**BERICHTE AUS DEM PSYCHOLOGISCHEN INSTITUT DER UNIVERSITÄT BONN**

**Band 21 (1995) Heft 3**

# Process Dissociation Measurement Models: Good versus Better

Edgar Erdfelder & Axel Buchner

# Table of Contents

# Abstract

Several methods to account for response bias in the process dissociation procedure have recently been proposed. A. P. Yonelinas and L. L. Jacoby (1995b) favor a dual-process signal-detection model (DPSDM) and claim that threshold-based models such as the extended measurement model (EMM) suggested by A. Buchner, E. Erdfelder, and B. Vaterrodt-Plünnecke (1995 should be rejected because threshold models are inconsistent with nonlinear receiver operating characteristics (ROCs) as obtained from confidence ratings. Their claim is shown to be incorrect. An EMM variant for confidence ratings is developed which accounts perfectly for nonlinear ROCs. It is demonstrated that, in contrast, the DPSDM cannot fit the ROC data of at least two of the three experiments reported by A. P. Yonelinas (1994). Further, it is argued that experimental manipulations of response biases result in much more thorough tests of process dissociation measurement models than confidence ratings. We close by suggesting a new two-high threshold extended measurement model which fits the Buchner et al. data better than both the EMM and the DPSDM.

# Zusammenfassung

In der neueren Literatur zur Prozeß-Dissoziations-Prozedur werden unterschiedliche Vorschläge zur Berücksichtigung von Antworttendenzen gemacht. A. P. Yonelinas und L. L. Jacoby (1995b) favorisieren ein Dual-Process-Signal-Detection-Modell (DPSDM) und kritisieren High-Threshold-Modelle wie z.B. das Extended-Measurement-Modell (EMM) von A. Buchner, E. Erdfelder und B. Vaterrodt-Plünnecke (1995), weil derartige Modelle mit nichtlinearen Receiver Operating Characteristic Curves (ROCs) - wie sie für Konfidenzratings beobachtet wurden - unvereinbar seien. Es wird gezeigt, daß diese Behauptung falsch ist. Eine Variante des EMM für Konfidenzratings, die nichtlineare ROCs fehlerfrei erklären kann, wird vorgestellt. Es wird gleichzeitig nachgewiesen, daß das DPSDM mit den Daten von mindesten zweien der drei Experimente von A. P. Yonelinas (1994) nicht vereinbar ist. Weiterhin wird die These vertreten, daß experimentelle Manipulationen von Antworttendenzen strengere Tests von Prozeß-Dissoziations-Meßmodellen erlauben als Konfidenzratings. Im abschließenden Teil der vorliegenden Arbeit wird ein Two-High-Threshold EMM vorgestellt, das mit den Buchner et al.-Daten besser als das EMM und das DPSDM vereinbar ist.

# Process Dissociation Measurement Models: Good versus Better

The process dissociation measurement model originally proposed by Jacoby (1991) aims at measuring the contributions of controlled ("conscious") and automatic ("unconscious") cognitive processes to task performance without accounting for possible effects of response biases. As a consequence, measures of controlled and automatic processes may be contaminated by response biases. Although this problem is not new and several correction methods have been proposed by various authors (cf. Jacoby, Toth & Yonelinas, 1993; Komatsu, Graf & Uttl, 1995; Roediger & McDermott, 1994) , it was not until recently that the problem was addressed by new measurement models that can account for simultaneous effects of controlled processes, automatic processes, and response biases on task performance. Buchner, Erdfelder, and Vaterrodt-Plünnecke (1995) have developed an extended measurement model (EMM) which is based on threshold theory (cf. Krantz, 1969; Luce, 1963a) whereas Yonelinas, Regehr, and Jacoby (in press) suggested a dual-process signal-detection model (DPSDM) using the framework of signal-detection theory (cf. Green & Swets, 1966) .

In their interesting and stimulating comment on Buchner et al. (1995), Yonelinas and Jacoby (1995b) argued that their DPSDM was superior to the EMM because the former, but not the latter, could account for nonlinear receiver-operating characteristics (ROCs) as obtained from confidence ratings (cf. Yonelinas, 1994) . They also argued that the experimental data used to validate the EMM by Buchner et al. (1995) were inappropriate, because the experimental manipulations might have affected response bias and memory processes simultaneously.

This article aims at refuting both claims. We begin by describing the EMM as applied to the process dissociation procedure using yes-no recognition judgments, and we discuss its relation to the measurement model variant suggested by Jacoby (1991). Next, we describe the DPSDM and show that it has some methodological disadvantages compared to the EMM. Nevertheless, both the EMM and the DPSDM seem to be superior to alternative methods of response bias correction that have

We will then present an appropriate extension of the EMM to confidence rating scales in order to show that nonlinear ROCs do not falsify the threshold concept underlying the EMM. Moreover, we maintain that the experimental data provided by Buchner et al. (1995) are appropriate validation hurdles that must be cleared by an adequate process dissociation measurement model. We agree with Yonelinas and Jacoby (1995b) that both the EMM and DPSDM fit these data quite well but not perfectly. We differ from Yonelinas and Jacoby (1995b) in that we attribute the less-than-perfect fit to limitations of both models.

We close by discussing generalizations of the EMM and the DPSDM that might help to overcome these limitations. A generalization of the DPSDM that does not require the normal distribution assumption is developed, and we show that this generalized version and, hence, the DPSDM as a submodel of it does not fit the data of Yonelinas' (1994) Experiments 1 and 2 at least. Also, a correlated processes signal detection model is developed that does not require the assumption of stochastic independence of controlled and automatic processes presumed by the DPSDM. Finally, we suggest two variants of a new two-high threshold extended measurement model and show that each of them fits the Buchner et al. (1995) data even better than both the EMM and the DPSDM.

## Process Dissociation Measurement Models

In order to avoid terminological confusions it is useful to distinguish between the *process dissociation procedure* and *process dissociation measurement models*. The process dissociation procedure is a class of experimental paradigms which is characterized by two test conditions—the inclusion test condition and the exclusion test condition—and three types of items to which both test conditions are applied. A process dissociation measurement model, in contrast, is a stochastic model which 'explains' the probabilities of participants' responses in the process dissociation procedure in terms of parameters that represent different types of cognitive processes.

Depending on the specific instructions defining the inclusion and exclusion test conditions and on the three types of items to which participants have to respond, Anas & Farinacci, 1992; Debner & Jacoby, 1994; Jennings & Jacoby, 1993; Lindsay & Jacoby, 1994; Toth, Reingold & Jacoby, 1994; Wippich, 1994; Yonelinas & Jacoby, 1995a) . However, to keep the presentation simple in this article, we will follow Yonelinas and Jacoby (1995b) and refer to recognition paradigms only. Nevertheless, almost all of our arguments apply also to other implementations of the process dissociation procedure.

In the recognition variant of the process dissociation procedure as introduced by Jacoby (1991), participants process a list of critical items in Phase 1 and another list of items (possibly differing in sensory modality) in Phase 2. In the test phase, participants are confronted with items from Phase 1 and Phase 2, and with distractors which were not presented before. Participants in the *inclusion test condition* are instructed to respond *old* to all items presented in Phase 1 or Phase 2, and they are to respond *new* to all items that were presented neither in Phase 1 nor in Phase 2. Participants in the *exclusion test condition* receive the same instructions for the Phase 2 items and for the new items, but they are told to 'exclude' Phase 1 items, that is, they must respond *new* to items which were presented in Phase 1.

***The Extended Measurement Model***

The EMM suggested by Buchner et al. (1995) explains the probabilities of *old* responses for different types of items and test conditions in terms of the *memory parameters* $c$, $u_{C+}$, and $u_{C-}$ and the *response bias parameters* $g_i$ and $g_e$. Parameter $c$ represents the unconditional probability of recollecting a Phase 1 item. Parameter $u_{C+}$ denotes the conditional probability of uncontrolled processes given that a Phase 1 item has already been recollected. This parameter can be ignored for the present purposes because it does not influence participants' responses.[1] Further, a Phase 1 item cannot be recollected with probability $1-c$, and parameter $u_{C-}$ represents the conditional probability of automatic processes leading to a cognitive state in which a

[1] Technically $u_{C+}$ is an unidentifiable parameter which means that an assumption must be made about this parameter in order to render the model identifiable (e.g., $u_{C+} = u_{C-}$, $u_{C+} = 1$, or $u_{C+} = 0$ which yield, respectively, model variants in which controlled and automatic processes are independent, redundant, or mutually exclusive; see Jones, 1987) . Which assumption is used, however, is relevant only if one is interested in the unconditional probability $u$ of automatic

Phase 1 item is accepted as *old*, given that it cannot be recollected. In the context of recognition tasks, this cognitive state has been described phenomenologically as a state of "familiarity-based responding" (e.g., Jacoby, 1991) which is why we adopt this terminology here. Finally, $g_i$ and $g_e$ denote the conditional probabilities of guessing *old* in the inclusion and the exclusion test conditions, respectively, given that the item neither can be recollected nor seems sufficiently familiar.

Four model equations expressing the probabilities of *old* responses as functions of the model parameters $c$, $u_{c-}$, $g_i$, and $g_e$ for different types of items and test conditions can be derived quite easily by summing up the probabilities of mutually exclusive cognitive processes that may lead to an *old* response. For instance, the probability of an *old* response to Phase 1 items in the inclusion test condition, $p_{1i}$, equals

$$p_{1i} = c + (1 - c) \cdot u_{c-} + (1 - c) \cdot (1 - u_{c-}) \cdot g_i, \tag{1}$$

because *old* responses can result (a) if the item is recollected which occurs with probability $c$, (b) if the item is not recollected but appears familiar which occurs with probability $(1 - c) \cdot u_{c-}$, or (c) if the item is neither recollected nor seems familiar but it is guessed that the item is from Phase 1; this occurs with probability $(1 - c) \cdot (1 - u_{c-}) \cdot g_i$. According to the EMM, distractor items cannot be recollected and cannot seem familiar which is why answers to distractors are conceived of as guesses. This implies the assumption that distractors can never be detected as *new* with certainty. Thus, the probability $p_{di}$ of an *old* response to distractor items (i.e., a "false alarm") in the inclusion test condition equals

$$p_{di} = g_i. \tag{2}$$

In the exclusion test condition, by contrast, the probability $p_{1e}$ of an *old* response to Phase 1 items equals

$$p_{1e} = (1 - c) \cdot u_{c-} + (1 - c) \cdot (1 - u_{c-}) \cdot g_e, \tag{3}$$

because these responses are assumed to occur only (a) if the item is not recollected but appears familiar which happens with probability $(1 - c) \cdot u_{c-}$, or because of guessing when an item is neither recollected nor seems familiar; this occurs with are to respond *new* if they can recollect a Phase 1 item. Again, distractor items are assumed to be responded to on the basis of a guessing process which implies that the probability $p_{de}$ of a false alarm in the exclusion test condition equals

$$p_{de} = g_e. \tag{4}$$

Note that the guessing probabilities $g_i$ and $g_e$ and, hence, the false-alarm probabilities may differ between the inclusion and exclusion test conditions.

Buchner et al. (1995) showed that all four model parameters $c$, $u_{c-}$, $g_i$, and $g_e$ are identifiable and demonstrated how to arrive at parameter estimates, confidence intervals for the estimates, and goodness-of-fit tests for models with parameter restrictions. In fact, because the EMM is formally a multinomial processing-tree (GPT) model (cf. Hu & Batchelder, 1994; Riefer & Batchelder, 1988), no new statistical work is needed, and easy-to-use software for statistical data analyses is also available (Hu, 1993; see Hu & Batchelder, 1994).

Quite a few other process dissociation measurement models can be derived from the EMM in a rather straightforward way (cf. Buchner et al., 1995). For example, the *original independence measurement model* (IMM) suggested by Jacoby (1991) is obtained

(a) by restricting the model equations to Phase 1 items (i.e., by omitting Equations 2 and 4 for the distractor items),

(b) by assuming that $g_i = g_e = 0$ (i.e., by positing that *old* responses never occur as a consequence of guessing), and

(c) by imposing the restriction that $u_{c+} = u_{c-}$, that is, by assuming that controlled recollections and automatic, familiarity-based judgments are stochastically independent.

As a consequence of assumption (c), the *unconditional* probability of a feeling of familiarity, $u_I$, must be equal to both $u_{c+}$ and $u_{c-}$.

Yonelinas et al. (in press) and Yonelinas and Jacoby (1995b) acknowledge that the IMM leads to invalid parameter estimates when response biases differ between inclusion and exclusion test conditions, but they hold on to the IMM for data sets with equal false-alarm rates in the inclusion and exclusion test conditions and—for certain purposes—also for data sets in which these false-alarm rates differ. We do not agree to this proposal for three reasons.

The first reason is that conclusions concerning familiarity effects based on the IMM are problematic even if the false-alarm rates in the inclusion and exclusion test conditions do not differ (i.e., if $g_i = g_e = k$). The parameter $c$ as determined by the IMM (henceforth referred to as $c_{IMM}$) is indeed not contaminated by response biases if $g_i = g_e = k$, that is, it is equal to the parameter $c$ as determined by the EMM for all values of $k$. The $u_{c-}$ parameter of the IMM (henceforth $u_{c-, IMM}$), in contrast, *is* contaminated by response biases for all values of $k > 0$, that is, it reduces to the $u_{c-}$ parameter of the EMM only if $k = 0$. To see this, one has to compare the equations representing $c$ and $u_{c-}$ as functions of the response probabilities for the IMM (see Buchner et al., 1995, Equations 3 and 8) and for the EMM (see Buchner et al., 1995, Equations 13 and 14). While Equation 13 reduces to Equation 3 for all values of $k$, Equation 14 reduces to Equation 8 for $k = 0$ only. This is obvious after inserting Equation 8 of Buchner et al. (1995) into their Equation 14 and observing that

$$u_{c-, IMM} = u_{c-} + k \cdot (1 - u_{c-}). \tag{5}$$

Thus, the IMM parameter $u_{c-, IMM}$ (and, hence, $u_I$ as obtained within the IMM) is not a meaningful measure of familiarity effects because it is artificially raised by guessing relative to the EMM parameter $u_{c-}$, and this artifact becomes more serious the smaller the effects of familiarity and the larger the influence of guessing processes. Moreover, whenever $k$ differs between two groups or experimental manipulations to which the process dissociation procedure is applied, any difference between these groups or manipulations in $u_{c-, IMM}$ or the nonexistence of such a difference can be an artifact of differences in response biases between the conditions. For instance, Roediger and McDermott (1994) have criticized Verfaellie and Treadwell (1993) for comparing the recognition performance of amnesics and normals using the process dissociation procedure and Jacoby's (1991) original measurement model despite the amnesics exhibiting a much higher false-alarm rate than the controls (i.e., $k$ differed between groups). In Phase 1 of this study, participants read words and solved anagrams. They heard words in Phase 2 and then

Among other things, Verfaellie and Treadwell (1993) reported that for the measure representing familiarity-based processes there was no difference between the groups and no interaction between presentation (read vs. anagram) and the group factor. For the purpose of the present illustration, we will focus on the recognition performance for the anagram words. In order to compare the IMM and the EMM we had to recover the response frequencies from the response probabilities reported by Verfaellie and Treadwell (1993). Table 1 presents the estimates for $c_{\mathrm{IMM}}$ and $u_{\mathrm{c-,\,IMM}}$ which we obtained from these recovered response frequencies using the IMM, and it presents the estimates for $c$ and $u_{\mathrm{c-}}$ as well as $g_{\mathrm{i}}$ and $g_{\mathrm{e}}$ of the EMM. Most interestingly, the estimates of $u_{\mathrm{c-,\,IMM}}$ are very similar for the amnesics and the controls. In fact, the IMM with the restriction that $u_{\mathrm{c-,\,IMM}}$(amnesics) = $u_{\mathrm{c-,\,IMM}}$(controls) fits the data well, $G^2(1) = 0.87$, given a critical $\chi^2_{(df = 1,\, \alpha = .05)} = 3.84$.[2] However, because the false-alarm rate was much larger for the amnesics than for the controls (see $\hat{g}_{\mathrm{i}}$ and $\hat{g}_{\mathrm{e}}$), we may suspect that $u_{\mathrm{c-,\,IMM}}$ for the amnesics may be artificially inflated to a larger degree than $u_{\mathrm{c-,\,IMM}}$ for the controls. Thus, the equality of $u_{\mathrm{c-,\,IMM}}$ for both groups may be an artifact. Indeed, an analysis using the EMM shows that when differences in response biases are taken into account, the estimates of the parameter representing automatic, familiarity-based memory effects, $u_{\mathrm{c-}}$, are no longer similar for the amnesics and the controls. Indeed, the EMM with the restriction that $u_{\mathrm{c-}}$(amnesics) = $u_{\mathrm{c-}}$(controls) does

---

[2] The log-likelihood goodness-of-fit statistic $G^2$ is asymptotically chi-square distributed when the null hypothesis holds true with degrees of freedom indicated in parentheses (see Hu & Batchelder, 1994, for details). All model-based statistical analyses reported in this article were conducted using the MBT program by Hu (1993; see Hu & Batchelder, 1994).

not fit the data, $G^2(1) = 4.15$, and has to be rejected, supporting the conclusion that there are differences between the participant groups in the automatic, familiarity-based memory effects.

An example of the reverse effect can be found in recognition data presented by Komatsu et al. (1995) . In their Experiment 1, for instance, participants counted the syllables of low and high frequency words in Phase 1. The estimates of $c_{\mathrm{IMM}}$ and $u_{\mathrm{c-,\,IMM}}$ as well as the estimates of $c$, $u_{\mathrm{c-}}$, $g_{\mathrm{i}}$, and $g_{\mathrm{e}}$ are also presented in Table 1. Komatsu et al. (1995) did not perform statistical tests on these data. However, $\hat{u}_{\mathrm{c-,\,IMM}}$ is evidently much larger for high than for low frequency words, and an analysis using the IMM shows that the model with the restriction that $u_{\mathrm{c-,\,IMM}}$(low frequency words) = $u_{\mathrm{c-,\,IMM}}$(high frequency words) does indeed not fit the data, $G^2(1) = 13.94$, which would support the conclusion that automatic, familiarity-based effects of memory are larger for high than for low frequency words. The problem with this result is that participants were more than twice as likely to accept high (.34) than low frequency distractors (.16), raising the suspicion that $u_{\mathrm{c-,\,IMM}}$ for the high frequency words may have been artificially inflated to a larger degree than $u_{\mathrm{c-,\,IMM}}$ for the low frequency words. Sure enough, when analyzing the data using the EMM, the difference in the estimates of the parameters representing automatic, familiarity-based memory effects largely disappeared, and the EMM with the equality restrictions on $u_{\mathrm{c-}}$ fitted the data very well, $G^2(1) = 0.74$.

These two examples show that even if the false-alarm rates do not differ between inclusion and exclusion test conditions (i.e., $g_{\mathrm{i}} = g_{\mathrm{e}} = k$), using the IMM is dangerous because it may lead to erroneous conclusions if $k$ differs between conditions. However, the IMM should also be avoided *if the false-alarm rates are constant* across all test conditions and across all groups or experimental manipulations, and *even if the absolute size of the familiarity effect is irrelevant* (i.e., one is interested only in possible differences of the familiarity effects between two conditions). Equation 5 implies that, given constant $k$, treatment effects on $u_{\mathrm{c-}}$ are mirrored in $u_{\mathrm{c-,\,IMM}}$ only qualitatively, and not quantitatively. If $u_{\mathrm{c-}}(\mathrm{A}) - u_{\mathrm{c-}}(\mathrm{B})$ denotes the difference in $u_{\mathrm{c-}}$ between treatments or groups A and B, and $u_{\mathrm{c-,\,IMM}}(\mathrm{A}) - u_{\mathrm{c-,\,IMM}}(\mathrm{B})$ denotes the corresponding difference in $u_{\mathrm{c-,\,IMM}}$, then the exact relation as derived from Equation

**Table 1.**
*Maximum-likelihood parameter estimates for data reported by Verfaellie and Treadwell (1993) and by Komatsu, Graf, and Uttl (1995, Experiment 1) according to the IMM (Jacoby, 1991) and the EMM (Buchner et al., 1995).*

| Verfaellie and Treadwell (1993) | | Amnesics | Normal Controls |
|---|---|---|---|
| IMM | $c_{\mathrm{IMM}}$ | .00 | .32 |
| | $u_{\mathrm{c-,\,IMM}}$ | .44 | .51 |
| EMM | $c$ | .00 | .33 |
| | $u_{\mathrm{c-}}$ | .21 | .40 |
| | $g_{\mathrm{i}}$ | .22 | .17 |
| | $g_{\mathrm{e}}$ | .28 | .19 |
| **Komatsu, Graf, and Uttl (1995)** | | **Low Frequency** | **High Frequency** |
| IMM | $c_{\mathrm{IMM}}$ | .48 | .30 |
| | $u_{\mathrm{c-,\,IMM}}$ | .48 | .63 |
| EMM | $c$ | .48 | .26 |
| | $u_{\mathrm{c-}}$ | .38 | .43 |
| | $g_{\mathrm{i}}$ | .15 | .38 |
| | $g_{\mathrm{e}}$ | .16 | .29 |

*Note*. There are slight differences between the estimates reported by Verfaellie and Treadwell (1993) and those based on the recovered response frequencies using the IMM (the largest such difference being.016). Verfaellie and Treadwell (1993, p. 8) excluded 7% of the responses to anagrams because these anagrams had not been solved in Phase 1. We reduced the recovered response frequencies uniformly by 7%, but our estimates must differ slightly from the 'correct' estimates to the degree to which the number of anagrams solved correctly differed between groups and test conditions.

$$u_{\mathrm{c-,\,IMM}}(\mathrm{A}) - u_{\mathrm{c-,\,IMM}}(\mathrm{B}) = (1 - k) \cdot (u_{\mathrm{c-}}(\mathrm{A}) - u_{\mathrm{c-}}(\mathrm{B})). \tag{6}$$

In other words, the familiarity treatment effect is diluted when it is analyzed in terms of $u_{\mathrm{c-,\,IMM}}$ to the degree to which the (constant) false-alarm rate $k$ increases, which reduces the power of statistical tests of the familiarity effect. Therefore, treatment effects which would turn out to be statistically significant when analyzed with respect to $u_{\mathrm{c-}}$ may be statistically insignificant when analyzed with respect to $u_{\mathrm{c-,\,IMM}}$. Note that the finding of 'invariances' in the effects of automatic memory considered important evidence in support of the "independence assumption" (i.e., the assumption that controlled and automatic processes make independent contributions to performance) favored by some researchers (e.g., Jacoby & Begg, 1995; Jacoby, Toth, Yonelinas & Debner, 1994; Jacoby, Yonelinas & Jennings, in press; Toth et al., 1994) . One might thus argue that finding such invariances is easier when using $u_{c-,\,IMM}$ rather than $u_{c-}$. Notwithstanding that, we believe that such invariances in the measure of automatic processes given differences in the measure of controlled processes cannot be counted as evidence in favor of the independence assumption, and we will give reasons for this further on.

Our second reason to reject the IMM is that the situation becomes even more tricky when $g_i$ is larger than $g_e$. This case occurs quite often in practice and it has the unfortunate consequence that not only $u_{c-,\,IMM}$ but also $c_{IMM}$—the IMM's recollection parameter—is artificially inflated by effects of response biases. Although Yonelinas et al. (in press) agree with this statement, they nevertheless recommend using Jacoby's (1991) IMM whenever "... the goal of the study is to examine the qualitative effects of a variable on recollection and familiarity" (Yonelinas et al., in press, p. 22 of the preprint) . Yonelinas and colleagues seem to assume that the disturbing effects of $g_i - g_e$ differences are *additive* so that they are canceled out when (a) mean differences between groups or experimental manipulations are analyzed and (b) the $g_i - g_e$ differences are nonzero but constant across groups or experimental manipulations. This is not quite correct. To see this, one has to write Jacoby's (1991) recollection measure $c_{IMM}$ (i.e., the difference between inclusion and exclusion hit rates, $p_{1i} - p_{1e}$) as a function of the EMM parameters. By subtracting the EMM model equations for $p_{1i}$ and $p_{1e}$ (i.e., Equations 1 and 3) we obtain

$$c_{IMM} = c + (1 - c) \cdot (1 - u_{c-}) \cdot (g_i - g_e). \tag{7}$$

Obviously, $c_{IMM}$ is not only contaminated by the effects of response biases but also by the effects of familiarity whenever $g_i$ and $g_e$ differ. Thus, contrary to what has been claimed by Yonelinas et al. (in press) , across-groups comparisons of $c_{IMM}$ may cause misleading conclusions even when both inclusion and exclusion false-alarm rates are constant across treatments. Assume, for instance, that two groups do but in their familiarity parameters $u_{C-}$. When recollection is measured in terms of $c_{IMM}$ then, according to Equation 7, this must lead to the erroneous conclusion that there is a group effect in the contributions of recollection to performance.

In summary, it is quite dangerous to use the IMM whenever the false-alarm rates differ between groups or experimental manipulations, between test conditions, or between both, and it is also rather dangerous to do so whenever false-alarm rates are actually constant across all conditions.

Our third objection to the IMM is one which is at the same time an objection to the DPSDM to be discussed below. The objection concerns the independence assumption formally expressed as $u_{C+} = u_{C-}$. This assumption can be added to the EMM without sharing the other problematic assumptions of the IMM (cf. Buchner et al., 1995). We will therefore discuss it as a possible supplement to the EMM.

Jacoby and colleagues have presented arguments in defense of the independence assumption (e.g., Jacoby & Begg, 1995; Jacoby et al., 1994; Jacoby et al., in press; Toth et al., 1994) . However, the assumption has also been criticized on various grounds and remains problematic the more so because alternative assumptions can be defended with good reasons, too (e.g., Curran & Hintzman, 1995; Joordens & Merikle, 1993; Richardson-Klavehn, Gardiner & Java, 1995; Russo & Andrade, 1995) . Unfortunately, empirical tests of different assumptions about the relation between $u_{C+}$ and $u_{C-}$ are impossible within the traditional process dissociation framework because $u_{C+}$ is not an identifiable parameter, that is, it is not uniquely determined by process dissociation data (cf. Buchner et al., 1995). We therefore cannot establish that $u_{C-} = u_{C+}$ (see Assumption (c) in the discussion of the extended measurement model above). This also implies that the finding of invariances in the effects of automatic memory processes across different experimental manipulations does not help us in deciding whether the independence assumption holds true or not. In terms of the EMM, showing that $u_{C-}$ is insensitive to certain experimental manipulations that affect other model parameters reveals no information about $u_I$ as long as we have no information about $u_{C+}$. As Russo and Andrade (1995, p. 421) have pointed out, it would be inappropriate to use invariances in $u_{C-}$ across conditions A and B to demonstrate independence (i.e., $u_I(A) = u_{C-}(A) = u_{C+}(A) = u_I(B) = u_{C-}(B) = u_{C+}(B)$) by

We can see two ways to solve this dilemma: (a) One could extend the traditional process dissociation procedure such that new measurement models can be formulated in which both $u_{C+}$ and $u_{C-}$ are identifiable or (b), one could rely upon the traditional process dissociation framework and abstain from any assumptions about $u_{C+}$. We prefer the second option because the parameter $u_{C+}$ does not appear to be necessary in order to assess automatic, familiarity-based memory processes. The identifiable parameter $u_{C-}$ provides a much safer ground for meaningful statements, and it will necessarily mirror the numerical results as would be obtained with $u_I$: It is important to keep in mind that $u_{C-}$ and $u_I$ differ only in interpretation, not in numerical value. Parameter $u_{C-}$ is a conditional probability that is uniquely determined by response probabilities whereas $u_I$ is an unconditional probability which rests on a questionable assumption about $u_{C+}$.

To put it in a nutshell, there are a number of serious arguments against the assumptions underlying the IMM, and we see none in favor of it. Pragmatic reasons such as computational simplicity cannot challenge this conclusion because statistical analyses in the EMM framework are conducted as easily as in the IMM framework (cf. Buchner et al., 1995).

***The Dual-Process Signal-Detection Model***

While quite a few process dissociation measurement models that have been suggested in the literature can be derived as special cases of the EMM (cf. Buchner et al., 1995), alternatives to the EMM framework are of course conceivable. Elaborating on prior work of Jacoby et al. (1993) and Yonelinas (1994) , Yonelinas et al. (in press) have recently suggested a particularly attractive alternative which also accounts for simultaneous effects of controlled processes, automatic processes, and response biases on task performance. Their DPSDM is similar to the EMM insofar as recollection is conceived of as a discrete cognitive state which is either present (with probability *c*) or absent (with probability $1-c$) in both the inclusion and the exclusion test conditions. It differs from the EMM, however, in conceptualizing familiarity as a continuous latent random variable rather than a discrete cognitive state. In analogy to standard signal-detection theory, Yonelinas et al. (in press) assume that the deviation which for convenience are taken to be $-d'/2$ and 1, respectively (cf. Macmillan & Creelman, 1991, p. 35) . The familiarity $U_1$ of Phase 1 items, in contrast, is increased by some (additive) amount $d'$ relative to distractor items due to prior processing, so that $U_1$ is also normally distributed but with mean $+d'/2$ and standard deviation 1. The parameter $d'$, therefore, may serve as a measure of the *familiarity increase* caused by studying the Phase 1 items. Building upon Jacoby's (1991) IMM, recollection and familiarity are assumed to be stochastically independent, so that the *conditional* familiarity distribution of Phase 1 items, given a failure of recollection, must mirror their *unconditional* familiarity distribution. As we will discuss below, this last assumption is absolutely crucial in the DPSDM context, because the DPSDM equations cannot be derived without it.

The assumptions of the DPSDM about consciously recollected Phase 1 items are the same as those of the EMM. The DPSDM differs from the EMM in that it implies that items which cannot be consciously recollected are judged *old* whenever their familiarity value exceeds some response criterion $k_i$ in the inclusion test condition or $k_e$ in the exclusion test condition. Let $\Phi(u) := p(U \leq u)$ denote the distribution function of a standard normal random variable $U$ with mean 0 and standard deviation 1. Then the probability that the familiarity value of a Phase 1 item exceeds the response criterion $k_i$ can be written as $p(U_1 > k_i) = 1 - p(U_1 \leq k_i) = 1 - \Phi(k_i - d'/2) = \Phi(d'/2 - k_i)$. Because the conditional familiarity distribution, given a failure of recollection, is assumed to match the unconditional distribution, Phase 1 items in the inclusion test condition are judged *old* with probability

$$p_{1i} = c + (1 - c) \cdot \Phi(d'/2 - k_i), \tag{8}$$

where the first term of the sum corresponds to a state of recollection and the second term to a familiarity value exceeding the response criterion *not* accompanied by a conscious recollection. In a completely analogous manner, the remaining three model equations can be derived:

$$p_{di} = \Phi(-d'/2 - k_i), \tag{9}$$

$$p_{1e} = (1 - c) \cdot \Phi(d'/2 - k_e), \tag{10}$$

and

$$p_{de} = \Phi(-d'/2 - k_e). \tag{11}$$

As we show in the Appendix, the four response probabilities $p_{1i}$, $p_{di}$, $p_{1e}$, and $p_{de}$ uniquely determine the two memory parameters $c$ and $d'$ as well as the two response bias parameters $k_i$ and $k_e$. Thus, the model is identifiable. Unfortunately, however, satisfactory solutions to the statistical problems of parameter estimation, computation of confidence intervals, and goodness-of-fit testing do not seem to exist at present. We acknowledge that Yonelinas et al. (in press, Footnote 1) took a first step by offering an algorithm which computes estimates of the four model parameters that exactly predict observed hit rates ($\hat{p}_{1i}$ and $\hat{p}_{1e}$) and false-alarm rates ($\hat{p}_{di}$ and $\hat{p}_{de}$) when inserted into the Equations 8 to 11.[3] However, this algorithm does not apply to restricted versions of the model which will often be needed in practice. For instance, one might be interested in estimating the parameters of a DPSDM for two experimental manipulations A and B assuming that $d'$(A) in Group A equals $d'$(B) in Group B. In this case, parameter estimates which predict the observed data perfectly will most likely not exist. Yonelinas et al. (in press) seem to suggest least-squares solutions for problems like this one, but this procedure will lead to estimates with unsatisfactory statistical properties.

To our knowledge, methods to compute confidence intervals or goodness-of-fit tests for restricted and unrestricted versions of the DPSDM have not been proposed so far. In practice, therefore, statistical data analyses within the framework of the DPSDM must proceed as follows:

(1) The inclusion versus exclusion test conditions are manipulated *within-subject*, such that each participant generates hit rates and false-alarm rates for both the inclusion and exclusion test conditions.

[3] Alternatively, these estimates can be derived by inserting relative frequencies as estimates of the response probabilities into Equations A1 to A5 in the Appendix. In a first step, the

(2) For each participant, the hit and false-alarm rates are then transformed into estimates of the model parameters using the algorithm suggested by Yonelinas et al. (in press) .

(3) These estimates are treated as dependent variables in ANOVAs or MANOVAs, and the usual *F*-tests are performed in order to assess treatment effects with respect to recollection, familiarity, or response bias.

Buchner et al. (1995, p. 141) have already criticized an analogous procedure which has often been used to analyze data in the framework of Jacoby's (1991) IMM. The same arguments apply here and there: Within-subject manipulations of the test conditions and single-participant estimates are problematic because (a) participants performing perfectly on Phase 1 items (i.e., their individual $\hat{p}_{1i} = 1$ and $\hat{p}_{1e} = 0$) have to be dropped from the data analyses as a consequence of undefined parameter estimates, (b) quite a few participants may have difficulties to follow the instructions to both types of test in succession (Graf & Komatsu, 1994; Richardson-Klavehn et al., 1995) and thus might also have to be dropped from the analyses, (c) systematic biases may result as a consequence of selective loss of participants in different experimental groups, (d) test order problems (e.g., carry-over effects across test conditions) may arise, and (e) parameter estimates for single participants will be relatively unreliable which adds error variance to the ANOVAs and MANOVAs and thus reduces the power of the *F*-tests.

For these reasons, a more elaborated statistical theory of the DPSDM and corresponding software for data analyses is highly desirable. This will take more effort than in case of the EMM, however, because formally the DPSDM neither is a multinomial GPT model (as defined and analyzed statistically by Hu & Batchelder, 1994) nor can it be reparameterized so as to make it a GPT model. (Interestingly, however, the DPSDM can be *generalized* to a distribution-free DPSDM which formally *is* a GPT model. We will return to this issue later.)

Besides the statistical problems involved, the DPSDM has the disadvantage of relying heavily upon the independence assumption with respect to recollection and the automatic, familiarity-based processes. Whereas this problematic assumption is framework in order to derive the conditional familiarity distribution, given a failure of recollection: If (a) Phase 1 items are recollected with a probability *c* which is larger than 0 and less than 1, (b) the familiarity distribution of distractors is a normal distribution, (c) prior processing increases the familiarity of Phase 1 items by amount *d*' relative to the distractors, and (d) recollection and familiarity are *not* stochastically independent, then the conditional familiarity distribution of nonrecollected Phase 1 items *cannot* be a normal distribution. As long as the exact nature of the dependency between recollection and familiarity is not specified further, this is all we know about the conditional familiarity distribution. As a consequence, the probability of a nonrecollected Phase 1 item exceeding the familiarity response criterion is an unknown function of *d*' and the response criterion. Thus, the model equations can no longer be written as functions of only four model parameters $c$, $d$, $k_i$, and $k_e$. As a result, the model becomes nonidentifiable.

In this sense, then, the DPSDM forms a community of fate with the independence assumption whereas the EMM does not, that is, the validity of the EMM is independent on whether the independence assumption holds true or not. In view of the serious criticisms raised against the independence assumption, this must be regarded as an advantage of the EMM. However, the DPSDM is not invalidated at this point because the independence assumption *may* hold true. In our opinion, therefore, the choice of the model framework (EMM versus DPSDM) should primarily depend on results of empirical model evaluations, not on a priori arguments raised against specific assumptions of the models. We will return to this issue below after discussing some alternative methods of correcting for response bias in the process dissociation procedure.

### *Alternative Methods to Correct for Response Bias*

As stated in the introduction, alternative methods of correcting for response bias have been proposed by Jacoby et al. (1993) , Komatsu et al. (1995) , and Roediger and McDermott (1994). In contrast to the EMM and DPSDM, these methods were not derived from process dissociation measurement models. Rather they were based on assumptions about the nature of response biases taken from other sources.

The additive $u_I$ adjustment as suggested by Jacoby et al. (1993) has already been discussed by Buchner et al. (1995, p. 142) and has been found to be inappropriate because it does not eliminate the response-bias contamination of $u_I$. Yonelinas et al. (in press) provide a good discussion of the correction methods suggested by Komatsu et al. (1995) and by Roediger and McDermott (1994) . These methods are sequential in nature: In a first step, one tries to decontaminate the probabilities of *old* responses in the inclusion and exclusion test conditions from the effects of response bias. In a second step, these corrected probabilities are entered into the model equations for the IMM as suggested by Jacoby (1991). The idea behind these sequential correction methods is that response biases can be eliminated from the data in a way that is model-independent. However, any generation or transformation of measurement values necessarily implies a model, at least an implicit one (cf. Gigerenzer, 1981) . This raises the question as to whether or not the models underlying the two steps are compatible. They are not, and this is so because the implicit models underlying the corrections in the first step are high-threshold models for simple yes-no recognition tasks (cf. Macmillan & Creelman, 1991, chap. 4) which do not take into account that the response probabilities are affected simultaneously by controlled processes, automatic processes, and response biases. As a consequence, some new problems emerge (cf. Yonelinas et al., in press) . For instance, the correction methods cannot account for false-alarm rates that are larger than the hit rates in the exclusion test condition, which are often observed in practice. Moreover, negative familiarity estimates result quite often as a consequence of the correction procedures. These problems do not occur when the EMM or the DPSDM are used. Therefore, we will be concerned with these two model frameworks in the remainder of this article.

## Evaluation of Process Dissociation Measurement Models

The EMM and the DPSDM currently seem to be the most valuable measurement tools for the process dissociation procedure. Which of these two tools should be used in practice? As outlined above, there are some pragmatic arguments in favor of the EMM, because the statistical analysis of the DPSDM needs some compared to the overriding *validity problem*: Do the models' parameters measure what they are supposed to measure?

We agree with Yonelinas and Jacoby (1995b) that this question should be answered empirically and not on a priori grounds. Because both the EMM and the DPSDM aim at measures of recollection and familiarity which are uncontaminated by effects of response bias, it seems natural to begin evaluating the models by testing whether the memory parameters remain unaffected when response bias varies. Both within and outside the process dissociation framework, two different techniques have been used in such tests: (a) experimental manipulations of response bias and (b) use of confidence ratings. How do the process dissociation measurement models come off when evaluated in these ways? We will first study evaluations based upon confidence ratings, and turn to response bias manipulations subsequently.

***Confidence Ratings***

Rather than responding simply *old* or *new* to every item, participants in a typical confidence rating experiment also state how sure they are of their recognition judgment on a $n$-point rating scale, $n > 2$. The problem is that both the EMM and the DPSDM have been introduced as measurement models for dichotomous *old-new* (or *yes-no*) judgments. Before these models can be applied to confidence rating data, they need to be extended to incorporate judgments on $n$-point rating scales. We will demonstrate that there is no unique way to do this. The results of the evaluations will therefore depend on exactly *how* the models are extended.

*Evaluation of the DPSDM.*

One way to extend the DPSDM to confidence rating data is to assume that recollected Phase 1 items are always responded to with the most extreme confidence ratings, that is, with $X = n$ in the inclusion test condition (*sure old*) and with $X = 1$ in the exclusion test condition (*sure new*). Familiarity-based responses to distractors and Phase 1 items that were not recollected, in contrast, are assumed to be based on response criteria associated with each of the $n$ rating categories: Participants will report the particular confidence level for which the familiarity value of the item just exceeds the response criterion. This is the DPSDM version for confidence ratings preferred by Yonelinas (1994) and by Yonelinas et al. (in press) .

If $k_{i(j)}$ and $k_{e(j)}$ denote the response criteria associated with rating category $X = j$ in the inclusion and exclusion test conditions, respectively, and given all other assumptions of the DPSDM are preserved, then the probability $p_{1i}(X \geq j)$, $j = 2, \ldots, n$, that the confidence rating $X$ is not less than $j$ for Phase 1 items in the inclusion test condition can be written in complete analogy to Equation 8 as

$$p_{1i}(X \geq j) = c + (1 - c) \cdot \Phi(d'/2 - k_{i(j)}). \tag{12}$$

The same holds true for both types of distractor items and Phase 1 items in the exclusion test condition:

$$p_{di}(X \geq j) = \Phi(-d'/2 - k_{i(j)}), \tag{13}$$

$$p_{1e}(X \geq j) = (1 - c) \cdot \Phi(d'/2 - k_{e(j)}), \tag{14}$$

$$p_{de}(X \geq j) = \Phi(-d'/2 - k_{e(j)}). \tag{15}$$

These model equations define a pair of parallel ROCs, both of which depend on parameters $c$ and $d'$. The first ROC curve expresses the hit rate as a function of the false-alarm rate for the inclusion test condition,

$$p_{1i}(X \geq j) = c + (1 - c) \cdot \Phi(\Phi^{-1}(p_{di}(X \geq j)) + d'), \tag{16}$$

and the second curve does that for the exclusion test condition,

$$p_{1e}(X \geq j) = (1 - c) \cdot \Phi(\Phi^{-1}(p_{de}(X \geq j)) + d'). \tag{17}$$

Yonelinas (1994) evaluated these ROC curves using a list discrimination version of the process dissociation procedure. This paradigm differs from the recognition paradigm described above because in each of the two test conditions one type of items must be excluded: Under *list1?* instructions, participants must respond *old* to Phase 1 items and *new* to Phase 2 items whereas under *list2?* instructions judgments must be reversed. Under both instructions distractors must be called *new*. Instead of just responding *old* or *new*, participants were asked to express the *sure old* (6). For purposes of data analyses, the data were collapsed across instructions as follows: Responses to Phase 1 items under *list1?* instructions and to Phase 2 items under *list2?* instructions served as inclusion data, responses to Phase 1 items under *list2?* instructions and to Phase 2 items under *list1?* instructions as exclusion data. Responses to distractors were also collapsed across instructions so that there was only one class of distractors instead of two. Note that in analyzing these data the assumption is made that memory parameters do not differ between Phase 1 and Phase 2 items. Because the retention interval is shorter for Phase 2 items as compared to Phase 1 items it might well be expected that recollection and perhaps even familiarity effects are more pronounced for Phase 2 items. In fact, there is some support for this hypothesis in Yonelinas' (1994) data.[4] However, we will follow Yonelinas (1994) and refer to the aggregated data only.

Qualitatively, the shape of the ROC curves (16) and (17) can account rather well for the data patterns obtained by Yonelinas (1994, Experiments 1 to 3), but formal goodness-of-fit tests have not been conducted which is why a *definite* answer about this model's fit to Yonelinas' data awaits to be given. To anticipate, however, we can *infer* that quantitatively the misfit of the DPSDM must be considerable, at least with respect to Yonelinas' (1994) Experiments 1 and 2. More precisely, we can infer the misfit of the DPSDM from the misfit of its generalized version (i.e., a distribution-free DPSDM) which will be described in the next section. The DPSDM is a submodel of the distribution-free DPSDM, and the misfit of the latter therefore implies the misfit of the former.

Note also that it is very important to fit both inclusion and exclusion ROCs simultaneously using *the same* parameters $d'$ and $c$. It is not sufficient—as has been done by Yonelinas and Jacoby (1995b, Figure 1) —to fit only an inclusion ROC without also fitting the corresponding exclusion ROC. When both inclusion and exclusion ROCs are considered in one figure, the fit of the DPSDM is less impressive than suggested by Figure 1 of Yonelinas and Jacoby (1995b) .

There are, of course, alternative ways to extend the DPSDM to rating scales. For example, tendencies to avoid extreme response categories could prevent some participants from choosing confidence levels 1 or 6 for recollected items in the inclusion or exclusion test conditions. If the possibility of intermediate confidence ratings for recollected items is implemented into the model, a more complicated pair of (not necessarily parallel) ROC curves results. In fact—as already discussed by Yonelinas (1994)—pronounced deviations of the empirical data from the ROC curves (16) and (17) at the extreme confidence levels suggest that such a model variant might provide a significantly better fit to the data. However, this has not been explored thus far.

*Evaluation of the EMM.*

Yonelinas and Jacoby (1995b) extended the EMM to confidence ratings by conceptualizing familiarity as a continuous random variable for this model, too. Although this deviates from the EMM assumptions as stated by Buchner et al. (1995), it is indeed possible to derive the EMM model equations within a modified DPSDM framework by assuming (a) that familiarity is a continuous random variable with density functions $f_d(u)$ for distractor items and $f_1(u)$ for Phase 1 items, (b) that there is some familiarity value $s$ such that $f_d(u) = 0$ for all values $u > s$, (c) that the likelihood ratio $f_1(u)/f_d(u)$ is constant for all values $u \leq s$, (d) that the response criteria $k_i$ and $k_e$ are always smaller than $s$, and (e) that all other DPSDM assumptions (except the normal distribution assumptions) hold true. The EMM model equations follow from these alternative assumptions because the EMM reduces to a simple one-high threshold (1HT) model for yes-no recognition tests when only nonrecollected items are considered as target items. As is well known, the model equations corresponding to the 1HT model can be derived by either conceptualizing recognition (or familiarity) as a discrete state or by conceptualizing it as a continuous random variable, coupled with the distribution assumptions mentioned above (see Macmillan & Creelman, 1991, chap. 4, for a very instructive discussion of these issues).

When the EMM is rephrased in terms of continuous underlying familiarity distributions, a straightforward extension to rating scales can be arrived at in the and Jacoby (1995b) did: They adopted the modified DPSDM framework for the EMM model equations, assumed each rating category to be associated with a unique response criterion, and posited that recollected Phase 1 items are always responded to with the most extreme confidence ratings. This way they arrived at a pair of parallel, linear ROC curves for confidence ratings, both with slope $(1-c) \cdot (1-u_{c-})$ and with intercepts $c + (1-c) \cdot u_{c-}$ for the inclusion ROC and $(1-c) \cdot u_{c-}$ for the exclusion ROC.

As illustrated by Figure 1 of Yonelinas and Jacoby (1995b) , this EMM extension obviously cannot account for the data obtained by Yonelinas (1994) and therefore must be rejected. Note, however, that this only shows the EMM extension preferred by Yonelinas and Jacoby (1995b) to be inadequate and is not damaging to the EMM itself. To demonstrate this, we suggest an alternative extension of the EMM to rating scales which is illustrated in Figure 1. Following Buchner et al. (1995), this extension is based on the notion of familiarity as a cognitive state rather than familiarity as a continuous random variable. Note that in its structural part, this EMM version for confidence ratings is completely isomorphic to the EMM as presented by Buchner et al. (1995, Figure 2). As in Buchner et al. (1995), we included the nonidentifiable parameter $u_{c+}$ in order to illustrate that it is possible to derive independence, redundancy, and exclusivity variants of this model by assuming $u_{c+} = u_{c-}$, $u_{c+} = 1$, and $u_{c+} = 0$, respectively. However, we prefer to make no assumptions about $u_{c+}$ and, therefore, simply to drop that parameter from the model.

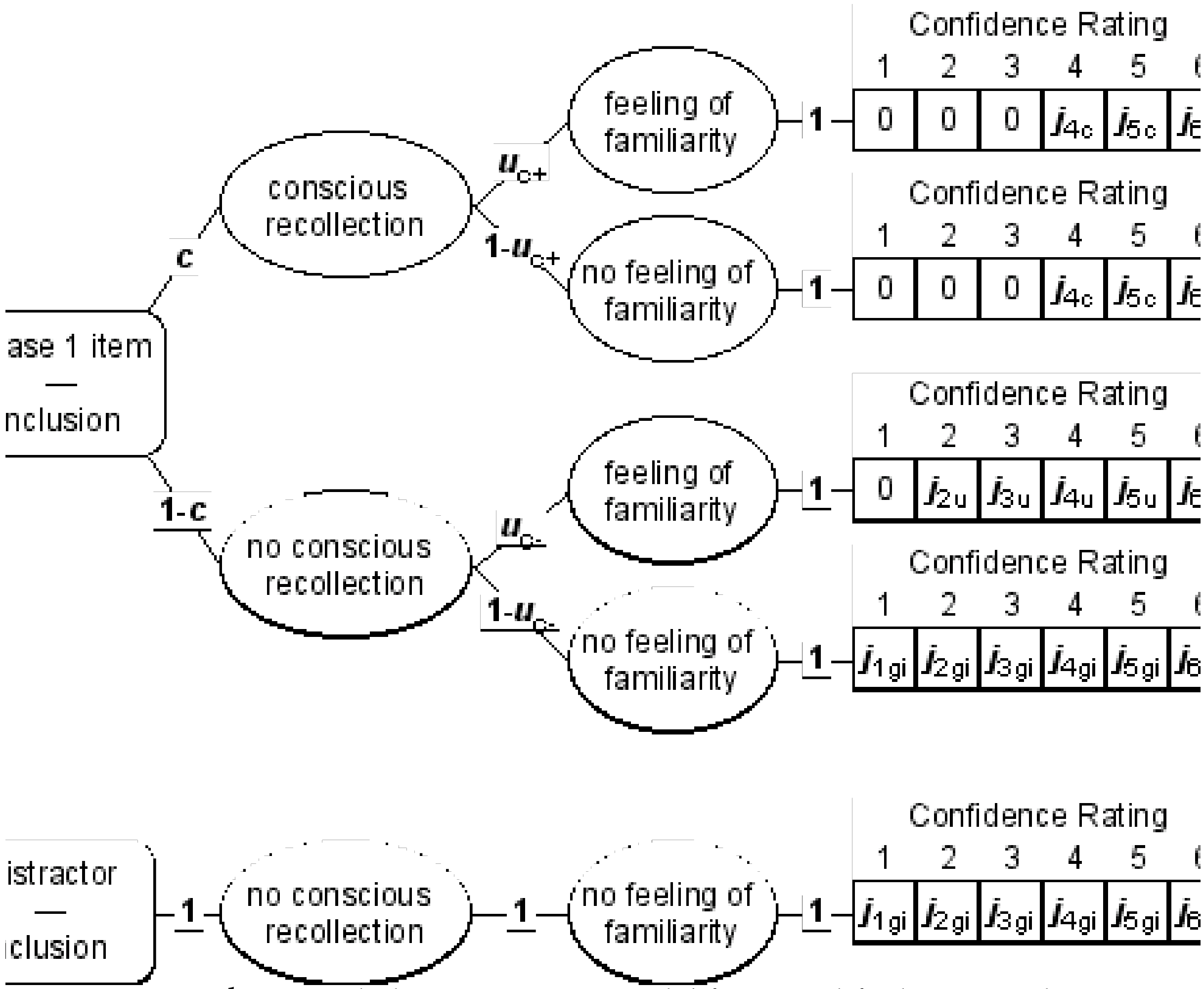



*Figure 1a.* The extended measurement model for a modified process dissociation procedure based on confidence ratings (inclusion test condition). Parameter $c$ denotes the unconditional probability of conscious recollections. Parameters $u_{c+}$ and $u_{c-}$ denote the conditional probabilities of automatic, familiarity-based memory effects if an item is and is not recollected, respectively. The $j$-parameters denote the conditional probabilities given a conscious recollection ($j_{1c}$ to $j_{6c}$), given an unconscious memory effect if an item is not recollected ($j_{1u}$ to $j_{6u}$), and given that guessing occurs ($j_{1g}$ to $j_{6g}$).

The model illustrated in Figure 1 differs from the EMM in that assumptions about conditional *old* versus *new* response probabilities were replaced by assumptions about conditional probabilities of confidence ratings given a certain cognitive state. The model parameters denoted by the letter $j$ represent these conditional probabilities. For instance, the model assumes that consciously recollected Phase 1 items are given the confidence ratings 4, 5, or 6 in the inclusion test condition and 1, 2, or 3 in exclusion test condition.

Thus, in order to account for possible biases against extreme ratings, the model provides for some small proportion of intermediate ratings, given a conscious

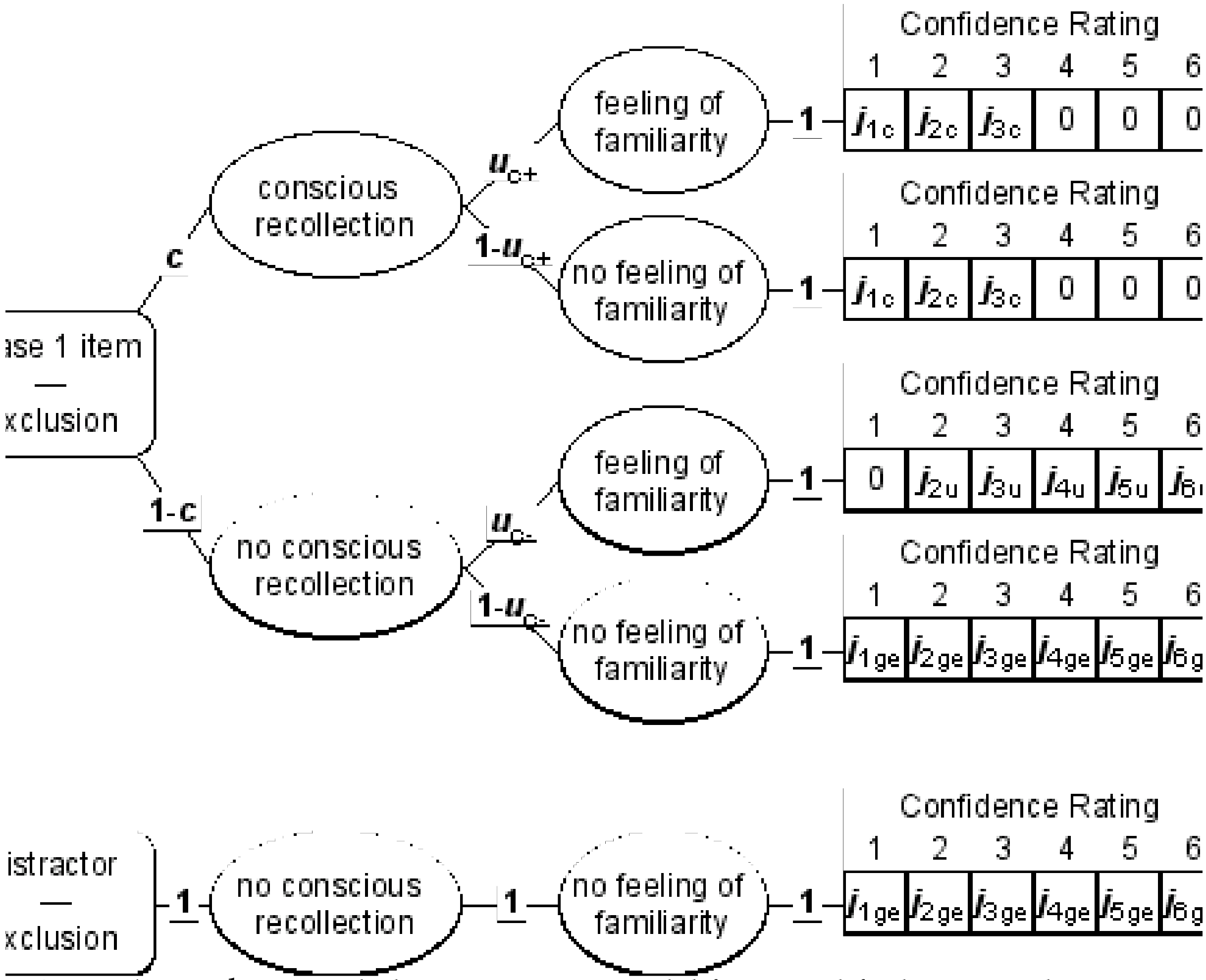



*Figure 1b.* The extended measurement model for a modified process dissociation procedure based on confidence ratings (exclusion test condition). Parameter *c* denotes the unconditional probability of conscious recollections. Parameters $u_{c+}$ and $u_{c-}$ denote the conditional probabilities of automatic, familiarity-based memory effects if an item is and is not recollected, respectively. The *j*-parameters denote the conditional probabilities given a conscious recollection ($j_{1c}$ to $j_{6c}$), given an unconscious memory effect if an item is not recollected ($j_{1u}$ to $j_{6u}$), and given that guessing occurs ($j_{1g}$ to $j_{6g}$).

are responded to with any of the response categories 2 to 6; only *sure new* judgments are excluded by the model. Finally, guessing may result in any of the six confidence rating categories and, therefore, the guessing parameters $j_{lgi}$, $l = 1,..., 6$, and $j_{lge}$, $l = 1,..., 6$, for the inclusion and exclusion test conditions, respectively, are not restricted at all.

Despite the large number of free parameters, this EMM extension to rating scales can be shown to be identifiable. Model equations expressing the probabilities of each of the six confidence ratings for each of the item types and test conditions as functions of the model parameters are easily derived from the processing tree into a particular response category for a specific tree. Also, a pair of ROC curves is easily obtained by successively computing $p_{1i}(X \geq 6) = p_{1i}(X = 6)$, $p_{1i}(X \geq 5) = p_{1i}(X = 6) + p_{1i}(X = 5)$, $p_{1i}(X \geq 4) = p_{1i}(X \geq 5) + p_{1i}(X = 4)$, and so forth, for both test conditions and item types. Due to space limitations, we will omit the extensive but easy-to-derive formulas here.

The pair of ROC curves implied by the EMM extension illustrated in Figure 1 consists of two functions which are noncontinuous and not necessarily parallel. They fit exactly the empirical ROC data reported by Yonelinas (1994, Experiments 1 to 3). Thus, the model's fit is perfect for all data sets.[5] This should not be too surprising because the goodness-of-fit test has $df = 0$, that is, there are as many free parameters as there are independent category probabilities to fit.

Much more important than the perfect fit is whether the parameter estimates obtained for this model make sense when it is applied to the raw frequencies of Yonelinas' Experiments 1 to 3. To examine this, we will first focus on the parameters representing states of varying confidence before we inspect the parameters representing controlled recollections and automatic, familiarity-based processes.

As one would expect, the preferred response categories for recollected items are indeed 6 and 1 in the inclusion and exclusion test condition, respectively (see Table 2). However, up to about 30% of the recollected items receive intermediate confidence ratings, probably as a consequence of biases against extreme ratings.

It should also be expected that response probabilities for recollected items (a) are symmetric in the inclusion and exclusion test conditions such that $j_{6c} = j_{1c}$, $j_{5c} = j_{2c}$, and $j_{4c} = j_{3c}$ and (b) are independent from the experimental treatment under which they were obtained. The first of these expectations is obvious, and the second follows from the reasoning that when participants consciously recollect items, their high level of confidence in *old* judgments (in the inclusion test condition) or *new* judgments (in the exclusion test condition) should not depend on whether the items stem from short or long lists (Yonelinas, 1994, Experiments 1 and 2) or were

[5] To be precise, the fit to the data from Experiment 3 of Yonelinas (1994) was not completely perfect; the log-likelihood goodness-of-fit statistic was $G^2 = 0.21$. This was due to the fact that all

**Table 2.**
*Maximum-likelihood parameter estimates for the extended measurement model depicted in Figure 1 when applied to the data reported by Yonelinas (1994).*

| | Maximum-likelihood estimate for | | | | | |
|---|---|---|---|---|---|---|
| | Experiment 1 | | Experiment 2 | | Experiment 3 | |
| Parameter | short lists | long lists | short lists | long lists | weak | strong |
| $c$ | .565 | .394 | .615 | .391 | .111 | .218 |
| $u_{c-}$ | .785 | .776 | .739 | .754 | .479 | .505 |
| $j_{1c}$ [a] | .753 | .722 | .714 | .642 | .860 | .739 |
| $j_{2c}$ [a] | .221 | .234 | .217 | .287 | .101 | .153 |
| $j_{3c}$ [a] | .026 | .044 | .069 | .071 | .039 | .108 |
| $j_{4c}$ [b] | .025 | .028 | .038 | .024 | .000 | .064 |
| $j_{5c}$ [b] | .172 | .180 | .196 | .280 | .009 | .193 |
| $j_{6c}$ [b] | .803 | .792 | .766 | .696 | .999 | .743 |
| $j_{1u}$ [c] | — | — | — | — | — | — |
| $j_{2u}$ | .154 | .169 | .187 | .188 | .159 | .073 |
| $j_{3u}$ | .231 | .229 | .239 | .191 | .147 | .119 |
| $j_{4u}$ | .306 | .299 | .229 | .314 | .175 | .158 |
| $j_{5u}$ | .200 | .232 | .224 | .227 | .289 | ..278 |
| $j_{6u}$ | .109 | .071 | .127 | .080 | .230 | .372 |
| $j_{1g}$ | .415 | .256 | .442 | .266 | .379 | .379 |
| $j_{2g}$ | .328 | .361 | .430 | .460 | .225 | .225 |
| $j_{3g}$ | .167 | .233 | .085 | .167 | .157 | .157 |
| $j_{4g}$ | .066 | .101 | .026 | .069 | .110 | .110 |
| $j_{5g}$ | .019 | .045 | .014 | .034 | .084 | .084 |
| $j_{6g}$ | .004 | .004 | .003 | .004 | .045 | .045 |

[a] Exclusion test condition. [b] Inclusion test condition. [c] *Sure new* judgments given no conscious recollection are excluded by the model.

presented for one second (weak items) or for three seconds (strong items) (Yonelinas, 1994, Experiment 3).

The appropriate restrictions on the $j_{lc}$ parameters of the model ($l = 1, \ldots, 6$) yield 6 degrees of freedom for a test of this compound hypothesis. The log-

(Experiment 2), and $G^2(6) = 4.42$ (Experiment 3). Given a critical value of $\chi^2_{(df = 6, \alpha = .001)} = 22.46$, there is obviously no reason to reject the hypothesis.[6]

One might also expect that confidence judgments for nonrecollected but familiar items do not depend on the experimental treatments. Our hypothesis is thus that the $j_{lu}$ parameters ($l = 2, \ldots, 6$) are equal across the experimental conditions. The goodness-of-fit test for this hypothesis has $df = 4$, and a reasonable critical value is $\chi^2_{(df = 4, \alpha = .001)} = 18.47$. Although the fit is slightly worse compared to the above results, this hypothesis is also tenable for Experiment 1, $G^2(4) = 5.99$, Experiment 2, $G^2(4) = 11.51$, and Experiment 3, $G^2(4) = 13.67$.

In contrast to confidence judgments for recollected or familiar items, confidence judgments based on guessing should largely depend on the experimental context. For example, given that participants cannot recollect items and do not find them familiar, then they should show more confidence in their *new* judgments if the items were presented on short lists rather than on long lists in the acquisition phase. In case of short lists, participants might reason that it must be easy to remember items from the list. Hence, if they cannot recollect an item and the item also does not seem familiar, participants might guess that the item must very likely be *new* (Strack & Bless, 1994) . In case of long lists, in contrast, participants might expect to forget a significant proportion of items from the list. As a result, when they cannot recollect an item and it also does not seem familiar, they might more often guess that the item is nevertheless old. There is strong evidence favoring this hypothesis in the data of Yonelinas' Experiments 1 and 2. The $G^2$ statistics testing the equality of the $j_{lg}$ parameters ($l = 1, \ldots, 6$) across the experimental conditions are $G^2(5) = 98.73$ and $G^2(5) = 115.62$ for Experiments 1 and 2, respectively. Both $G^2$-statistics are considerably larger than the critical value $\chi^2_{(df = 5, \alpha = .001)} = 20.52$. Table 1 shows that these significant effects are indeed due to higher confidence in guessed *new* judgments in case of short lists. The present hypothesis cannot be evaluated for

---

[6] Given sample sizes (i.e., participants · sessions · items) ranging from $N = 6912$ to $N = 8640$ per experiment and $df < 10$ for the set of hypothesis tests to be reported in this section, the power to

Experiment 3 because the same set of distractors was used for both weak (1 s presentation duration) and strong (3 s presentation duration) items.

An interesting feature of our results is that they confirm Yonelinas' (1994) conclusions for Experiments 1 and 2, but not for Experiment 3. Like Yonelinas (1994), we found conscious recollections (as measured by $c$) to be more frequent for short lists than for long lists in Experiment 1, $G^2(1) = 56.98$, and in Experiment 2, $G^2(1) = 78.79$. Also, $c$ was found to be larger for strong than for weak items (3 s vs. 1 s presentation duration) in Experiment 3, $G^2(1) = 14.98$. For all three experiments, the $G^2$-statistics for the tests that $c$ is identical in the two encoding conditions exceed the critical value $\chi^2_{(df = 1, \alpha = .001)} = 10.83$. Familiarity effects as reflected in $u_{c-}$ were found neither for the list length manipulation in Experiment 1, $G^2(1) = 0.04$, and in Experiment 2, $G^2(1) = 0.08$, nor for the long versus short presentation duration manipulation in Experiment 3, $G^2(1) = 0.49$. This latter conclusion contradicts the one reached by Yonelinas (1994). According to our results, there is either no effect or a very tiny effect of the presentation duration manipulation on $u_{c-}$ (see Table 2). Currently, we are unable to decide whether $u_{c-}$ really is unaffected (or almost unaffected) by item presentation duration or whether the difference between 1 and 3 seconds was too small to reveal the presentation time effect in this study.

To summarize, it is quite easy for an extended EMM to account for the data of Yonelinas (1994) in a psychologically meaningful and reasonable way. The least we can say is that the fit of this model is not worse than that of the DPSDM. What this demonstrates is that the structural part of the EMM is not limited to *yes-no* recognition tasks. The model can be extended successfully to experimental situations in which confidence ratings are required following the recognition judgments in the process dissociation procedure. In fact, the basic underlying conceptual ideas can also be (and have already been) successfully transferred to other paradigms such as, for instance, lexical decision (cf. Vaterrodt-Plünnecke, 1994, who presented a two-high threshold model separating effects of conscious perceptions, implicit memory, and response bias) and Baars' (1992) slip technique to induce errors in speech (cf. Bröder & Bredenkamp, 1995, who extended the EMM to the slip paradigm) .

***Experimental Manipulations of Response Bias***

Confidence ratings provide a simple and economical way to arrive at empirical ROC curves. However, as shown in the last section, it is quite difficult to decide empirically between process dissociation measurement models on the grounds of rating-based ROC curves only. In fact, various types of models based on completely different theoretical rationales can account quite well for the ROC data of Yonelinas (1994), not only those based on two-factor theories of memory but also global memory models based on single process theories such as Gillund and Shiffrin's (1984) SAM model. This has recently been demonstrated by Ratcliff, Van Zandt, and McKoon (in press) .

Past research on measurement models for simple detection and recognition tasks has shown that it is almost impossible to test models based on threshold theory against signal-detection models if one uses ROC curves only. Some threshold models, for instance those suggested by Luce (1963b) or Krantz (1969) , are so flexible that they can mimic almost perfectly the ROC curves predicted by various versions of signal-detection theory. Other threshold models such as the plain one-high or the two-high threshold model (cf. Macmillan & Creelman, 1991, chap. 4; Snodgrass & Corwin, 1988) do indeed predict simple linear ROCs for *yes-no* detection or recognition tasks, but they can of course account for nonlinear ROCs based on confidence ratings when extended in a way analogous to that illustrated in Figure 1 for the EMM.

Our view converges with that of other authors. Lockhart and Murdock (1970) , for instance, concluded about attempts to differentiate between finite state and signal-detection models of memory that a "routine examination of operating characteristics is, in general, inconclusive, *especially if generated by the use of confidence ratings*" (p. 105, emphasis added; see Banks, 1970, for a similar conclusion) .

For these reasons, we cannot go along with Yonelinas and Jacoby (1995b) and Yonelinas et al. (in press) when they claim that ROC curves derived from confidence ratings are powerful tools for assessing model validity. A more thorough test of process dissociation measurement models can be arrived at by using *yes-no* affect response bias. The list of examples includes (a) manipulations of the proportion of targets relative to distractors (e.g., Dusoir, 1983; Kintsch, 1967; Marken & Sandusky, 1974; Parducci & Sandusky, 1965; Parks, 1966; Ratcliff, Sheu & Gronlund, 1992; Tanner, Haller & Atkinson, 1967) , (b) manipulations of payoffs for hits and false alarms (e.g., Banks, 1969; Galanter & Holman, 1967; Hume, 1974; Levine, 1966; Smith, 1969, 1970; Snodgrass & Corwin, 1988; Swets, Tanner & Birdsall, 1961; Wender, 1975), or (c) instructional manipulations (e.g., Colquhoun, 1967; Egan, Greenberg & Schulman, 1961). These experimental tests are more powerful than tests based on confidence ratings because the former refer to the original versions of measurement models while the latter refer to extended versions which are based on additional assumptions. Obviously, fits or misfits of these extended versions may heavily depend on the nature of the *additional assumptions*. Fit or misfit of the original models, however, only depends on the validity of the *core assumptions* of the models.

To our knowledge, experimental manipulations of response biases in the process dissociation procedure have only been investigated by Buchner et al. (1995) which is why we refer to these data only. Buchner et al. (1995) influenced response biases by manipulating (a) the proportion of targets relative to distractors in the recognition test (Experiment 1), (b) payoffs for hits and false alarms (Experiment 2), and (c) instructions that did versus did not point to the base-rates of required *old* versus *new* responses (Experiment 3). All three methods can be considered standard methods in response bias research, the first two being used more often than the third.

*Evaluation of the DPSDM.*

When the DPSDM is applied to *yes-no* recognition paradigms, the predicted pair of ROC curves can be inferred directly from the Model Equations 8 to 11. In fact, this pair of curves is identical to that specified in Equations 16 and 17, except that the cumulative rating probabilities have to be replaced by probabilities of *old* responses. Note, however, that deviations from these curves can no longer be attributed to intermediate confidence ratings for recollected items. This is why DPSDM evaluations based on *yes-no* recognition tasks are more thorough than tests based on

Unfortunately, due to the lack of techniques to compute confidence intervals and goodness-of-fit tests, no formal statistical evaluation of the DPSDM's fit to the Buchner et al. (1995) data can be performed at this point. Even ANOVA analyses based on single-participant estimates of the model parameters are impossible because inclusion versus exclusion test conditions were manipulated between subjects for the reasons specified before. The only thing that can be done is to compute parameter estimates for the aggregated data of single experimental conditions. However, nothing is known about the standard errors of these estimates. This is also true for the estimates averaged across experiments and item classes which are reported in Table 1 of Yonelinas and Jacoby (1995b) . Therefore, it is impossible to assess, for example, whether a difference of.07 in average recollection estimates between bias conditions is or is not a reason to reject the model. It is important to note that the same difference of.07 may be significant when referring to one model and nonsignificant when referring to another model. Whether or not significance is reached depends on the model-specific standard errors of the estimates which are unknown in the present case. Thus, the data presented in Table 1 of Yonelinas and Jacoby (1995b) are of no help in evaluating the model.

However, the most serious problem in Table 1 of Yonelinas and Jacoby (1995b) is that they do not refer to estimates of $d'$ as measures of familiarity. Instead, they report the average estimate of the probability of accepting a nonrecollected Phase 1 item on the basis of familiarity, *given an average false-alarm rate* (.18 in case of the Buchner et al. data). Obviously, this is not a meaningful indicator of familiarity effects. According to the DPSDM, the probability to accept nonrecollected Phase 1 items on the basis of familiarity must increase with the false-alarm rate in a way predicted by ROC curves corresponding to standard (i.e., equal variance) signal-detection theory. Hence, a difference between bias conditions with respect to familiarity ($d'$) corresponds to a difference between two of these ROC curves. Almost irrespective of the size of $d'$, all these ROC curves are close to each other near the points (0, 0) and (1, 1) in the ROC diagram. They differ noticeably only in the midrange of the diagram. Thus, differences between the DPSDM familiarity average false-alarm rate happens to be either small—as is the case in the Buchner et al. (1995) data—or large. Therefore, even if there were large response bias effects in *d'* it would be difficult to see them in the familiarity indicator used by Yonelinas and Jacoby (1995b) .

For these reasons, little can be inferred from Yonelinas and Jacoby's (1995b) Table 1 except that the hypothesis of no differences between DPSDM memory parameters across bias conditions does not fit the Buchner et al. (1995) data perfectly. Yonelinas and Jacoby's (1995b) try to explain this less-than-perfect fit by speculating that the experimental manipulations used by Buchner et al. (1995) might have affected response bias and memory processes simultaneously. However, they give no reason for this supposition with respect to Experiments 2 and 3, but only for the target-distractor ratio manipulation used in Experiment 1. Yonelinas and Jacoby (1995b) suspect that participants' motivation to engage in recollection might increase with the proportion of required *old* responses in the inclusion and exclusion recognition tests. However, even if this hypothesis should turn out to be correct—which is rather doubtful from our point of view—it remains to be shown that recollection motivation actually influences recollection performance in memory experiments. For the time being, the empirical evidence tends to favor the opposite hypothesis (e.g., Nilsson, 1987; O'Dekirk, Wyatt, & Ellis, 1993) .

*Evaluation of the EMM.*

As in case of the DPSDM, the ROC curves implied by the EMM when applied to *yes-no* recognition tasks can be inferred directly from the Model Equations 1 to 4. These happen to be parallel, linear ROCs with slope $(1-c) \cdot (1-u_{C-})$. Their intercepts are $c + (1-c) \cdot u_{C-}$ for the inclusion ROC and $(1-c) \cdot u_{C-}$ for the exclusion ROC. Note that although these simple linear functions have already been shown to be unable to account for the rating-based ROC curves observed by Yonelinas (1994) , they may nevertheless provide acceptable approximations to ROC curves based on yes-no recognition tasks. This is so because the failure of the EMM's ROCs when applied to rating scales may be due to the failure of the additional (in fact, unreasonable) distribution assumptions which are necessary to derive them within a modified

A series of experimental evaluations of the EMM has already been reported by Buchner et al. (1995). These evaluations were performed by testing whether the memory parameters $c$ and $u_{c\text{-}}$ really remain stable when response bias varies. Although no reference to ROC curves was made by Buchner et al. (1995), these tests can be shown to evaluate the hypothesis that hit and false-alarm rates observed under different test and response bias conditions lie on a pair of parallel, linear ROC curves as described above.

We do not want to reiterate the Buchner et al. (1995) results here. However, because Yonelinas and Jacoby (1995b) seem to argue that the Buchner et al. (1995) reasons to favor the EMM over the IMM are largely an artifact of the significance levels used, some clarifying comments are in order. To account for the different sample sizes involved, Buchner et al. (1995) in fact preferred different significance levels when testing hypotheses concerning the IMM and EMM parameters. However, it should be clear from their Table 3 (Buchner et al., 1995, p. 153) that *irrespective* of the significance levels used, the hypothesis of no differences in memory parameters across bias conditions fitted better for the EMM parameters than for the IMM parameters in six out of six tests. The average difference in $G^2$ goodness-of-fit statistics is as large as 9.7. If the IMM significance level had also been used for the EMM, only one of the six statistical decisions would have changed. However, the conclusions would remain exactly the same: Obviously, the EMM is a significant improvement over the IMM. Nevertheless, the EMM is less than perfect, which is why the Buchner et al. (1995) paper was entitled "*Toward* unbiased measurement of conscious and unconscious memory processes."

Table 1 of Yonelinas and Jacoby (1995b) cannot reveal the EMM advantage over the IMM because the figures displayed were based on average estimates across experiments and across item types. As outlined in detail by Buchner et al. (1995), we did not expect both of the IMM parameters to be equally affected by the response bias manipulations in each of the three experiments. This holds true particularly for $c_{IMM}$. As shown in Equation 7, the bias in $c_{IMM}$ does not depend on the absolute magnitude of the false-alarm rates in the inclusion and exclusion test conditions but manipulation on $c_{IMM}$ can only be expected when the difference $g_i - g_e$ varies across the bias conditions. Experiment 3 of Buchner et al. (1995) was designed to produce this pattern of false-alarm rates by a simple instructional manipulation. Half of the participants were informed of the differences in base-rates of required *old*:*new* responses, and half were not. The results indeed showed a strong response bias effect on $c_{IMM}$. The difference in $\hat{c}_{IMM}$ between the standard and the base-rate instruction groups was.23 and.11 for words read and for those solved as anagrams, respectively, in Phase 1. With the EMM there was a sizable reduction in these group differences. The $\hat{c}_{IMM}$ differences were.15 and.07 for read and the anagram words, respectively.

To summarize, the EMM can be shown to be a significant improvement over the IMM. However, the ROCs corresponding to the EMM provide only an approximation to the experimental data which is not completely satisfactory.

## Improvement of Process Dissociation Measurement Models

According to the results reported in the last section, both the DPSDM and the EMM seem to be improvements over the IMM. However, when evaluated on the basis of experimental data which provide more thorough tests than confidence ratings, none of these models appears to be completely convincing. This points to limitations of both models. Thus, we find ourselves encouraged to re-examine the critical assumptions underlying the models and to look for alternative assumptions that might help to overcome the limitations. In the sequel, we will first consider possible generalizations of the DPSDM that avoid some of the difficulties associated with the version presented by Yonelinas (1994) . Next, we turn to generalizations of the EMM.

### *Generalizations of the Dual-Process Signal-Detection Model*

In our opinion, two of the assumptions referred to when deriving the DPSDM model equations are particularly critical: the normal distribution assumption and the independence assumption. We know of no a priori reasons which would justify the normal distribution assumption with respect to familiarity. For instance, it does not seem too difficult to construct distractor material—intentionally or incidentally—such that this assumption is violated.

Also, as already mentioned, we are not aware of any a priori reasons justifying the independence assumption, and the arguments put forward so far in favor of and against assuming independence between recollection and familiarity seem to have about equal weight (cf. Cowan & Stadler, in press; Curran & Hintzman, 1995; Jacoby & Begg, 1995; Jacoby et al., 1994; Jacoby et al., in press; Joordens & Merikle, 1993; Richardson-Klavehn et al., 1995; Russo & Andrade, 1995; Toth et al., 1994) . The least one can say at the moment is that the issue is not yet settled. It appears thus desirable not to have to rely on a debatable assumption. We will first consider how to dispense with the normal distribution assumption and then turn to the independence assumption.

*The distribution-free DPSDM.*

When the normal familiarity distributions assumed by the DPSDM are replaced by arbitrary but identically shaped familiarity distributions, a distribution-free DPSDM results. The parameters of this model are illustrated in Figure 2. To avoid unnecessary complexity, Figure 2 refers to the inclusion test condition only. The picture for the exclusion test condition would be identical, except that the exclusion response criterion $k_e$ may differ from the inclusion response criterion $k_i$.

According to the distribution-free DPSDM, the proportion of nonrecollected Phase 1 items with familiarity values exceeding the response criterion $k_i$ can be written as the sum of $f_i \cdot (1 - g_i) + g_i$. The parameter $g_i$ again denotes the false-alarm rate in the inclusion test condition, that is, the proportion of distractor items for which familiarity exceeds the response criterion $k_i$. As can be seen in Figure 2, the product $f_i \cdot (1 - g_i)$ is the proportion of nonrecollected Phase 1 items exceeding the familiarity response criterion $k_i$ *in addition* to those that would exceed it, given that the familiarity increase due to prior processing was $d' = 0$. Thus, $f_i = 0$ if and only if $d' = 0$. For a fixed response criterion $k_i$ and given a fixed shape of the density functions, $f_i$ increases monotonically with $d'$, so that $f_i$ can be regarded as an ordinal measure of familiarity increase due to prior processing of Phase 1 items.

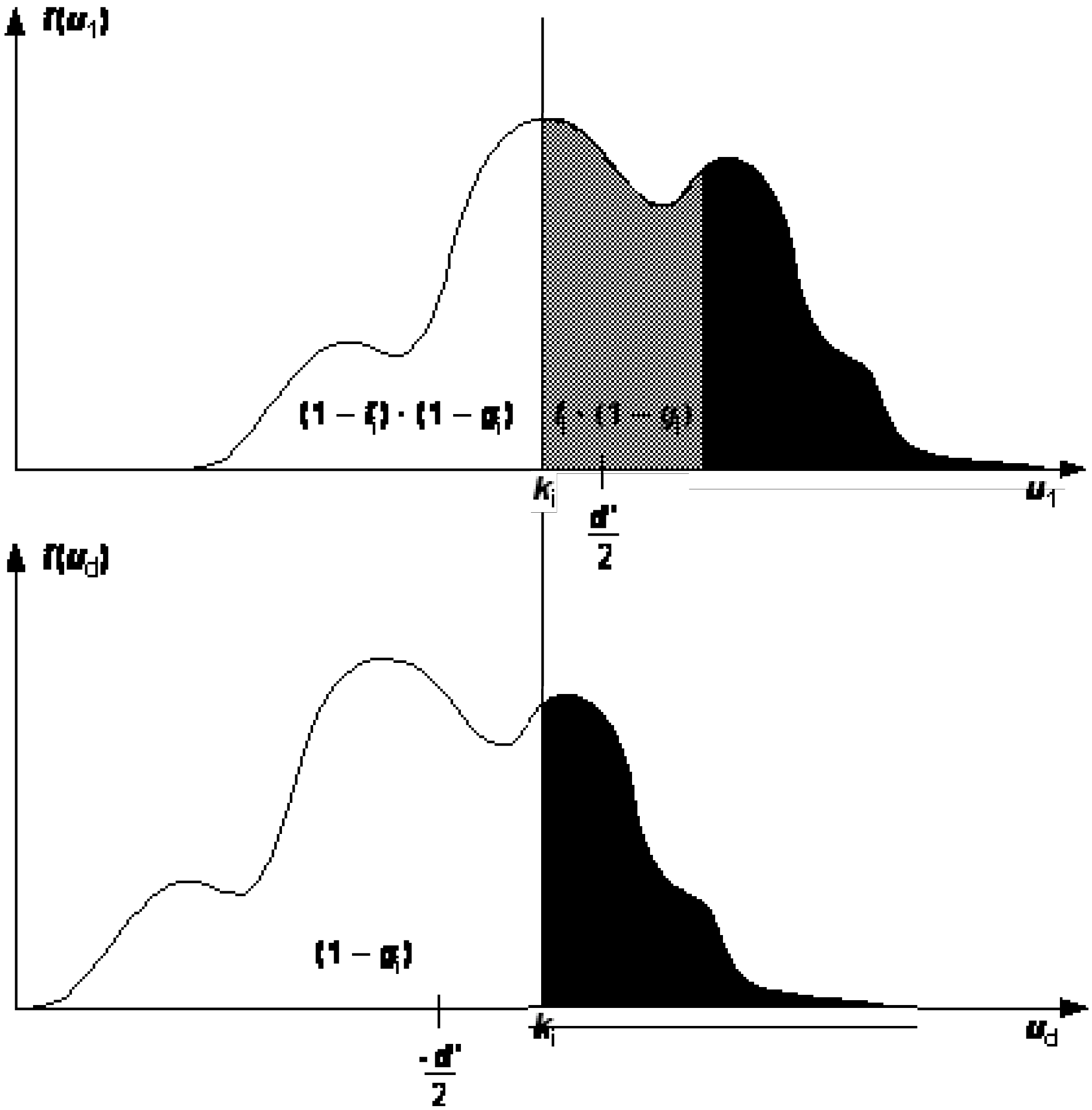



*Figure 2*. Illustration of the distribution-free dual-process signal-detection model. The familiarity distributions of the Phase 1 items (upper panel) and the distractors (lower panel) have identical but arbitrary shapes.

Unfortunately, $f_i$ does not only depend on $d'$ but also on the response criterion and on the nature of the density functions. For instance, if both $d'$ and the distribution shapes remain constant but the response criterion $k_i$ varies, then, in general, $f_i$ will also vary. Similarly, if the exclusion test condition is characterized by a familiarity response criterion $k_e$ which differs from $k_i$, then, in general, the exclusion familiarity parameter $f_e$ will also differ from $f_i$. Thus, there is no unique familiarity measure in the distribution-free DPSDM. The model must provide for as many familiarity parameters as there are different response criteria. This is the price to pay for dropping the distribution assumption: If a specific distribution assumption is rewritten as functions of a single parameter $d'$. If distribution assumptions are omitted, however, the familiarity parameters cannot be reduced to a single $d'$.

Although this fact renders the model somehow less attractive, important results can be inferred by applying it to the data of Yonelinas (1994). Two features of the distribution-free DPSDM are responsible for this:

(1) The DPSDM favored by Yonelinas and Jacoby is a proper submodel of the distribution-free DPSDM, so that the goodness-of-fit of the former can only be as good or worse than the goodness-of-fit of the latter.

(2) The distribution-free DPSDM is formally a GPT model, so that it is easy to analyze it statistically using the methods described by Hu and Batchelder (1994) .

By replacing the normal integrals $\Phi(\cdot)$ in Equations 8 to 11 with the corresponding integrals of the distribution-free DPSDM, and by reparameterizing these integrals in terms of $g_i$, $g_e$, $f_i$, and $f_e$ (as illustrated in Figure 2), we arrive at the following set of model equations:

$$p_{1i} = c + (1 - c) \cdot (g_i + (1 - g_i) \cdot f_i), \tag{18}$$

$$p_{di} = g_i, \tag{19}$$

$$p_{1e} = (1 - c) \cdot (g_e + (1 - g_e) \cdot f_e), \tag{20}$$

$$p_{de} = g_e. \tag{21}$$

Obviously, this is a nonidentifiable GPT model because four response probabilities cannot uniquely determine five model parameters. When inclusion and exclusion false-alarm rates do not differ, however, the model becomes identifiable, because $g_i = g_e = g$ implies $f_i = f_e = f$. For this reason, the model can be applied to the data of Yonelinas (1994) who used only one distractor category.

The distribution-free DPSDM must of course be extended before it can be applied to rating scale data. The extension is derived easily because we can proceed along the lines already discussed for the DPSDM. By doing so, one arrives at the Model Equations 22, 23, and 24 for the cumulated rating probabilities in the inclusion test condition, the exclusion test condition, and the distractor item condition, respectively:

$$p_{1\mathrm{i}}(X \geq j) = c + (1 - c) \cdot (g_{(\mathrm{j})} + (1 - g_{(\mathrm{j})}) \cdot f_{(\mathrm{j})}). \tag{22}$$

$$p_{1\mathrm{e}}(X \geq j) = (1 - c) \cdot (g_{(\mathrm{j})} + (1 - g_{(\mathrm{j})}) \cdot f_{(\mathrm{j})}). \tag{23}$$

$$p_{\mathrm{d}}(X \geq j) = g_{(\mathrm{j})}. \tag{24}$$

Interestingly, these model equations turn out to be equivalent to a union of *n*-1 restricted EMMs (i.e., one for each rating category $j \in \{2, \ldots, n\}$), if one assumes that the recollection probability $c_{(\mathrm{j})}$ is a constant parameter *c* for all *j* while the false-alarm rates $g_{(\mathrm{j})}$ and the familiarity parameters $f_{(\mathrm{j})}$ may differ between rating categories.

We fitted this model to the cumulated raw frequencies of each of Yonelinas' (1994) experimental conditions. The likelihood-ratio goodness-of-fit test has $df = 4$ and the sample sizes vary between $N = 3456$ and $N = 4840$. Therefore, a reasonable significance level is again $\alpha = .001$. Using the corresponding critical value $\chi^2_{(df = 4, \alpha = .001)} = 18.47$ for the model test we have a power of about .97 for effects of size $w = .10$.

Somewhat to our surprise, the distribution-free DPSDM clearly did not fit the data of Yonelinas' Experiment 1 (short lists: $G^2(4) = 77.77$; long lists: $G^2(4) = 43.12$) and Experiment 2 (short lists: $G^2(4) = 92.15$; long lists: $G^2(4) = 61.40$). For Experiment 3, in contrast, the fit was quite good (weak items: $G^2(4) = 0.80$; strong items: $G^2(4) = 7.35$). Taken together, these results are rather disappointing for both the DPSDM and it's distribution-free generalization, because the misfit of the latter implies the misfit of the former.

In fact, Yonelinas (1994) detected the source of the misfit when he noticed that his recollection estimates (which, as a consequence of $g_{\mathrm{i(j)}} = g_{\mathrm{e(j)}} = g_{(\mathrm{j})}$, happen to be $\hat{c}$, estimates) turned out not to be constant as predicted but rather decreased at the extremes of the false-alarm dimension. Yonelinas (1994) attributed these decreases to "bottom effects" and to "ceiling effects." However, there are no statistical reasons to estimates of the recollection parameters are unbiased estimates of the underlying response probabilities in the entire [0, 1] interval. Moreover, a marked curvilinearity in the recollection estimates can be observed across the full range of false-alarm rates, not only at the extremes (cf. Yonelinas, 1994, Figures 5 and 7). Therefore, attributing this curvilinearity to bottom and ceiling effects is not completely convincing.

We argue that the only explanation of the misfit which leaves the core assumptions untouched is that recollected items did not always receive the most extreme confidence ratings. This explanation is reasonable because biases against extreme ratings are a frequent phenomenon, and that would explain the observed curvilinearity in the $\hat{c}_{(j)}$ estimates across the false-alarm dimension. Note also that the curvilinearity is reduced, but not completely eliminated, in the recollection estimates for Experiment 3 (cf. Yonelinas, 1994, Figure 9). A reduced curvilinearity must be expected when (a) intermediate confidence ratings for recollected items do indeed occur, but (b) the recollection probability *c* is rather small. If there are few recollected items—as seems to be the case in Yonelinas' (1994) Experiment 3—, then obviously the ratings preferred for recollected items cannot affect the data structure significantly.

Of course, it is also conceivable that the misfit is not due to intermediate confidence ratings for recollected items. Alternatively, it might be due to violations of one of the core assumptions underlying the distribution-free DPSDM. How are we to decide between these two possibilities? Again, using *yes-no* recognition tasks in combination with experimental manipulations of response bias may be helpful. If the misfit of the rating scale model really is a consequence of intermediate confidence ratings for recollected items, then there should be no such misfit when the original distribution-free DPSDM is applied to *yes-no* recognition tasks. In contrast, if the misfit is caused by violations of one of the core assumptions, then the original distribution-free DPSDM should not fit *yes-no* recognition data either.

Unfortunately, however, the distribution-free model cannot be evaluated statistically by referring to the data of Buchner et al. (1995). This is implied by the fact identifiable. Also, identifiability cannot be achieved by adding the restrictions that (a) the familiarity parameters of anagram and read items do not differ and (b) the recollection parameters do not differ between the bias conditions.

For the time being, therefore, nothing can be said about the performance of the distribution-free DPSDM when applied to *yes-no* recognition tasks. What is needed is a set of experimental data comparable to those published by Buchner et al. (1995), but satisfying the restriction that $g_i = g_e$ for each of the bias conditions.

*The correlated-processes signal-detection model.*

Another possibility of generalizing the DPSDM is to drop the independence assumption while leaving the normal distribution assumption unchanged. The model equations corresponding to this correlated-processes signal-detection model (CPSDM) can most easily be derived by first considering the *unconditional* familiarity distribution of Phase 1 items, that is, the familiarity distribution of recollected and nonrecollected Phase 1 items combined. This *must* be a normal distribution with mean $d'/2$ and standard deviation 1 when (a) the distractor familiarity distribution is a normal distribution with mean $-d'/2$ and standard deviation 1, and (b) processing of Phase 1 items increases their familiarity additively by amount $d'$ relative to distractor items.

According to the CPSDM, the proportion $c$ of recollected Phase 1 items can be decomposed into two additive components with respect to *any* response criterion $k$, namely a proportion $q(k)$ of recollected items with familiarity values exceeding $k$ and another proportion $1 - q(k)$ of recollected items with familiarity values not exceeding $k$. Therefore, there is a proportion $q(k) \cdot c$ of recollected items among Phase 1 items with familiarity values exceeding $k$ and another proportion $(1 - q(k)) \cdot c$ of recollected items among Phase 1 items with familiarity values not exceeding $k$.

In the inclusion test condition, a proportion $\Phi(d'/2 - k_i)$ of Phase 1 items has familiarity values larger than the response criterion $k_i$. All of these items will be judged *old*, irrespective of whether they were recollected or not. In addition to these items, there is a proportion $(1 - q(k_i)) \cdot c$ of recollected items with familiarity values not exceeding $k_i$ which will also be judged old. Therefore, In the exclusion test condition, in contrast, only those Phase 1 items will be judged *old* that (a) exceed the familiarity response criterion $k_e$ and (b) are *not* consciously recollected. Hence,

$$p_{1e} = \Phi(d'/2 - k_e) - q(k_e) \cdot c. \tag{26}$$

The inclusion and exclusion distractor equations are identical to those already derived for the DPSDM (cf. Equations 9 and 11).

The DPSDM as a proper submodel of the CPSDM is obtained by setting $q(k_i) = \Phi(d'/2 - k_i)$ and $q(k_e) = \Phi(d'/2 - k_e)$. If these two restrictions hold true, then recollection and familiarity are said to be *uncorrelated* with respect to $k_i$ and $k_e$. If, in contrast, the relations $q(k_i) > \Phi(d'/2 - k_i)$ and $q(k_e) > \Phi(d'/2 - k_e)$ turn out to be correct, then recollection and familiarity are said to be *positively correlated* relative to $k_i$ and $k_e$, respectively. Finally, *negatively correlated* recollection and familiarity processes correspond to the relations $q(k_i) < \Phi(d'/2 - k_i)$ and $q(k_e) < \Phi(d'/2 - k_e)$. If desired, an *index of correlation*,

$$R_k := \ln(q(k) \,/\, \Phi(d'/2 - k)), \tag{27}$$

may be defined with respect to any response criterion $k$. Positive, zero, and negative values of $R_k$ correspond to positive, zero, and negative correlations between recollection and familiarity, respectively.

As a consequence of the additional parameters $q(k_i)$ and $q(k_e)$, the CPSDM is of course nonidentifiable in its most general form. However, identifiable submodels exist, especially when applied to several groups or experimental manipulations simultaneously. For example, submodels that correspond to the redundancy and exclusivity model variants as discussed by Buchner et al. (1995) can be defined. The redundancy variant, for instance, posits a perfect positive correlation between recollection and familiarity so that the recollected items are those with the largest familiarity values. Thus,

$$q(k_i) = \begin{cases} 1, & \text{if } \Phi(d'/2 - k_i) \geq c, \\ \Phi(d'/2 - k_i)\,/\,c, & \text{if } \Phi(d'/2 - k_i) < c \end{cases} \tag{28}$$

and

Inserting these restrictions into Equations 25 and 26 shows that the inclusion and exclusion ROC curves implied are flat across some part of the false-alarm dimension. For hit rates $p_{1i} \geq c$, the *inclusion* ROC curve corresponds to standard (equal-variance) signal-detection theory. However, because $p_{1i}$ cannot drop below $c$, a constant value $p_{1i} = c$ is implied for the remainder of the inclusion ROC curve. The corresponding *exclusion* ROC curve must be parallel to the inclusion ROC. It is obtained by subtracting $c$ from each point on the inclusion curve.

An exclusivity variant, in contrast, is based on assuming a perfect negative correlation between recollection and familiarity so that the recollected items are those with the lowest familiarity values. Therefore,

$$q(k_i) = \begin{cases} 0, & \text{if } \Phi(d'/2 - k_i) \leq 1 - c, \\ (\Phi(d'/2 - k_i) - (1 - c))/c, & \text{if } \Phi(d'/2 - k_i) > 1 - c \end{cases} \tag{30}$$

and also

$$q(k_e) = \begin{cases} 0, & \text{if } \Phi(d'/2 - k_e) \leq 1 - c, \\ (\Phi(d'/2 - k_e) - (1 - c))/c, & \text{if } \Phi(d'/2 - k_e) > 1 - c. \end{cases} \tag{31}$$

By inserting these terms into Equations 25 and 26 we obtain partially flat inclusion and exclusion ROC curves, too. In this case, however, the *exclusion* ROC curve follows standard signal-detection theory for hit rates of $p_{1e} \leq 1 - c$. Again, because the exclusion ROC cannot increase beyond $p_{ie} = 1 - c$, the remainder of this curve corresponds to the constant value $p_{1e} = 1 - c$. The parallel *inclusion* ROC is obtained by adding $c$ to each point on the exclusion ROC.

However, these redundancy and exclusivity variants of the CPSDM seem to perform quite badly compared to the standard DPSDM. According to the results of Yonelinas (1994), no part of the inclusion or exclusion ROCs corresponds to flat lines. Instead, the ROCs appear to increase strictly monotonically across the entire false-alarm dimension. These qualitative results seem to be typical of recognition ROC curves in general. Therefore, no statistical evaluation is necessary in order to reject both the redundancy variant and the exclusivity variant of the CPSDM.

Note, however, that the rejection of these models does not imply anything about the empirical adequacy of the DPSDM. In particular, independence of recolmany other submodels in addition to the DPSDM, the redundancy variant, and the exclusivity variant. Currently, we cannot decide whether any of these submodels is an improvement over the standard DPSDM favored by Yonelinas et al. (in press) and Yonelinas and Jacoby (1995b) . Methods to analyze statistically any submodels of the CPSDM model are much to seek.

### *Generalizations of the Extended Measurement Model*

As in case of the DPSDM, there are several possible and plausible ways of generalizing the EMM. One could think of more than one familiarity state, for example, each corresponding to a different degree of familiarity. Even more dramatic modifications of the basic assumptions are conceivable. In this section, however, we will try to retain all of the basic assumptions underlying the EMM and nevertheless improve its goodness-of-fit by generalizing the model, with a focus on the processes involved in responding to distractor items.

One seemingly minor, but in fact relatively important assumption implied by the EMM model equations is that distractor items are never detected as *new*. Following this assumption, the false-alarm rates can be equated with the probabilities of guessing *old* in the state of recognition uncertainty. From a statistical viewpoint, this renders the EMM quite simple. However, as has already been stated by Buchner et al. (1995, p. 143), conducting appropriate validation experiments can be tricky because the distractor material must be selected and presented such that the probability of participants detecting distractors becomes a negligible quantity. Salient distractors, for instance, must be excluded from the distractor list because participants might reason that they would have recognized a particularly salient item as old had it been presented earlier, and infer that the item must certainly be new (cf. Strack & Bless, 1994).

Although Buchner et al. (1995) have tried to avoid salient items, some of their distractor items could nevertheless have been salient enough to encourage judgment strategies similar to the one just described. The EMM would have been misspecified for the validation experiments reported by Buchner et al. (1995) to the degree to which distractor detection had actually occurred.

It should be noted that prior research on threshold models for standard *yes-no* recognition tasks and source monitoring tasks has pointed to similar problems for other models which were also based on the presupposition that distractor items are always responded to by guessing. The one-high threshold model (1HTM), for example, was clearly found to be inappropriate for simple *yes-no* recognition tasks (see e.g., Kintsch, 1977; Macmillan & Creelman, 1991; Murdock, 1974; Snodgrass & Corwin, 1988) . However, these problems can be solved or at least reduced by adding an additional parameter to the 1HTM that represents the probability of detecting distractors as new. This model, which is known as the two-high (Snodgrass & Corwin, 1988) or double-high (Macmillan & Creelman, 1990, 1991) threshold model (2HTM), provides a reasonably good linear approximation to ROC curves based on *yes-no* recognition tasks. Moreover, Snodgrass and Corwin (1988) and Macmillan and Creelman (1990) showed that sensitivity and bias measures based on the 2HTM compare favorably with many alternative measures based on various other model frameworks.

Batchelder and Riefer (1990) developed a rather complex GPT model for source monitoring tasks (see Johnson, Hashtroudi & Lindsay, 1993, for a recent review) . This one-high threshold source monitoring model (1HTSM) reduces to the 1HTM if there exists only one source of items. For this reason, Kinchla (1994) has criticized Batchelder and Riefer's model. Although the inadequacy of the 1HTM for simple *yes-no* recognition tasks does not necessarily imply the inadequacy of the 1HTSM for source monitoring tasks, Kinchla's critique stimulated Batchelder, Hu, and Riefer (1994) , Batchelder, Riefer and Hu (1994) , and Bayen, Murnane and Erdfelder (in press) do develop alternative GPT source monitoring models that were intended to cope with possible problems of the 1HTSM when evaluated against empirical data. In fact, Bayen et al. (in press)
have recently shown empirically that a two-high threshold extension of the 1HTSM which reduces to the 2HTM if there exists only one item source outperforms both the 1HTSM and a low-threshold source monitoring model variant when evaluated by using experimental manipulations of item detection and source identification.

detection manipulation in their item memory parameters, the same effect was captured properly by the appropriate parameter of the two-high threshold source monitoring model (2HTSM).

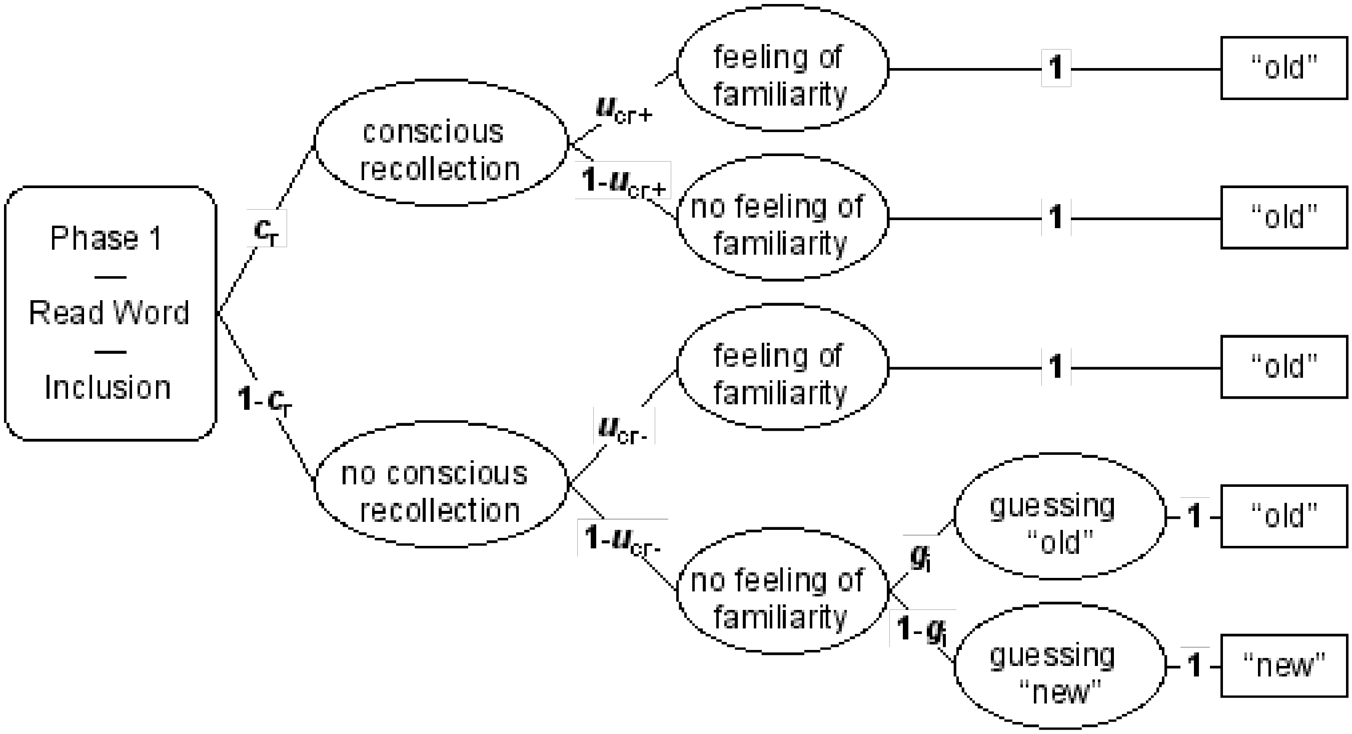
The EMM suggested by Buchner et al. (1995) is related to Batchelder and Riefer's 1HTSM because it also reduces to the 1HTM when only nonrecollected Phase 1 items are considered as target items. According to the results of Bayen et al. (in press) , the 2HTSM appears to be an improvement over the 1HTSM. Therefore, one might suspect that a two-high threshold generalization of the EMM is also an improvement over the EMM.

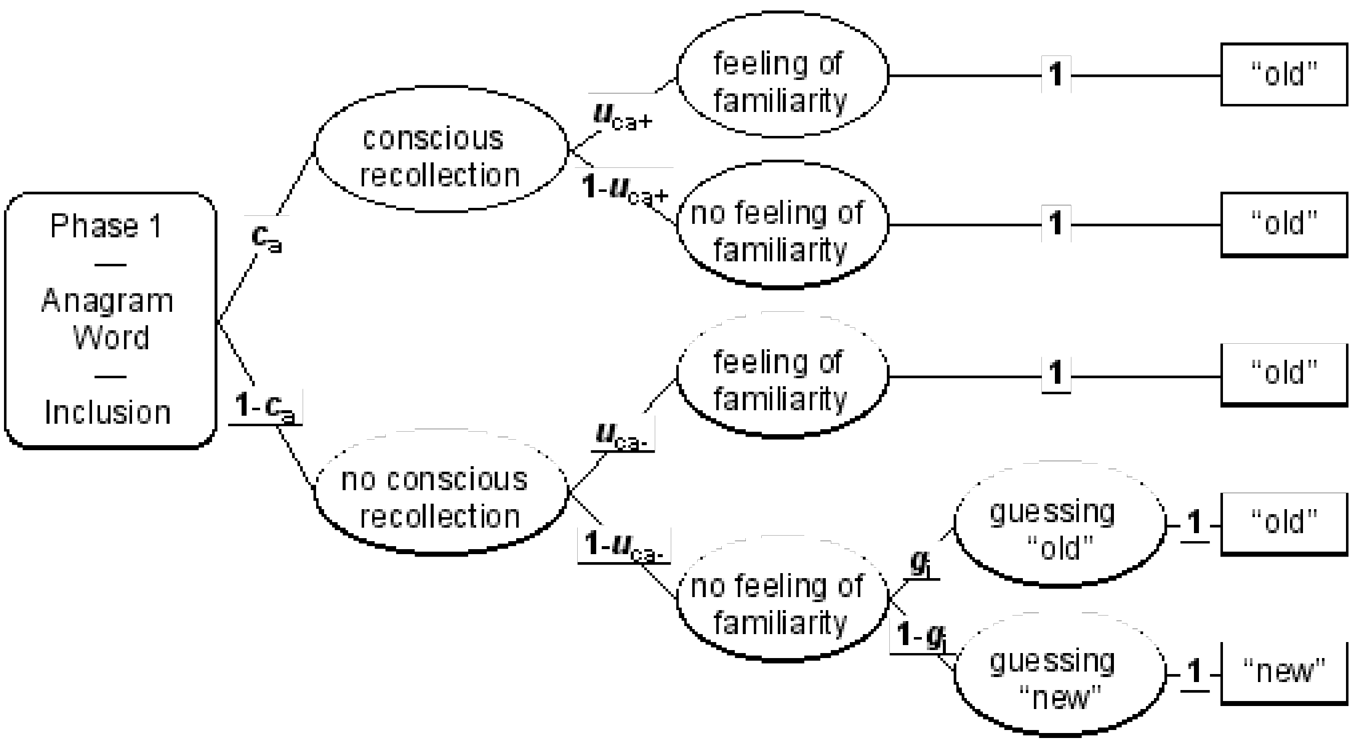
The two-high threshold extended measurement model (2HT-EMM) we suggest is illustrated in Figure 3. There is only one important difference between this model and the EMM shown in Figure 2 of Buchner et al. (1995). The 2HT-EMM, but not the EMM, provides for the possibility of *consciously rejecting* distractor items with probability $d$ in both the inclusion and exclusion test conditions. Participants are assumed to guess only if distractors are not detected (which occurs with probability $1-d$). Note that the EMM as a proper submodel of the 2HT-EMM is obtained by setting $d = 0$.

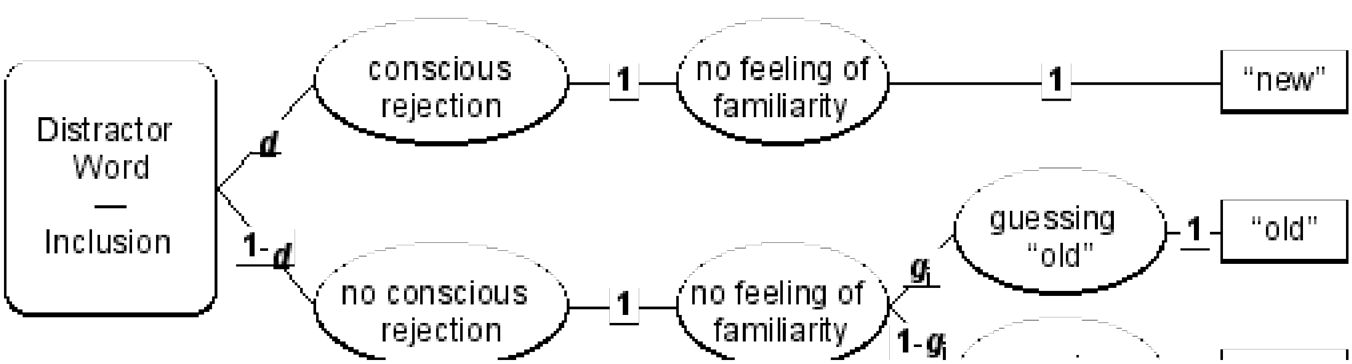
Another minor difference between the present Figure 3 and Figure 2 in Buchner et al. (1995) is that Figure 3 has already been tailored to the validation experiments conducted by Buchner et al. (1995). Therefore, two types of Phase 1 items—read items and anagram items—are distinguished. The recollection parameters corresponding to read and anagram items are labeled $c_r$ and $c_a$, the corresponding familiarity parameters $u_{cr-}$ ($u_{cr+}$) and $u_{ca-}$ ($u_{ca+}$), respectively. Following Buchner et al. (1995), the nonidentifiable parameters $u_{cr+}$ and $u_{ca+}$ were included into the model so that independence, redundancy, and exclusivity variants of the 2HT-EMM—among others—can be defined. However, in the analyses reported below these nonidentifiable parameters were dropped.

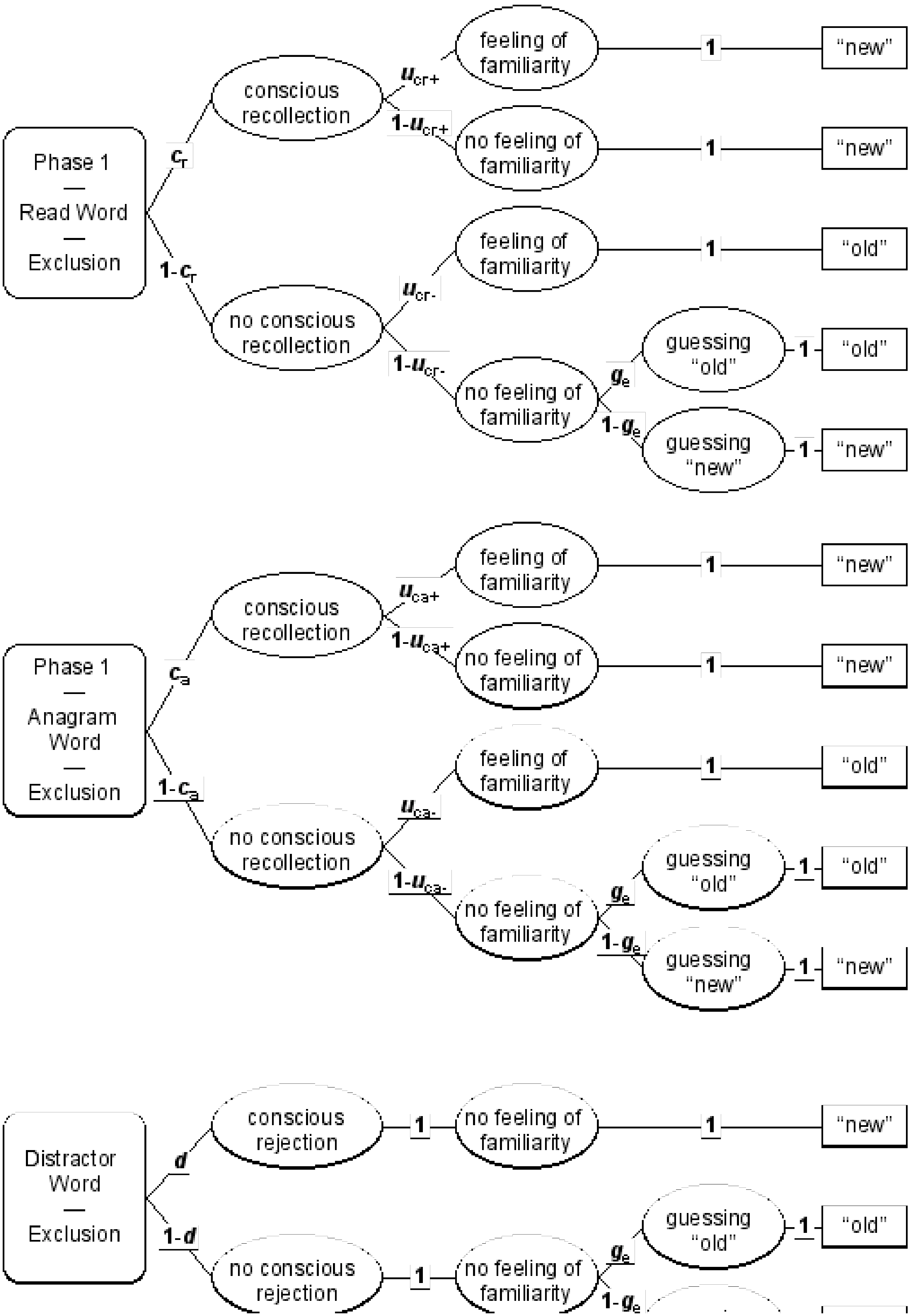
Despite the elimination of $u_{cr+}$ and $u_{ca+}$, the 2HT-EMM is not identifiable without imposing further restrictions. This is a consequence of the additional parameter $d$. We will consider two possible ways of achieving identifiability, and we validation data of Buchner et al. (1995). Although it is easy to extend the 2HT-EMM to rating scale data, too, we will not consider this possibility here for the reasons already outlined above.

*Figure 3* (See the ollowing 2 pages.) The two-high threshold extended measurement model (inclusion condition). Parameters $c_r$ and $c_a$ denote the unconditional probabilities of controlled recollections of read and anagram words, respectively. Parameters $u_{c+}$ and $u_{c-}$ denote the conditional probabilities of automatic, familiarity-based memory effects if a read or anagram word is and is not recollected, respectively. Parameter $d$ represents the unconditional probability of controlled rejections of distractors. Parameters $g_i$ and $g_e$ represent the probabilities (a) of guessing that a distractor is old if it was not consciously rejected, and (b) of guessing that a read or anagram word is old given that it has been neither


Phase 1
—
Read Word
—
Inclusion
cr
1-cr
conscious recollection
no conscious recollection
ucr+
1-ucr+
ucr-
1-ucr-
feeling of familiarity
no feeling of familiarity
feeling of familiarity
no feeling of familiarity
gi
1-gi
guessing "old"
guessing "new"
1
"old"
"old"
"old"
"old"
"new"


Phase 1
—
Anagram Word
—
Inclusion
ca
1-ca
conscious recollection
no conscious recollection
uca+
1-uca+
uca-
1-uca-
feeling of familiarity
no feeling of familiarity
feeling of familiarity
no feeling of familiarity
gi
1-gi
guessing "old"
guessing "new"
1
"old"
"old"
"old"
"old"
"new"


Distractor Word
—
Inclusion
d
1-d
conscious rejection
no conscious rejection
1
no feeling of familiarity
no feeling of familiarity
gi
1-gi
guessing "old"
"new"
"old"


Phase 1 — Read Word — Exclusion
$c_r$
$1-c_r$
conscious recollection
no conscious recollection
$u_{cr+}$
$1-u_{cr+}$
$u_{cr-}$
$1-u_{cr-}$
feeling of familiarity
no feeling of familiarity
$g_e$
$1-g_e$
guessing "old"
guessing "new"
1
"new"
"old"
Phase 1 — Anagram Word — Exclusion
$c_a$
$1-c_a$
$u_{ca+}$
$1-u_{ca+}$
$u_{ca-}$
$1-u_{ca-}$
Distractor Word — Exclusion
$d$
$1-d$
conscious rejection
no conscious rejection

*The multiple-groups two-high threshold EMM.*

The most simple way to arrive at an identifiable 2HT-EMM version is to apply the model to at least two experimental groups simultaneously and to assume that only the response bias parameters $g_i$ and $g_e$ but not the core parameters $c_r$, $c_a$, $u_{cr-}$, $u_{ca}$-, and $d$ differ between groups. This assumption is reasonable for Experiments 1 to 3 of Buchner et al. (1995) because different groups correspond to different levels of response bias in these experiments.

As depicted in Figure 3, we applied this multiple-groups 2HT-EMM to read and anagram items simultaneously. Therefore, the sample size was $N$ = 4800 in each of the three experiments (cf. Buchner et al., 1995). Given this large sample size, $df = 3$ , and $\alpha$ =.001, the power of the $G^2$ goodness-of-fit test is about.99 even for "small" deviations from the model (i.e., Cohen's $w$ =.1). Therefore, we decided again to use $\chi^2_{(df = 3, \alpha = .001)}$ = 16.27 as a critical value for our statistical decisions.

Fitting the multiple-groups 2HT-EMM with the restrictions that parameters $c_r$, $c_a$, $u_{cr-}$, $u_{ca}$-, and $d$ do not differ between groups resulted in the goodness-of-fit statistics $G^2(3) = 2.25$, $G^2(3) = 8.47$ and $G^2(3) = 2.99$ for Experiments 1, 2, and 3, respectively. Obviously, the fit is quite good in each of the three tests, and there is no reason to reject the model. In addition, the multiple-groups 2HT-EMM performs better than the EMM for which the corresponding goodness-of-fit statistics are $G^2(4) = 9.57$, $G^2(4) = 20.74$, and $G^2(4) = 10.01$ for Experiments 1, 2, and 3, respectively.[7] These $G^2$ statistics can be compared directly to the parallel 2HT-EMM statistics, because both the degrees of freedom and the sample sizes underlying the tests are identical. Although the fit of the EMM is acceptable for Experiments 1 and 3 at least, it is clearly worse than the fit of the 2HT-EMM in each of the three tests.

*The single-group two-high threshold EMM.*

Unfortunately, the multiple-groups 2HT-EMM cannot be applied whenever there is only a single experimental group of participants or whenever there are several groups *not* satisfying the restriction that group membership does not affect

---

[7] These goodness-of-fit statistics deviate from those published by Buchner et al. (1995)

the memory parameters of the model. In order to render the model identifiable, some a priori assumptions concerning the $d$ parameter are unavoidable in these cases.

The problem can be solved by imposing equality restrictions on $d$. For instance, one may assume that the probability of detecting a distractor item as new is equal to the probability of detecting a target item as old. In fact, both the 2HTM evaluated by Snodgrass and Corwin (1988) and the 2HTSM evaluated by Bayen et al. (in press) were based on this presupposition. However, if an analogous procedure is to be applied to the 2HT-EMM, and if there is more than one type of old target items, then one may run into conceptual problems. Whenever the memory parameters corresponding to different types of old target items differ, which of these parameters should be equated with the probability of detecting a distractor? This is exactly the problem we face in the Buchner et al. (1995) data. Throughout all three experiments, the probability of consciously recollecting a target was higher for anagram than for read words. That pattern of results is also mirrored in the parameter estimates obtained for the multiple-groups 2HT-EMM: The maximum-likelihood estimates for $c_r$ and $c_a$ were.11 and.62 (Experiment 1),.28 and.67 (Experiment 2) and.26 and.57 (Experiment 3), respectively.

A possible solution to this dilemma is to assume that $d$ is a *weighted average* of the different recollection parameters. If we have only two types of target items—read and anagram items—this means that $d = \square \cdot c_r + (1 - \square) \cdot c_a$, where $\square$ is a (fixed) weighting factor that depends on the proportions of different item types in the recognition test ($\square$ =.5 in case of the Buchner et al. experiments). This approach is relatively pragmatic. However, it has the advantages (a) to be applicable to any number of target item types without necessarily assuming homogeneity of the memory parameters and (b) to reduce to what others have suggested if there is only one class of target items or if several items types are available but do not differ in their memory parameters.

The maximum-likelihood parameter estimates obtained by applying this single-group 2HT-EMM to the data of Buchner et al. (1995) are illustrated in Figure 4 and the IMM, the corresponding parameter estimates and confidence intervals obtained for these models are displayed, too.

Clearly, the parameter estimates corresponding to different response bias conditions differ less in the 2HT-EMM than in case of the other two models. If recollection estimates are averaged across item types and experiments we obtain means of.438 for the liberal response bias conditions and.424 for the conservative conditions. Thus, if measurement bias would be assessed as in Yonelinas and Jacoby's (1995b) Table 1, the bias with respect to $\hat{c}$, would be found to be only.014. The same computations for $\hat{u}_{c-}$ would result in average estimates of.193 for the liberal response bias conditions and.207 for the conservative conditions across all three experiments. Again, the residual bias is as small as.014. Thus, even if we were to rely solely on these statistics, it would seem that the 2HT-EMM outperforms not only the IMM quite clearly but also the EMM and the DPSDM (see Table 1 in Yonelinas & Jacoby, 1995b) .

However, for the reasons already outlined above, these descriptive indices are less important than the results of more thorough statistical evaluations. The fact that the 95% confidence intervals in Figure 4 overlap in each case suggests that the response bias manipulation did not affect the 2HT-EMM parameters $c_r$, $c_a$, $u_{cr-}$, $u_{ca-}$, and $d$ significantly. The goodness-of-fit test for this hypothesis has $df$ = 4 instead of $df$ = 3 because parameter $d$ in the single-group 2HT-EMM is no longer a free parameter as in the multiple-groups 2HT-EMM. Thus, a reasonable critical value is $\chi^2_{(df = 4, \alpha = .001)}$ = 18.47. Testing the above hypothesis statistically results in goodness-of-fit indices $G^2$ of 2.91, 8.63, and 3.02 for Experiments 1, 2, and 3, respectively. Thus, the fit of the single-group 2HT-EMM is about as good as the fit of the multiple-groups 2HT-EMM for the Buchner et al. (1995) data at least. In the light of these results, both variants of the 2HT-EMM seem to be the best available process dissociation measurement models at the moment.

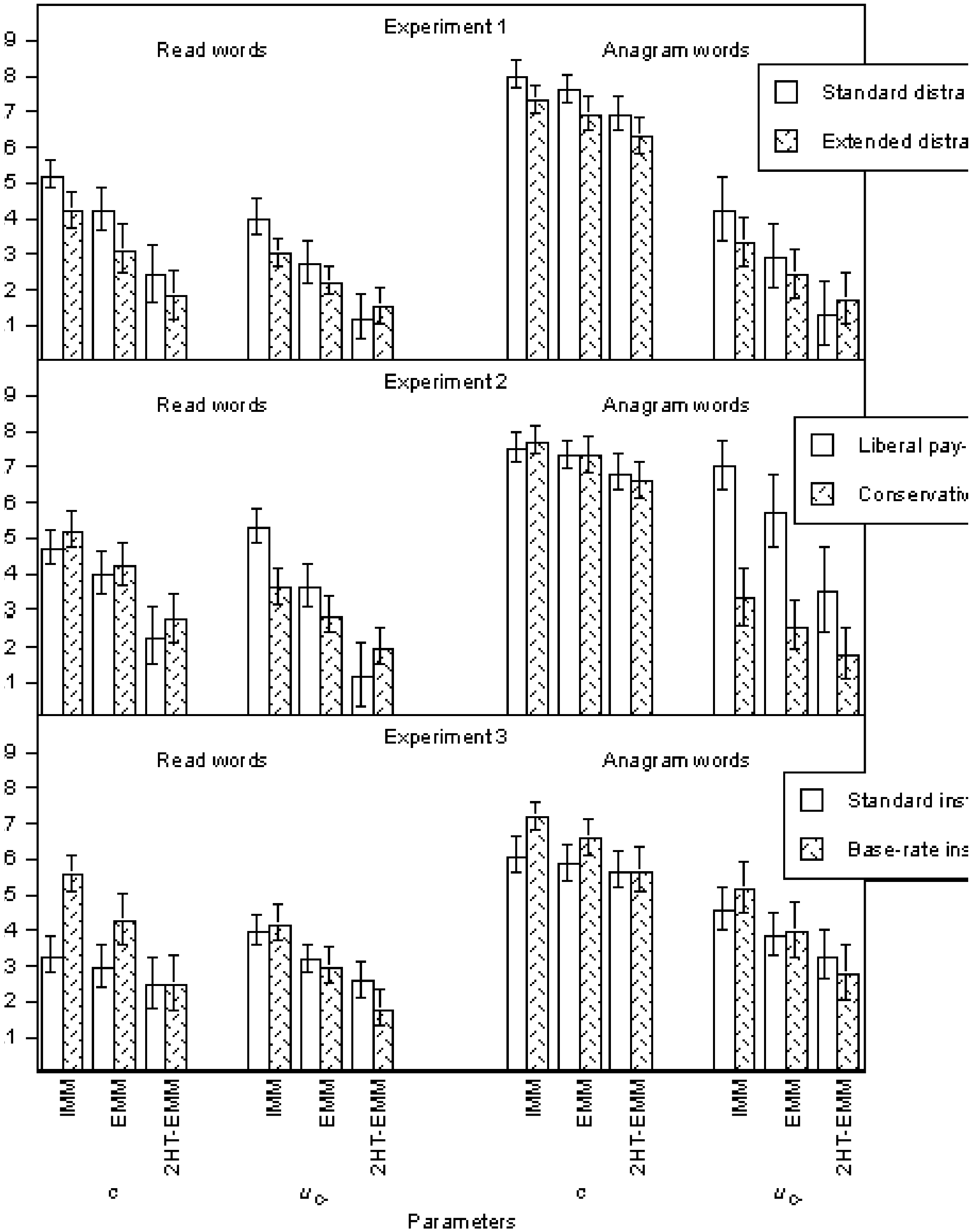



*Figure 4.* Estimates for the parameters $c$ and $u_{c\text{-}}$ according to the IMM, the EMM, and the single-group 2HT-EMM for the data of Experiments 1 to 3 of Buchner et al. (1995). The error bars represent the 95 % confidence intervals.

## Discussion

Buchner et al. (1995) presented the EMM as a measurement model that takes guessing into account for the process dissociation procedure. They used exstrated that it is superior to the IMM presented by Jacoby (1991). Yonelinas, Regehr, and Jacoby (in press) and Yonelinas and Jacoby (1995b) proposed, as an alternative, the DPSDM which differs from the EMM in that familiarity is conceived of as a continuous latent random variable rather than a discrete cognitive state. Yonelinas and Jacoby (1995b) went on to argue that this DPSDM provided a better fit than the EMM to the ROC data obtained from a confidence rating procedure by Yonelinas (1994) , and a slightly better fit than the EMM to the experimental data reported by Buchner et al. (1995). As we have shown, this analysis has two weaknesses. For one thing, Yonelinas and Jacoby (1995b) argue only on a descriptive basis and compare simple parameter estimates without considering confidence intervals for these estimates. Thus, the EMM may well "look" as if it could fit the Buchner et al. (1995) data less well than the DPSDM, but without a formal goodness-of-fit test for the models we remain in some sort of limbo and cannot reach a definite conclusion.

Aside from this, it is important to note that the EMM must be extended before it can be applied to confidence rating ROCs, and that this can be done in several possible ways. Thus, when an extended EMM fails to fit confidence rating data, we do not know whether the core of the model assumptions was inadequate or whether only the particular extension was inappropriate.

This latter ambiguity was resolved by our developing an extension of the EMM to confidence rating data which fits exactly the empirical ROC data reported by Yonelinas (1994, Experiments 1 to 3). Much more importantly, we demonstrated in a series of detailed goodness-of-fit tests for this extended EMM that the parameter estimates obtained were psychologically meaningful and reasonable when we applied it to the data from Yonelinas' (1994) Experiments 1 to 3. In other words, when extended appropriately, the EMM can account very well for ROCs generated from confidence rating data, which means that the model's core assumptions do not need to be rejected. This result refutes claims to the contrary by Yonelinas and Jacoby (1995b) , and it also confirms anew the conclusion reached by Banks (1970) and by Lockhart and Murdock (1970) that ROC curves are in general uninformative for distinguishing between threshold and signal-detection models, especially if the

A few other points need to be considered when evaluating the EMM against the DPSDM. First, we know of no satisfactory solutions to the statistical problems of parameter estimation, computation of confidence intervals, and goodness-of-fit testing for restricted models within the DPSDM framework. We have given reasons for why within-subject manipulations of the test conditions and single-participant estimates are problematic. These problems do not exist for the EMM because it is formally a GPT model for which an elaborated statistical framework exists (cf. Hu & Batchelder, 1994; Riefer & Batchelder, 1988) . Second, as we have shown, the validity of the DPSDM is tied to the independence assumption. Considering the serious criticisms of the independence assumption, it seems wise to stand away from it. The EMM does that. It does not need *any* assumption about the relation between recollection and automatic, familiarity-based processes to be identifiable. Finally, when a formal model test is used, it turns out that the DPSDM does not fit the ROC data of Yonelinas (1994) which it was designed to fit. This can be inferred by generalizing the DPSDM to a model that does no longer need the normal distribution assumption and by performing goodness-of-fit tests for the generalized version. These tests can be conducted because the generalized version is formally a GPT model. The generalized, distribution-free DPSDM does not fit the data of Yonelinas (1994). Hence, the DPSDM as its proper submodel cannot fit them either.

Returning to the EMM, there seems to be little disagreement about the superiority of that measurement model for the process dissociation procedure over the IMM originally suggested by Jacoby (1991) when it comes to correcting for response biases. In the series of experiments reported by Buchner et al. (1995) it was possible to show that the hypothesis of no differences in the memory parameters across bias conditions fitted consistently better for the EMM than for the IMM parameters. However, we felt compelled to analyze in this article the fact that Yonelinas et al. (in press) and Yonelinas and Jacoby (1995b) nevertheless endorsed using the IMM for data sets with equal false-alarm rates in the inclusion and exclusion test conditions and—for certain purposes—also for data sets in which these false-alarm rates differ.

only when false-alarm rates differ between inclusion and exclusion test conditions, but also when false alarms are the same in the inclusion and exclusion test conditions and differ between groups or experimental manipulations. In fact, as we have shown, examples exist in the literature of exactly these sorts of contaminations in the parameters when the IMM is used (Komatsu et al., 1995; Verfaellie & Treadwell, 1993) . On top of that, further analyses showed that even if the false-alarm rates are constant across all test conditions and across all groups or experimental manipulations, and even if the absolute size of the familiarity effect is not of interest to the researcher, the IMM still leads to contaminated estimates of the contributions of automatic, familiarity-based processes when compared to the EMM.

To summarize so far, we conclude that (a) the EMM is clearly superior to the IMM in that the former, but not the latter, was able to account for data in which response biases were manipulated experimentally (see Buchner et al., 1995); (b) it can be rather dangerous to use the IMM even in situations for which it was recommended as being acceptable by Yonelinas et al. (in press) and by Yonelinas and Jacoby (1995b) ; (c) the EMM has the advantage of being embedded into the statistical theory of GPT models (cf. Hu & Batchelder, 1994; Riefer & Batchelder, 1988) which allows for satisfactory methods of parameter estimation, computation of confidence intervals, and goodness-of-fit testing for restricted models; (d) the EMM can be generalized to confidence rating data, and it fits perfectly such data provided by Yonelinas (1994) whereas the DPSDM as favored by Yonelinas et al. (in press) and by Yonelinas and Jacoby (1995b) can be shown to not be able to fit these very same data; and (e) the EMM is superior to the version of the DPSDM favored by Yonelinas et al. (in press) and by Yonelinas and Jacoby (1995b) in that it does not depend on the questionable assumption of independence between recollection and familiarity-based processes.

Although the EMM shows up to advantage in this model comparison, it is clear that its fit to the data presented by Buchner et al. (1995) was good but still less than perfect. Instead of following Yonelinas and Jacoby (1995b) in attributing this imperfect fit to influences of participants' willingness to engage in recollection

1987; O'Dekirk et al., 1993) we looked for generalizations of both the DPSDM and the EMM that might provide a better fit to the available data. From all options considered, the 2HT generalization of the EMM appears to be the most promising. The motivation for this generalization was that the EMM's assumption of distractors never being detected as new may be too restrictive for many experimental settings. We introduced two variants of a new 2HT-EMM. The multiple-groups variant of the 2HT-EMM is applicable whenever there is more than one experimental manipulation or group and the assumption is reasonable that the parameters representing the memory processes do not vary across groups or experimental manipulations. The single-group variant of the 2HT-EMM assumes that the probability of a conscious rejection is equal to the probability of recollecting a target or, more pragmatically, the "average" target if there is more than one class of targets and the recollection parameters for these differ. Both model variants can be evaluated using the data provided by Buchner et al. (1995). In fact, the fit of both model variants to these data was not only clearly better than the IMM's fit but also better than the fit of the EMM, indicating that at least on a certain proportion of trials participants indeed consciously identified distractors as new in those experiments. This does not mean that the 2HT-EMM will be superior to the EMM (i.e., 2HT-EMM assuming $d = 0$) in all applications. Whether the 2HT-EMM or the EMM is more adequate in practice will largely depend on the peculiarities of the experimental situation. Finally, an informal analysis based only on a comparison of parameter estimates as suggested by Yonelinas and Jacoby (1995b, Table 1) demonstrated that the 2HT-EMM was also superior to the DPSDM.

In sum then, the 2HT-EMM does seem to provide the best available framework for constructing measurement models for the process dissociation procedure at this stage.

# References

Baars, B.J. (1992). A dozen competing-plans techniques for inducing predictable slips in speech and action. In B.J. Baars (Ed.), *Experimental slips and human error: Exploring the architecture of volition. Cognition and language.* (S. 129-150). New York: Plenum Press.

Banks, W.P. (1969). Criterion change and response competition in unlearning. *Journal of Experimental Psychology*, *82*, 216-223.

Banks, W.P. (1970). Signal detection and human memory. *Psychological Bulletin*, *74*, 81-99.

Batchelder, W.H., Hu, X. & Riefer, D.M. (1994). Analysis of a model for source monitoring. In G.H. Fischer & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics, and methodology* (S. 51-65). New York: Springer.

Batchelder, W.H. & Riefer, D.M. (1990). Multinomial processing models of source monitoring. *Psychological Review*, *97*, 548-564.

Batchelder, W.H., Riefer, D.M. & Hu, X. (1994). Measuring memory factors in source monitoring: Reply to Kinchla. *Psychological Review*, *101*, 172-176.

Bayen, U.J., Murnane, K. & Erdfelder, E. (in press). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Begg, I.M., Anas, A. & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General*, *121*, 446-458.

Bröder, A. & Bredenkamp, J. (1995). *SLIP-Technik, Prozeßdissoziationsmodell und multinomiale Modellierung: Neue Werkzeuge zum experimentellen Nachweis „Freud'scher Versprecher"?* Manuscript submitted for publication.

Buchner, A., Erdfelder, E. & Vaterrodt-Plünnecke, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *Journal of Experimental Psychology: General*, *124*, 137-160.

Buchner, A., Faul, F. & Erdfelder, E. (1992). *G•Power: A priori, post-hoc, and compromise power analyses for the Macintosh* [Computer-Programm]. Bonn, FRG: Bonn University.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd. ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Colquhoun, W.P. (1967). Sonar target detection as a decision process. *Journal of Applied Psychology*, *51*, 187-190.

Cowan, N. & Stadler, M.A. (in press). Estimating unconscious processes: Implications of a general class of models. *Journal of Experimental Psychology: General*.

Curran, T. & Hintzman, D.L. (1995). Violations of the independence assumption in process dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 531-547.

Debner, J.A. & Jacoby, L.L. (1994). Unconscious perception: Attention, awareness, and control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 304-317.

Dusoir, T. (1983). Isobias curves in some detection tasks. *Perception and Psychophysics*, *33*, 403-412.

Egan, J.P., Greenberg, G.Z. & Schulman, A.I. (1961). Operating characteristics, signal detectability, and the method of free response. *Journal of the Acoustical Society of America*, *33*, 993-1007.

Galanter, E. & Holman, G.L. (1967). Some invariances of the isosensitivity function and their implications for the utility function of money. *Journal of Experimental Psychology*, *73*, 333-339.

Gigerenzer, G. (1981). *Messung und Modellbildung in der Psychologie*. München: Ernst Reinhardt Verlag.

Gillund, G. & Shiffrin, R.M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*, 1-67.

Graf, P. & Komatsu, S.-I. (1994). Process-dissociation procedure: Handle with caution! *European Journal of Cognitive Psychology*, *6*, 113-129.

Green, D.M. & Swets, J.A. (1966). *Signal detection theory and psychophysics*. New York: John Wiley & Sons.

Hu, X. (1993). *Statistical inference program for multinomial binary tree models* [Computer-Programm]. Irvine, CA: University of California, Irvine.

Hume, A.L. (1974). Optimal response biases and the slope of ROC curves as a function of signal intensity, signal probability, and relative payoff. *Perception & Psychophysics*, *16*, 377-384.

Jacoby, L.L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513-541.

Jacoby, L.L. & Begg, I.M. (1995). *In defense of functional independence: Violations of assumptions underlying the process dissociation procedure?* Manuscript submitted for publication.

Jacoby, L.L., Toth, J.P. & Yonelinas, A.P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General*, *122*, 139-154.

Jacoby, L.L., Toth, J.P., Yonelinas, A.P. & Debner, J.A. (1994). The relationship between conscious and unconscious influences: Independence or redundancy? *Journal of Experimental Psychology: General*, *123*, 216-219.

Jacoby, L.L., Yonelinas, A.P. & Jennings, J.M. (in press). The relation between conscious and unconscious (automatic) influences: A declaration of independence. In J. Cohen & J.W. Schooler (Eds.), *Scientific approaches to the question of consciousness*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Jennings, J.M. & Jacoby, L.L. (1993). Automatic versus intentional uses of memory: Aging, attention, and control. *Psychology and Aging*, *8*, 283-293.

Johnson, M.K., Hashtroudi, S. & Lindsay, D.S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3-28.

Jones, G.V. (1987). Independence and exclusivity among psychological processes: Implications for the structure of recall. *Psychological Review*, *94*, 229-235.

Joordens, S. & Merikle, P.M. (1993). Independence or redundancy? Two models of conscious and unconscious influences. *Journal of Experimental Psychology: General*, *122*, 462-467.

Kinchla, R.A. (1994). Comments on Batchelder and Riefer's multinomial model for source monitoring. *Psychological Review*, *101*, 166-171.

Kintsch, W. (1967). Memory and decision aspects of recognition learning. *Psychological Review*, *74*, 496-504.

Kintsch, W. (1977). *Memory and cognition* (2nd. ed.). New York: John Wiley & Sons.

Komatsu, S.-I., Graf, P. & Uttl, B. (1995). Process-dissociation procedure: Core assumptions fail, sometimes. *European Journal of Cognitive Psychology*, *7*, 19-40.

Krantz, D.H. (1969). Threshold theories of signal detection. *Psychological Review*, *76*, 308-324.

Levine, J.M. (1966). The effects of values and costs on the detection and identification of signals in auditory vigilance. *Human Factors*, *8*, 525-537.

Lindsay, D.S. & Jacoby, L.L. (1994). Stroop process dissociations: The relationship between facilitation and interference. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 219-234.

Lockhart, R.S. & Murdock, B.B. (1970). Memory and the theory of signal detection. *Psychological Bulletin*, *74*, 100-109.

Luce, R.D. (1963a). Detection and recognition. In R.D. Luce, R.R. Bush & E. Galanter (Eds.), *Handbook of mathematical psychology. Vol. 1* (S. 245-308). New York: John Wiley & Sons.

Luce, R.D. (1963b). A threshold theory for simple detection experiments. *Psychological Review*, *70*, 61-79.

Macmillan, N.A. & Creelman, C.D. (1990). Response bias: Characteristics of detection theory, threshold theory, and "nonparametric" indexes. *Psychological Bulletin*, *107*, 401-413.

Macmillan, N.A. & Creelman, C.D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.

Marken, R.S. & Sandusky, A.J. (1974). Stimulus probability and sequential effect in recognition memory. *Bulletin of the Psychonomic Society*, *4*, 49-51.

Murdock, B.B. (1974). *Human memory: Theory and data*. Potomac, MD: Lawrence Erlbaum Associates.

Nilsson, L.G. (1987). Motivated memory: Dissociation between performance data and subjective reports. *Psychological Research*, *49*, 183-188.

O'Dekirk, J.M., Wyatt, B.S. & Ellis, N.R. (1993). Individual differences in location memory. *Bulletin of the Psychonomic Society*, *31*, 66-68.

Parks, T.E. (1966). Signal-detectability theory of recognition-memory performance. *Psychological Review*, *73*, 44-58.

Ratcliff, R., Sheu, C.f. & Gronlund, S.D. (1992). Testing global memory models using ROC curves. *Psychological Review*, *99*, 518-535.

Ratcliff, R., Van Zandt, T. & McKoon, G. (in press). Two factor theory, single process theories, and recognition memory. *Journal of Experimental Psychology: General*.

Richardson-Klavehn, A., Gardiner, J.M. & Java, R.I. (1995). Memory: Task dissociations, process dissociations, and dissociations of consciousness. In G. Underwood (Ed.), *Implicit cognition* (S. 85-158). Oxford: Oxford University Press.

Riefer, D.M. & Batchelder, W.H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*, 318-339.

Roediger, H.L. & McDermott, K.B. (1994). The problem of differing false-alarm rates for the process dissociation procedure: Comment on Verfaellie and Treadwell (1993). *Neuropsychology*, *8*, 284-288.

Russo, R. & Andrade, J. (1995). The directed forgetting effect in word fragment completion: An application of the process dissociation procedure. *The Quarterly Journal of Experimental Psychology*, *48A*, 405-423.

Smith, P.T. (1969). The non-monotonicity of the psychometric function in recognition memory. *Perception & Psychophysics*, *5*, 329-337.

Smith, P.T. (1970). Decision processes in long-term memory. *Acta Psychologica*, *33*, 315-325.

Snodgrass, J.G. & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*, 34-50.

Strack, F. & Bless, H. (1994). Memory for non-occurrences: Metacognitive and presuppositional strategies. *Journal of Memory and Language*, *33*, 203-217.

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

Tanner, T.A., Haller, R.W. & Atkinson, R.C. (1967). Signal recognition as influenced by presentation schedules. *Perception & Psychophysics*, *2*, 349-358.

Toth, J.P., Reingold, E.M. & Jacoby, L.L. (1994). Toward a redefinition of implicit memory: Process dissociations following elaborative processing and self-generation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 290-303.

Vaterrodt-Plünnecke, B. (1994). Multinomiale Modellierung impliziter Gedächtnisprozesse: Ein alternativer Ansatz. *Zeitschrift für experimentelle und angewandte Psychologie*, *41*, 295-314.

Verfaellie, M. & Treadwell, J.R. (1993). Status of recognition memory in amnesia. *Neuropsychology*, *7*, 5-13.

Wender, K.F. (1975). Signal identification: The comparison of two competitive theories. *Perception & Psychophysics*, *18*, 250-254.

Wippich, W. (1994). Unbewußte Effekte und Voreingenommenheiten bei Urteilen zu Personennamen. *Zeitschrift für experimentelle und angewandte Psychologie*, *41*, 154-172.

Yonelinas, A.P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1341-1354.

Yonelinas, A.P. & Jacoby, L.L. (1995a). Dissociating automatic and controlled processes in a memory search task: Beyond implicit memory. *Psychological Research*, *57*, 156-165.

Yonelinas, A.P. & Jacoby, L.L. (1995b). *Response bias in the process dissociation procedure*. Manuscript submitted for publication:.

Yonelinas, A.P., Regehr, G. & Jacoby, L.L. (in press). Incorporating response bias in a dual-process theory of memory. *Journal of Memory and Language*.

# Appendix

## Proof of the Identifiability of the DPSDM

The first two model equations of the DPSDM can be solved for $d'/2 - k_i$ and $-d'/2 - k_i$ by making use of the inverse $\Phi^{-1}$ of the standard normal distribution function $\Phi$. Equation 8 implies

$$d'/2 - k_i = \Phi^{-1}((p_{1i} - c)/(1 - c)) \tag{A1}$$

and Equation 9 implies

$$-d'/2 - k_i = \Phi^{-1}(p_{di}). \tag{A2}$$

By subtracting (A2) from (A1) we obtain

$$d' = \Phi^{-1}((p_{1i} - c)/(1 - c)) - \Phi^{-1}(p_{di}). \tag{A3}$$

Applying an analogous procedure to the remaining two Model Equations 10 and 11 yields

$$d' = \Phi^{-1}(p_{1e}/(1 - c)) - \Phi^{-1}(p_{de}). \tag{A4}$$

All model parameters but $c$ are eliminated by subtracting Equation A4 from Equation A3:

$$\Phi^{-1}((p_{1i} - c)/(1 - c)) - \Phi^{-1}(p_{1e}/(1 - c)) = \Phi^{-1}(p_{de}) - \Phi^{-1}(p_{di}). \tag{A5}$$

This equation uniquely determines $c$ as a function of $p_{1i}$, $p_{di}$, $p_{1e}$, and $p_{de}$, because (a) the term $\Phi^{-1}((p_{1i} - c)/(1 - c))$ is strictly monotonically decreasing in $c$, (b) $\Phi^{-1}(p_{1e}/(1 - c))$ is strictly monotonically increasing in $c$, and, thus, the complete left side of Equation A5 is a strictly monotonically decreasing function

$$f(c) := \Phi^{-1}((p_{1i} - c)/(1 - c)) - \Phi^{-1}(p_{1e}/(1 - c)). \tag{A6}$$

Therefore, an inverse function $f^{-1}$ exists. When $f^{-1}$ is applied to both sides of Equation A5 we obtain

$$c = f^{-1}(\Phi^{-1}(p_{de}) - \Phi^{-1}(p_{di})). \tag{A6}$$

Since the right side of Equation A6 only depends on the four response probabilities, $c$ must be identifiable.

Looking back at Equation A3 or A4 it is obvious that $d'$ must be identifiable when $c$ is identifiable and $c \neq 1$. Finally, when $d'$ is identifiable, then the identifiability of both $k_i$ and $k_e$ follows immediately from Equations 9 and 10. This completes the proof.

# Authors' Notes

Edgar Erdfelder, Department of Psychology, University of Bonn, Bonn, Germany, and Axel Buchner, Department of Psychology, University of Trier, Trier, Germany. We would like to thank Melanie Steffens for valuable comments on an earlier draft. The research reported in this article was supported by grants from the Deutsche Forschungsgemeinschaft to Edgar Erdfelder (Er 224/1-1) and to Axel Buchner (Bu 945/1–1). Correspondence concerning this article should be addressed to Edgar Erdfelder, Psychologisches Institut, Universität Bonn, Römerstraße 164, D-53117 Bonn, Germany. Electronic mail may be sent via Internet to erdfelder@uni-bonn.de.

## Berichte aus dem Psychologischen Institut der Universität Bonn

Die „Berichte aus dem Psychologischen Institut der Universität Bonn“ (ISSN 0931-024X) gibt es seit 1975. Ab Jahrgang 5 (1979) beginnt die Heftzählung in jedem Jahr bei Heft 1. Von 1984 bis 1994 wurden die „Bonner Berichte“ von Dr. Edgar Erdfelder und Dr. Joachim Funke herausgegeben. Seit 1995 fungieren Dr. Edgar Erdfelder und Dr. Michael Kavšek als Herausgeber. Eine Übersicht über die zuletzt publizierten Hefte gibt nachfolgende Liste.

**<u>Band 17 (1991)</u>**

Heft 1: Funke, J. (1991). Der Rückfall des Alkoholabhängigen: Versagen der Therapie oder Bestandteil erfolgreicher Behandlung?

Heft 2: Buchner, A., & Funke, J. (1991). Transfer of associations in finite state automata.

Heft 3: Buchner, A., Schmitt, L., Funke, J., & Nikelowski, B. (1991). MacFAUST: Program for constructing finite automata as instruments in problem solving research.

Heft 4: Pörzgen, B. (1991). Weiterentwicklungen in der psychologischen Erforschung der Geschlechtsrollendifferenzierung.

Heft 5: Regenwetter, M. (1991). Multinomial modeling of cognitive processes in the „Memory“ game.

**<u>Band 18 (1992)</u>**

Heft 1: Pörzgen, B. (1992). Das Leib-Seele-Problem und einige seiner Konsequenzen und Lösungsansätze in der Psychologie.

Heft 2: Erdfelder, E. (1992). Parameter estimation and goodness-of-fit tests for Batchelder and Riefer´s storage-retrieval model: A manual for users of the SRM program.

Heft 3: Faul, F. (1992). Determinanten des Kontexteffekts bei geometrisch-optischen Täuschungen: Ein Experiment zur Tiefensignalverarbeitungstheorie.

Heft 4: Fischer, V. (1992). Wie beurteilen deutsche Studenten Ausländer?

**<u>Band 19 (1993)</u>**

Heft 1: Beckmann, J.F., Funke, J. & Guthke, J. (1993). Zum Zusammenhang von komplexem Problemlösen, Intelligenztest- und Lerntest-Indikatoren.

Heft 2: Meiser, T. (1993). Loglinear Rasch models: Theory and applications in longitudinal research.

Heft 3: Faul, F. & Erdfelder, E. (1993). Vergleich verschiedener experimenteller Verfahren zum Test additiver Informations-Integrations-Modelle.

Heft 4: Hinnersmann, H. (1993). Evaluation eines Trainingansatzes zum Schlußfolgern und Urteilen.

Heft 5: Faul, F. (1993). NCM: Numerisches Conjoint Measurement. Eine Programmdokumentation.

Heft 6: Meiser, T., Fichtel, M., Rompe, P. & Rudinger, G. (1993). Mixture distribution Rasch models for the analysis of longitudinal discrete data.

**<u>Band 20 (1994)</u>**

Heft 1: Erdfelder, E., Faul, F. & Buchner, A. (1994). GPOWER: A general power analysis program.

**<u>Band 21 (1995)</u>**

Heft 1: Ruoß, M. (1995). Schmerz und Kognition. Kognitive Faktoren bei der Entstehung, Aufrechterhaltung und psychologischen Behandlung chronischer Schmerzen.

Heft 2: Scholz, O.B. & Adler, F. (1995). Atopische Dermatitis und ihre Beziehungen zu dermatologischen, allergologischen, ökologischen und psychosozialen Einflußfaktoren. Selbstberichte davon Betroffener.

Heft 3: Erdfelder, E. & Buchner, A. (1995). Process dissociation measurement models: Good versus better.