsychology findings: How can they be linked? *Neuroimage, 20,* S146–S154.

Shaw, M. L. (1980). Identifying attentional and decision-making components in information processing. In R. S. Nickerson (Ed.), *Attention and performance VIII* (pp. 277–296). Hillsdale, NJ: Erlbaum.

Shwartz, S. P., Pomerantz, J. R., & Egeth, H. E. (1977). State and process limitations in information processing: An additive factors analysis. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 402–410.

Sidtis, J. J., Strother, S. C., Anderson, J. R., & Rottenberg, D. A. (1999). Are brain functions really additive? *Neuroimage*, *9*, 490–496.

Siebner, H. R., Hartwigsen, G., Kassuba, T., & Rothwell, J. C. (2009). How does transcranial magnetic stimulation modify neuronal activity in the brain? Implications for studies of cognition. *Cortex*, *45*, 1035–1042.

Sigman, M., & Dehaene, S. (2005). Parsing a cognitive task: A characterization of the mind's bottleneck. *PLoS Biology*, *3*, 334–349.

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, *106*, 467–482.

Simon, H. A. (2005). The structure of complexity in an evolving world: The role of near decomposability. In W. Callebaut & D. Rasskin-Gutman (Eds.), *Modularity: Understanding the development and evolution of natural complex systems* (pp. ix–xiii). Cambridge, MA: MIT Press.

Smulders, F. T. Y., Kok, A., Kenemans, J. L., & Bashore, T. R. (1995). The temporal selectivity of additive factor effects on the reaction process revealed in ERP component latencies. *Acta Psychologica*, *90*, 97–109.

Soltysik, D. A., Peck, K. K., White, K. D., Crosson, B., & Briggs, R. W. (2004). Comparison of hemodynamic response nonlinearity across primary cortical areas. *Neuroimage*, *22*, 1117–1127.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and performance II, Acta Psychologica*, *30*, 276–315.

Sternberg, S. (1984). Stage models of mental processing and the additive-factor method. *Behavioral and Brain Sciences*, *7*, 82–84.

Sternberg, S. (1998a). Discovering mental processing stages: The method of additive factors. In D. Scarborough & S. Sternberg (Eds.), *An invitation to cognitive science, Volume 4: Methods, models, and conceptual issues* (pp. 703–863). Cambridge, MA: MIT Press.

Sternberg, S. (1998b). Inferring mental operations from reaction-time data: How we compare objects. In D. Scarborough & S. Sternberg (Eds.), *An invitation to cognitive science, Volume 4: Methods, models, and conceptual issues* (pp. 365–454). Cambridge, MA: MIT Press.

Sternberg, S. (2001). Separate modifiability, mental modules, and the use of pure and composite measures to reveal them. *Acta Psychologica*, *106*, 147–246.

Sternberg, S. (2003). Process decomposition from double dissociation of subprocesses. *Cortex*, *39*, 180–182.

Sternberg, S. (2004). Separate modifiability and the search for processing modules. In N. Kanwisher & J. Duncan (Eds.), *Attention and performance XX: Functional neuroimaging of visual cognition* (pp. 125–139). Oxford, UK: Oxford University Press.

Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review*, *68*, 301–340.

Teller, D. (1984). Linking propositions. *Vision Research*, *24*, 1233–1246.

Tiplady, B., Bowness, E., Stien, L., & Drummond, G. (2005). Selective effects of clonidine and temazepam on attention and memory. *Journal of Psychopharmacology*, *19*, 259–265.

Ulrich, R., Mattes, S., & Miller, J. (1999). Donders's assumption of pure insertion: An evaluation on the basis of response dynamics. *Acta Psychologica*, *102*, 43–75.

Uttal, W. R. (2001). *The new phrenology: The limits of localizing cognitive processes in the brain*. Cambridge, MA: MIT Press.

Walsh, V., & Pascual-Leone, A. (2003). *Transcranial Magnetic Stimulation: A Neurochronometrics of Mind*. Cambridge, MA: MIT Press.

Wan, X., Riera, J., Iwata, K., Takahashi, M., Wakabayashi, T., & Kawashima, R. (2006). The neural basis of the hemodynamic response nonlinearity in human primary visual cortex: Implications for neurovascular coupling mechanism. *Neuroimage*, *32*, 616–625.

Woodman, G. F., Kang, M.-S., Thompson, K., & Schall, J. D. (2008). The effect of visual search efficiency on response preparation: Neurophysiological evidence for discrete flow. *Psychological Science*, *19*, 128–136.