.

Separate Modifiability and the Search for Processing Modules

Saul Sternberg*

Abstract

One approach to understanding a complex process or system starts with an attempt to divide it into *modules*: parts that are independent in some sense, and functionally distinct. In this chapter I discuss a method for the modular decomposition of neural and mental processes that reflects recent thinking in psychology and cognitive neuroscience. This process-decomposition method, in which the criterion for modularity is *separate modifiability*, is contrasted with task comparison and its associated subtraction method. Four illustrative applications of process decomposition and one of task comparison are based on the event-related potential (ERP), transcranial magnetic stimulation (rTMS), and functional magnetic resonance imaging (fMRI).

---

* E-mail: saul@psych.upenn.edu
Tel: 215-898-7162
Fax: 215-898-7210
Address:
Department of Psychology
University of Pennsylvania
3815 Walnut Street
Philadelphia, PA 19104-6196, USA

Separate Modifiability and the Search for Processing Modules

Saul Sternberg

1. Modules and Modularity

The first step in one approach to understanding a complex process or system is to attempt to divide it into *modules*: parts that are independent in some sense, and functionally distinct.[1] [2] In the present context the complex entity may be a *mental process*, a *neural process,* the *brain* (an anatomical processor), or the *mind* (a functional processor). Four corresponding senses of 'module' are:

Module$_1$: A part of a *mental process*, functionally distinct from other parts, and investigated with behavioral measures, supporting a functional analysis.

Module$_2$: A part of a *neural process*, functionally distinct from other parts, and investigated with brain measures, supporting a neural-process analysis.

Module$_3$: A *neural processor*[3] (part of the brain), a population P_α of neurons that is functionally specialized to implement a particular neural process $\boldsymbol{\alpha}$. If it is also *localized* (the sole occupant of a delimited brain region) and the only population that implements $\boldsymbol{\alpha}$, one may find selective task impairment from localized damage of P_α, and selective activation of P_α by tasks that require $\boldsymbol{\alpha}$.[4]

Module$_4$: A *mental processor* or *faculty* (part of the mind), functionally specialized ('domain specific'), informationally isolated from (some) other processors ('encapsulated'), and a product of evolution (see Spelke, Chapter 2, this volume).

Most of the present discussion will be concerned with the first two senses. In what follows I describe and illustrate an approach to the decomposition of mental and neural processes into *Modules*$_1$ and *Modules*$_2$ that reflects recent thinking in psychology and cognitive neuroscience.[5] This *process-decomposition* method and three illustrations based on the event-related potential (ERP) are described in Section 2. I contrast process decomposition with the more familiar *task-comparison* method in Section 3, describe an example of the latter based on the effects of repetitive transcranial magnetic stimulation (rTMS), and discuss the subtraction method as an embodiment of task comparison. Unlike task comparison, which is often used in a way that requires modularity to be assumed without test (Shallice 1988, Ch. 11; Sternberg 2001, Appendix A.1.), the process-decomposition method incorporates such a test. I consider the use of fMRI activation maps in process decomposition for the discovery of *Modules*$_2$ in Section 4, list some

---

1. Heuristic arguments for the modular organization of complex biological computations have been advanced by Simon (1962) and, in his 'principle of modular design', by Marr (1976).

2. A module may itself be composed of modules.

3. *Processes* occur over time; their arrangement is described by a flow-chart. They are often confused with *processors* (parts of a machine), whose arrangement can be described by a circuit diagram.

4. By 'selective impairment' I mean impairment of tasks that require $\boldsymbol{\alpha}$ and not of tasks that don't. By 'selective activation' I mean activation of P_α by tasks that require $\boldsymbol{\alpha}$ and not by tasks that don't. In practice, *selective* activation is sometimes taken to mean the weaker *differential* activation (Kanwisher *et al.* 2001), akin to the tuning curves for simple features.

5. In Sternberg (2001) I discuss and defend the method, describe its antecedents, illustrate it with a dozen applications to mental and neural processes, and explicate its inferential logic.

desiderata in such applications, and present an example. I briefly consider the question of task-general modules (Section 5) and the relation between *Modules*$_2$ and *Modules*$_3$ (Section 6), and close with a few questions (Section 7).

2. The Process-Decomposition Method

2.1 Separate Modifiability

Much thinking by psychologists and brain scientists about the decomposition of complex processes appeals implicitly to *separate modifiability* as a criterion for modularity: two (sub)processes **A** and **B** of a complex process (mental or neural) are modules iff each can be changed independently of the other. To demonstrate separate modifiability of **A** and **B**, we must find experimental manipulations (factors) F and G that influence them 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.[6] Such double dissociation of subprocesses should be distinguished from the more familiar double dissociation of tasks (Sternberg 2003).

2.2 Processes and Their Measures, Pure and Composite

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 *measures* M_A of process **A**, not about the process as such. Depending on the available measures, there are two ways to assess separate modifiability of **A** and **B**. Suppose we have *pure measures* M_A and M_B of the hypothesized modules: A pure measure of a process is one that reflects changes in that process only. Examples include the sensitivity and criterion parameters of signal-detection theory (reflecting sensory and decision processes), and the durations of two different neural processes. To show that F and G influence **A** and **B** selectively, we must demonstrate their selective influence on M_A and M_B. The influence and invariance requirements are both critical. Unfortunately, it is seldom appreciated that persuasive evidence for invariance cannot depend solely on failure of a significance test of an effect: such a failure could merely reflect variability and low statistical power.[7]

Instead of pure measures, suppose we have a *composite measure* M_{AB} of the hypothesized modules — a measure to which they both contribute. To demonstrate selective influence in this case we must also know or confirm a *combination rule* — a specification of how the contributions of the modules to the measure combine. Examples of composite measures are the ERP at a particular point on the scalp (which may reflect several ERP sources), and mean reaction time, $\overline{RT}$ (which may depend on the durations of several processes). Whereas factorial experiments are desirable for pure measures[8], they are essential with a composite measure; unfortunately they are rare.

A given measure may be pure or composite, depending on the hypothesized modules of interest. However, rather than being determined a priori, this attribute of a measure is one of the components of a theory that is tested as part of the process-decomposition method.[9]

---

6. Separate modifiability of **A** and **B** is also evidence for their functional distinctness (Sternberg 2001, p. 149); information about what a process does is provided by the sets of factors that do and don't influence it.

7. In evaluating a claim that an effect is null, it is important to have at least an index of precision (such as a confidence interval) for the size of the effect. An alternative is to apply an *equivalence test* (Berger and Hsu 1996; Rogers, Howard, and Vessey 1993) that reverses the asymmetry of the standard significance test. 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.

8. See Sternberg (2001), Appendix A.9.

2.3 Three Examples of Decomposition of Neural Processes with ERPs

Here I provide brief summaries of three applications of these ideas, in which the brain measures are derived from ERPs.[10]

2.3.1 Parallel Modules for Selecting a Response and Deciding Whether to Execute It.

Osman *et al.* (1992) investigated a task in which a speeded response was required to two of four equiprobable stimuli. The location of the stimulus (left versus right) indicated the correct response (left hand versus right hand); its category (letter vs digit) determined whether that response should be executed (Go versus NoGo trials). The two factors were the stimulus-response mapping (*SRM*, spatially compatible versus incompatible), and Go-NoGo (letter-digit) discriminability (*GND*, easy versus hard). The two hypothesized pure measures depend on the lateral asymmetry of the motor-cortex voltage versus time for Go and NoGo trials, $A_{MC}(t, Go)$ and $A_{MC}(t, NoGo)$.[11] One measure (M_α) is the time interval from stimulus onset to when the sum $A_{MC}(t, Go) + A_{MC}(t, NoGo)$ exceeds zero, which reflects the duration of $\boldsymbol{\alpha}$, the hypothesized response-selection module. The other measure (M_β) is the interval from stimulus onset to when the difference $A_{MC}(t, Go) - A_{MC}(t, NoGo)$ exceeds zero, which reflects the duration of $\boldsymbol{\beta}$, the hypothesized execution-decision module. They found that M_α is influenced by *SRM* (effect = $\Delta = 121 \pm 17$ ms, $n = 6$) but negligibly by *GND* ($\Delta = 2.5 \pm 5.0$ ms, $n = 6$)[12], whereas M_β is influenced by *GND* ($\Delta = 43 \pm 14$ ms) but negligibly by *SRM* ($\Delta = 3.3 \pm 8.8$ ms), evidence for the separate modifiability of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.[13] Other aspects of the data indicate that $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ operate in parallel.

2.3.2 Serial Modules for Interpreting a Stimulus and Initiating the Response.

Smulders *et al.* (1995) investigated a task in which a digit stimulus indicated which hand had to execute a speeded response. The two factors were stimulus quality (*SQ*, two levels) and response complexity (*RC*, a single keystroke versus a string of three keystrokes). The two hypothesized pure measures were M_α, the duration of process $\boldsymbol{\alpha}$ (from the stimulus to the onset of motor cortex asymmetry), and M_γ, the duration of process $\boldsymbol{\gamma}$ (from the onset of motor cortex asymmetry to the response). They found that M_α is influenced by *SQ* ($\Delta = 34 \pm 6$ ms, $n = 14$) but negligibly by *RC* ($\Delta = 4 \pm 8$ ms), whereas M_γ is influenced by *RC* ($\Delta = 21 \pm 7$ ms) but negligibly by *SQ* ($\Delta = 1 \pm 8$ ms), evidence for two neural processing modules arranged as stages.[14]

2.3.3 Two Modules in Word Classification.

Kounios (1999, 2002) required subjects to classify each of a sequence of spoken nouns by meaning. Most of the words required no response, while 5% were targets (names of body parts)

---

9. See Sternberg (2001), Sections 2 and 3 and Appendix A.2.3.

10. These examples are treated in detail in Sternberg (2001) in Sec. 6, Appendix A.6, and Sec. 14.

11. A_{MC} is the amplitude difference between the scalp ERPs associated with the parts of the motor cortex that control the left and right hands, taken in the direction that favors the response signaled by the stimulus location; an increase in A_{MC} from baseline is sometimes called the 'lateralized readiness potential'.

12. A cautionary note on the meaning of 'negligible', using this example. With SE = 5.0 ms, 5 *df*, and a mean of 2.5 ms, a 95% confidence interval based on the *t*-statistic indicates that the true *GND* effect may be as large as 15 ms. Another way of indicating the precision of the data is that a *GND* effect would have to be as large as ±13 ms to be detected, in the sense of differing significantly from zero at the $p = 0.05$ level. See Note 7.

13. SE estimates are based on between-subject variability. However, the SE values provided for the second and third examples are likely to be overestimates because balanced effects, such as those of condition order, were treated as error variance.

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. For present purposes, 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:

(a) *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.

(b) *Selective Influence*: $\boldsymbol{\alpha}$ is influenced by *SAT* but not *REL*, whereas $\boldsymbol{\beta}$ is influenced by *REL* but not *SAT*.

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

It can be shown that this theory implies that the effects of *SAT* and *REL* on $M_{\alpha\beta}$ will be additive at all scalp locations.[15] Kounios found such additivity (mean main effects of *REL* and *SAT* were $1.3 \pm 0.2 \mu$V and $2.1 \pm 0.4 \mu$V, respectively, while the mean interaction contrast was $0.01 \pm 0.3 \mu$V, $n = 36$), supporting the above theory and hence the modularity of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ during the 600 to 800 ms epoch.[16] 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.[17]

3. Process Decomposition versus Task Comparison

The cases above exemplify a process-decomposition method whose goal is to divide the complex process by which a particular task is accomplished into modular subprocesses, a method that has been used to find $Modules_1$ and $Modules_2$. The factor manipulations are not intended to produce 'qualitative' changes in the complex process (such as adding new operations, or replacing one operation by another), which may be associated with a change in the task, just 'quantitative' ones that leave it invariant.[18] The task-comparison method is a more popular approach to understanding the structure of complex processes. Here one determines the influence of factors

---

14. *SQ* and *RC* also had additive effects on concurrently measured $\overline{RT}$ (a composite measure), consistent with their selectively influencing two functional modules, **A** and **C**, that are arranged as stages. (Main effects of *SQ* and *RC* were 34 ± 3 ms and 25 ± 7 ms, respectively; their interaction was a negligible 2 ± 5 ms.) Together with the similarity of effect sizes in the neural and behavioral analyses, this suggests that **A** and **C** are implemented by $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, respectively.

15. In the present context, the effect of a factor on a measure is the change in the measure produced by changing the level of that factor. Letting (i, j) indicate the levels of factors REL_i and SAT_j, additivity (non-interaction) of their effects means that $M_{\alpha\beta}(2,2) - M_{\alpha\beta}(1,1) = [M_{\alpha\beta}(2,1) - M_{\alpha\beta}(1,1)] + [M_{\alpha\beta}(1,2) - M_{\alpha\beta}(1,1)]$.

16. 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 (c).

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

18. Qualitative task changes should be avoided because they reduce the likelihood of discovering modules. Evidence is required to assert qualitative task invariance. One kind of evidence is the pattern of factor effects: for each factor, each change in level should influence the same operations and leave the same other operations invariant. The usefulness of such evidence is one of several reasons for using factors with more than two levels. Unfortunately, few studies (and none of the three examples above) have done so. See Sternberg (2001), Appendices A.2.1 and A.9.2.

on performance in different tasks, rather than on different parts of the complex process used to carry out one task. The data pattern of interest is the selective influence of factors on tasks, i.e. the single and double dissociation of tasks. (A classical factor used in brain studies is the amount, usually presence versus absence, of damage in a particular brain region.) Although it may achieve other goals, task comparison is inferior to process decomposition for discovering the modular subprocesses of a complex process: The interpretation of task comparison often requires assuming a theory of the complex process in each task (specification of at least the set of subprocesses) and a theory of their relationship (which subprocesses are identical across tasks); the method includes no test of such assumptions. In contrast, process decomposition requires a theory of only one task, and, as illustrated by the examples above, incorporates a test of that theory.

Insert Fig. 1 about here

3.1 An Example of Task Comparison: Effects of Magnetic Brain Stimulation

An elegant example of task comparison is provided by Merabet et al. (2003) in their experiment on the effects of rTMS on subjective numerical scaling of two tactile perceptual dimensions, based on palpation of a set of tactile dot arrays by the fingers of one hand. The two dimensions were distance (between dots), and roughness. Where rTMS had an effect, it reduced the sensitivity of the obtained scale values to the differences among dot arrays. One measure of *relative sensitivity* is the slope, b, of the linear regression of post-rTMS scale values on non-rTMS scale values. If there were no effect we would have $b = 1.0$; the effect of rTMS is measured by $1 - b$. The data (Fig. 1) indicate that performance in the roughness-judgement task is influenced by rTMS of the contralateral somatosensory cortex (rTMS_s, Panel A1), but negligibly by rTMS of the contralateral occipital cortex (rTMS_o, Panel A2), while performance in the distance-judgement task is influenced by rTMS_o (Panel B2), but negligibly by rTMS_s (Panel B1), a double dissociation of the two tasks.[19] Plausible modular theories of the two tasks might include modules for control of stimulus palpation in each task ($\boldsymbol{\alpha}_d$, $\boldsymbol{\alpha}_r$), for generation of a complex percept ($\boldsymbol{\beta}_d$, $\boldsymbol{\beta}_r$), for extraction of the desired dimension ($\boldsymbol{\gamma}_d$, $\boldsymbol{\gamma}_r$), and for conversion of its value into a numerical response ($\boldsymbol{\delta}_d$, $\boldsymbol{\delta}_r$). Any or all of these processes might differ between tasks. The striking findings indicate that the members of one or more of these pairs of processes depend on different regions of the cortex. A weak pair of task theories might assert that $\boldsymbol{\alpha}_d$ and $\boldsymbol{\alpha}_r$ depend on occipital and somatosensory cortex, respectively[20] , but say nothing about the other processes. A stronger pair of task theories might include the assumptions that $\boldsymbol{\alpha}_d$ and $\boldsymbol{\alpha}_r$ are identical, ($\boldsymbol{\alpha}_d = \boldsymbol{\alpha}_r = \boldsymbol{\alpha}$), that $\boldsymbol{\beta}_d$ and $\boldsymbol{\beta}_r$ are identical, ($\boldsymbol{\beta}_d = \boldsymbol{\beta}_r = \boldsymbol{\beta}$), and that $\boldsymbol{\delta}_d$ and $\boldsymbol{\delta}_r$ are identical ($\boldsymbol{\delta}_d = \boldsymbol{\delta}_r = \boldsymbol{\delta}$). Given the results we could then conclude that it is $\boldsymbol{\gamma}_d$ and $\boldsymbol{\gamma}_r$ that must be implemented by processors in the different regions. And the results would then also suggest that none of processes $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, or $\boldsymbol{\delta}$ is sensitive to either rTMS_s or rTMS_o, perhaps indicating that they are implemented by processors in neither of the stimulated regions. But unfortunately the findings do not bear on the validity of such hypothesized task theories, weak or strong, or even on

---

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

20. Palpation for a distance judgement, but not a roughness judgement, might be associated with covert eye movements.

the question whether the operations in either task can be decomposed into modular subprocesses.

3.2 The Subtraction Method: Task Comparison with a Composite Measure

One variety of task comparison, devised by Donders (1868) for the RT measure, has also often been used with brain activation measures (e.g. Petersen, Fox, Posner, Mintun, and Raichle 1988). Suppose we are interested in studying a subprocess $\boldsymbol{\beta}$ of a complex process. If $\boldsymbol{\beta}$ were implemented by a localized neural processor P_β in region R_β, then the level of activation of R_β might be a pure measure of the subprocess. However, suppose instead that P_β is not localized (Haxby, Chapter 3, this volume), and what we have is a composite measure that reflects contributions from more than one subprocess. Under these conditions the subtraction method is sometimes used. This method requires three hypotheses. In a simple case they are: *H1* (Task Theory 1): Task 1 is accomplished by process $\boldsymbol{\alpha}$; *H2* (Task Theory 2): Task 2 is accomplished by $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$; *H3* (Combination Rule): Contributions u_α of $\boldsymbol{\alpha}$ and u_β of $\boldsymbol{\beta}$ to measure $M_{\alpha\beta}$ combine by *summation*. (Possible justifications of this combination rule include, for a brain-activation measure, an assumption that $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are implemented by different populations of neurons that contribute independently to the measure; and for an RT measure, an assumption that $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are arranged as stages.) Let the $M_{\alpha\beta}$ measures in the two tasks be M_1 and M_2. The hypotheses imply that M_1 and $M_2 - M_1$ are estimates of u_α and u_β, respectively, and can thus play the roles of pure measures of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. But having these measures provides no test of the hypotheses.[21] If summation proves to be incorrect as the combination rule, then other strategies may be available. For example, suppose measured activation were shown to be a decelerating function of the amount of neural activity, in particular, a logarithmic function. Then we would have $M_{\alpha\beta} = log(u_\alpha + u_\beta)$, and the subtraction method could be applied to the transformed activation measure $M'_{\alpha\beta} = exp(M_{\alpha\beta}) = u_\alpha + u_\beta$.

4. Neural Processing Modules Inferred from Activation Maps

Modular neural subprocesses can be discovered by applying process decomposition to the kinds of activation measures provided by PET and fMRI. Suppose localization of function, such that two such subprocesses, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, are implemented by different processors, P_α and P_β, in non-overlapping regions R_α and R_β. Then activation levels in R_α and R_β are pure measures of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and, with sufficiently precise data and factors that influence the subprocesses selectively, separate modifiability is easy to test.[22] However, if $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ 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 $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and to test separate modifiability we must know or show how their contributions to the activation measure are combined.[23]

---

21. One way to test the set of hypotheses is to extend it by finding two additional tasks that satisfy *H4* (Task 3 is accomplished by $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$) and *H5* (Task 4 is accomplished by $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$) and to extend *H3* by including $\boldsymbol{\gamma}$. The extended set of hypotheses can then be tested by confirming its prediction that $M_4 - M_3 = M_2 - M_1$.

22. 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 $Modules_3$ (Kanwisher *et al.* 2001), that stimuli more prototypical of those for which a processor is specialized will produce greater activation.

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

4.1 Some Desiderata for Process Decomposition using fMRI

1. The subject should be *performing a task* while measurements are taken. Even in sensory studies enough evidence has emerged favoring task effects at early levels of cortical processing so it is no longer appropriate merely to present stimuli to a passive observer.[24]

2. To increase the likelihood of discovering modules, the subject should be performing the *same* task as factor levels are varied. By 'same task' I mean that a persuasive argument can be made that for all combinations of factor levels, the same set of processing operations is involved, varying only 'quantitatively'.[25]

3. Because the *invariance* of one measure across levels of factor F is at least as important in our inferences as the *influence* of F on another measure, it is critical that we have some index of precision (such as a confi dence interval) for the size of an effect when that effect is claimed to be null.[26]

4. Because the effects of factors on activation levels of selected voxels are the quantities of interest, the (mean) activation levels should be reported, rather than only quantities (such as the t-values of 'statistical maps') which amalgamate the means and variances of such activation levels.

5. While factorial experiments are not required with pure measures, they are desirable, to assess the generality —hence persuasiveness —of the pattern of effects, as in the example below.

6. Each factor should be studied at more than two levels. The resulting tests of generality protect against being misled by patterns fortuitously associated with particular levels, and the data provide evidence about qualitative task invariance.

7. Especially if process decomposition can be based on behavioral data, such data should be taken concurrently with the fMRI data, to permit comparisons and mutual validation of the two kinds of decomposition, and to investigate the relations among $Modules_1$, $Modules_2$, and $Modules_3$.

4.2 An Example of Process Decomposition using fMRI: Number Comparison

Several of these desiderata are satisfi ed in a study by Pinel *et al.* (2001) that appears to involve pure measures. Subjects had to classify visually displayed numbers, k, as being greater or less than 65. One factor was notation (N), which could be Arabic numerals (e.g. '68') or number names (e.g. 'SOIXANTE-HUIT'). The other was numerical proximity (P), $|k - 65|$, with three levels. A similar study (Dehaene, 1996) had shown additive effects of N and P on $\overline{RT}$ (a composite measure); this was interpreted to indicate two modular subprocesses arranged as stages: encoding (**E**), influenced by N, w hich determines the meaning of the stimulus and is slower for number names than numbers, and comparison (**C**), influenced by P, w hich performs the comparison and is slower for greater (closer) proximities. In the new study, most of the brain regions reported whose activation is influenced by N or by P are influenced signifi cantly by only one of them, consistent with two separately modifi able neural processes $\boldsymbol{\varepsilon}$ and $\boldsymbol{\gamma}$ that are implemented by separately localized processors.[27] Averaging absolute effect sizes and SEs over

24. For example, with passive observing, different stimuli may attract attention differentially, which could influence activation measures.

25. See Note 18.

26. See Note 7 and Sternberg (2001), Sec. 1.5 and Appendix A.11.2. The same desideratum applies when an interaction (the modulation by one factor of the effect of another) is claimed to be null.

the regions of each type, for the nine N-sensitive regions the N effect was $0.17 \pm 0.05\%$ (median p-value = 0.01), while the P effect was $0.06 \pm 0.08\%$; for the seven P-sensitive regions the P effect was $0.32 \pm 0.10\%$ (median p-value = 0.01), while the N effect was $0.04 \pm 0.04\%$.[28] fMRI data from three well-behaved regions are shown in Fig. 2, Panels B, C, and D. The concurrently collected RT data (Fig. 2, Panels A) replicated the earlier study, suggesting that we associate the neural modules ($Modules_2$) $\boldsymbol{\varepsilon}$ and $\boldsymbol{\gamma}$ with the functional modules ($Modules_1$) **E** and **C**, respectively; it is important that the functional and neural modules were selectively influenced by the same factors. However, while the direction of the P effect was the same in all the brain regions it influenced, the direction of the N effect was not: the change from numeric to verbal notation increased activation in some regions (e.g. Fig. 2, Panels C) and decreased it in others (e.g. Fig. 2, Panels D).[29] This is consistent with different neural populations implementing the $\boldsymbol{\varepsilon}$ process, depending on the level of N.[30]

Insert Fig. 2 about here
----

5. Task-general processing modules

One plausible expectation is that different tasks are accomplished by different subsets of a small set of ‘basic’ modular processes. To test this expectation we need a reasonable number of tasks for which persuasively successful decompositions have been achieved.[31] On the other hand, to get adequate data we require subjects to learn a task to a point of stable performance. With such intensive practice, it seems possible that the brain is sufficiently flexible that special-purpose routines would be developed that are specific to that task. Thus, an alternative plausible expectation is that at least some modular subprocesses are task-specific rather than task-general. In that sense, perhaps there is no ‘fundamental architecture of the mind’, but rather a flexible

---

27. The way in which regions were selected may have contributed to this finding: regions found to be sensitive to P appear to have been selected only if the effect of P was not modulated by N. However, among the nine N-sensitive regions, the $N \times P$ interaction was significant in two, and the P effect significant in a third, in my analyses. How to interpret the coexistence of some regions showing selective influence with others showing joint influence (especially in combination with such persuasive RT data) is an important unresolved issue.

28. Effects are on the peak response over time in an event-related design (Pinel *et al.* 2001) for the ‘best’ voxel in each region, measured as a percentage of the intertrial activation level. (The mean peak response for the 16 voxels was 0.28%.) The ‘best’ voxel in an N (P) -sensitive region is the one whose N (P) effect is most significant. Such voxel selection can introduce unknown biases. SEs are based on between-subject variability over the nine subjects.

29. Two difficulties are created when an effect can have either sign, as in this case: (a) there may be voxels within which there is a mixture of effects in two directions, such that the effect appears to be null, and (b) even if such cancellation can be assumed not to occur, claims of null effects are harder to support statistically.

30. Without requiring it, this finding invites us to consider that there are two qualitatively different encoding processes, $\mathbf{E}_v$ and $\mathbf{E}_n$, one for each notation, rather than ‘one’ process whose settings depend on N. 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 3), but evidence for modularity emerges nonetheless. Whereas activation data (multidimensional) from such a simple (two-factor) experiment can support a claim of operations replacement, based on the idea that the processes implemented by different processors are probably different, RT data (unidimensional) that might support such a claim would require a more complicated experiment.

31. For speeded tasks in which processes are arranged as stages, Sanders (1998, Chapter 3) has amassed some suggestive evidence for a small set of functional modules.

architect, who has some stylistic tendencies worth studying.

6. Processes and Processors

What is the relation between $Module_3$ (a localized neural processor in region R that implements a particular process) and $Module_2$ (a modular subprocess of a complex neural process)? To answer this, consider one kind of evidence used to establish a $Module_3$: T_a and T_b are two classes of tasks, such that brain region R_α is activated during T_a, but not during T_b (or is activated *more* during one than the other), and such that we are willing to assume that all tasks T_a require a particular process $\boldsymbol{\alpha}$ to be carried out, whereas none of tasks T_b do. While it may seem plausible, such task-specificity of R_α does not imply that the process $\boldsymbol{\alpha}$ that it implements in a given task is a modular subprocess in the sense of being modifiable separately from other subprocesses in that task.[32]

7. Some Questions

1. Is separate modifiability too strong or too weak to be a useful criterion for partitioning a process? What are the relative merits of alternative criteria for modularity, and alternative approaches to module identification? Is the weaker differential modifiability[33] more useful than separate modifiability?

2. Consider modular functional processes ($Modules_1$) in a task (supported by behavioral evidence) and modular neural processes ($Modules_2$) 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 Pinel, et al. 2001 and Smulders *et al.* 1995, summarized above) in which both brain and behavioral measures are taken, both directed at process decomposition.[34]

3. Is the encapsulation of $Module_4$ equivalent to separate modifiability? Given a mental faculty ($Module_4$), must there be a corresponding specialized neural processor ($Module_3$)?

Acknowledgements

I thank Janice Hamer, Nancy Kanwisher, Allen Osman, Seth Roberts, two anonymous reviewers, and especially Teresa Pantzer for comments on the manuscript, and Nancy Kanwisher, David E. Meyer, and Allen Osman for helpful discussions. For providing unpublished details of their data I thank Stanislaus Dehaene, John Kounios, Lotfi Merabet, Alvaro Pascual-Leone, Philippe Pinel, and Fren Smulders.

REFERENCES

---

32. Suppose, for example, that $\boldsymbol{\alpha}$ provides a motivational or attentional resource that is required by one or more other processes $\boldsymbol{\gamma}$ that differ across tasks T_a. A change in $\boldsymbol{\alpha}$ would then induce a change in $\boldsymbol{\gamma}$, so they would not be separately modifiable.

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

34. One starting point would be to take cases where behavioral data already exist that persuasively favor a modular decomposition, as was done by Pinel *et al.* 2001, 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.

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

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.

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

Haxby, J. V. (2003). Analysis of topographically organized patterns of response in fMRI data: Distributed representations of objects in ventral temporal cortex. This volume, AAA-BBB.

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

Kounios, J. (1999). The additive-amplitude method —Neural modules revealed by ERPs. *International Journal of Psychophysiology 33*, 42-43.

Kounios, J. (2002). Functional neurocognitive modularity revealed by event-related brain potentials. Manuscript submitted for publication. [Change year to 2003; Manuscript under review?]

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

Merabet, L., Thut, G., Murray, B., Andrews, J., Hsiao, S., and Pascual-Leone, A. (2003). Feeling by sight or seeing by touch? Manuscript submitted for publication.

Osman, A., Bashore, T. R., Coles, M. G. H., Donchin, E., and 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., Mintun, M., and 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., and LeBihan, D. (2001). Modulation of parietal activation by semantic distance in a number comparison task. *NeuroImage 14*, 1013-1026.

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

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

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

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

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

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

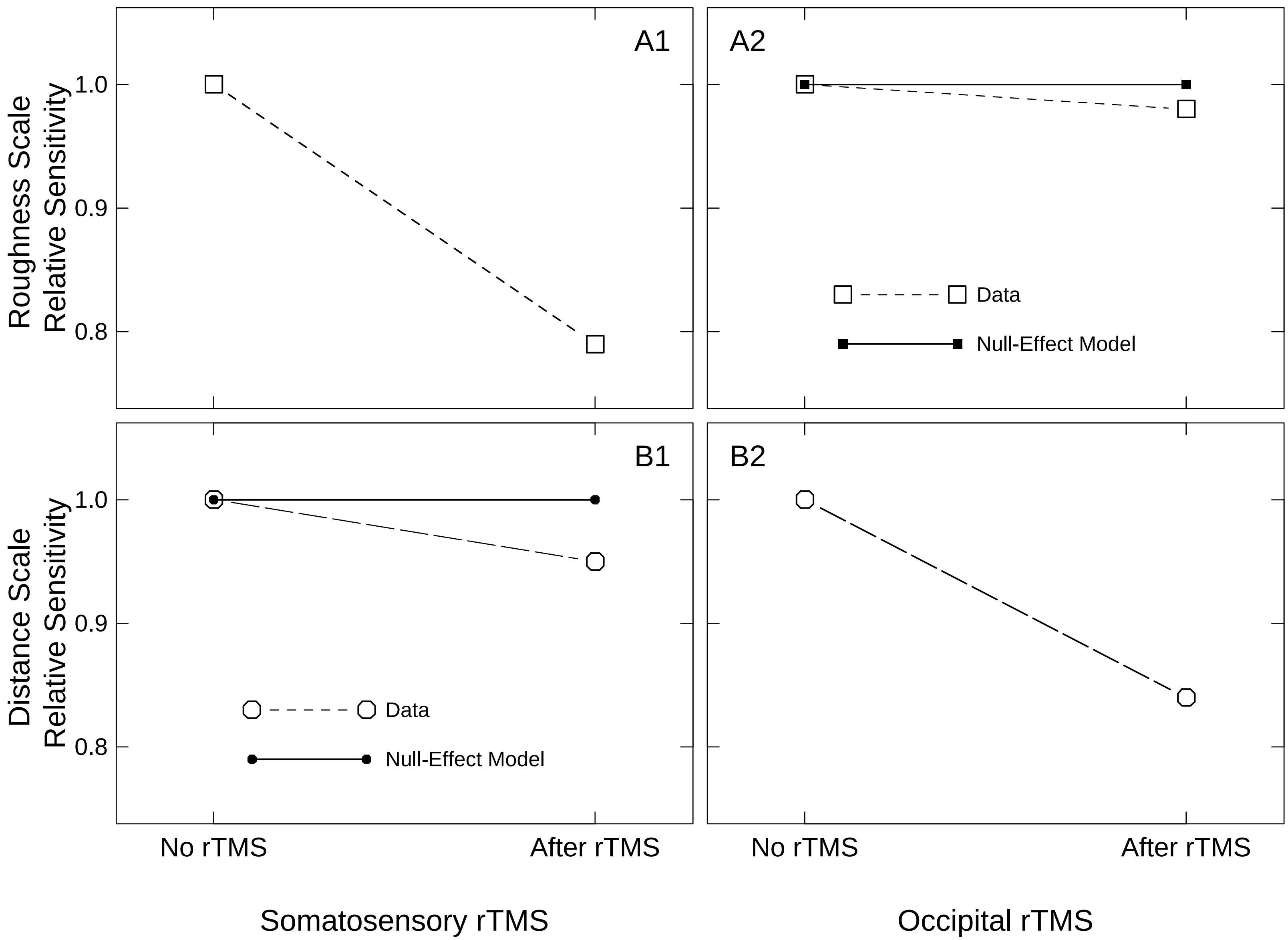
Sternberg, S. (2003). Process decomposition from double dissociation of subprocesses. *Cortex, 39*, 180-182.

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

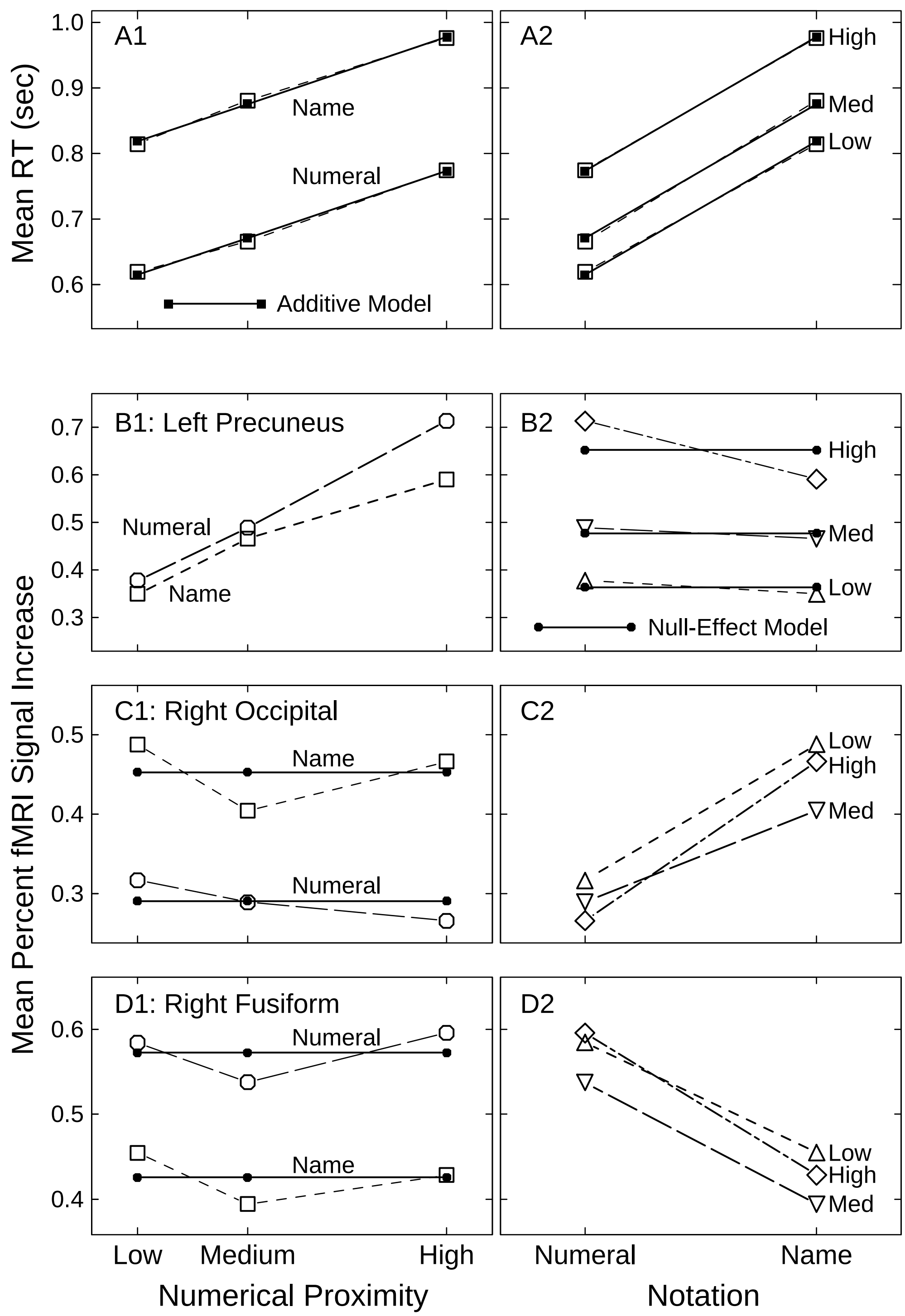
FIGURE CAPTIONS

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

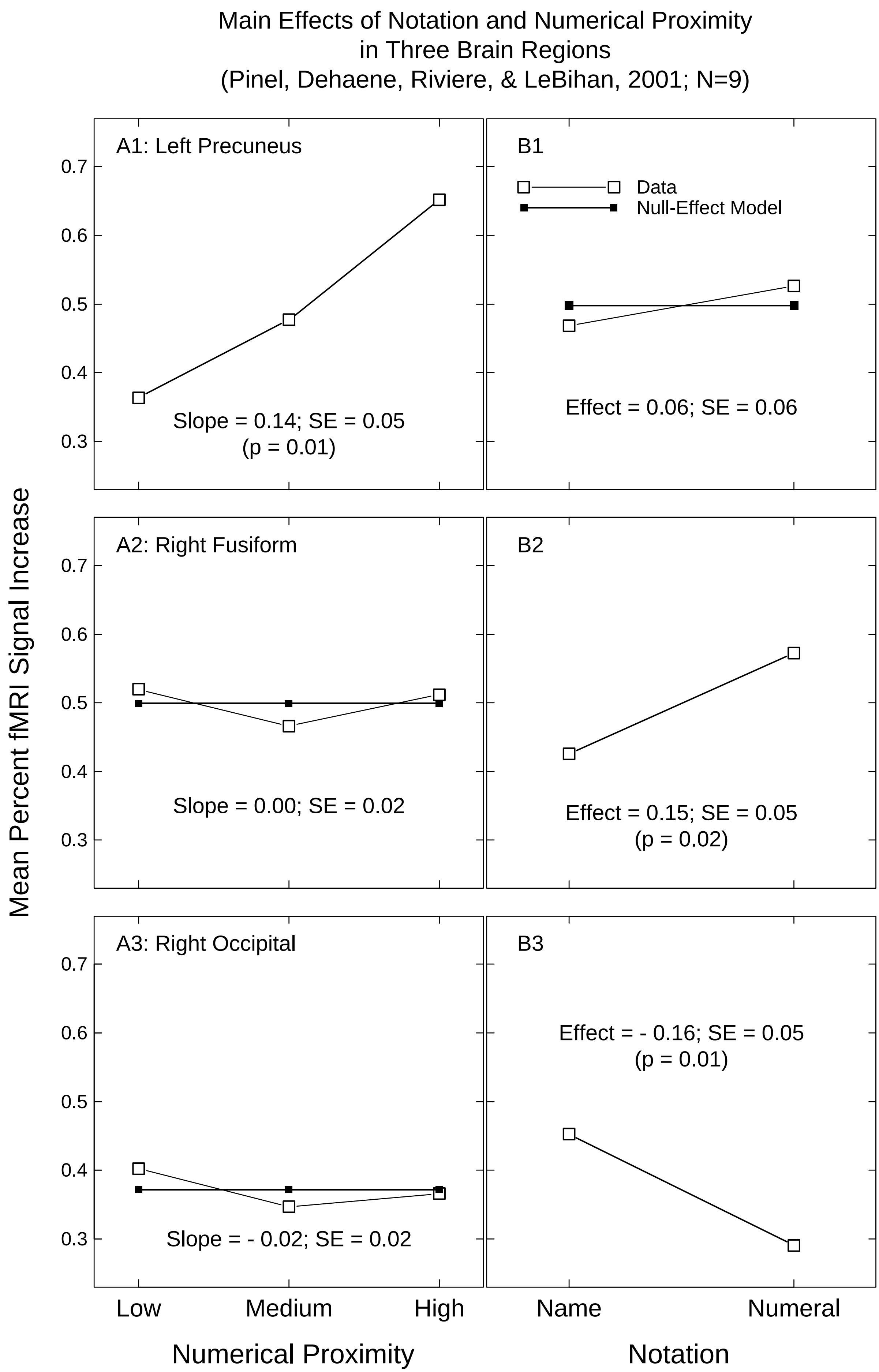
Fig. 2 Reaction-time and selected brain-activation data from Pinel *et al.* (2001). The same data are plotted on the left as functions of P (proximity), with N (notation) the parameter, and on the right as functions of N, with P the parameter. Means over subjects of median RTs for correct responses are 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.) Activation measures from three sample brain regions are shown in Panels B, C, and D, accompanied by fitted null-effect models in Panels B2, C1, and D1. Shown in Panels B1, C1, and D1, the main effects of P (from low to high, using fitted linear functions) are $0.29 \pm 0.09\%$ ($p \approx 0.01$), $-0.03 \pm 0.03\%$, and $0.00 \pm 0.04\%$, respectively. Shown in Panels B2, C2, and D2, the main effects of N are $-0.06 \pm 0.06\%$, $0.16 \pm 0.05\%$ ($p \approx 0.01$), and $-0.15 \pm 0.05\%$ ($p \approx 0.02$), respectively.



A1
A2
B1
B2
Roughness Scale
Relative Sensitivity
Distance Scale
Relative Sensitivity
1.0
0.9
0.8
Data
Null-Effect Model
No rTMS
After rTMS
Somatosensory rTMS
Occipital rTMS



A1
A2
Mean RT (sec)
1.0
0.9
0.8
0.7
0.6
Name
Numeral
Additive Model
High
Med
Low
B1: Left Precuneus
B2
0.7
0.6
0.5
0.4
0.3
Numeral
Name
High
Med
Low
Null-Effect Model
C1: Right Occipital
C2
0.5
0.4
0.3
Name
Numeral
Low
High
Med
D1: Right Fusiform
D2
0.6
0.5
0.4
Numeral
Name
Low
High
Med
Mean Percent fMRI Signal Increase
Low
Medium
High
Numerical Proximity
Numeral
Name
Notation


Main Effects of Notation and Numerical Proximity
in Three Brain Regions
(Pinel, Dehaene, Riviere, & LeBihan, 2001; N=9)
A1: Left Precuneus
Slope = 0.14; SE = 0.05
(p = 0.01)
B1
Data
Null-Effect Model
Effect = 0.06; SE = 0.06
A2: Right Fusiform
Slope = 0.00; SE = 0.02
B2
Effect = 0.15; SE = 0.05
(p = 0.02)
A3: Right Occipital
Slope = - 0.02; SE = 0.02
B3
Effect = - 0.16; SE = 0.05
(p = 0.01)
Mean Percent fMRI Signal Increase
0.7
0.6
0.5
0.4
0.3
Low
Medium
High
Name
Numeral
Numerical Proximity
Notation