Confidence and Recognition Dissociations in Recognition Memory: Evidence for Familiarity and Recollective Processes

Thomas A. Busey Anne Arici
Indiana University, Bloomington

Submitted Manuscript. Please do not quote.
Send Correspondence to: Tom Busey Anne Arici
Department of Psychology Department of Psychology
Indiana University Indiana University
Bloomington, IN, 47405 Bloomington, IN, 47405
busey@indiana.edu aarici@indiana.edu
Abstract

In four recognition memory experiments, scale-invariant dissociations between confidence and recognition judgments to novel faces are explored. Composite distractor faces were created by morphing two studied faces. These morph distractors are identified as 'old' more often than studied faces, yet observers give the morph distractors lower confidence ratings when they say 'old'. Similar dissociations are found in forced-choice experiments, which rule out explanations based on unequal-variance signal detection theory and properties of morphs. A candidate model based on sampling has properties of a recollective process, and can account for these dissociations. This model, however, fails to account for recognition data from an additional experiment in which the number of similar faces is varied at study. Together the results of all experiments support a combination of recollective and familiarity-based processes in both confidence and recognition memory judgments. Properties of these two processes are explored and extended to other domains.
Consider the following scenario. While riding an elevator on campus, a student catches your eye. The face appears familiar and reminds you of one of your class members. How do you decide whether to trust your memory and engage the student in conversation? The answer to this question requires knowledge of both the results of a probe of memory as well as how these results should be evaluated. If you can place more confidence in the results of your query of memory that lead to the initial feeling of familiarity, you may be more likely to risk a mistaken identity and begin a conversation.

In this article we will argue that an evaluation of the metacognitive aspects of recognition memory reveals not only how we monitor the outcome of memory but also how memory representations are accessed during the recognition of novel faces. In particular, we will argue that a complete account of recognition and confidence data from four experiments require processes that in the literature have been detailed as recollective and familiarity-based mechanisms. Through our discussion, we offer specific definitions of these processes and describe why neither alone is sufficient to account for the entire set of data.

For our studies, we have chosen a recognition memory paradigm using novel faces. We do this not only for obvious everyday and eyewitness applications, but also because faces have received relatively little attention in the memory modeling literature (but see Yonelinas, et al. 1999; Criss & Shiffrin (submitted) for exceptions). It is our intent to ground our accounts of both our recognition and our confidence data in the existing models of memory. One benefit of this approach is that it extends computational memory models into the domain of metacognition (see Van Zandt, 2000; and Clark, 1997, for similar approaches). We have subjects study novel faces, and then perform recognition judgments followed by confidence judgments. The use of novel faces provides a degree
of control over the similarity of studied items and allows us to create stimuli with known similarity properties, which would be more difficult with words.

**Existing Models of Memory**

The existing models of memory differ on many dimensions, but one possible grouping considers what information is used to make a recall or recognition judgment. This ordering groups models into those that rely on a comparison of the test item with a single item in memory to produce a recognition judgment, and those that combine information across multiple items in memory to make a recognition judgment. This distinction seems to capture the essential difference between recollective processes, in which a single item is recalled from memory and compared with the test probe, and a familiarity-based process, where information is combined across multiple items to produce a single value, where no one item is singled out for special treatment.

The class of models known as Global Memory models (e.g. SAM (Raaijmakers & Shiffrin, 1980; 1981, Gillund & Shiffrin, 1984), MINERVA2 (Hintzman, 1986, 1988), CHARM (Eich, 1982; 1985), Todam (Murdock, 1982), REM (Shiffrin & Steyvers, 1997)) combine or sum activations across all items in memory to compute a scalar that represents global familiarity of a test item. Models of categorization and recognition that rely on summed similarity (e.g. the Generalized Context Model, Nosofsky (1986)) share this property of combination over items to produce a single summary score.

In contrast to summing information across items in memory, several models have implemented a version in which memory is sampled and a single item is recovered for comparison with the test item. This is analogous to a recollective process because in the final step information from one item is singled out for comparison with the test item. Many of the global memory models described above include an optional recall mechanism, in which a sampling and recovery process can be used to recover individual
item information from memory. Only a few models have argued for a single sampling process. For example, the SimSample model (Busey & Tunnicliff, 1999) conjoins the sampling mechanism of SAM ((Raaijmakers & Shiffrin, 1980; 1981, Gillund & Shiffrin, 1984)) with an MDS representation to predict recognition of novel faces. Unlike SAM, SimSample does not repeatedly sample items. The model was able to successfully account for the behavior of very distinctive and very typical faces in a novel recognition paradigm (Busey & Tunnicliff, 1999). The neural net model of McClelland & Chappell (1998) shares many features with the REM model, differing only in that a Max rule rather than a sum is used to compute recognition responses.

One reason these models with such differing architectures can account for data in similar memory paradigms may be that the outputs from familiarity-based and recollection-based mechanisms are often highly correlated. This can occur if one item produces an activation that is much greater than the others, which results in a sum rule (given by a familiarity-based process) producing a result similar to a max rule (given by a recollection process). Thus evidence demonstrating the existence of both mechanisms has had to contend with the possible correlation of these processes (Curran & Hintzman, 1997), and several authors have argued that a single-process account is sufficient to explain extant data (Ratcliff, Van Zandt & McKoon, 1995).

More recent models have included elements of both recollection and familiarity-based processes. Yonelinas (1994, 2001) has proposed a dual-process model for recognition using signal-detection theory to differentiate between familiarity and recollection. In his model, familiarity is seen as a graded signal-detection process that sums activity over multiple items, whereby participants make responses on the basis of the familiarity of the test item in its context. Recollection, however, acts as a high-threshold all-or-none process, where participants are either successful in retrieving information about the study
event, or they fail to retrieve such information. High-threshold recollective processes are associated with higher confidence in the recollection response, while responses based on familiarity result in a wider range of confidence.

Rotello, Macmillan & Reeder (submitted) have proposed a similar model in which specific memory (associated with a recollection process) and general memory (analogous to a familiarity process) both contribute to old/new and remember/know judgments. However, neither old/new nor remember/know judgments are assumed to rely only on a single memory source; instead, the old/new decision is the sum of the outputs from specific and general memory systems, while the remember/know judgment is computed from the difference between the two systems. Thus if the outputs of either (or both) system are high, the subject will say 'old', but the subject will only give a 'remember' response if the output of the specific memory dominates the general memory output.

Recently Norman and O'Reilly (submitted) have suggested a neural instantiation of recollection and familiarity-based mechanisms. In their model, the hippocampus recalls specific details, while the neocortex extracts a scalar value that is akin to a global familiarity signal. An advantage of this approach is that it attempts to remain biologically plausible, although this does make the model somewhat complex. The model's strength is that it defines the interactions between the hippocampal and neocortical systems, and the authors have described a number of testable predictions that relate similarity of test items to the functions of the two memory components.

Existing Models of Metacognition

While much of the work on memory models has stressed the nature of the memory representation and the interactions between studied words, test probes and context, the work on metacognition and confidence judgments has stressed the nature of the test item. Metacognition researchers have acknowledged the contributions of the match between
the test item and studied items (known as trace access theory, Burke, MacKay, Worthley, & Wade, 1991; Hart, 1967; King, Zechariester, & Shaughnessy, 1980), but most of the work has addressed those factors at test that might influence confidence judgments beyond trace access. Variables associated with the test probe such as perceptual fluency (Kelley & Lindsay, 1993), ease of encoding (Benjamin, Bjork, & Schwartz, 1998; Begg, Duft, Lalonde, Melnick, & Sanvito, 1989), and test luminance (Busey & Tunnicliff, 1999) have all been shown to selectively influence confidence. These have been formalized by Koriat (1993, 1995, 1997) into what he terms the accessibility hypothesis, which assumes that learners have no knowledge of their own memory beyond what they can retrieve or what they can infer about what they cannot access. Thus, feelings of confidence stem from a combination of the overall accessibility of information, including the completeness, speed, and ease of retrieval. If an item is easily retrieved, the resulting confidence is high. However, a combination of more subtle cues such as partial recall and contextual information can cause a subjective feeling of knowing, even when retrieval is not possible.

The relationship between recognition and confidence for novel faces has received much attention in the applied literature, where the focus has been on the reliability of eyewitness testimony (see Wells, 1993, for a recent discussion). Here, much of the research has pointed to relatively low correlations between confidence and recognition accuracy, although this correlation improves if one focuses on within-subject correlations provided about several different events by one eyewitness (Lindsay, Read, & Sharma, 1998). Formal models of the nature of the memory processes for these effects have been developed in only a few cases (Clark, 1997; Van Zandt, 2000).

In summary, models of both memory and metamemory focus on the nature of processes by which information in the test item is used to probe memory. However,
memory models stress the nature of the representation and the process by which test item information is compared against memory. Several authors have models that have been extended to suggest how confidence might be determined from this memory probe operation (e.g. Clark, 1997; Dobbins, Kroll, & Liu, 1998; Van Zandt, 2000, and Yonelinas, 1994, 2001), but much of the work on face recognition in eyewitness testimony as address the size of the relation and the variables that affect the correlation, rather than taking a formal modeling approach. This leaves open the question of how memory is accessed for novel faces, and how confidence is determined through this process. Against this background, our first goal is to determine how confidence and recognition are related in a recognition memory paradigm with novel faces. We will use this data to constrain candidate models of recognition memory and confidence.

**General Methods for Experiments**

Rather than pick a random selection of faces, we chose to build in particular relations into our stimulus set, while still maintaining photorealistic images. This creates stimuli with known relations between existing images in the space, which is a compromise between completely structured spaces (as might be done with colors) and complete unstructured spaces (as might be done by choosing faces at random). We built in similarity relations by using morphed stimuli, which combine the features of two faces. This produces a new, photorealistic face with several important properties. First, this stimulus is similar to at least two other faces (which we call the parent faces). Second, the morph tends to lie near the midpoint of the two parent locations in the multidimensional space (Busey, 1998). Finally, due to the geometry of the space, the location of the morph tends to fall near the center of the space, making the morph more likely to be viewed as more typical than either parent face (Busey, 1998). This may have consequences for the memory processes, such that a familiarity mechanism might be more active for typical
faces. The morphing process does introduce certain visual artifacts, which we control for in Experiments 3 and 4.

Our central concern in Experiments 1a and b is the relation between confidence and recognition when parents are compared to their associated morphs. Parent faces were studied along with additional target faces (which we call filler faces) and at test the morph faces were added as distractors in an old-new face recognition paradigm. After each recognition judgment, participants indicated their confidence in the accuracy of their just-previous response. Hit rates and false-alarm rates were calculated across participants to measure the probability of saying "old" given a target or a distractor.

Experiment 1a measured confidence and recognition for parents, morphs, and filler faces. By comparing the hit rates, false-alarm rates, and confidence ratings we were able to identify a confidence/recognition inversion that suggests the two measures may reflect different processes. Experiment 1b was designed to replicate Experiment 1a with the additional feature of a speeded task.

**Experiment 1a**

The goal of experiment 1a was to measure the recognition and confidence for parents, morphs, and filler faces. By examining the relationship between morphs and parent faces, we can see how confidence and recognition vary as a function of the similarity relations between the two types of faces, and use this information to identify possible memory processes at work. When the morphs are distractors, we anticipate that these faces will attract a large number of false alarms, based on their high similarity to the studied parents and many other faces. Morph distractors have no match in memory, thus, false alarms to these faces could reflect the contribution of a familiarity mechanism.
Retrospective confidence ratings were taken after each response, which, when conjoined with the recognition rates in a formal model, provide additional information about the nature of the memory processes as described below.

Method

Participants

Participants were 263 undergraduates from Indiana University who participated in the experiment as part of course requirements. Participants were run in small groups with approximately 4-6 subjects participating at a time in 50 total groups. Note that in an attempt to create a robust dataset that could be used to test and reject models, all of our experiments have a large number of participants, in some cases over 500. This is required by our choice of experimental design, which has many different conditions and allows us to compute hit and false alarm rates for individual faces where this information is useful for model testing. The large number of participants should not be interpreted as an indication that subjects were not tested properly or that studies were run in a classroom setting. All participants were carefully run in small groups in the laboratory in experiments that lasted between 30 and 60 minutes.

Stimuli

The stimuli consist of 60 photographs of bald white men (Kayser, 1985). All photographs show faces at the same frontal view and angle, under the same lighting conditions. All the photographs show men who are clean-shaven and have similar expressions. Sixteen parent faces were used to create 8 morphs, and were included in the 60-face total.

The photographic images were displayed on a 21-in. Macintosh gray-scale monitor using luminance control and gamma correction from a Video Attenuator and the VideoToolbox software library (Pelli & Zhang, 1991). The background luminance was
set to 5 cd/m². The gray-scale values in the images were scaled to cover the range between 5 cd/m² for black to 80 cd/m² for white. Data were collected by using one numeric keypad per participant and a single PowerMac computer.

Design and Procedure

Two study sets of faces were constructed so that each group included 8 parents, 4 morphs and 18 filler faces, for a total of 30 faces. The design is counterbalanced so that one group of participants studies one set of 30 faces (Set A) and the next group studies a similar set of 30 faces (Set B). At test, both sets A and B are presented to measure the hit rates and false alarms to the targets and distractors. Of particular importance, the morphs and associated parents are not all targets at the same time. That is, if the parents are studied, their corresponding morphs are not studied, and instead are included as distractors at test. Conversely, the morphs that are studied have their associated parents included in the test session as distractors. Across subjects, hit rates and false alarm rates are recorded for all faces.

At the start of the study session, participants were told to study each face in preparation for a following memory test. The 30 faces were shown one at a time, in randomized order. Each face was presented for 1200 ms, followed by a 2-second delay before presentation of the next face. Immediately after the study phase, participants were tested in an old/new picture recognition paradigm. The 60 faces were presented in random order and participants were asked to identify each as 'old' or 'new', for faces they had studied or had never seen, respectively. Retrospective confidence ratings were taken immediately after each test face. Participants were asked to respond how confident they were in the accuracy of their just previous response and assign a corresponding confidence value from a scale of 1 to 5 (1=0% confident, 2=25% confident, 3=50% confident, 4=75% confident, and 5=100% confident).
Results and Discussion

Figure 1 shows the results for Experiment 1a. The mean hit rate to the filler faces (any face that is not a parent or a morph) across subjects was 0.659, with a mean confidence rating of 72.25%, showing a fairly high hit rate and confidence rating for these faces. The false alarm rate to filler faces was low at 0.209. The critical comparison is between parents and morphs. The parent hit rate was 0.657 and false-alarm rate was 0.370. The high parent false alarm rate is due to the fact that the associated morph was studied as dictated by the design of the experiment.

The morph faces provided both a higher hit rate (0.753) and a higher false-alarm rate (0.712) than the parent hit rate. However, despite being more likely to say 'old' to the morph distractor than to the parent targets, participants were more confident when saying 'old' to the parents than when saying 'old' to the morph distractors. Confidence when participants said 'old' to parent targets is 68.3%, and for morph distractors is 64.1%. This reveals a confidence and recognition inversion between parent targets and morph

Figure 1. Experiment 1a data. Right panel: Probability of saying 'old' to different experimental conditions. Left panel: Mean confidence for 'old' responses. Error bars represent ±1 SEM.
distractors: In the conditions where participants are more likely to say 'old' (morph distractors), they are least confident when they say 'old'. This effect replicates in Experiment 1b.

Inversions of recognition and confidence suggest that several different processes might be at work, influencing recognition and confidence in different ways. One possibility is that recognition relies on the output of a recollective mechanism with the possible contribution from familiarity-based mechanisms (Yonelinas, 1997). Morphs, in particular, may promote responses based on a familiarity mechanism due to their high similarity to parents and other faces. However, confidence may be influenced more by the outputs of a recollective process such that when a recollection process is used, participants may realize this and give higher confidence values as a result (Yonelinas, 2001; Yonelinas, Kroll, Dobbins, & Soltani, 1999). Clearly the morph distractor cannot be recollected because it was not studied, and this might lead to the lower confidence for the morph distractors. Thus, to investigate the possibility of familiarity-based processes at work with the morphs, we designed a condition to selectively enhance a familiarity mechanism.

**Experiment 1b**

The goal of Experiment 1b was to replicate the findings in Experiment 1a, while also promoting reliance on a familiarity-based mechanism by decreasing the ability to use recollective processes. To this end, Experiment 1b was identical to Experiment 1a, except for the addition of a speeded response requirement. Past research in the verbal learning domain has suggested that familiarity-based recognition has an earlier onset and shorter time course in comparison to slower recollective processes that involve searches through memory (Curran, 1999, 2000; Rotello & Heit, 2000). Requiring participants to respond more quickly may force them to rely more on the results of familiarity mechanisms.
before recollective processes have time to complete. If the confidence and accuracy inversion seen with the morphs in Experiment 1a is related to familiarity-based mechanisms, we expect the speeded task to selectively affect morph distractors.

Method

The stimuli, design and procedures were identical to those in Experiment 1a, with the following exceptions:

Participants

Participants were 209 Indiana University undergraduates who participated in the experiment for course credit. Participants completed the experiment in small groups with approximately 4-6 individuals per group in a total of 50 groups.

Design and Procedure

Participants were instructed that their reaction times would be recorded for each test response. They were encouraged to respond "old" or "new" as quickly as possible, while

Figure 2. Experiment 1b data. Right panel: Probability of saying 'old' to different experimental conditions. Left panel: Mean confidence for 'old' responses. Error bars represent ±1 SEM.
maintaining a relatively high level of accuracy. There were no questions from participants about the implementation of these instructions.

Results and Discussion

Figure 2 shows the data from Experiment 1b. Similar to Experiment 1a, this experiment resulted in a high hit rate (0.669) and mean confidence rating (72.25%) for filler faces. The false-alarm rate to filler distractors was low at 0.232. Parent faces have a hit rate of 0.682 and false-alarm rate of 0.394. Of particular interest, this experiment replicated the confidence and recognition inversion for the morphs found previously in Experiment 1a. The morph false-alarm rate of 0.767 was higher than the parent hit rate (0.682), while mean confidence to morph distractors (62.38%) remained lower than that of the parent targets (70.1%). As with Experiment 1a, participants were least confident with the stimuli (morph distractors) that they were most likely to give an 'Old' rating.

Across-Experiment Comparisons

To compare across experiments, Figure 3 shows a scatterplot of the different conditions. Recognition and confidence are on different axes, and the individual points come from different experimental conditions. Filled symbols represent the unspeeded
task in Experiment 1a, and open symbols represent the speeded Experiment 1b. Asking
the participants to make speeded responses in Experiment 1b has a selective effect on the
morph distractors, as shown in Figure 3. Speeding observers’ responses has the effect of
slightly increasing both the probability of saying 'Old' and confidence for the condition,
which can be seen by comparing filled vs. open circles in Figure 3. Most of these
increases are within the standard error bars. However, the morph distractors show a much
larger effect, such that the subjects who responded quickly treated morph distractors as
essentially identically to Fast and Slow morph targets (gray arrow in Figure 3). This large
shift does not occur with morph targets, indicating that this effect is not due to the
physical properties of the morphs, but depends instead on the study status of the morph
distractors faces (that is, they were not studied).

One possible explanation for this selective increase is that a familiarity process
underlies much of the false recognition of the morphs. When observers are asked to make
speeded responses, they may rely more heavily on a familiarity-based response which
may be larger for morph distractors than in the unspeeded case. This would lead to a
higher false alarm rate for morphs. The familiarity process may also contribute to
confidence, and when the familiarity process is stronger, confidence may increase as
well. This leads to both an increase in the probability of 'old' and confidence in the
speeded Experiment 1b. Other factors, such as a lack of confirmation from a recollective
process, may account for the confidence/recognition inversion. Note that this explanation
requires an overall increase in familiarity response (all else being equal) in the speeded
task, especially for the morphs, and a relatively minimal contribution from a recollection
process for the morphs. Because morph distractors are not studied, they cannot be
recollected to any great extent, and thus the false alarm rates are likely to result mainly
from a familiarity-based process.
This tentative explanation requires much more supporting evidence from Experiments 2-4. However, the fact that morph distractors show a selective dissociation in the speeded task suggests that more than one process may be at work for these stimuli. We will return to this argument later in the paper, but first a number of additional explanations are ruled out first.

**Signal Detection Interpretation of Experiments 1a and 1b**

The confidence/recognition inversions seen in Experiments 1a and 1b are difficult to reconcile with a single-dimensional model such as an equal-variance Signal Detection Theory account, which would predict a monotonic relation between confidence and recognition. However, an unequal variance version can account for this inversion, and there may be reasons to expect differences in the variances of the different target and distractor types. Below we explore this model and describe both why the equal-variance version fails and why the unequal variance version of the model might account for our data.

Predictions for the equal-variance SDT model were generated by assuming that a particular class of faces (e.g. morph distractors) can be represented as a distribution along a single dimension that represents the strength of the match to memory or global familiarity. The exact memory process is unspecified in signal detection theory; the only constraint is that recognition and confidence judgments are based on the same unidimensional value. Here we refer to this value by the neutral term 'strength'. The targets (or distractors) are assumed to be normally distributed with variance set to 1.0. The mean of this distribution (i.e. its position along the strength axis) is a free parameter, and we assume that it reflects the strength of the match or familiarity value, but also might include stimulus specific effects such as smoothing/averaging effects of the morphs as discussed in later sections. The mean of the distractor distribution was fixed at
0 to establish the scale, and the means of the remaining distributions were free parameters. In Experiments 1a and 1b, observers first made an old/new decision and then gave a 0-100% confidence rating using one of 5 values. Thus our strength axis has 1 decision criterion for old/new and 4 confidence criterion on either side. These 9 criterion values were also free parameters, giving means (two for morphs and two for parents) and 9 criterion values for a total of 13 free parameters to fit 36 data points. Data from filler faces was also included in the modeling, although the inclusion of these data places only slight constraints on the placement of the decision and confidence criterion. We report only mean confidence in Table 1, but the entire ROC functions were fit to the data. The best-fitting values of the free parameters were found using the Solver function in Excel to minimize RMSE.

The confidence/recognition dissociations in Experiment 1a and 1b are such that observers are more likely to say 'old' to morph targets and morph distractors than to parent targets, but were more confident when they

Figure 4. **Top panel:** Equal-variance signal-detection representation of the data from Experiment 1a from parents and morphs. **Bottom panel:** Unequal-variance signal detection model of the same data. The parent targets have a higher variance, and this allows the unequal variance version to account for the confidence/recognition inversion seen in Experiments 1a and 1b.
said 'old' to parent targets. We can infer from the recognition data (assuming equal variance) that the parent target distribution is shifted to the left of the morph target and distractor distributions. However, mean confidence is a monotonic function of the location of the distribution: shifting a distribution to the right will always increase confidence for 'Old' responses. Thus the equal-variance SDT account predicts higher confidence for morph targets and distractors than for parent targets, and as shown in Table 1, this model cannot account for the confidence/recognition inversion seen in Experiments 1a and 1b.

The unequal variance model has fewer constraints and as a result might account for the confidence/recognition dissociation. We fit the unequal variance version by fixing the filler distractor distribution variance at 1.0 and allowing the other variances to vary freely. As shown in Table 1, this model does a much better job fitting the data, and can account for the confidence/recognition dissociation as shown in the Unequal Variance section. The bottom panel of Figure 4 shows the signal detection distributions for parents and morphs, and demonstrates how this model fits the confidence/recognition dissociation: morph target and distractor distributions are still shifted (slightly) to the right of the parent target distribution, but the morph distributions have smaller variances than parent targets, which causes their confidence values to cluster in the lower confidence values. Thus a smaller variance can put more area to the right of the decision criterion (and therefore produce a high probability of saying 'old'), but still produce lower confidence values when saying 'old'.

The fact that the unequal variance version of signal detection theory can account for the confidence/recognition dissociation does not necessarily provide an explanation for the results. SDT is usually mute with regard to the value of the variances, with the exception that target distribution variances typically increase as the mean increases. Our
data violate that relation; morph target and distractor distributions are higher than either parent distribution and the morphs have the smallest variances. However, there may be other reasons for the relatively small variances for morph distributions. Although the morphs were created from the parents, the morph distributions may be more homogenous. Busey (1998) demonstrated with a similar set of faces that morphs tend to cluster near the center of face space because the average of any two points produces a point that tends to fall near the center of the space. If the morphs tend to cluster together in the center of this space, they might produce a more regular set of values on the strength axis. This would lead to a distribution with a smaller variance and makes for a logical explanation of the confidence/recognition dissociation that we will have to test.

To summarize, the confidence/recognition dissociation seen in Experiments 1a and 1b may simply reflect the fact that the morph faces are more homogenous and therefore produce a distribution along the familiarity axis that has a smaller variance. As a result, this distribution can be to the right of the parent distribution at the level of the means, but

| Table 1. Data from Experiment 1 and predictions from the Equal Variance and Unequal Variance versions of SDT. |
|---|---|---|
| Experiment 1A | Data | P("Old") | Conf When "Old" |
| Morph Distractors | 0.714 | 64.0 |
| Parent Targets | 0.658 | 69.3 |
| Signal Detection Theory- Equal Variance Predictions | P("Old") | Conf When "Old" |
| Morph Distractors | 0.686 | 69.5 |
| Parent Targets | 0.658 | 68.3 |
| Signal Detection Theory- Unequal Variance Predictions | P("Old") | Conf When "Old" |
| Morph Distractors | 0.704 | 65.0 |
| Parent Targets | 0.655 | 69.8 |
| Experiment 1B | Data | P("Old") | Conf When "Old" |
| Morph Distractors | 0.767 | 66.8 |
| Parent Targets | 0.682 | 70.0 |
| Signal Detection Theory- Equal Variance Predictions | P("Old") | Conf When "Old" |
| Morph Distractors | 0.738 | 71.5 |
| Parent Targets | 0.683 | 69.0 |
| Signal Detection Theory- Unequal Variance Predictions | P("Old") | Conf When "Old" |
| Morph Distractors | 0.759 | 67.3 |
| Parent Targets | 0.678 | 70.5 |
have fewer high-confidence responses. This accounts for the confidence/recognition inversion seen with parents and morphs. This hypothesis is tested (and ruled out) in Experiments 2 and 3, which also serve to demonstrate aspects of the recognition process that produce the confidence/recognition dissociation.

Although the unequal variance SDT account of the confidence/recognition inversion is successful, this model remains mute concerning the selective influence that the speeded task had on the morph distractors (see Figure 3). The increase in both probability of saying 'old' and confidence can easily be accounted for by sliding the morph distractor distribution to the right while leaving the other three distributions relatively fixed. However, nothing in signal detection theory would predict why we would expect only the morph distractors to show such a shift. This leaves open the possibility for an additional explanation such as a familiarity/recollection distinction, which might explain why morph distractors are selectively affected by the speeded response requirement.

Experiment 2

To rule out the unequal variance version of signal detection theory as the only explanation of our confidence/recognition dissociation, we turn to a forced-choice design. In the signal detection theory account of forced choice data, the strength value is computed as difference between target and distractor distributions. Assuming no bias between left and right test items on the screen in a two-alternative forced-choice task (and we expect and find none with our simultaneous presentations), this subtraction eliminates any difference in variance between parent and morph distributions. Thus if we continue to find confidence/recognition dissociations we can disconfirm even an unequal-variance version of signal detection theory.

To directly compare parents and morphs in Experiment 2, we study only parent and filler faces and then test one of the parents against its morph (Figure 5 illustrates this
experimental design). This strength of this design is that it addresses another limitation of Experiment 1a and 1b, specifically, that at test we included both parents and morphs. One might argue that in Experiments 1a and 1b the high false alarm rate seen with the morph distractors results from the fact that participants mistake the morph for the parent. The forced-choice design eliminates this difficulty by testing the morph against one of its parents (the subject is never tested on the other parent). The subjects are told (truthfully) that each test trial contains exactly one previously studied face. Thus on a parent/morph trial there is no likelihood that subjects will confuse a morph with the parent it is tested against. This also allows side-by-side comparison between a parent and the morph, thus providing additional information to the participants such as which features are most diagnostic for discriminating between faces, and the range of possible confusions, which helps participants adjust their confidence scales.

Recall that in Experiments 1a and 1b, the morph false alarm rate was approximately equal to (in fact, slightly higher than) the parent hit rate. Based on this finding, in forced choice we expect that morphs and parents will be chosen approximately equally often, despite the fact that morphs are never studied\(^1\). To address the adequacy of a signal detection theory account, we look at the distribution of confidence judgments. A dual process account suggests different levels of confidence associated with the two processes (Yonelinas, 2001). A recognition judgment may result from contributions of both a familiarity-based process and a recollective process. When viewing a studied parent,

\(^{1}\) Based on the Experiment 1 recognition data we might predict that observers choose the morph more often than the parent. While this situation can occur (see Busey, 2001), for our purposes it is sufficient that they are simply equal. The increase in the diagnostic information provided by the 2AFC paradigm could lead to the morphs and parents being chosen approximately equally often.
participants may recall specific features from memory, and we expect they will be more confident than when they are globally matching information on the basis of familiarity. However, on other trials such as with the morphs, a high output from a familiarity process could compensate for a low recollection output in recognition. This could lead to a situation where participants are equally likely to choose the parent and the morph (one on the basis of high recollection, the other on the basis of high familiarity), but confidence could be higher for the parent because of its larger contribution from recollection processes.

Method

The experimental conditions were similar to Experiments 1a and 1b, with the following exceptions. At test we paired each filler face with a randomly chosen filler distractor. For parent targets, we tested only one parent (counterbalanced in design) and paired it with the associated morph distractor at test. We presented the two faces side-by-side on the computer monitor in a two alternative forced-choice paradigm, and combined the confidence scale with the recognition response.

Figure 5. Experiment 2 design.
Participants

Participants were 259 undergraduates attending Indiana University fulfilling a course requirement. The experiment was completed in small groups with an average of 4 people participating at a time in a total of 50 groups.

Design and Procedure

This experiment differs from Experiment 1 in that a two-alternative forced-choice paradigm was used. Of particular interest, parents were always studied and tested against their associated morphs (distractors) at test.

The study session procedure was the same as in the first experiment. However, during the study session, only the 16 parent faces and 14 filler faces were presented for a total of 30 target faces. None of the morph faces were studied, and instead served as distractors at test. Participants were asked to study each target face and remember the faces for a subsequent recognition test.

Immediately following the study phase, participants were tested using a two-alternative forced-choice format. Participants were asked to choose one of two faces presented that was previously studied. Test pairs were constructed so that participants either chose between filler targets and filler distractors, or between a target parent and its associated distractor morph. There were a total of 22 trials in the test phase: 8 parent/morph pairs and 14 target/distractor pairs were presented in random order.

Instead of rating confidence after each recognition response, recognition and confidence judgments were taken simultaneously in the form of a single key press. The numbers 1, 4, and 7, which are vertically located on the left side of the keypad, corresponded to choosing the left test face. Confidence was measured by the magnitude of the number chosen; 1= low confidence, 4= more confident and 7=very confident. Similarly, the numbers 3, 6, and 9 on the vertical right side of the keypad signified the
selection of the test face to the right, and confidence corresponded to the number selected; 3 = low confidence, 6 = more confident and 9 = very confident. These were converted to values of 1, 2 or 3 for purposes of analyses, with 3 associated with the very confident condition.

Results and Discussion

Of central interest is the probability of choosing the parent face over the morph. Averaged across all subjects and all 8 morph-parent triplets, the probability that the subjects choose the parent over the morph is 0.53, which is not significantly different from the chance value of 0.5. ($t(31)<1$, $p=0.77$). Thus we conclude that subjects in our experiment were equally likely to choose the parent or the morph. However, for confidence we find a difference between morphs and parents. When choosing the parents, subjects gave larger values of confidence. The mean confidence judgment for parents was 2.09, and for morphs was 1.84. This difference is significant using a two-tailed test ($t(31) = 4.24$, $p<0.001$).

Figure 6 shows the scatter plot for the parents and morphs, demonstrating that although the subjects were equally likely to choose the parent or the morph, parent faces are given higher confidence ratings, regardless of the choosing rates. This dominance extends over the entire range of overlap, as summarized by the linear regression fits of the two data. Thus it is not the case that this effect is carried by a few morphs or a few parents.
SDT fails to account for Experiment 2

The standard version of the signal detection model, even an unequal variance version, cannot account for the effects of Experiment 2. The top row of Figure 7 shows typical old/new and forced choice representations in signal detection theory. The top-left panel shows the distractor and target distributions on the familiarity axis for old/new recognition, with target items distributed along higher values (since they were actually studied and will therefore seem more familiar or have a better match to memory). As shown in the top-right panel of Figure 7, the signal detection model of forced-choice data has a decision axis that is the difference between the strength values of the two images on the screen. Assume for the moment that the target item (the parent) is shown on the right side of the screen. In forced choice, the designation of hits and false alarms is arbitrary, and we will define our decision variable \( W \) as the difference between the strength of the right face minus the strength of the left face:

\[
W = \text{Strength}_{\text{Right Face}} - \text{Strength}_{\text{Left Face}}
\]

which will tend to be positive if the target face appears on the right, and negative if it appears on the left. The top-right panel of Figure 7 shows the two distributions of target-on-right and target-on-left trial types. Note that assuming no left/right bias (and we found none), the two distributions are symmetric, equally distant from zero, and have equal variance despite the unequal variances of the original familiarity distributions of targets and distractors. Thus under an equal-bias situation, the subject will say 'right' if the \( W \) is greater than zero, and 'left' if \( W \) is less than zero. Confidence is typically viewed as finer gradations along the decision axis, and thus is also based on the difference of familiarity values. For example, in the upper-right panel of Figure 7, if the target is shown on the right, the subject will only rarely say 'left', and almost never give a high confidence
Figure 7. **Top Row:** Traditional signal detection representation of targets and distractors places the target distribution to the right of the distractors, and targets typically have a larger variance (top-left panel). When this experiment is repeated as a 2AFC design, the variance differences subtract away, leaving a difference distribution for trials with the target on the right and target on the left (top-right panel). **Bottom Row:** From Experiment 1, we expect that the morph distractors will have a similar mean but smaller variance (lower-left panel). This could lead to the situation illustrated in the bottom-right graph, with the target-on-left and target-on-right distributions falling on top of each other. Note that in this situation, the model predicts confidence will be identical for both parents and morphs when they are chosen.

response when they do make this error.
The bottom row of Figure 7 illustrates the situation for parents and morphs in Experiment 2. Based on the results of Experiment 1, we expect that parents and morphs will have distributions with means that are approximately equal, although the morph distribution will have a smaller variance. In 2AFC, this translates into two distributions with identical means and variances, centered on zero. This leads to the situation where subjects are equally likely to choose the parent or the morph in forced choice, as was found in Experiment 2. However, the signal detection model makes a very strong prediction for confidence in this situation: For parents and morphs, confidence should be identical when each is chosen. This can be seen most clearly in Panel D. Under an equal-bias situation, the confidence criterion are placed symmetrically around zero. In addition, the two curves will be on top of each other, which is required to account for the fact that parents and morphs are chosen equally often. This produces a situation in which the distribution of confidence values is equal for parents and morphs because of the symmetry of the model. Thus the model predicts that confidence will be equal when the subject chooses the parent or chooses the morph. However, confidence was significantly higher when subjects chose the parent, thus ruling out signal detection theory as the only explanation of our confidence/recognition dissociation, even an unequal-variance version².

² Note that a version of signal detection theory that included non-symmetric (i.e. skewed) distributions might be able to account for the dissociation between confidence and recognition in Experiment 2. However, this model would lack an explanation for why the distributions should be skewed, and barring evidence of skewed distributions we reject SDT as an account of the Experiment 2 data.
Having ruled out the unequal variance version of signal detection theory, below we consider several alternative explanations for the confidence/recognition dissociations. Table 2 summarizes these accounts, along with the experiment that ultimately disconfirms each account as the only explanation of the confidence/recognition dissociations.

### Global Density Properties

One possible explanation for the confidence/recognition dissociation could be related to the location of the morphs in psychological space. Morphs tend to be very similar to other faces, and thus appear more typical than the parent faces. Participants may understand the structure of the 'face space' when they are being tested and may be
intuitively aware that more confusions may occur in dense regions of highly similar and
typical faces. Participants might infer that they should be cautious when a test face comes
from a dense region, and conversely believe they will be more accurate when the face is
from a sparse area. Thus while they are more likely to say 'old' to a morph than to a
parent based on the morphs high similarity to many faces in memory, they might realize
that the morph is in a dense region and reduce their confidence accordingly.

**Surface Features of Morphs**

Another explanation for the confidence/recognition inversion may reside in the morph
face itself. Due to the nature of the morphing process, certain artifacts may be present
that could influence subject confidence. Busey (1998) found that morph facial features
and wrinkles are less distinct due to blending techniques that average out the differences
between the two parents. The smoothness of the morph may somehow attract participants
to respond 'old' to this face, but the absence of distinctive features in the morph may
make participants reluctant to use a high confidence response (Zaki & Nosofsky, 2001).
This variable will be investigated in Experiment 3.

Two additional explanations will be considered in later studies. The first is a sampling
model, which can account for the Experiment 2 finding that subjects were more confident
when choosing parents than when choosing morphs, despite the fact that they choose the
two equally often. To anticipate our results, we will rule this out as the only account in
Experiment 4. The second account of our confidence/recognition dissociation to be
discussed later is based on a dual-process account in which familiarity and recollection-
based mechanisms affect confidence in different ways. As discussed previously, a
familiarity-based process may provide a sense of acquaintance with the test face. An
additional recollection-based process that accesses a particular trace in memory may also
contribute. If the recollection mechanism is successful in finding a match, the resulting
confidence is high. When familiarity drives a response, however, the result is lower confidence. We will return to this idea in the General Discussion after ruling out both a sampling model and a familiarity-based model as the sole explanations of the confidence/recognition dissociation.

**Experiment 3**

While Experiment 2 successfully ruled out an unequal variance model, it did not control for other possible explanations of the confidence/recognition dissociation. Morph stimuli have certain peculiarities inherent in their design. The benefit of using morphs in these studies is that these photorealistic blends introduce known relations between target and distractor stimuli. This creates a distractor stimulus that might produce a strong response in a familiarity-based process. However, these same benefits could become detrimental if they are not systematically controlled. As discussed in Table 2, the confidence/recognition dissociation might be an artifact of the morphs’ lack of surface features or their location in a dense region of face space. To control for the morphs’ characteristics, we modified our experimental design in Experiment 3 to include morphs at both study and test. Figure 8 illustrates our design. Half of the morphs were studied targets, while the other half had their parents studied and became distractors at test in a forced-choice paradigm. This manipulation pits the studied morph against the unstudied morph at test. This controls for global density, which

![Figure 8.](image)
is provided by filler faces and parents of other morphs. On average the studied and unstudied morphs will have the same overall global density, and thus this account cannot be used to explain confidence and recognition differences between studied and unstudied faces. In addition, both the studied and unstudied morphs will share the smooth surface features common to morphed images, and so this explanation also cannot be used to explain a confidence/recognition dissociation.

Experiment 3 is similar in design to Experiment 2, and although the unstudied morph is not studied, its parents are, and thus we expect the choosing rates to be approximately equal. This results from the fact that the studied morph is very similar to one item in memory (itself), while the unstudied morph is moderately similar to two items in memory (its parents). As with Experiment 2, this will provide an additional test of the unequal variance signal detection model. The test will again measure confidence, and the critical question is whether confidence will be higher for the studied morph than for the unstudied morph.

Method

Participants

A total of 502 participants from Indiana University completed this experiment for partial fulfillment of course requirements. Each subject provided half as many replications of the critical conditions as in Experiment 2, and thus we need more observers that we typically use in our experiments. Participants were run in small groups of approximately 4-6 people, with a total of 90 groups.

Procedure

The procedure was the same as Experiment 2 with the exception that we compared morphs against each other in a two-alternative forced-choice paradigm. One morph (the
target) was shown at study; the other morph (the distractor) was not studied, but its parents were and thus we refer to these two faces as studied and unstudied morphs.

Half of the morphs were presented at study, in addition to the parents and filler faces. The other half of the morphs were not studied and served as distractors at test. Neither parent of a target morph was studied. At test, a studied morph was compared against an unstudied morph in a forced-choice test. There were a total of 18 test trials in the forced-choice recognition test: 4 morph/morph pairs, and 14 target/distractor pairs were presented in random order. Morphs were paired prior to all data collection and thus the two morphs in a pair alternated as target and distractor. The left/right presentation of studied and unstudied morphs was counterbalanced across participants.

Results and Discussion

The results mirror those of Experiment 2. The observers were equally likely to choose the studied or unstudied morphs: The probability of choosing the studied morph over the unstudied morph is 0.4995. This is clearly not significantly different from .5, and even if it were, it would be in the wrong direction since a number less than 0.5 implies that participants on average would choose the unstudied morph over the studied morph. However, when choosing the studied morph, observers were more confident than when choosing an unstudied morph: The average confidence rating when the observers chose a
studied morph is 2.11, and when they chose the unstudied morph was 1.99. As in Experiment 2, this difference is significantly different by a 2-tailed t-test ($t(31) = 2.38; p=0.023$). Thus we find a further dissociation of confidence and recognition, and in the process attribute this dissociation to memorial processes, because we have controlled for visual artifacts associated with the morphing process.

As with Experiment 2, we need to rule out the possibility that this confidence/recognition dissociation is due to just a few faces. Figure 9 shows the scatter plot of the studied and unstudied morphs. Each point represents the choosing rate of a particular morph, graphed against its confidence when it is chosen. There are 16 points of each type because we graph left and right presentations separately. The probability of choosing a particular face is balanced across the graph, but studied morphs are consistently given higher confidence ratings when they are chosen. This occurs even for parents that are chosen only rarely, and thus this effect occurs over the entire range of data, as shown by the linear regression lines fit to the studied and unstudied morph data.

The results of Experiment 3 not only again rule out an unequal variance signal detection model, but also the surface features and global density issues discussed in Table

![Figure 10. Sampling model explanation of Experiment 3 results. The studied morph is more similar to one item in memory (itself) while the unstudied morph is less similar to 2 items in memory (its parents).]
2. Similar to Experiment 2, our participants were equally likely to choose the studied morph or the unstudied morph, but were more confident when choosing a studied morph.

**Sampling Model Account of Experiments 2 and 3**

The confidence/recognition dissociations seen in Experiments 2 and 3 are inconsistent with a signal-detection account, including variants in which the distributions of morphs and parents have different variances. This would seemingly contradict any single-dimensional model in which confidence and recognition are both based on the same information. However, there is an alternative model that captures the spirit of a single-dimensional model, yet still might account for the Experiment 3 data. In addition, this model has elements of what might be viewed as a recollective process, and evidence in favor of the sampling model account could be taken as evidence in favor of a recollective process involved in the recognition process of our novel faces.

Previously, Busey & Tunnicliff (1999) argued for the existence of a sampling model for face recognition, in which the test face is used to sample faces in memory. Both recognition (the probability of saying 'old') and confidence are related to the similarity between the test and sampled faces. If the two are more similar, the observer is more likely to say 'old' and will be more confident when doing so. The sampling process is related to the similarity between the test face and other faces such that similar faces were more likely to be sampled; thus the name SimSample. Although the SimSample model applied the sampling process to faces, many of the model assumptions are based on the Search of Associative Memory (SAM) model (Gillund & Shiffrin, 1984) and the Generalized Context Model (GCM; Nosofsky, 1986). Both of these models have been extremely successful in accounting for various phenomena in their respective areas, and thus many of the assumptions that underlie SimSample have been validated in previous work.
The SimSample model accounts for a number of effects, most notably the fact that distinctive faces are more likely to be correctly recognized, and distinctive distractors correctly rejected. SimSample is a member of a class of sampling models that can be viewed as process-model instantiations of a recollective process. This results from the fact that the final decision is based on the sampling and recovery of a single item from memory, which is compared against the test item.

The sampling model might account for the findings of Experiment 3 according to the following logic, as illustrated in Figure 10. The studied and unstudied morphs were chosen equally often, which implies that in designing the stimuli we were able to adjust the similarity of the parents so that having two nearby parents for the unstudied morph was equivalent to having actually studied the morph in the memory experiment. The sampling model accounts for this by having one very similar item to sample for the studied morph, and two somewhat similar faces (the parents) to sample for the unstudied morph. This leads to equal choosing rates in our particular design. However, the sampling model assumes that confidence is based only on the similarity between the sampled face and the test face. Other models such as a summed similarity model would predict that confidence is determined by the summed similarity to all faces in memory, rather than just to the sampled face. This would lead to a close correspondence between confidence and recognition rates, and could not account for the dissociations seen in Experiments 1-3.

Because the sampling model assumes confidence to be a function of only the similarity between the test and sampled faces, this predicts lower confidence for the unstudied morphs in Experiment 3. This results from the fact that unstudied morphs have a lower similarity to their parents than studied morphs have to themselves. Thus the studied morph will have higher confidence when chosen, which is due to the greater
similarity to its own image in memory. A similar line of logic applies to the dissociation in Experiment 2.

The sampling model is not a single-dimensional model in the strict sense, because different sources of information affect recognition and confidence. However, there is only one underlying representation (the ‘face space’). The choosing rates are affected by both the number of nearby faces and their similarity, while confidence is a function only of the similarity between the sampled face and the test face. This leads to a prediction of a confidence/recognition dissociation by the model.

**Testing the Sampling Model**

The sampling model can be directly tested using an extension of the paradigm used in Experiments 1a and 1b. In these experiments, both parents of each morph were studied, and the morph was tested as a distractor. Suppose that instead of studying both parents, only one of the two was studied. Most models, including a sampling model, will predict an increase in false alarm rates to the morph when both parents are studied compared with studying only one parent. This results from the possibility of sampling the second parent. Models that rely on summed similarity for confidence predict an increase in confidence as well when the second parent is added. However, there is a quantitative difference in terms of how much confidence will increase for the two models. A summed similarity model will predict a fairly large increase in confidence, on the order of that added by the first parent. However, the sampling model predicts only a very small increase in confidence when a second parent is added to the study list, which is due to a statistical effect described below. Intuitively, this small increase results from the fact that mean confidence *when a parent is sampled* does not change when the second parent is added, while the false alarm rate will increase due to the likelihood of sampling a second parent. The right panel of Figure 10 illustrates this logic, where the length of the arrows
indicates the degree of similarity between the test face (the morph) and the sampled face (either parent). The two-parent condition surrounds the morph with more faces, but the average similarity between the morph and parents doesn't change as the second parent is added to the study list.

Note that for purposes of exposition we have ignored the contributions of faces other than the two parents. These filler faces do contribute, and will tend to be sampled less when the second parent is added. These filler faces will tend to be associated with lower confidence levels, and thus adding the second parent with its higher confidence values will result in a small increase in confidence. Our experimental design and modeling includes the contributions of the filler faces, and we will use our experimental conditions to predict how much confidence should increase when the second parent is added to the study list. If we find that confidence increases more than predicted, we will reject the sampling model as the only explanation of the confidence/recognition dissociation. This will lead to the suggestion that both recollective and familiarity-based processes contribute to the recognition of our novel faces.

**Experiment 4**

In Experiment 4 we return to an old/new recognition paradigm and vary the number of parents shown at study for each morph, as illustrated in Figure 11. To test the sampling model we will have to introduce an extra condition that estimates the contributions of the filler faces through a multinomial version of the sampling model, as described in a subsequent section. Below we describe the logic behind the design.

When parent faces are studied and the associated morph is tested, the parent faces are very likely to be sampled. However, filler faces might also be sampled, and their contribution to confidence must also be considered. The filler faces tend to be less similar to the morph than its parents, and as a result when filler faces are sampled and result in an
'old' response they tend to produce a lower confidence response. Thus, in addition to the two- and one-parent conditions, we also include a no-parents condition in which neither parent is studied, which allows us to estimate both the baseline false alarm rate for the morphs, as well as confidence when observers say 'old'.

In the condition in which both parent faces are studied, mean confidence when observers say 'old' to the morph is a mixture of contributions from sampling filler faces, sampling Parent A, and sampling Parent B. When only one parent is studied, observers are more likely to sample filler faces (since the model assumes they sample exactly one face). These filler faces tend to have lower confidence associated with them, and as a result we might expect confidence to increase in the Both Parents condition due to the fact that other faces are not being sampled as often. This fact could allow the sampling model to produce an increase in confidence in the Both Parents condition over the One Parent condition. To solve this problem we will use a quantitative model to estimate how
much confidence should increase from the one parent to the two parent condition if a sampling model is correct.

This leads to a critical prediction we test in Experiment 4: If confidence increases in the Both Parents condition more than the sampling model says it should, we reject the sampling model as the only account of the confidence/recognition dissociation.

Method

Participants

A total of 238 participants from Indiana University completed this experiment for partial fulfillment of course requirements. Participants were run in small groups of approximately 4-6 people, with a total of 50 groups.

Procedure

Experiment 4 is identical to Experiments 1a and 1b, with two exceptions. First, no morphs were in the study lists. Second, the number of parents of each morph placed in the study list was systematically varied. Thus we measured false alarm rates for the morphs for four conditions: 1) no parents studied, 2) parent A studied, 3) parent B studied, 4) both parents studied.

For the purposes of the modeling below, we will refer to faces that are neither morphs nor parents as filler faces with the abbreviation $f$.

Results

Data for the essential conditions are found in Table 3, which shows the probability of saying 'old' for the morphs and the mean confidence when observers said 'old'. We found no significant difference between the Parent A and Parent B conditions ($t(15)<1.0$). We would not expect a difference given that the two parent faces were randomly assigned to be an A or B face, and because we found no differences we collapse the data into a single
One-Parent condition in Table 3. However, we keep these two conditions separate for purposes of model testing.

As more parents were studied, observers were more likely to say 'old', which is consistent with both a familiarity-based model and a sampling model. More importantly for the predictions of the sampling model, confidence increases when the observer says 'old' as the second parent is added to the study list (bold numbers in Table 3). The sampling model may not be able to account for the relatively large increase in confidence seen between the One- and Both-Parents conditions. However, to test this we need to estimate the contributions of filler faces to the sampling process using a quantitative model, as described below.

**Sampling Model Predictions**

The model we adopt is a multinomial version of the sampling model, and relies on the following assumptions. First, exactly one item is sampled from memory. This sampling process is assumed to be done in parallel and relies on processes that are not available to conscious recollective processes. The contents of memory are assumed to be activated.

<table>
<thead>
<tr>
<th>Number of Parents in the Study List</th>
<th>Data</th>
<th>Theory</th>
<th>Data</th>
<th>Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Parents</td>
<td>0.412</td>
<td>0.407</td>
<td>58.2</td>
<td>57.8</td>
</tr>
<tr>
<td>One Parent</td>
<td>0.577</td>
<td>0.591</td>
<td><strong>62.9</strong></td>
<td>64.8</td>
</tr>
<tr>
<td>Both Parents</td>
<td>0.683</td>
<td>0.659</td>
<td><strong>70.1</strong></td>
<td>66.8</td>
</tr>
</tbody>
</table>
simultaneously, but short-term memory does not have the capacity to hold all faces for comparison with the test image. Instead, an individual is sampled and brought into short term memory and compared to the test face. The similarity between the sampled face and the test face determines both the probability of saying 'old' and the confidence associated with the old/new judgment. Thus the sampling model assumes that the information that underlies the confidence response comes from a single face, which makes it similar to other recollective processes (Yonelinas, 2001). We assume that the sampling process occurs only once, rather than the repeated cycle in the original SAM model. In previous work the multi-sampling version was attempted without improvement to the fits (Busey & Tunnicliff, 1999).

As more parent faces are added at study for a given morph, other faces are less likely to be sampled. One mechanism that captures this property is a ratio rule that assumes that the different classes of faces (filler faces and parents) have strengths associated with them, and the likelihood of sampling a particular face is given by its relative strength. Thus, the probability of saying 'old' to a morph if neither parent is studied is given by:

\[ p(\text{old} \mid \emptyset) = \frac{S_f}{S_f + O_f} = O_f \]

where \( S_f \) is the strength associated with the filler faces in terms of the sampling process, and \( O_f \) is the probability of saying 'old' given one of the Filler faces was sampled. Panel A of Figure 12 illustrates this single path.

---

3 An extension to this model formulation might include a separate path for a guessing response. For example, if none of the filler faces were sampled the subject might be forced to guess, in which case they would have some probability of saying 'old', and some confidence distribution when saying 'old'. We have not included this guessing component into our modeling. For tasks such as cued recall it is reasonable to
If Parent A is added to the study list, observers have the opportunity to sample either a Filler face or the Parent face, as illustrated by the two paths in Panel B of Figure 12. Because exactly one face is sampled in the sampling model, these are mutually exclusive and exhaustive events. Once a face has been sampled, the observer says 'old' with some probability $O_f$ or $O_a$. Thus the probability of saying 'old' when Parent A is studied is:

$$p(\text{old} \mid A) = \frac{S_f}{S_f + S_a} O_f + \frac{S_a}{S_f + S_a} O_a$$

where $S_a$ is the strength associated with Parent A, and $O_a$ is the probability of saying 'old' when Parent A is sampled. The first half of this equation gives the probability of sampling a filler face, multiplied by the probability of saying ‘old’ when a filler face is sampled. The second half is the probability of sampling parent A, multiplied by the probability of saying ‘old’ when parent A is sampled. The equation for Parent B follows directly:

$$p(\text{old} \mid B) = \frac{S_f}{S_f + S_b} O_f + \frac{S_b}{S_f + S_b} O_b$$

The probability of an 'old' respond in the Both Parents condition includes contributions from Filler faces, Parent A and Parent B:

$$p(\text{old} \mid AB) = \frac{S_f}{S_f + S_a + S_b} O_f + \frac{S_a}{S_f + S_a + S_b} O_a + \frac{S_b}{S_f + S_a + S_b} O_b$$

assume that sampling failures occur. However, the mechanism in SimSample uses similarity and the Luce Choice Rule as the bases for the sampling. We used relatively homogenous faces in our experiments, and it is likely that this high degree of similarity will produce at least one face as a result of the sampling process. This is especially true for the morphs, who have an extremely high false alarm rate of .412 in the no-parents condition, suggesting that similar faces were available for sampling and were often used for the basis of a false alarm. Previous work with these faces (Busey, 2001) has demonstrated that the morphs are extremely similar to many faces in the set, making sampling failures unlikely.
The three parts of Equation 4 come from the three parts of the tree in panel C of Figure 12, where the probability of going down each branch of the tree is given by 

\[ p(\text{old} \mid AB) = \frac{S_x}{S_f + S_a + S_b}, \quad x \in \{f, a, b\}. \]

Once a particular face has been sampled, the observer will say old with some probability \( O_f, O_a \) or \( O_b \).
The predictions from the sampling model for confidence pertain to confidence when the observer says 'old', and are therefore conditional confidence values. These are derived from a confidence parameter \( C_x, x \in \{a, b\} \) which gives the mean confidence \textit{given} the observer gave an 'old' response to either a filler face, Parent A or Parent B\(^4\). With neither parent presented at study, the only way the observer can respond 'old' according to the sampling model is by sampling a filler face. Thus filler faces provide the only contribution to the probability of saying old and therefore the expectation of mean confidence for the no-parent condition is reduced to the free parameter \( C_f \):

\[
E(c \mid old, \emptyset) = \frac{S_f}{S_f + S_b} O_f C_f = C_f
\]

The expectation for confidence in the Parent A condition is the weighted sum of the expectations of confidence when the observer says 'old' to either filler faces or Parent A. This can be seen in Panel B of Figure 12, where the probability of saying 'old' on the basis of sampling a filler face or Parent A give the weightings in the weighted sum:

\[^4\] The model presented in the text predicts mean confidence for a particular condition, rather than the entire distribution of confidence values as would be represented by an ROC curve. However, we have performed the identical modeling using predictions for the entire confidence distribution, including confidence when say 'new', and found equivalent results. The model gives an increase in confidence from 1 parent to two parents of 64.8 to 66.6, which is much smaller than that observed in the data (\(t(15) = 2.29, p<0.05\)). Thus for simplicity we present predictions only for mean confidence.
\[
E(c \mid \text{old}, A) = \sum \frac{S_f}{S_f + S_a} O_f C_f + \sum \frac{S_a}{S_f + S_a} O_a C_a
\]

where, for example, \(\frac{S_f}{S_f + S_f} O_f\) gives the probability of saying 'old' on the basis of having sampled a filler face.

The equation for the expectation of confidence for the condition in which Parent B is studied follows directly:

\[
E(c \mid \text{old}, B) = \sum \frac{S_f}{S_f + S_b} O_f C_f + \sum \frac{S_b}{S_f + S_b} O_b C_b
\]

When both parents are studied (along with the filler faces, of course), the confidence equation requires three terms to correspond to the three paths seen in Panel C of Figure 12. Each part gives the probability of making an 'old' response on the basis of a particular type of face (filler, Parent A or Parent B), which in turn gives the weightings for the weighted sum that gives a prediction for the overall confidence when both parents are studied:
\[ E(c \mid \text{old}, AB) = \]

\[
\frac{S_f}{S_f + S_a + S_b} O_f + \frac{S_a}{S_f + S_a + S_b} O_a + \frac{S_b}{S_f + S_a + S_b} O_b + C_f + 8
\]

\[
\frac{S_f}{S_f + S_a + S_b} O_f + \frac{S_a}{S_f + S_a + S_b} O_a + \frac{S_b}{S_f + S_a + S_b} O_b + C_a + 8
\]

\[
\frac{S_f}{S_f + S_a + S_b} O_f + \frac{S_a}{S_f + S_a + S_b} O_a + \frac{S_b}{S_f + S_a + S_b} O_b + C_b
\]

**Summary of Predictions**

Equations 1-4 provide predictions from the sampling model for the probability of saying 'old' when various parents are studied. Equations 5-8 provide predictions for mean confidence when the observer says 'old'. These predictions are derived by fitting 5 free parameters: \( S_a, O_f, O_a, C_f \) and \( C_a \) fit to 8 datapoints. The parameters \( S_f, S_a \) and \( S_b \) are ratio strengths, and so we fixed \( S_f \) at 1.0 to establish the scale. As stated above, no significant differences were found between those faces in the Parent A condition and those in the Parent B condition (\( t < 1 \)), and none was expected, given the random assignment of faces to categories. This provides a justification to set \( S_b = S_a, O_b = O_a \), and \( C_b = C_a \). Without this simplification we would have as many free parameters as data points. However, we can relax these restrictions if we fit the entire ROC curve rather than mean confidence (see Footnote 4), and when we did this we obtained essential the same results as those reported below.

Data from 8 conditions was obtained for all 8 morph pairs, which consists of the mean probability of saying 'old' in the 4 conditions and mean confidence for the four
conditions. The values of the free parameters were obtained using the Solver minimization procedure in Microsoft Excel, which has been shown to provide good minimizations in recognition memory modeling (Dodson, Prinzmetal, & Shimamura, 1998). The predictions were compared to the data via minimization of the root mean squared error, although maximum likelihood techniques were used to fit the entire ROC function in a separate simulation that gave equivalent results. The predicted values were then averaged and summarized in Table 3.

The predictions reveal a critical deficiency in the model: the sampling model cannot predict the large increase in confidence seen when the second parent is added to the study list. Across the morphs in the one parent and two parent conditions, the model cannot account for the spread between the one- and both-parent conditions. The model predicts a 2% increase in confidence due to the reduced sampling of filler faces in the both-parent condition, and we observe a 7.2% increase. To test whether the observed increase is significantly higher than the predicted increase, we fit the model to each morph separately, fitting 5 parameters to 8 datapoints. We then entered the data from the individual morphs into a t-test and found that the 7.2% increase is indeed larger than the 2% increase predicted by the sampling model (two tailed t-test, t(15) = 2.16; p<0.05).

This large increase in confidence disconfirms the sampling model. Thus we reject the sampling model as the only account of our confidence/recognition dissociations in Experiment 1-3.

**Adding noise to the sampling process**

Despite this disconfirmation, there is a plausible extension of the sampling model that might in principle account for the increased confidence seen in the Both parent condition. Suppose that noise is added to the sampling process that has the effect of making one parent more similar to the morph that it usually appears. In the face-space approach, this
is equivalent to moving the parent closer to the morph in psychological space. If this were to occur, this parent would be more likely to be sampled, and if so, would tend to produce a higher confidence if the observer says 'old'. These conclusions result from the fact that both the probability of saying 'old' and confidence when an 'old' response is made are determined by the similarity between the test face and the sampled face.

This mechanism could increase the confidence in the Both parent condition more than the original version of the sampling model says it should using the following logic. In the one-parent condition, only one parent could potentially benefit from being moved closer to the morph. However, with both parents in memory, either parent face might be moved closer to the morph, and whichever is moved closer will be the more likely one to be sampled (and therefore produce a higher confidence rating than a no-noise model would predict). Thus the Both parents condition produces a situation with two chances to produce a higher confidence value, while the One parent condition has only one chance. This mechanism therefore selectively boosts confidence in the Both parent condition over the no-noise version of the sampling model, which is exactly what would be required to account for the empirical data of Experiment 4. Thus the addition of noise has the potential to save the sampling model.

This version of the sampling model is difficult to disconfirm quantitatively, because it requires assumptions about how much noise is in the distances in psychological space. However, this model can be at least tentatively disconfirmed on qualitative grounds. Consider the effects of adding noise to parents that are either similar to each other (and therefore both similar to the morph), versus parents that are dissimilar. In the latter case, the parents are further away from the morph in psychological space. Distance is typically converted to similarity according to an exponentially-decreasing function (e.g. Thurstone, 1927; Shepard, 1962a, 1962b, 1974), which implies that noise added to locations of
dissimilar parents will have only a marginal, if any, effect on the sampling process. The sampling process is much more likely to benefit from noise for those parents that are similar to each other (and therefore similar to their morph). Thus we expect the no-noise version of the sampling model to fail for similar morphs, and fit best for morphs with dissimilar parents.

To summarize the prediction of adding noise to the location in ‘face space’, we expect the no-noise version of the sampling model to fail for morphs with similar parents and fit better for morphs with dissimilar parents, since adding noise has little effect for far-away faces in face-space. The relevant comparisons are illustrated in Figure 13, which plots the goodness of the fit for the 8 morphs against the similarity between the parents. The similarity between the parents was obtained in a large-scale similarity rating experiment in which all possible pairs of faces were rated for similarity. The similarity values were converted to z-scores, which accounts for the negative values on the abscissa of Figure 13.

Inspection of Figure 13 illustrates that the model is performing poorest for the parents that are dissimilar, while the noise mechanism will provide the most benefit for morphs whose parents are similar. That is, adding noise to the model would provide help for those faces where it is needed the least. We expect a negative relation and we find a slight (but non-significant) positive relation. There is nothing in Figure 13 that suggests that
adding noise will assist the sampling model in regions of the face space where it is performing poorly. Thus we return to our original conclusion from Experiment 4: the sampling model simply cannot account for the increases in confidence for the morph distractors when the second parent is added to the study list.

**Variability of Strength Parameters Across Conditions**

The strength parameters $S_f, S_a$ and $S_b$ are estimates of the influence that each type of face has on the sampling process initiated by the presentation of a morph distractor at test. We use the different experimental conditions to estimate these influences. It is important to point out that we assume that the $S_f$ from equation 1 is the same value as in later equations. There is a subtle issue underlying this assumption that merits discussion. All observers were tested with the same set of 60 faces, but some of the 30 studied faces varied across observers according to the conditions assigned to each of the 8 morphs. For example, an observer might not study either parent for morph 1, but study both parents for morph 2, and so on. When estimating $S_f$ for morph 1, the parent faces for the other morphs are considered filler faces since they are not related to morph 1. However, this implies that the set of filler faces will change slightly from condition to condition for a particular morph. A cleaner design would have held constant the set of filler faces (including parents of other morphs) and vary just the number of parents studied for one morph. Unfortunately this logic leads to a huge explosion of conditions: we would have had to replicate the experiment 8 times, and we would no longer have a within-subjects design for the number-of-parents-studied manipulation.

As a compromise we chose two morphs for each of the four parent conditions (no parents studied, parent A studied, parent B studied, or parents A and B studied) for the first subject, and then rotated these conditions through the morphs in a counterbalanced schedule for different subjects. Thus across all observers the set of filler faces for a
particular morph included on average 7 parents from other morphs chosen randomly according to the conditions assigned to the 7 other morphs. These seven parents represent a small minority of the 30 faces in the study list and may not bias the sampling process in a systematic way. However, to test the possibility that \( S_f \) might change across conditions, we fit a version of the model for each morph that included a separate value of \( S_f \) for the both-parents condition. Across the 8 morphs, the estimated values for \( S_f \) for the both-parents condition were not significantly different from the values of \( S_f \) from the no-parent, parent A and parent B conditions (\( t(7)=1.71, p=.13 \)) and most of the estimated values were very close to 1.0 (the fixed value of \( S_f \)). Thus we have no evidence that \( S_f \) changes systematically as more parents are studied, and justifies our assumption that the set of filler faces is essentially the same across parent conditions for a given morph.

**General Discussion**

The sampling model fails a critical test: when the second parent is added to the study list, confidence increases more than the sampling model predicts it should. This disconfirms the sampling model as the only mechanism at work in the recognition memory paradigm, and suggests that the sampling model by itself is insufficient to explain the confidence/recognition dissociations seen in the forced-choice paradigms of Experiments 2 and 3. This disconfirms the fourth candidate model in Table 2 (the sampling model) that might explain the confidence/recognition dissociations.

Having disconfirmed candidate models 1-4 in Table 2, what explanation would be consistent with all aspects of the data from Experiment 4? We would like to suggest that some combination of recollective and familiarity-based processes could provide a unified account. This combination has been suggested by a number of researchers with regard to faces, most notably Yonelinas (2001) and Vokey & Read (1992). Related work with verbal materials also supports this distinction, including the work of Jacoby (Jacoby,
1991; Jones & Jacoby, 2001), Rotello (Rotello & Heit, 2000; Rotello, Macmillan, & Reeder, submitted), Hintzman & Curran (1994), as well as most recently with ERP data (Curran, 2000). Care must be taken when adopting the two-process model, since much of this work has been controversial (for example, Ratcliff, Van Zandt, & McKoon (1995) demonstrated that a single process model (SAM) was identified by process dissociation to have two processes). However, we have disconfirmed a number of plausible single- and multi-dimensional models as the only account of our confidence/recognition dissociation, and at this point offer the following multiple-process account as one possible candidate.

Our primary distinction between familiarity and recollection processes depends upon what information is used to make a particular decision. In a familiarity-based model, evidence is accumulated across items to produce a single feeling of familiarity. Recollection considers evidence from multiple items in memory, but then selects one item for retrieval and comparison with the target. The sampling model closely follows this account. The clearest evidence against the familiarity-based model as a complete account of recognition and confidence judgments comes from the data of Experiments 2 and 3. Familiarity-based models cannot produce equal choosing rates for targets and distractors, yet have lower confidence when distractors are chosen. The sampling model account suggests that subjects recall an item (usually one quite similar to the test item, because the sampling depends on the similarity between each item and the test item) and then uses properties of that recalled item to make recognition and confidence judgements. However, this cannot be the entire story, because the sampling model cannot account for the Experiment 4 data. Since both models fit different aspects of the data, yet either alone is insufficient to account for the entire dataset, this leaves us with the possibility that both processes are active concurrently. Thus we might term this the Sampling plus Global Familiarity account.
Comparison with Other Models

The above account is not inconsistent with existing dual-process and two-process models of memory. It does, however, rule out even an unequal variance version of signal detection theory and the sampling model, both of which could be viewed as single-process models.

Work by Banks (2000) has proposed a two-dimensional model to account for source judgements and item recognition. This work is similar to research by Jacoby using process dissociation and remember/familiar judgements. Both accounts argue that memory is multi-dimensional, although they differ on the nature of the two processes. The present account differs from the other paradigms in that source monitory paradigms require subjects to make an additional judgement about context (i.e. which list did this item come from). Thus while these models have argued for the existence of two distinct processes, the present work demonstrates that two processes are at work in traditional recognition memory experiments and provides evidence for the nature of the two processes.

Yonelinas (2001) has proposed a model in which recollection is viewed as a high-threshold process: an item is either recalled in its entirety, or it is not. This model differs from the sampling model, because in the sampling model the recognition process (the probability of saying 'old') is a continuous function of the similarity between the sampled item and the test item. Thus the sampling model does not predict linear ROC functions.

Perhaps the most well-developed model is the two-process model of Rotello et al (submitted). This model includes global strength and specific strength components which are stated to correspond to familiarity and recollection processes. However, process model elements are not included in the instantiation of the model. The advantage of the Rotello model, however, is that is very specific about how the two strengths are
combined. The recognition process sums the output of the two processes, while the remember/know judgements are made on the basis of the difference between the two strengths. Both the Rotello model and the present work acknowledge that observer responses are not process-pure. That is, remember and know judgements are unlikely to cleanly dissociate explicit and implicit memory.

Unlike the Rotello et al (submitted) model, the present account does not consider the method by which the outputs of the two processes combine to produce recognition and confidence judgments. Instead, we have chosen to provide evidence for a two memory system by showing that either candidate process by itself is insufficient to account for the entire dataset. We do not have to consider the potential consequences of correlations between the two systems (see Curran & Hintzman (1997) for more details).

One way to consider the combination rule (as well as the impact of correlations) is to conduct quantitative model fits that use multi-dimensional scaling spaces derived from similarity ratings between all pairs of faces. This approach was adopted by Busey & Tunnicliff (1999) for familiarity-based models and sampling models separately, and has both strengths and weaknesses. If two models make qualitatively similar predictions but different only quantitatively, then this is the only approach. However, the results of these quantitative fits relies on the goodness of the multidimensional scaling solution. We have collected such data but have chosen to first analyze our data at the qualitative level before proceeding with quantitative fits. The two types of models could be fit with an assumed combination rule such as a weighted sum. The number of parameters does increase, but the number of datapoints is very large, since hit and false alarm rates (as well as confidence) are predicted for each face. The complexity of this approach leaves it outside the scope of the present work.
Additional Evidence for Familiarity and Recollection

Evidence from a number of domains has converged to suggest separate processes are at work in recognition memory. Recent electrophysiological work by Curran (2000) and Rugg et al (1998) have independently shown cases in which familiarity and recollection-like processes can be dissociated on the basis of EEG/ERP data. Rugg et al (1998) examined the ERP traces for hits, misses and correct rejections. In a frontal electrode in a time interval of approximately 300-600 ms after test stimulus onset, he found that the traces for misses and correct rejections where indistinguishable, but hits produced a larger deviation from baseline. He took this as evidence that the ERPs in this region reflected the decision the subjects would eventually make. However, during the same time interval but in a parietal electrode, the traces for hits and misses were indistinguishable, but differed from the correct rejections. Rugg took this as evidence that the parietal electrode knew about the status of the misses (i.e. that they had been seen previously), and associated this pattern of responses with a familiarity mechanism. The frontal response appeared to correspond with a more controlled or conscious process, not unlike recollection.

A similar conclusion was drawn by Curran using a different paradigm. He presented singular and plural nouns at study, and then used the alternative form at test for some items (i.e. if coats is studied, the similar distractor would be coat). He found a similar dissociation, such that a frontal site distinguished between similar and new items but grouped studied and similar items together. This is consistent with a familiarity mechanism that is not overruled by a recollection mechanism that recalled the original form of the noun. In a similar time interval a parietal site distinguished between studied and unstudied items (i.e. similar and new items). This was attributed to a recollection mechanism that overrules the familiarity engendered by the similar distractor. This
dissociation is qualitatively similar to the Rugg conclusion, but the sites attributed to the different processes are opposite. This may reflect differences in the recording techniques and/or methodologies.

Evidence from speeded responses suggests that the familiarity process has an earlier onset. Work by Rotello & Heit (2000) and Curran (2000) supports this idea. In associative recognition, the recollection process may overrule an early familiarity response given sufficient time, a process known as recall-to-reject. We have evidence from the Experiments 1a-1b comparison that supports a similar idea in face recognition. In the state-trace plot of Figure 3, we found that only the morph distractors showed a large increase in both confidence and false alarm rate when observers made a speeded response. Thus the familiarity engendered by a single presentation of an item (or in this case of the morph's parents at study) seems sufficient to selectively affect recognition and confidence. That we don't see a similar large shift for morph targets suggests that familiarity plays less of a role with these stimuli, which, after all, could be recollected since they were studied.

The present work is consistent with a recent connectionist model proposed by Norman and O'Reilly (submitted), which includes both recollection and familiarity-based mechanisms. Aggleton & Brown (1999) have proposed a similar model. In the current approach we have argued for the existence of the two mechanisms and described the nature of their contributions for recognition and confidence. The Norman & O'Reilly approach also describes the nature of the interactions between the two memory systems. In particular, the hippocampus may serve to serve to separate similar patterns to avoid blending. This may suggest that the responses we get from our morph stimuli may be primarily due to contributions from a familiarity-based system. A fruitful research line may be to conjoin the connectionist model with a similarity-based face space to test some
of the model predictions of target-distractor similarity. This may indicate whether the model can correctly recognize distinctive faces while still producing high false alarm rates to morph distractors. The model has not been extensively applied to confidence ratings beyond standard ROC analyses, and so it is less clear how the model would account for data from Experiments 2 and 3. One possibility would be to suggest that the output of the hippocampal system is associated with higher confidence values when it drives the recognition process.

**Applications to Eyewitness Testimony**

One advantage of the present approach over prior work is that we are working with a paradigm that is not that different from traditional lineup experiments and therefore could potentially have an impact on the legal setting. In principle the studied/unstudied morph paradigm could easily be adapted to a lineup setting (with less experimental power, since fewer stimuli are involved). Anecdotal evidence from classroom demonstrations suggests that fewer faces are required, and it may be possible to demonstrate that both familiarity and recollection (as exemplified by the sampling model) are at work in eyewitness testimony.

Should this be the case, it would be important to delineate the conditions under which a familiarity mechanism is at work, because even though it is associated with low confidence, it can lead to very high errors with the morph distractors (some as high as 70%). This is clearly an important line of future work for those working in eyewitness identification.

**Evidence for Prototypes and Blending in Memory**

Somewhat orthogonal to the question of separate familiarity and recollection processes is the question of the fidelity of the memory representation upon which the two processes act. Previously, Busey & Tunnicliff (1999) used a similar set of faces and
design to provide evidence for the Sampling model (termed SimSample). However, they could not rule out the possibility of additional blending between exemplars in memory to create prototypes that generated very high false alarm rates to the morphs. More recent work by Zaki & Nosofsky (2001) suggested that some of these unaccounted-for effects were due to the properties of the morphs. They introduced a clever experiment in which they told subjects that they were part of a subliminal experiment and were told to watch for faces embedded in visual noise. In fact, no faces were studied. A 'recognition' test was then administered, and the authors found that observers gave more 'old' ratings to the morphs than to the parents. They argued that these elevated ratings were due to visual artifacts such as smoothing or symmetry in the morphs. Other explanations for their data are possible; for example, the morphs tend to be more typical than the parents, and observers may be influenced by typicality when giving recognition judgments. However, their point is still valid, and they went on to describe other properties of prototype stimuli that might affect performance in recognition and categorization experiments, and yet remain outside the consideration of the models. These properties include symmetry, compactness, and other properties of stimuli that lie in the central tendency of a set.

The designs of the present Experiments 3 and 4 control for the effects of morphs. Experiment 3 supports the sampling model, but does not address the blending issue. The design of Experiment 4 does allow for the testing of a blending hypothesis, using the existing models in Figure 12 and equations 1-8. Note first that the sampling model does not include a blending mechanism. However, if a blending mechanism is at work, then we can find evidence for it by examining the false alarm rate in the two-parent condition, and compare this to what is predicted by the sampling model based on the false alarm rates for the one-parent conditions. If blending does occur, this will produce more false alarms than the sampling model predicts. We computed the obtained and predicted false alarm rates using the same model fits used to fit the confidence data. The predicted false
alarm rate is 0.689 and the obtained false alarm rate is 0.683. The predicted value is actually higher than the obtained value, and thus we conclude that we have no evidence for blending in the present paradigm.

One difference between the present paradigm and the previous work is the choice of stimuli. The original work used a larger set of the bald faces, included men with facial hair and ethnic minorities. The present set consisted entirely of Caucasian clean-shaven men. This more homogeneous set may have altered the behavior of subjects, allowing them to be more discriminative in areas with more faces. This may have worked against a blending mechanism. However, this is speculation, and at this point the simplest explanation is to accept the conclusions drawn by Zaki & Nosofsky (2001) and suggest that much if not all of the prototype effects seen with the morphs (and with other prototype stimuli such as dot patterns) is a result of extra-model properties such as smoothing, symmetry or compactness. Researchers who use these stimuli should consider controlling for potential artifacts using methodologies similar to those in Experiments 3 and 4.

Summary and Conclusions

In four experiments we have shown several different kinds of recognition/confidence dissociations, and used the results to rule out a number of different models as the only explanation of the dataset. These include unequal-variance signal detection, visual artifacts, typicality (global density) and separate familiarity and recollection-based models. We conclude that a cogent account of the entire dataset requires a model that includes elements of both a familiarity-based process and a recollective process.

This supports a view in which participants view a test face in a memory paradigm, and use it to activate the contents of memory in parallel to obtain a global familiarity measure. In addition, the participant may also sample individual items in memory, bring
them into working memory for comparison with the test stimulus, and use the results of this comparison process to refine their recognition and confidence judgments.
References


