A Hierarchical Process-Dissociation Model

Jeffrey N. Rouder
University of Missouri—Columbia

Jun Lu
American University

Richard D. Morey, Dongchu Sun, and Paul L. Speckman
University of Missouri—Columbia

In fitting the process-dissociation model (L. L. Jacoby, 1991) to observed data, researchers aggregate outcomes across participant, items, or both. T. Curran and D. L. Hintzman (1995) demonstrated how biases from aggregation may lead to artificial support for the model. The authors develop a hierarchical process-dissociation model that does not require aggregation for analysis. Most importantly, the Curran and Hintzman critique does not hold for this model. Model analysis provides for support of process dissociation—selective influence holds, and there is a dissociation in correlation patterns among participants and items. Items that are better recollected also elicit higher automatic activation. There is no correlation, however, across participants; that is, participants with higher recollection have no increased tendency toward automatic activation. The critique of aggregation is not limited to process dissociation. Aggregation distorts analysis in many nonlinear models, including signal detection, multinomial processing tree models, and strength models. Hierarchical modeling serves as a general solution for accurately fitting these psychological-processing models to data.

Keywords: process dissociation, human memory, Bayesian hierarchical models, aggregation bias

In many mid-level cognitive domains, such as memory and attention, researchers rely on relatively simple measurement models. Examples include the theory of signal detection (Green & Swets, 1966), the model of conjoint recognition (Brainerd, Reyna, & Mojardin, 1999), process dissociation (Jacoby, 1991), the power law of cognitive learning (Newell & Rosenbloom, 1981), and the similarity-choice model (Luce, 1963). These measurement models are nonlinear—that is, data are assumed to come from a more complex process than the addition of true scores and noise. The advantage of using nonlinearity is that the researcher can faithfully model psychological process and measure constructs inaccessible to analyses of variance and regression models.

In most analyses, researchers need to aggregate outcomes over people, items, or both to produce summary measures. This aggregation is not overly problematic for analysis with linear models (Rouder & Lu, 2005). Unfortunately, this sanguine state does not apply for nonlinear models. Over the past 50 years, several authors have critiqued applications of nonlinear measurement models that require aggregation of outcomes across participants, items, or both (e.g., Ashby, Maddox, & Lee, 1994; Curran & Hintzman, 1995; Estes, 1956; Estes & Maddox, 2005; Heathcote, Brown, & Mewhort, 2000; Luce, 1959; Rouder & Lu, 2005). Although these critiques appear in different domains and sometimes without reference to one another, they are part of a more general critique: Aggregation in nonlinear models often distorts the measurement of psychological process.

These critiques have not been well heeded in practice, and we suspect it is because of a lack of alternatives to aggregation. Aggregation seems necessary, especially in fields such as memory and psycholinguistics. In most experiments, a participant is tested once on an item; each participant–item combination is unrepli-
cated. The resulting performance for any participant–item combination is often dichotomous. For example, in stem completion, the stem is either completed with a studied word or not; in recognition memory, the item is judged either new or old. For these paradigms, the estimation of parameters in memory models relies on estimating probabilities of events. For example, process-dissociation parameters are functions of stem-completion probabilities; signal detection parameters are functions of old-response probabilities.

To generate proportions that estimate these probabilities, researchers aggregate over dichotomous outcomes; they aggregate across items, participants, or both. The critique that aggregation distorts analysis is problematic in many paradigms because aggregation seems necessary.

We have recently advocated nonlinear hierarchical models as an alternative to aggregation in experimental designs (Rouder & Lu, 2005; Rouder, Lu, Speckman, Sun, & Jiang, 2005; Rouder, Lu, et al., 2007; Rouder, Morey, Speckman, & Pratte, 2007). These hierarchical models account for participant and item variability. Unlike conventional hierarchical linear models (e.g., Raudenbush

Jeffrey N. Rouder and Richard D. Morey, Department of Psychological Sciences, University of Missouri—Columbia; Jun Lu, Department of Statistics, American University; Dongchu Sun and Paul L. Speckman, Department of Statistics, University of Missouri—Columbia.

We thank Mike Pratte, Courtney Gilbert, and Andrew Kent for help in running the reported experiments. We thank Timothy Curran and Jeffrey Toth for providing us with stimuli. We thank Larry Jacoby and Jeff Toth for supportive comments. This research was supported by National Science Foundation Grant SES-0351523 and National Institute of Mental Health Grant R01-MH071418.

Correspondence concerning this article should be addressed to Jeffrey N. Rouder, Department of Psychological Sciences, 210 McAlester Hall, University of Missouri, Columbia, MO 65211. E-mail: rouderj@missouri.edu
& Bryk, 2002), ours are nonlinear and are tailored for specific psychological-processing models. In this article, we present a hierarchical version of Jacoby’s (1991) process-dissociation model. To account for item and participant variability, there are separate recollection and automatic activation parameters for every participant–item combination. By imposing a hierarchical structure on these participant–item parameters, the model provides the needed constraint for accurate estimation without recourse to aggregation.

We divide the article as follows. In the next section, we provide examples of how aggregation compromises analysis with the Jacoby process-dissociation model. Then, we introduce the hierarchical process-dissociation model and show how it does indeed provide for accurate measurement. We analyze two experiments with the hierarchical model. Although the interpretation requires a caveat, the results provide support for the basic tenets of the Jacoby process-dissociation model.

The Jacoby Process-Dissociation Model

In the past 25 years, research in memory has been greatly influenced by multiple-systems approaches. These theories are supported, in part, by studies on anterograde amnesias. Amnesias exhibit greatly degraded performance on tasks that directly probe memory, such as cued recall. Yet they show somewhat intact performance on indirect tasks, such as priming (Cave & Squire, 1992) and stem completion (Graf & Schacter, 1985). These results provide evidence that there may be an implicit form of memory preserved in amnesias that is separate from conscious recollection (e.g., Schacter, 1992; Tulving & Schacter, 1990). Although there are several different descriptions of this implicit form (Schacter & Tulving, 1994), it is common to consider it automatic and based on familiarity.

The goal of Jacoby’s process-dissociation model is to separately measure conscious recollection and automatic activation. We describe its application to a stem-completion task. In the standard version of the task, participants study a list of words and are then tested on them. At test, items appear as to-be-completed stems (e.g., the stem br_ may be completed to form the word bread). In the process-dissociation procedure, there are two testing conditions: an include and an exclude condition. In the include condition, participants are instructed to complete the stem with a previously studied word. In the exclude condition, participants are instructed to complete the stem with a word not studied. In the include condition, either successful conscious recollection or successful automatic activation is sufficient for successful stem completion. In the exclude condition, successful automatic activation and a failure of conscious recollection are sufficient and necessary for successful stem completion. By contrasting these conditions, an estimate of the probability of successful recollection and successful automatic activation may be obtained. The process-dissociation model is formally given as follows: Let \( I \) and \( E \) be the probabilities of completing a stem with a studied word in the include and exclude conditions, respectively. For simplicity, recollection and automatic activation are assumed to be all-or-none, dichotomous processes. Let \( R \) and \( A \) be the probability of conscious recollection and automatic activation, respectively. The Jacoby process-dissociation model is given by

\[
I = R + (1 - R)A
\]  

and

\[
E = (1 - R)A.
\]

Estimators may be obtained by solving the above equations for \( R \) and \( A \):

\[
R = \frac{I - E}{1 - (I - E)}
\]

and

\[
A = \frac{E}{1 - (I - E)}.
\]

where \( I \) and \( E \) are the proportion of stems completed with studied words in the include and exclude conditions, respectively.

One subtle aspect of the process-dissociation equations is that they are compatible with many psychological models (Buchner, Erdfelder, & Vaterrodt-Plunneck, 1995). Figure 1 shows a general model in which the probability of automatic activation depends on whether recollection occurred (\( A^p \) denotes the probability of automatic activation given retrieval success; \( A \) denotes the probability of automatic activation given retrieval failure). The probabilities of stem completion in the include and exclude conditions are \( R + (1 - R)A \) and \( (1 - R)A \), respectively, which are the same as in Equations 1 and 2. The parameter \( A^p \) does not enter into the equations. Consequently, only automatic activation conditioned on recollection failure is measured in the process-dissociation framework (see Buchner et al., 1995).

The measurement of automatic activation conditioned on retrieval failure is undesirable. Most psychologists conceptualize automatic processes as quicker than deliberative ones (Kahneman, 1973), implying that automatic activation could not be dependent on conscious recollection. Moreover, early memory-system theorists explicitly championed the independence of implicit and explicit systems (e.g., Schacter, 1992). To make the measurement of \( A^p \) interpretable and psychologically plausible, it is widely assumed that automatic activation and recollection are independent (Jacoby, 1998). This independence assumption implies that \( A^p = A \) and that measurement of \( A \) applies to cases in which recollection succeeds as well as fails. Importantly, independence, in this case, means that

\[
\begin{align*}
\text{Include Trial} & \quad \frac{R}{1 - A} + 1 - A \\
\text{Exclude Trial} & \quad \frac{R}{1 - A} + 1 - A
\end{align*}
\]

Figure 1. A process-dissociation model for stem completion in which automatic activation depends on whether recollection succeeds or fails. The parameter \( R \) denotes the probability of successful conscious recollection. The parameters \( A^p \) and \( A \) denote the probability of successful automatic activation contingent on recollection success and failure, respectively, + = stem completed with studied item; 0 = stem completed with item not studied.
the model does not imply any specific time course of processing. Recollection could come after automatic activation or be performed simultaneously with it. All that is implied is that the probability of successful automatic activation does not depend on the state of recollection and vice versa.

Aggregation in the Process-Dissociation Model

Our main concern is about the effects of aggregation in applying the process-dissociation model. Curran and Hintzman (1995) first explored this concern, and our development follows theirs. The first step is expanding the notation of the model such that it can account for participant and item effects. Let \( I_{ij} \) and \( E_{ij} \) be the probability that the \( i \)th participant completes the \( j \)th stem with the corresponding studied item in the include and exclude conditions, respectively. The expanded process-dissociation model is

\[
I_{ij} = R_{ij} + (1 - R_{ij})A_{ij}
\]

and

\[
E_{ij} = (1 - R_{ij})A_{ij},
\]

where \( R_{ij} \) and \( A_{ij} \) are the probabilities that the \( i \)th participant consciously recollects and elicits automatic activation from the \( j \)th item, respectively. To allow the measurement of \( A_{ij} \) to be interpretable and psychologically plausible, we assume that recollection and automatic activation are independent but only at the level of participant–item combinations. That is, for the \( i \)th participant observing the \( j \)th item, the success of recollection is independent of the success of automatic activation. This form of independence refers to the independence of the underlying psychological processes. Consequently, we term it process independence.

Because each participant–item combination is unreplicated, researchers do not estimate \( I_{ij} \) and \( E_{ij} \). Instead, they aggregate. We consider here aggregation of items to compute participant-specific proportions and denote these by \( \hat{I} \) and \( \hat{E} \). The goal, then, is to recover participant-specific averaged parameters \( \hat{R} = \frac{\sum R_{ij}}{J} \) and \( \hat{A} = \frac{\sum A_{ij}}{J} \), where \( J \) is the number of items. The question at hand is whether the estimation procedure of Equations 3 and 4 yield good estimates of \( \hat{R} \) and \( \hat{A} \).

Curran and Hintzman (1995) argued that participants who are particularly good at conscious recollection may also be particularly good at eliciting automatic activation. Likewise, items that are particularly easy to consciously recollect may elicit a greater automatic activation. Additionally, they argued that even interactions may be correlated: Greater conscious recollection of particular participant–item combinations may be accompanied by greater automatic activation of those combinations. None of these correlations violate process independence. Process independence describes what happens conditioned on a particular participant and item. These correlations are at a secondary level and describe the distribution of item and participant effects rather than the underlying psychological processes.

Curran and Hintzman (1995) showed that these secondary correlations lead to systematic distortions in automatic activation estimates from aggregation. In our notation, estimates of automatic activation derived from \( \hat{I} \) and \( \hat{E} \), do not converge in the large-sample limit to \( \hat{A} \). Curran and Hintzman provided a cogent informal explanation of how secondary correlations distort estimates of automatic activation (see also Curran & Hintzman, 1997; Jacoby, Begg, & Toth, 1997). We provide a simple example.

Consider the case of a single participant, and suppose that half of the studied items are easy and the other half are hard. Assume easy items are consciously recollected with probability \( R = .5 \) and automatically activated with probability \( A = .5 \). Hard items have corresponding probabilities \( R = .1 \) and \( A = .1 \). Process independence may be assumed; that is, recollection for easy items is independent of automatic activation for easy items and the same for hard items. Following Equations 1 and 2, the true values of stem completion for easy items are \( I = .75 \) and \( E = .25 \) for the include and exclude conditions, respectively. The true values for hard items are \( I = .19 \) and \( E = .09 \), respectively. Aggregating these probabilities across items yields stem-completion probabilities of \( I = .47 \) and \( E = .17 \). The goal is to estimate item-averaged conscious recollection and automatic activation. True values of these item averages are \( R = .3 \) and \( A = .3 \). Estimates are obtained by substituting probabilities \( \hat{I} = .47 \) and \( \hat{E} = .17 \) into Equations 3 and 4. The resulting estimates are \( \hat{R} = .3 \) and \( \hat{A} = .243 \).

Although the estimate of conscious recollection is accurate, the estimate of automatic activation is too low. The underestimation reflects the fact that automatic activation is measurable only when recollection fails. If the two processes are correlated, instances of recollection failure will be greatly associated with instances of automatic activation failure. Importantly, this bias is asymptotic as we used true probabilities as data.

Perhaps the most powerful element of the Curran and Hintzman (1995) critique is that they assumed correlations that do not violate process independence. These correlations violate independence where it is secondary, that is, across the distributions of items and participants. Therefore, the Curran and Hintzman critique shows how secondary sources of correlation, which are often not of general interest, play havoc in analysis.

Curran and Hintzman (1995) showed how this bias threatens the gold-standard evidence for the veracity of Jacoby’s process-dissociation model: selective influence. In a selective influence test, the researcher manipulates a select variable. In the model, there is a prespecified set of parameters that should be affected by this manipulation and a complementary set that should not (estimates should be the same across different levels of the manipulation within sampling error). If selective influence holds, the results are interpreted as support for the model (see Rouder, 2004).

One popular selective influence test is to assess whether study duration affects recollection but not automatic activation. The general finding is that study duration does indeed affect recollection but not automatic activation (Jacoby et al., 1997; Jacoby, Toth, & Yonelinas, 1993; cf. Curran & Hintzman, 1995). This selective influence finding has been crucial: It provides evidence for the validity of the model with the independence assumption (see Cowan & Studler, 1996, for extended discussion).

Curran and Hintzman (1995) argued that the systematic bias undermines tests of the invariance of \( A \). Here, we provide a simple example: Consider the case in which both easy and hard items are presented for short and long study times. Suppose that the true values in the short study-time condition are \( R = .5 \), \( A = .5 \), for easy items and \( R = .1 \), \( A = .1 \), for hard items. As discussed before, the true averages are \( \hat{R} = .3 \), \( \hat{A} = .3 \), while the estimates are
\( \hat{R} = .3, \hat{A} = .243 \). Suppose that true values in the long study-time condition are \( R = .9, A = .9 \), for the easy items and \( R = .15, A = .15 \), for the hard items. The true averages for the long study-time condition are \( \bar{R} = .525, \bar{A} = .525 \). The effect of study time, on average, is to raise both \( R \) and \( A \) to .525 from .3. The invariance of \( A \) should be rejected, for \( A \) varies as much as \( R \) does.

The estimates in this long study-time condition, obtained via Equations 3 and 4, are \( \hat{R} = .525, \hat{A} = .229 \). Unfortunately, the estimates of \( \bar{A} \) are nearly equal in the two study-time conditions (.243 vs. .229), even though the true difference is large. In this case, this estimated invariance is an artifact of aggregation. As Curran and Hintzman explained, the underestimation of \( \bar{A} \) may be an artifact of aggregation. We addressed this problem by developing a model for which aggregation is not necessary. Afterward, we applied the model to the results of a selective influence experiment (Experiment 1) to decide whether selective influence truly holds.

A Hierarchical Process-Dissociation Model

Model Specification

Our approach is to extend Jacoby’s process-dissociation model with parameters that account for participant and item variability. We follow the development in Lu, Speckman, Sun, and Rouder (2007), who provided the following three-level hierarchical model. The first level is the process-dissociation model:

\[
I_{ijk} = R_{ijk} + (1 - R_{ijk})A_{ijk},
\]

and

\[
E_{ijk} = (1 - R_{ijk})A_{ijk},
\]

where as before \( i \) and \( j \) index participants and items, respectively, and \( k \) indexes study-duration condition. For each participant \( i \), each item \( j \) appears in only one of the study conditions. Therefore, we cannot obtain all combinations of participants, items, and study durations, necessitating further constraint.

The second level describes how \( R_{ijk} \) and \( A_{ijk} \) are constrained to reflect only main effects of participants, items, and conditions. Parameters \( R_{ijk} \) and \( A_{ijk} \) are probabilities that range between zero and one. It is convenient and common to place linear models on parameters that span the reals, and one popular method is to model probit-transformed probabilities. This transform and the closely related logit transform are the main staple in psychometrics (Nunnally & Bernstein, 1994). Let \( a_{ijk} \) and \( r_{ijk} \) denote the transformed parameters, that is, \( A_{ijk} = \Phi(a_{ijk}) \) and \( R_{ijk} = \Phi(r_{ijk}) \), where \( \Phi \) is the standard normal cumulative distribution function. The curve in Figure 2 shows the probit transform.

The second level is a main effects model on transformed parameters \( a \) and \( r \):

\[
r_{ijk} = \alpha^{(\text{rec})}_i + \beta^{(\text{rec})}_j + \mu^{(\text{rec})}_k,
\]

and

\[
a_{ijk} = \alpha^{(\text{act})}_i + \beta^{(\text{act})}_j + \mu^{(\text{act})}_k.
\]

Parameters \( \alpha^{(\text{rec})}_i \) and \( \alpha^{(\text{act})}_i \) denote the \( i \)th participant’s conscious recollection and automatic activation abilities, respectively; parameters \( \beta^{(\text{rec})}_j \) and \( \beta^{(\text{act})}_j \) denote the \( j \)th item’s propensity to elicit conscious recollection and automatic activation, respectively; and parameters \( \mu^{(\text{rec})}_k \) and \( \mu^{(\text{act})}_k \) denote the overall level of conscious recollection and automatic activation in the \( k \)th condition, respectively. Figure 2 shows an example of the model on conscious recollection for values of a condition effect \( (\mu^{(\text{rec})}_k = .5) \), a participant effect \( (\alpha^{(\text{rec})}_i = -1) \) and an item effect \( (\beta^{(\text{rec})}_j = 1.5) \). After adding these main effects and transforming, the corresponding value of parameter \( R_{ijk} \) is .84. To maintain identifiability, participant and item effects are treated as zero-centered while condition effects are unconstrained. This approach of using additive models on appropriately transformed variables is analogous to the Rasch model in item response theory (Lord & Novick, 1968).

The models in Equations 9 and 10 are not identifiable without additional constraint. For example, a constant \( C \) may be added to \( \alpha^{(\text{rec})}_i \) and subtracted from \( \beta^{(\text{rec})}_j \) with no effect on \( r_{ijk} \). If people and items were fixed effects, then the constraint would be \( \Sigma\beta^{(\text{rec})}_j = 0 \), \( \Sigma\alpha^{(\text{rec})}_i = 0, \Sigma\beta^{(\text{act})}_j = 0, \) and \( \Sigma\beta^{(\text{act})}_j = 0 \). In memory experiments, both participants and items are sampled from larger populations to which researchers wish to generalize findings. In this case, it is appropriate to consider participant and item effects as random (Clark, 1973). A natural choice in this application is to model participant and item effects as draws from bivariate normals with arbitrary covariance matrices:
The approach we take to analyze the model is Bayesian. Although many researchers prefer the Bayesian approach on philosophical grounds (e.g., Berger & Wolport, 1988; Edwards, Lindman, & Savage, 1963; Wagenmakers, in press), we have far more pragmatic reasons for this choice. Simply put, while Bayesian implementation is relatively straightforward, a more classical (frequentist) implementation is not. Moreover, Bayesian modeling is in the mainstream in statistics, economics, epidemiology, and psychometrics. To help bring Bayesian analysis to the mainstream of experimental psychology, we have provided a tutorial on Bayesian hierarchical analysis for experimental psychologists (see Rouder & Lu, 2005).

In Bayesian analysis, the researcher provides a prior distribution for the parameters. In many applications, the prior may be made noninformative or nearly noninformative, as they are in this application. In this model, priors are needed for grand means (\(\mu_k\)) and \(\mu^{(r)}\) and covariance matrices (\(\Sigma_a\) and \(\Sigma_B\)). The priors for these parameters, as well as a general outline of model analysis, are provided in the Appendix. Software in the R language is provided at http://web.missouri.edu/~umcaspsychpcl/code. Lu et al. (2007) provided more complete coverage of the model, including derivations of conditional posterior distributions. Lu et al. also studied the behavior of the model under a number of alternative priors.

Model Assumptions

There are three new assumptions in the hierarchical process-dissociation model: the probit link; the bivariate normal distribution on random effects; and the additivity between participants, items, and conditions. The first two of these, the probit link and the bivariate normal, are common choices in models of this type. The third assumption about additivity deserves more scrutiny. True underlying additivity implies that there are no interactions. An interaction in this context would imply that a particular participant-item combination is different from the sum of main effects. Curran and Hintzman (1995) argued that these interactions are not only plausible but may be positively correlated across recollection and automatic activation. Such unmodeled correla-

\[
\begin{align*}
\alpha_i^{(r)} &\sim N(0, \Sigma_a), \quad i = 1, \ldots, I, \\
\beta_j^{(r)} &\sim N(0, \Sigma_B), \quad j = 1, \ldots, J,
\end{align*}
\]

where \(I\) and \(J\) reference the number of participants and items, respectively. The covariance matrices describe how the population of random effects are distributed. The covariance matrix \(\Sigma_a\) has three unique elements: the variance of participant effects on conscious recollection, the variance of participant effects on automatic activation, and the covariance among these participant effects. If the covariance is positive, then participants who have a higher degree of conscious recollection tend to have a higher degree of automatic activation. The model is flexible and may account for any degree of correlation: positive, negative, or none. These covariances are free parameters that are estimated rather than assumed.

Model Assumptions

There are three new assumptions in the hierarchical process-dissociation model: the probit link; the bivariate normal distribution on random effects; and the additivity between participants, items, and conditions. The first two of these, the probit link and the bivariate normal, are common choices in models of this type. The third assumption about additivity deserves more scrutiny. True underlying additivity implies that there are no interactions. An interaction in this context would imply that a particular participant-item combination is different from the sum of main effects. Curran and Hintzman (1995) argued that these interactions are not only plausible but may be positively correlated across recollection and automatic activation. Such unmodeled correla-

\[
\begin{align*}
\alpha_i^{(r)} &\sim N(0, \Sigma_a), \quad i = 1, \ldots, I, \\
\beta_j^{(r)} &\sim N(0, \Sigma_B), \quad j = 1, \ldots, J,
\end{align*}
\]

where \(I\) and \(J\) reference the number of participants and items, respectively. The covariance matrices describe how the population of random effects are distributed. The covariance matrix \(\Sigma_a\) has three unique elements: the variance of participant effects on conscious recollection, the variance of participant effects on automatic activation, and the covariance among these participant effects. If the covariance is positive, then participants who have a higher degree of conscious recollection tend to have a higher degree of automatic activation. The model is flexible and may account for any degree of correlation: positive, negative, or none. These covariances are free parameters that are estimated rather than assumed.

Model Assumptions

There are three new assumptions in the hierarchical process-dissociation model: the probit link; the bivariate normal distribution on random effects; and the additivity between participants, items, and conditions. The first two of these, the probit link and the bivariate normal, are common choices in models of this type. The third assumption about additivity deserves more scrutiny. True underlying additivity implies that there are no interactions. An interaction in this context would imply that a particular participant-item combination is different from the sum of main effects. Curran and Hintzman (1995) argued that these interactions are not only plausible but may be positively correlated across recollection and automatic activation. Such unmodeled correla-

\[
\begin{align*}
\alpha_i^{(r)} &\sim N(0, \Sigma_a), \quad i = 1, \ldots, I, \\
\beta_j^{(r)} &\sim N(0, \Sigma_B), \quad j = 1, \ldots, J,
\end{align*}
\]

where \(I\) and \(J\) reference the number of participants and items, respectively. The covariance matrices describe how the population of random effects are distributed. The covariance matrix \(\Sigma_a\) has three unique elements: the variance of participant effects on conscious recollection, the variance of participant effects on automatic activation, and the covariance among these participant effects. If the covariance is positive, then participants who have a higher degree of conscious recollection tend to have a higher degree of automatic activation. The model is flexible and may account for any degree of correlation: positive, negative, or none. These covariances are free parameters that are estimated rather than assumed.

Model Assumptions

There are three new assumptions in the hierarchical process-dissociation model: the probit link; the bivariate normal distribution on random effects; and the additivity between participants, items, and conditions. The first two of these, the probit link and the bivariate normal, are common choices in models of this type. The third assumption about additivity deserves more scrutiny. True underlying additivity implies that there are no interactions. An interaction in this context would imply that a particular participant-item combination is different from the sum of main effects. Curran and Hintzman (1995) argued that these interactions are not only plausible but may be positively correlated across recollection and automatic activation. Such unmodeled correla-

\[
\begin{align*}
\alpha_i^{(r)} &\sim N(0, \Sigma_a), \quad i = 1, \ldots, I, \\
\beta_j^{(r)} &\sim N(0, \Sigma_B), \quad j = 1, \ldots, J,
\end{align*}
\]

where \(I\) and \(J\) reference the number of participants and items, respectively. The covariance matrices describe how the population of random effects are distributed. The covariance matrix \(\Sigma_a\) has three unique elements: the variance of participant effects on conscious recollection, the variance of participant effects on automatic activation, and the covariance among these participant effects. If the covariance is positive, then participants who have a higher degree of conscious recollection tend to have a higher degree of automatic activation. The model is flexible and may account for any degree of correlation: positive, negative, or none. These covariances are free parameters that are estimated rather than assumed.

Model Assumptions

There are three new assumptions in the hierarchical process-dissociation model: the probit link; the bivariate normal distribution on random effects; and the additivity between participants, items, and conditions. The first two of these, the probit link and the bivariate normal, are common choices in models of this type. The third assumption about additivity deserves more scrutiny. True underlying additivity implies that there are no interactions. An interaction in this context would imply that a particular participant-item combination is different from the sum of main effects. Curran and Hintzman (1995) argued that these interactions are not only plausible but may be positively correlated across recollection and automatic activation. Such unmodeled correla-

\[
\begin{align*}
\alpha_i^{(r)} &\sim N(0, \Sigma_a), \quad i = 1, \ldots, I, \\
\beta_j^{(r)} &\sim N(0, \Sigma_B), \quad j = 1, \ldots, J,
\end{align*}
\]

where \(I\) and \(J\) reference the number of participants and items, respectively. The covariance matrices describe how the population of random effects are distributed. The covariance matrix \(\Sigma_a\) has three unique elements: the variance of participant effects on conscious recollection, the variance of participant effects on automatic activation, and the covariance among these participant effects. If the covariance is positive, then participants who have a higher degree of conscious recollection tend to have a higher degree of automatic activation. The model is flexible and may account for any degree of correlation: positive, negative, or none. These covariances are free parameters that are estimated rather than assumed.

Model Assumptions

There are three new assumptions in the hierarchical process-dissociation model: the probit link; the bivariate normal distribution on random effects; and the additivity between participants, items, and conditions. The first two of these, the probit link and the bivariate normal, are common choices in models of this type. The third assumption about additivity deserves more scrutiny. True underlying additivity implies that there are no interactions. An interaction in this context would imply that a particular participant-item combination is different from the sum of main effects. Curran and Hintzman (1995) argued that these interactions are not only plausible but may be positively correlated across recollection and automatic activation. Such unmodeled correla-
tions would exert downward bias on the estimation of automatic activation. The argument is analogous to that used to show downward bias with easy and hard items in the previous example. Therefore, it is necessary to evaluate the practical consequences of the additivity assumption. The results of Experiments 1 and 2 serve as a guide for evaluation; therefore, we defer discussion of the consequences of additivity until after presenting the experiments. To foreshadow, these consequences are not large and do not change our conclusions.

Experiment 1

Although selective influence has been previously established, it may be an artifact of aggregation. Experiment 1 was designed to assess whether selective influence holds even when the data are analyzed without recourse to aggregation via the hierarchical process-dissociation model.

Experiment 1 is modeled after Curran and Hintzman’s (1995) Experiments 4 and 5. Participants were asked to study a list of items and then asked to stem-complete these items in accordance with include and exclude instructions. Following Curran and Hintzman, we manipulated study duration through three levels: items not studied, items studied for 1 s, and items studied for 10 s. The predictions about the effect of study time on recollection are straightforward: More study time promotes greater recollection. More to the point, recollection should be better for items studied for 10 s than for items studied for 1 s. The predictions about automatic activation are more subtle. Automatic activation may rise with the duration of study until a point, after which additional study does not raise automatic activation further. On the basis of the relationship between priming and study duration, Neill, Beck, Bottalico, and Molloy (1990) concluded that additional study past 1 s does not increase automatic activation. Following previous process-dissociation researchers, we predicted that the automatic activation for the 1-s and 10-s study conditions would be about equal and greater than the no-study condition.

Method

The method for Experiment 1 closely follows that of Jacoby et al. (1993) and Curran and Hintzman (1995).

Participants. Sixty-six undergraduate students at the University of Missouri—Columbia received partial class credit for participation.

Materials. The items consisted of 116 five-letter words from Jacoby et al. (1993). The stems were the first three letters of the items followed by two blanks. Each stem yielded only two valid completions, one of which was the item. For example, item lapel yielded a stem lap___ which has valid completions of lapel and lapse. Twenty of these items were used as fillers; the remaining 96 comprised the critical set used in analysis. The 96 critical items were those used by Curran and Hintzman (1995, Experiments 4 and 5). All items were presented and tested on a PC. The size of each letter on the display was 5 mm × 4 mm. All items were presented in an uppercase type.

Design. The main independent variables were study duration and instructions at test. Study duration was manipulated through three levels (no study, 1-s presentation, 10-s presentation). Study duration was blocked—all items of the same study duration were presented together to prevent selective rehearsal of short-duration items during the longer study period of the long-duration items. The order of study condition blocks was counterbalanced across participants. Testing instructions were either the include or the exclude instructions. These were randomly intermixed throughout the test list. Altogether, there were six conditions obtained by crossing study duration and item instructions. These six conditions were crossed with participants in a balanced Latin square design.

Procedure. The experiment was divided into instructions, study, and test phases. In the instructions phase, participants were shown a sample item and a sample stem and were taught the include and exclude instructions. In the study phase, items were presented sequentially and in the center of the display. The order of items within a block was randomized, with the exception that each block began and ended with five filler items.

The test phase began on completion of the study phase. Participants were cued with the simple phrases “Use a word from the list” or “Do NOT use a word from the list” to indicate whether a stem was to be completed under the include or exclude conditions, respectively. To make these cues more salient, the former was presented in green, and the latter was presented in red. Participants typed in the two letters that completed the stem. They were allowed to complete the stem with “XX” if they could not think of a valid completion. The order of stems was randomized. Participants were self-paced at test. Total time to complete a session was approximately 45 min.

Results

The data were submitted to conventional (aggregation-based) analysis and to the hierarchical model analysis.

Conventional analysis. Stem-completion rates for include and exclude conditions were generated for each participant by aggregating over items. The resulting rates were submitted to Equations 3 and 4 to compute estimates of \( \hat{R}_{ik} \) and \( \hat{A}_{ik} \) for each participant in each condition. In some cases, especially those in which items were not studied, the proportion of stem completion was greater in the exclude than in the include condition. When data are ordered this way, the resulting estimates from Equation 3 for \( \hat{R}_{ik} \) are negatively valued. Whereas negative values are not interpretable in the process-dissociation model, we used maximum likelihood subject to the constraint that parameter estimates be positive. Constrained maximization was performed with the simplex algorithm (Nelder & Mead, 1965). Means of \( \hat{R}_{ik} \) and \( \hat{A}_{ik} \) are shown in the first row of Table 1. We refer to these estimates as the item-aggregation estimates.

<table>
<thead>
<tr>
<th></th>
<th>Recollection</th>
<th>Automatic activation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No study</td>
<td>1 s</td>
</tr>
<tr>
<td>Item aggregation</td>
<td>.100</td>
<td>.233</td>
</tr>
<tr>
<td>Double aggregation</td>
<td>.065</td>
<td>.216</td>
</tr>
<tr>
<td>Bayesian equivalent</td>
<td>.042</td>
<td>.220</td>
</tr>
</tbody>
</table>
In addition to item aggregation, we also computed estimates from aggregating stem-completion rates for each condition over both participants and items. These double-aggregated stem-completion rates were submitted to Equations 3 and 4 to compute estimates of \( R_{\lambda_{ik}} \) and \( A_{\lambda_{ik}} \) for each condition. These are shown in the second row of Table 1. The third row is from the hierarchical Bayesian model, which is discussed subsequently.

The aggregation-based analyses tell a consistent and plausible story. Greater study durations yielded increasing levels of recollection. The story for automatic activation is only slightly more complex. Studied items elicited greater automatic activation than unstudied items. The effect of duration was modest, with somewhat more automatic activation in the 1-s study condition than in the 10-s study condition. (Inferential statements are made with the hierarchical Bayesian model discussed subsequently.)

There are some differences between the two aggregation methods, especially for the no-study condition. The difference reflects a small-sample bias in estimation. As the true value of \( R \) becomes vanishingly small, half of all stem completions are better for the exclude than for the include condition. For this half, the estimate of \( R \) is forced to be positive by the constrained maximum-likelihood procedure. For the other half, the half in which stem completion is better for the include than the exclude condition, the estimate of \( R \) from Equation 3 is small but positive. The average of these two cases yields a small upward bias in estimating \( R \). Fortunately, these biases decrease with increasing sample size. Not surprisingly, then, the bias is more noticeable for item aggregation than for double aggregation as the sample sizes are smaller for the former than for the latter. Because item-aggregation estimates have some nonsymptotic bias for conditions with low recollection, they do not serve well in this application. Consequently, we use the double-aggregation estimates for comparison with the hierarchical Bayesian model.

Many previous researchers in process dissociation have not discussed the problem of negative recollection estimates when overall recollection estimates are very low. We strongly suspect researchers have admitted negative values of recollection with item aggregation; all reports simply stated that their estimates of \( R \) R were forced to be positive by the probit transform. Negative individual estimates were admitted. Fortunately, the hierarchical model allows researchers to avoid this issue as estimates are constrained to be positive by the probit transform.

There is an inconsistency in the literature with regard to whether automatic activation decreases with increasing study time. Jacoby and colleagues (see Jacoby et al., 1997, Table 1) reported that the average decrease in A with study time across eight experiments is .003. In contrast, Curran and Hintzman (1995) found more sizable declines. They found an average of .047 across three experiments, with a .10 decline in their Experiment 5, which they considered to be the most valid of their experiments. The size of the decline in our data (double aggregation) is .03, which is intermediate between those found by the two groups.

Hierarchical model analysis. We follow Lu et al. (2007) in analyzing the data of Experiment 1. We computed posterior distributions of parameters \( \mu_{1}^{(r)} \) and \( \mu_{2}^{(r)} \), the overall grand mean effect on recollection and automatic activation, for different study conditions. In frequentist analysis, such as the above aggregation analyses, the end product is a point estimate, such as those in Table 1. In Bayesian analysis, the posteriors themselves provide all information needed for estimation and inference. The mean of the posterior serves as a point estimate; an interval that contains 95% of the mass is called a 95% credible interval and is analogous to a frequentist confidence interval.

Posterior distributions of recollection (see Figure 3, top panels) vary considerably as a function of duration, whereas posteriors of automatic activation (see Figure 3, bottom panels) vary much less so. This graph serves as the main evidence for selective influence in the experiment. Parameters \( \mu_{1}^{(r)} \) and \( \mu_{2}^{(r)} \) are on the probit scale and are not directly comparable to double-aggregation estimates of \( R_{\lambda_{ik}} \) and \( A_{\lambda_{ik}} \). To provide for a direct comparison, we first computed Bayesian posterior mean point estimates of all \( R_{ijk} \) and \( A_{ijk} \), which are on the probability scale rather than the probit scale. These point estimates were then averaged over participants and items for each condition. The resulting values are displayed in Table 1 and are labeled Bayesian equivalent.

Selective influence. Selective influence is obtained if study duration affects only recollection. The Curran and Hintzman (1995) critique centered on the possibility that automatic activation increases with study duration, but this effect is masked by an increasing downward bias from aggregation. The hierarchical model provides for an accurate test of selective influence; the critical question is whether the hierarchical model estimators reveal an increase in automatic activation with study duration.

The left-hand column of Figure 3 reveals a strong effect of study on recollection and a weak one on automatic activation. We focus on the contrast between the 1-s and 10-s study conditions as this provides for the strongest test of selective influence as discussed previously. Let \( k = 1, 2 \) denote the 1-s and 10-s study conditions, respectively. The contrasts of interest are \( \mu_{2}^{(r)} - \mu_{1}^{(r)} \), for recollection, and \( \mu_{2}^{(a)} - \mu_{1}^{(a)} \), for automatic activation. Posterior distributions for these contrasts are shown in the right-hand column of Figure 3. The dotted vertical lines indicate the 95% credible interval between the 2.5 and 97.5 percentiles, the Bayesian analogue to the 95% confidence interval. In the top panel (for recollection), the value of zero difference is outside the 95% credible interval, indicating that the difference in recollection between the 1-s and 10-s conditions is statistically significant. The bottom-right panel shows the same plot for automatic activation. Automatic activation trends lower in the 10-s than in the 1-s study condition, indicating that there is a lack of evidence for the claim that automatic activation increases with study duration. Note that the value of zero is well within the 95% credible interval, indicating a lack of evidence for a reliable difference across the 1-s and 10-s conditions. The corresponding 95% confidence interval on the probability scale is (−.06, .02). Therefore, from a Bayesian point

\( \text{1 The current experiment is a Latin square design rather than a complete factorial design. Each participant observed each item in only one of the three study conditions. Fortunately, the additive structure of the model makes estimation possible in incomplete designs. The technical constraints for design matrices for linear models are provided in Christensen (1996), and the design of Experiment 1 meets these constraints.} \)
of view, there is less than a 2.5% chance that the automatic activation in the 10-s condition is more than .02 greater than that in the 1-s condition.

One of the statistical problems in testing selective influence is that invariances are assessed by failing to reject the null hypothesis. Such practices may be abused if the power of the test is not sufficient. When the null hypothesis is not rejected, it is useful to supplement the analysis with a relative comparison, and we compared the effects of study time on automatic activation with those on recollection. We constructed an effects-ratio statistic by dividing the difference in $A$ (see Figure 3, bottom-right panel) by the difference in $R$ (see Figure 3, top-right panel) and denote it by $\eta$. Because these differences are random in the Bayesian framework, the effects-ratio statistic has a distribution. The posterior mean of $\eta$ is $\bar{\eta} = -0.169$, indicating that the magnitude of the effect of study duration on automatic activation is less than a fifth of that on recollection and in the opposite direction. The Curran and Hintzman (1995) concern is that automatic activation increases with study duration. This concern may be addressed by computing the posterior probability that $\eta$ is positive and large. We decided to place a one-fifth criterion on the effect-size ratio to address Curran and Hintzman’s concern. The appropriate test statistic is the posterior probability that $\eta > .2$ given the data. This probability is less than .016, which indicates that Curran and Hintzman’s concerns about aggregation masking an increase in automatic activation are unwarranted for this data set.

Figure 3. The top-left and bottom-left panels show posterior distributions of overall recollection ($\mu^{\text{r}}_k$) and automatic activation ($\mu^{\text{a}}_k$), respectively. The top-right panel shows the contrast between 10-s and 1-s study conditions on recollection. The zero point is outside the 95% credible interval indicating a significant effect of study duration. The bottom-right panel shows the same for automatic activation. The zero point is well within the 95% credible interval, indicating a lack of evidence for an effect of study duration. Densities in left panels are obtained by Gaussian kernel smoothing of model-based outputs.
Participant and item random effects. The hierarchical analysis allows for the study of individual and item differences. The top row of Figure 4 shows the relevant effects as scatterplots. The top-left plot is for participants and shows participants’ automatic activation estimates as a function of their recollection estimates. The top-right plot shows the same for items. The scatterplot format is convenient as it shows the variability and correlation across random effects. The range of variation is modest. Standard deviations, measured in probit units, are on the order of .4. The correlations between recollection and automatic activation are clearly different for participants than for items, with there being only a small correlation for the former ($r = .11$) and a large one for the latter ($r = .89$).

Parameters $\Sigma_{\alpha}$ and $\Sigma_{\mu}$ account for correlations among the random effects for the population of participants and items. The bottom row of Figure 4 shows posterior distributions of correlation coefficients as calculated from the posteriors on $\Sigma_{\alpha}$ and $\Sigma_{\mu}$. The lack of evidence for a correlation for participant effects can be seen in the posterior distribution of the correlation coefficient derived from the covariance matrix $\Sigma_{\alpha}$ (see Figure 4, bottom-left panel; posterior $M = .03$). The substantial correlation across items is evident in the posterior distribution of the correlation coefficient derived from the population-level covariance matrix $\Sigma_{\mu}$ (see Figure 4, bottom-right panel; posterior $M = .67$). The correlations from the covariance matrices are more conservative than those from the individual random-effect estimates. This conservatism is expected. The correlations from the random effects are akin to sample correlations in that they describe properties of these particular items and participants. The correlations from parameters $\Sigma_{\alpha}$ and $\Sigma_{\mu}$ are population-level estimates. More evidence is needed to produce extreme estimates of these population-level parameters than to produce extreme estimates for a particular set of random effects.

![Random effects for Experiment 1](image)

Figure 4. Random effects for Experiment 1. Top left: Participant random effects on automatic activation as a function of those on recollection. Top right: Same plot for item random effects. Bottom: Plots of the posterior distributions of the correlation coefficients as derived from $\Sigma_{\alpha}$ and $\Sigma_{\mu}$, respectively.
Discussion

The process-dissociation model passes the selective influence test: Study duration affected recollection to a far greater extent than automatic activation. This result cannot be due to aggregation artifacts as the hierarchical model avoids aggregation. In the remainder of this discussion, we use the hierarchical model to assess how badly aggregation distorts estimation in conventional process-dissociation analysis. The results are fairly encouraging, indicating that many of the previous demonstrations of selective influence are reasonable.

Hierarchical estimates of overall automatic activation are greater in value than their double-aggregation counterpart (see Table 1). The difference between estimation methods is .022 in the 1-s condition and .029 in the 10-s condition, which suggests a small underestimation bias. The more important consideration is how this asymptotic bias varies with study duration. Curran and Hintzman (1995) predicted that aggregation estimates for automatic activation would be greater for the 1-s study condition than for the 10-s study condition. Their rationale is as follows. As discussed in the prior example with easy and hard items and selective influence, the amount of downward bias in automatic activation increases with greater true values of recollection. Because recollection is greater in the 10-s study condition than in the 1-s study condition, there is greater downward bias on the automatic activation estimate in the 10-s than in the 1-s study condition. Indeed, the double-aggregation estimate of automatic activation is .027 greater in the 1-s than in the 10-s condition. Yet a similar pattern was found for Bayesian estimates, which are not asymptotically biased. Automatic activation is .020 greater in the 1-s than in the 10-s study condition. The difference, .007, is an estimate of the amount of difference in downward bias across conditions. The significance of this amount is a function of the size of the effects, and in many applications, this degree of bias is fairly inconsequential.

We performed a small simulation study to assess the effects of aggregation more generally. The hierarchical model provides estimates of condition, participant, and item effects for recollection and automatic activation, with the item effects being highly correlated. We used these estimates as true values and then simulated data from the hierarchical model. Parameters were then estimated with the double-aggregation method. This process of simulating data from fixed true values and estimating parameters was repeated 500 times to provide a reasonable estimate of the expected degree of bias in double aggregation. This simulation provided the expected downward biases for the 1-s and 10-s conditions, which were .022 and .029, respectively.

The amount of downward bias increases with the degree of correlation and the amount of recollection. We used simulations to study the effect of different values of recollection on downward bias. Once again, estimates from Experiment 1 served as true values for simulation, with the exception of the overall recollection parameter $\mu^{\text{opt}}$. This parameter was manipulated through several values to simulate the effect of different study-duration conditions. For each value of $\mu^{\text{opt}}$, we repeated the simulation 500 times. Figure 5 shows the expected bias in the double-aggregation estimate of automatic activation as a function of overall recollection (plotted on a probability scale). Although the amount of bias increased with recollection, the rate of increase was not too great for smaller values of recollection. Aggregation becomes suspect either when recollection varies greatly across conditions or when recollection is of a high value such that small differences result in large differences in bias. The two vertical lines in Figure 5 show the recollection values for the 1-s and 10-s study conditions in Experiment 1, respectively.

Preexperimental Automatic Activation

The striking pattern of correlations across items and its absence across participants is a strong result. Curran and Hintzman (1995) found similar patterns, though not to the same degree. Their average sample correlation across items is about .5, whereas ours is about .8. Even though the results are concordant, ours is far easier to interpret, especially with regard to statistical significance. Their result is based on aggregation, and it is unknown what effect aggregation has on correlation estimates. Moreover, it is well known that their aggregation across participants increases the Type I error in subsequent statistical tests (Clark, 1973). Our estimates are not only greater in magnitude, they are made without recourse to aggregation, and statistical significance is easily assessed.

Curran and Hintzman (1995) provided a straightforward conjecture for the correlation across items. Stems themselves vary in how conducive they are to completion by the studied item. For example `hun__` is a better stem for `humor` than `hou__` is for `hound`. This variability induces a correlation in performance across include and exclude conditions. In this section, we describe a test of this and other explanations of the correlation. Unfortunately, we find little evidence for these explanations in our data.

In Experiment 1, there was a no-study baseline condition. Inclusion of this condition allows for the assessment of how much automatic activation is preexperimental and how much results
from the study list. We adapted an extension from Buchner et. al (1995) to separate these two contributions. On include trials where the item has been studied, either recollection, study-driven automatic activation, or preexperimental automatic activation is sufficient for stem completion. On exclude trials where the item has been studied, stem completion comes about when recollection fails and either preexperimental or stimulus-driven automatic activation succeeds. When an item is not studied, stem completion occurs only from preexperimental automatic activation. These statements yield the following model equations for a particular participant tested with a particular stem when items have been studied:

\[ I_{ijk} = R_{ijk} + (1 - R_{ijk})(A_{ijk} + (1 - A_{ijk})B_j), \]

and

\[ E_{ijk} = (1 - R_{ijk})(A_{ijk} + (1 - A_{ijk})B_j). \]

The corresponding equation in the no-study condition (denoted by \( k = 0 \)) is

\[ I_{i0j} = E_{i0j} = B_j. \]

Equations 9 and 10 were used to constrain \( R_{ij} \) and \( A_{ij} \). Parameter \( B_j \) was modeled as

\[ b_j = \mu^{(ij)} + \beta^{(ij)}_j, \]

where \( b_j = \Phi^{-1}(B_j) \). There are no participant effects in \( b_j \). Any participant effects would occur as an interaction, that is, given stems are completed at different rates by different participants. It is impossible to model such interactions without replicating participant–item pairings. We discuss the consequences of not modeling interactions after presenting all of the experimental data.

Participant random effects are modeled by extending Equation 11; a trivariate normal is used to model random item effects:

\[
\begin{pmatrix}
\beta^{(i)}_j \\
\beta^{(j)}_i \\
\beta^{(ij)}_k
\end{pmatrix} 
\sim N(0, \Sigma_{\beta}), \quad j = 1, \ldots, J,
\]

where \( \Sigma_{\beta} \) is a \( 3 \times 3 \) covariance matrix of six free parameters. We assume process independence in the model as previously discussed. We term this extended model with preexperimental automatic activation as the \textit{extended hierarchical process-dissociation model} and used it to reanalyze Experiment 1.

With this extended hierarchical model, the pattern of correlation among items observed in the previous hierarchical model-based analysis may be decomposed into correlations among three processes: recollection, study-driven automatic activation, and preexperimental automatic activation. Parameter \( \beta^{(ij)}_k \) serves as a relative index of how well a stem primes its completion in the absence of any study. Curran and Hintzman’s (1995) speculation about the correlation between recollection and automatic activation motivated two questions: (a) Is there a correlation between study-driven automatic activation and recollection, and (b) if so, is this correlation solely due to preexperimental automatic activation? Question (b) is answered by assessing whether the correlation between study-driven automatic activation and recollection is attenuated to zero when preexperimental automatic activation is controlled (partialed out).

Analysis with the extended hierarchical model reveals that much, though not all, of the automatic activation is preexperimental. Preexperimental automatic activation is .266 on average, whereas study-driven automatic activation is .051 on average. Almost all of the previous findings held: Figure 6A shows that study-driven automatic activation did not differ substantially across the 10-s and 1-s study conditions. Figure 6B shows a lack of covariation in study-driven automatic activation and recollection across participants.

The relationship between the three item-based processes (recollection, study-driven automatic activation, and preexperimental automatic activation) is shown as pairwise scatterplots in Figures 6D–6F. All three processes covary together to some degree. The scatterplots show that there is a moderate correlation between study-driven automatic activation and recollection (\( r = .68 \) for sample values and \( r = .42 \) for generalization to the population). To assess whether this correlation was driven by preexperimental automatic activation, we estimated a partial correlation. There are two approaches to doing so: one based on individual random effects \( \beta \) and another using the population-level covariance matrix \( \Sigma_{\beta} \). The latter is more appropriate for our purposes. If we had used the former, it could be claimed that a lack of attenuation in partial correlation is due to noisy individual random-effect estimates (this phenomenon is referred to as an \textit{errors-in-variables problem} in the statistics literature). The use of the population-level covariance matrix corrects for sampling noise in the individual item estimates. The posterior distribution of the population-level partial correlation of stimulus-driven automatic activation and recollection, when conditioned on preexperimental automatic activation, is shown in Figure 6C. The posterior mean partial correlation is .35, which is not that different from the population marginal correlation of .42. Moreover, there is little evidence that this population-level correlation is zero, a value that would indicate that all of the observed correlation is from variation in preexperimental automatic activation. In fact, the probability that the population-level partial correlation is greater than zero is .93.

Stem-completion baselines are not the only property that varies across items. We assessed whether commonly considered word properties, age of acquisition (Gilhooly & Logie, 1980), word frequency (Kucera & Francis, 1967), meaningfulness (Toglia & Battig, 1978), familiarity, and imagability (Coltheart, 1981) could account for the correlation between study-driven automatic activation and recollection. Consequently, we computed the partial correlation between study-driven and automatic activation and recollection conditioned on each of these properties. We make the assumption that these properties are measured without error. Hence, there is no errors-in-measures problem, and the sample partial correlation coefficient is not subject to distortion in attenuation. There was only a small degree of attenuation in partial correlation, the largest of which was from imagability. The partial correlation is \( r = .66 \), which is only slightly smaller than the marginal sample partial correlation of \( r = .68 \). The correlation of study-driven familiarity and recollection is not readily driven by common dependencies on word properties. Given that these word properties typically account for only a small fraction of overall variance in memory experiments, the result is not too surprising.
Figure 6. Reanalysis of Experiment 1 with the extended hierarchical process-dissociation model. A: Posterior of the difference in study-driven automatic activation between the 1-s and 10-s study conditions shows no evidence of an effect. B: The scatterplot of participant specific recollection and study-driven automatic activation effects shows little correlation. C: Posterior distribution of the partial correlation of recollection and study-driven activation for controlled preexperimental automatic activation. The vertical dashed line is the one-tailed 95% credible bound. D–F: Pairwise scatterplots of the item effects on recollection, study-driven automatic activation, and preexperimental automatic activation. Sample corr. = sample-level correlation; Pop. corr. = population-level correlation.
Experiment 2

Analysis of Experiment 1 provided support for the hierarchical process-dissociation model. The goal in Experiment 2 was to provide the following tests of the validity and usefulness of the model: (a) The pattern of correlations in Experiment 1, with no correlation in participant effects and sizable correlation in item effects, is noteworthy. We assessed its replicability under different study conditions. (b) If the model is accurately capturing systematic item-based effects, then these should be largely stable. Experiment 2 provided an opportunity to investigate the stability of these item-based estimates. (c) The usefulness of the model is its ability to provide accurate estimates in conditions where aggregation fails. In Experiment 2, we explored whether differences in recollection could be made sufficiently large between conditions such that aggregation led to consequential distortions.

In Experiment 2, we implemented a levels-of-processing manipulation, as this is known to dramatically affect performance (Craik & Lockhart, 1972). In previous process-dissociation analyses of stem-completion tasks, processing level has been shown to dramatically affect estimates of recollection (Bergerbest & Goshen-Gottstein, 2002; McBride & Dosher, 1999; Toth, Reingold, & Jacoby, 1994). The effect on automatic processing, on the other hand, has been inconsistent. Bergerbest and Goshen-Gottstein (2002) found more automatic activation for rating the pleasantness of items than for counting vowels or syllables. Toth et al. (1994) and McBride and Dosher (1999), in contrast, found no increase in automatic activation across similar deep and shallow study conditions. Whereas there is some disagreement in the literature on this point, our study provides needed evidence. In the shallow-processing condition, participants were asked to count vowels in items at study; in the deep-processing condition, participants were encouraged to use whatever strategies they believed best facilitated recall for a subsequent memory test.

Method

Participants. Forty-four undergraduate students at the University of Missouri—Columbia received partial class credit for participation.

Materials. Thirty-two words from Experiment 1 served as critical items. Fewer critical items were used in Experiment 2 to guarantee a high rate of recollectability for the deep-processing condition. The items were chosen to maximize variance of item effects. There were 16 additional filler items.

Design. The main independent manipulations were the processing level at study and the instructions at test. For the shallow level, participants were asked to count vowels, and the items were displayed for 1 s. For the deep level, participants were asked to study items for a later memory test, and the items were displayed for 12 s. Processing level was blocked to prevent selective rehearsal of shallow-condition items during the study period for deep-condition items. The order of study conditions was counterbalanced across participants. Testing instructions were either the include or the exclude instructions. These were randomly intermixed throughout the list. Altogether, there were four conditions obtained by crossing the two levels-of-processing conditions with the item instructions. These four conditions were crossed with participants in a balanced Latin square design.

Procedure. Procedure was identical to that of Experiment 1. Total time to complete Experiment 2 was approximately 15 min.

Results

Conventional analyses and the hierarchical model analyses were performed. We used the hierarchical model with a single source of automatic activation (Equations 7–12) instead of the extended hierarchical model with preexperimental and stimulus-driven automatic activation because the later needs a no-study control condition, which was not included in the design. Overall recollection and automatic activation estimates from both aggregation methods and from the hierarchical process-dissociation model are shown in Table 2.

There is a sizable effect of processing level on recollection and a more modest one on automatic activation for all methods of analysis. Item-aggregation estimates seem compromised in this case as evidenced by the substantially higher estimate of recollection in the shallow condition than in the other methods. As stated earlier, this high estimate is likely from nonasymptotic bias. According to the double-aggregation analysis, the difference in automatic activation across the different processing levels is .037. Unfortunately, there is no way of assessing the statistical significance of this difference from double aggregation unless one wishes to assume that there are neither item nor participant effects in stem completion.

To construct a test of this difference with aggregation, we reanalyzed the data with single aggregation in a dubious manner. Negative values of recollection for individuals were admitted where indicated by Equation 3. We do not recommend this practice and use it simply to compare analyses. The resulting item-aggregation automatic activation averages were .219 and .255 for the shallow and deep conditions, respectively. A paired t test of these estimates revealed no significant difference of study condition on automatic activation, t(42) = .91, p > .1. Hence, if we followed the standard practice of doing estimation solely through Equations 3 and 4, we would have concluded an invariance in automatic activation.

The hierarchical model analysis yields a higher estimate of automatic activation for the deep condition than did aggregation. Moreover, the difference across study conditions is .059, which is about 50% greater than that from aggregation (.037). Figure 7 shows the posterior distributions for overall recollection and automatic activation. The plots in the top row show a substantial levels-of-processing effect on recollection. The effect on automatic activation is more subtle. The bottom-left panel shows the posterior of the appropriate contrast. The 95% credible interval for this

<table>
<thead>
<tr>
<th></th>
<th>Shallow</th>
<th>Deep</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item aggregation</td>
<td>.128</td>
<td>.412</td>
</tr>
<tr>
<td>Double aggregation</td>
<td>.068</td>
<td>.398</td>
</tr>
<tr>
<td>Bayesian equivalent</td>
<td>.063</td>
<td>.393</td>
</tr>
</tbody>
</table>

Table 2

Estimates of Overall Recollection and Automatic Activation as a Function of Study Condition for Experiment 2
contrast excludes zero, although not by much. There is, therefore, some evidence that automatic activation is greater in the deep condition than in the shallow condition. This evidence would not have been available from aggregation analysis; distortions from aggregation resulted in a downward bias that differentially affected estimates of automatic activation in the deep condition. This bias masked the increase in automatic activation with study condition in exactly the manner described by Curran and Hintzman (1995).

Figure 8 shows participant and item random effects. There is little evidence for correlation of recollection and automatic activation across participants. Correlation across the individual participant effects is .081; the estimated population correlation from $\Sigma_r$ is .007. There is evidence for a sizable correlation, however, across items. Correlation from the item effects is .841; population correlation from $\Sigma_g$ is .590. This is the same pattern seen in Experiment 1.

The items used in Experiment 2 are a subset of those used in Experiment 1. If the model is well specified, then item effects should be stable across different participants and study conditions. Figure 9 is a scatterplot of item effects for those items that were used in both experiments. There is a high degree of correlation (.69 and .79 for recollection and automatic activation, respectively). Item-effect estimates are surprisingly similar across the two experiments even though these experiments entailed different participants and study instructions. This stability provides converging evidence that the hierarchical Bayesian model is well specified.

**Discussion**

Experiment 2 provides for three noteworthy conclusions: First, the pattern of correlations replicated. Items that elicited higher levels of recollection tended to elicit higher levels of automatic activation. Across participants, however, recollection and automatic activation appeared uncorrelated. Second, there is stability among item estimates. Items that led to good performance in Experiment 1 tended to do so in Experiment 2. Third, the model...
proved exceptionally useful when compared with aggregation methods. The hierarchical model shows that there is a levels-of-processing effect on automatic activation; conventional aggregation methods miss this effect due to increases in downward bias with recollection. Here is a case in which bias from aggregation has a practical consequence. Therefore, the need to consider bias from aggregation depends on the particular application at hand.

The levels-of-processing effect is a failure of selective influence. Bergerbest and Goshen-Gottstein (2002) proposed that this failure may result from exceedingly shallow encoding. The vowel-counting task may preclude even automatic encoding of the orthography of the word. Bergerbest and Goshen-Gottstein ran an additional experiment in which participants counted syllables instead of vowels in the shallow conditions. Although this manipulation was hypothesized to increase the likelihood of encoding the orthography of the word, these researchers still found greater automatic activation for deep study than for shallow study. It is unclear if the model fails or if the shallow-level manipulation focuses attention away from encoding the orthography of the whole word.

Caveat: Correlated Interactions

The hierarchical model is a vast improvement over conventional methods. Even so, one important issue remains unresolved. Like conventional methods, it does not account for interactions between participants and items. A particular participant may be able to recollect a particular word well because of recent experience. Curran and Hintzman (1995) gave the following hypothetical: The stem mus__ may be more often completed with music than muscle by most participants. Participants who are physical-therapy majors, however, may reverse the ordering. Interactions of this type are assumed not to exist in the hierarchical model.

Concern arises if these interactions exist and are correlated across recollection and automatic activation. For example, the physical-therapy major may both better recollect and elicit more automatic activation from studying MUSCLE than other participants do relative to other items. As pointed out by Curran and Hintzman (1995), ignoring the interactions will lead to asymptotic downward bias. In the hierarchical model, the lack of an interaction term means that correlated interactions will result in a down-
ward bias in automatic activation parameter $\mu^{(a)}$. Unfortunately, this downward bias will be greater for conditions with higher amounts of recollection, making it difficult to assess selective influence. The argument for this case is analogous to the previously provided simple example of the effects of selective example with easy and hard items.

It would be advantageous to allow arbitrary interaction terms in the model. Unfortunately, it is simply impossible to include these terms for most memory designs as these designs yield no replicates. Whereas interactions cannot be modeled, the best approach is to assess the consequences of these possible interactions with simulations, as we do below. These simulations reveal that the presence of interactions is of no practical consequence for the preceding analyses.

To assess the impact of correlated interactions, we performed a series of simulations with interactions perfectly correlated across recollection and automatic activation, which represents a worst case scenario. In these cases, the additivity assumption in the hierarchical model is misspecified. To simulate data, we used the following equations to generate true recollection and automatic activation probability parameters:

$$R_{ijk} = \Phi(\alpha_i^{(r)} + \beta_j^{(r)} + \mu_{ij}^{(r)} + \pi_{ijk}),$$

and

$$A_{ijk} = \Phi(\alpha_i^{(a)} + \beta_j^{(a)} + \mu_{ij}^{(a)} + \pi_{ijk}).$$

The equations are the second level of the hierarchical model, with the addition of an interaction term, $\pi_{ijk}$. Note that the same value of interaction, $\pi_{ijk}$, enters both equations. Hence, interactions are perfectly positively correlated. We sampled each $\pi_{ijk}$ from a zero-centered normal with variance $\sigma^2_{\pi}$.

The size of this variance, $\sigma^2_{\pi}$, determines the amount of covariance of the interaction and the degree of downward bias in estimating automatic activation. We picked a value for $\sigma^2_{\pi} = .2$, which is 1.25 times as large as the average variance of the random effects across both experiments.

The simulations are modeled after the 1-s and 10-s study conditions in Experiment 1. True values were the estimates obtained from the hierarchical model, with the exception that perfectly correlated interactions were added and grand means of automatic activation were set equal across both study conditions ($\mu^{(a)} = \mu^{(a)} = -.394$; this value is the average of the estimates from Experiment 1). For each trial in Experiment 1, values of $R_{ijk}$ and $A_{ijk}$ were computed according to Equations 13 and 14. If the trial was an include trial, the stem-completion probability was $p = R_{ijk} + (1 - R_{ijk})A_{ijk}$; if the trial was an exclude trial, $p = (1 - R_{ijk})A_{ijk}$. From these probabilities, dichotomous stem-completion events (successes or failures) were sampled. The resulting data set was then submitted for hierarchical model estimation with the assumption of additivity. This process of generating data with large correlated interactions and analyzing them with the hierarchical model with additivity was repeated 1,000 times. The critical question is the degree of asymptotic bias in automatic activation estimates.

As expected, there is some downward bias. For the 1-s condition, the bias is $-.032$ in probit units ($-.012$ in probability units). For the 10-s condition, the bias increases to $-.052$ in probit units ($-.020$ in probability units). The greater bias in the 10-s condition is expected as downward bias increases with increasing recollection. Comparisons of automatic activation across conditions entail subtraction, and the biases cancel to a large degree. The amount of bias in the difference is about $-.022$ in probit units and $-.0078$ in probability units.

Automatic activation estimates in Experiment 1 (see Figure 3, bottom-right panel) may be reexamined in light of this bias. The bias is negative, indicating that a small positive adjustment may be added to the posterior. When added, the posterior mean shifts to $-.036$ from $-.058$, which strengthens the claim of invariance. The effect-size ratio changes to $-.100$ from $-.168$. The probability that $\eta$, the effect-size ratio, is greater than $.2$ changes to $-.036$ from $-.016$. Therefore, the possibility of correlated interactions poses no practical threat to the previous conclusions.

Although interactions pose no practical threat to our data analysis, there are designs in which interactions could potentially be a problem, especially when there are large differences in recollection.
across conditions. The presence of interactions does not threaten the validity of the rejection of selective influence in Experiment 2; additional downward bias would make the effect larger. Perhaps the most important point of the exercise is that diligent researchers may assess the potential effects of correlated interactions through simulation.

Conclusions

Process Dissociation

The hierarchical Bayesian model of experimental data yields three lines of support for the process-dissociation model. First, the model passes a study-duration selective influence test. Increasing study duration from 1 s to 10 s affects recollection but not automatic activation. Second, the model reveals a striking pattern of correlations. Participants’ recollection and automatic activation abilities are uncorrelated; that is, participants with high recollection abilities tend to be no better at automatic activation. A clearly different pattern emerges for items. Items that are better recalled tend to elicit a greater degree of automatic activation. The fact that the model can dissociate these two opposing patterns provides informal evidence of support. Third, the model shows stability in item estimates across different participants and encoding instructions.

One contribution of the model is that it provides a means of assessing the detrimental effects of aggregating across participants and items in process-dissociation experiments. For Experiment 1, where the difference in recollection across conditions was moderate, there was no practical harm in aggregation. In Experiment 2, in which there was a large difference in recollection across conditions, the effects of aggregation were consequential. It may turn out that in many experiments, there are no practical disadvantages to aggregation. We emphasize that such knowledge is only accessible after hierarchical model analysis and not before it.

Modeling Mnemonic Processes

The dangers of aggregation are well known in several contexts. Even so, the problem has not been articulated in a sufficiently general manner. We cannot underestimate the following point: All nonlinear models are subject to distortion when data are aggregated. This problem holds not only for process dissociation but for most memory models in the field, including most multinomial processing tree models (Batchelder & Riefer, 1999), most strength models (e.g., similarity-choice model; Luce, 1963), and cognitive-skill-acquisition models (Heathcote et al., 2000; Rickard, 2004). Ignoring the problem is tantamount to assuming that there is no variation across items or participants.

We have recently proposed that aggregation may be avoided entirely by implementing hierarchical models (Rouder & Lu, 2005; Rouder, Lu, et al., 2007). In Rouder, Lu, et al. (2007), for example, we developed a hierarchical version of the theory of signal detection that avoids problems with aggregation. Fortunately, the skills and techniques encompassed in the modeling generalize well across many domains. For example, much of the modeling discussed here is similar to that for our signal detection models. The hierarchical elements of the models presented here will generalize to other multinomial and random utility type models. In this vein, while the focus has been on process dissociation, the modeling endeavor is far more general.

On a final note, we realize that the message that aggregation is cause for concern is not good news. The advocated modeling solution places burdens on analysts. These hierarchical models are based on a level of statistics that is beyond the training afforded in most psychology departments in the United States; however, we believe the time spent learning the requisite statistics is well worth the investment. There are now several texts on Bayesian modeling with coverage of hierarchical models (e.g., Congdon, 2006; Gelman, Carlin, Stern, & Rubin, 2004; Lee, 2004). We have provided a primer of Bayesian hierarchical modeling for experimental psychologists (Rouder & Lu, 2005). Bayesian hierarchical models of mnemonic processes are not implemented in commercial software. Nonetheless, they may be programmed in high-level languages such as R, Winbugs, and MATLAB.

References


Curran, T., & Hintzman, D. L. (1997). Consequences and causes of...


(Appendix follows)
We provide a complete specification of the priors and an overview of the method of analysis. A more detailed analysis of and justification for these priors may be found in Lu et al. (2007). Lu et al. also provided alternative analyses of process dissociation with other priors.

### Specification of the Prior

Priors are needed for $\mu_k^{(r)}$, $\mu_k^{(c)}$, $\Sigma_\alpha$, and $\Sigma_\beta$. Priors on grand means are

$$\mu_k^{(r)} \sim \text{Normal}(0, \sigma^2_{\mu_k}), \quad k = 1, \ldots, K,$$

(A1)

and

$$\mu_k^{(c)} \sim \text{Normal}(0, \sigma^2_{\mu_k}), \quad k = 1, \ldots, K.$$  

(A2)

The value of $\sigma^2_{\mu_k}$ must be chosen beforehand. Any large values will lead to an appropriately diffuse prior. We use $\sigma^2_{\mu_k} = 1,000$ in application.

Priors on covariance matrices are

$$\Sigma_\alpha \sim \text{Inverse Wishart}(m, \Omega)$$

(A3)

and

$$\Sigma_\beta \sim \text{Inverse Wishart}(m, \Omega).$$

(A4)

The density of the inverse Wishart distribution is

$$f(S|m, \Omega) = \frac{|\Omega|^{m/2} |S|^{-(m+3)/2}}{2^m \Gamma(m/2)} \exp\left(-\frac{1}{2} tr(\Omega S^{-1})\right),$$

(A5)

where $S$ is a positive definite matrix and $\Gamma_p$ is the multivariate gamma function given by $\Gamma_p(a) = \pi^{p(p-1)/2} \Gamma_p(\frac{p}{2}) \Gamma(\frac{p}{2}(a - \frac{1}{2}(j - 1)))$ (see Gelman et al., 2004, p. 574).

The Wishart prior is perhaps the most common choice for covariance matrices. It is conjugate for the covariance matrix of a multivariate normal distribution, and this simplifies analysis. Details of the inverse Wishart and its use as a prior for covariance matrices can be found in standard texts on statistics (e.g., Tanner, 1998). The inverse Wishart has two parameters, $m$ and $\Omega$, which

---

*Appendix*

### Analysis of the Hierarchical Process-Dissociation Model

---

**Figure A1.** Convergence in Markov chain Monte Carlo sampling. Top row: Values on each iteration for selected parameters (from left to right: grand mean $\mu_k^{(r)}$ for the deep condition, item effect $\beta_{ij}$, and the off-diagonal term of the participant precision matrix $\Sigma_{ij}$). Bottom row: Autocorrelation of the chains for the selected parameters. The modest degree of autocorrelation is offset by the long length of the chains.
must be set beforehand. The choice \( m = 2 \) leads to the least informative prior. Parameter \( \Omega \) is a \( 2 \times 2 \) matrix. An appropriate choice in this context is to set \( \Omega \) to the identity matrix. The inverse Wishart prior is informative; that is, the prior does provide information about the scale of random effects (Lu et al., 2007; Rouder, Lu, et al., 2007). To ensure the choice of \( \Omega \) did not overly bias estimates, we repeated the analyses with the alternatives \( \Omega = .2I \) and \( \Omega = 5I \). Results for these choices were highly similar to the reported ones for \( \Omega = I \).

Analysis

The target of analysis in Bayesian statistics is the derivation of the marginal posterior distribution for each parameter. Often, these marginal posteriors cannot be derived as closed-form expressions. We follow the common approach of deriving closed-form expressions for full-conditional posterior distributions and sampling these with Markov chain Monte Carlo (MCMC) techniques (Gelfand & Smith, 1990). The derivation of full-conditional posterior distributions for this application may be found in Lu et al. (2007). These full-conditional posteriors are inverse Wishart and multivariate normal, which may be sampled directly without recourse to the more computationally intensive Metropolis-Hastings algorithm. Details are found in Lu et al.

We sampled chains of 6,000 iterations, with the first 1,000 iterations serving as a burn-in period. Convergence was rapid, and chains exhibited only modest autocorrelation. The top panels of Figure A1 show chains for select parameters for Experiment 2. The bottom panels show the degree of autocorrelation in the chains for these parameters. As can be seen, mixing is sufficient.

All analyses were done in the R statistical language (http://cran.r-project.org/). Sampling from the inverse Wishart and multivariate normal was accomplished with the MCMCpack (MCMCpack.wustl.edu) and MASS package (Venables & Ripley, 2003), respectively. A full computer implementation of the analysis may be found at http://web.missouri.edu/~umcaspsychpcl/code.

Received March 16, 2007
Revision received August 21, 2007
Accepted August 23, 2007

Low Publication Prices for APA Members and Affiliates

Keeping you up-to-date. All APA Fellows, Members, Associates, and Student Affiliates receive—as part of their annual dues—subscriptions to the American Psychologist and APA Monitor. High School Teacher and International Affiliates receive subscriptions to the APA Monitor, and they may subscribe to the American Psychologist at a significantly reduced rate. In addition, all Members and Student Affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

Essential resources. APA members and affiliates receive special rates for purchases of APA books, including the Publication Manual of the American Psychological Association, and on dozens of new topical books each year.

Other benefits of membership. Membership in APA also provides eligibility for competitive insurance plans, continuing education programs, reduced APA convention fees, and specialty divisions.

More information. Write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242.