# Different Confidence-Accuracy Relationships For Feature-Based And Familiarity-Based Memories

Mark Tippens Reinitz
University of Puget Sound

William J. Peria
University of Washington

Julie Anne Séguin
Victoria University of Wellington

And

Geoffrey R. Loftus
University of Washington

Send correspondence to: Mark Tippens Reinitz
University of Puget Sound
1500 North Warner
Tacoma, WA, 98416
mreinitz@pugetsound.edu

* Supported by NIMH Grant MH41637 to Geoffrey Loftus. Authors thank Greg Bigelow, Krysta Brai Bouchard, Grace Chen, Katrina Elizabeth Jones, Trystan Larey-Williams, Jennifer Mao, Andrew Seliber, and Bryan Muzzall Wells for their help with data collection. John Dunn, Timothy Perfect, and two anonymous reviewers provided helpful comments on earlier versions of the manuscript.

## Abstract

Participants studied naturalistic pictures presented for varying brief durations and then received a recognition test on which they indicated whether each picture was old or new and rated their confidence. In one experiment they indicated whether each “old”/ “new” response was based on memory for a specific feature in the picture or instead on the picture’s “general familiarity,” while in another experiment, we defined pictures that tended to elicit “feature” versus “familiarity” responses; thus feature/familiarity was a dependent variable in one experiment and an independent variable in the other. In both experiments feature-based responses were more accurate than familiarity-based ones, and confidence and accuracy increased with duration for both response types. However, when confidence is controlled for, mean accuracy was higher for familiarity-based than for feature-based responses. The theoretical implication is that confidence and accuracy arise from different underlying information. The applied implication is that confidence differences should not be taken as implying accuracy differences when the phenomenal basis of the memory reports differ.

**Different Confidence-Accuracy Relationships for Feature-Based and Familiarity-Based Memories**

As confidence in recognition responses increases, accuracy also tends to increase (e.g., Brewer & Wells, 2006; Sauer, Brewer, Zweck, & Weber, 2009). However, the relationship between confidence and accuracy is far from perfect. Two classes of theories have recently been proposed to explain the underlying factors that give rise to confidence and accuracy in recognition experiments; we refer to these as *unidimensional models* and *multidimensional models*. By unidimensional models both confidence and accuracy are based on the same underlying dimension (memory strength). By multidimensional models confidence and accuracy arise at least in part from different underlying informational components.

Picture-memory experiments have demonstrated that recognition responses are sometimes based on memory for a *specific feature* and sometimes on *general familiarity*; for instance, a naturalistic scene may be identified because some item in the picture is remembered or simply because the picture seems familiar. Loftus and Bell (1975) and Loftus and Kallman (1977) showed that across exposure durations and generalized over various circumstances, participants were more accurate when their responses were based on memory for features rather than on familiarity. However, Loftus and his colleagues did not solicit confidence responses. Both feature-based and familiarity-based responses are likely to vary in their associated confidence. If this variation is large then there is likely overlap in the confidence ranges associated with the two response types. Consider two recognition responses where confidence is equal, but one is feature-based and the other is familiarity-based. Their relative accuracy will be determined by the specific functions relating confidence and accuracy for the two response types, which are currently unknown. Our goal is to map out these confidence-accuracy relationships as a basis for testing between unidimensional and multidimensional models.

## Tests of confidence-accuracy relationships

Most recent investigations of confidence-accuracy relationships have used between-subjects designs in which participants viewed simulated crimes and the recognition tests involved choosing a suspect from a photo lineup. Some of these studies have found only weak relationships between confidence and accuracy (e.g., Bothwell, Deffenbacher, & Brigham, 1987; Sporer, Penrod, Read, & Cutler, 1995); this has led many researchers to propose that confidence and accuracy arise from independent underlying causes that are differentially influenced by various factors; i.e., that multidimensional models are needed to describe confidence-accuracy relationships.

Brewer and Wells (2006) demonstrated that when analyses only include those participants who actually chose an individual from a lineup confidence tends to increase with accuracy. Additionally, when manipulations such as retention interval are used to increase variability of confidence and accuracy, stronger confidence-accuracy relationships are found (e.g., Sauer et al., 2009). These positive relationships have led some authors to propose that confidence and accuracy both result from decision processes that utilize the same underlying information in memory. Some of these models are based in signal detection theory and others take the form of accumulator models (for a discussion see Sauer, Brewer, & Weber, 2008, pp.528-529). Many of these are unidimensional models. For instance, consider the simplest application of signal detection theory to the confidence-accuracy relationship. Here accuracy is determined by the distance between the target and distracter distributions, and confidence is determined by the distance between the response criterion and the target distribution. In this case both confidence and accuracy are derived from the same underlying information (memory strength) and so the model is unidimensional.

It is clear that confidence and accuracy are not *always* based in identical information. For instance, Busey, Tunnicliff, Loftus, and Loftus (2000) showed that the effects of rehearsal and study-item luminance on accuracy and confidence in a recognition experiment were well accounted for by a unidimensional model. However, increasing test-item luminance increased confidence but not accuracy in a way that could only be accounted for by a multidimensional model. The focus of the research reported here is not to test whether confidence and accuracy must always be related by a unidimensional model, but instead whether a unidimensional model is sufficient to account for the effects of feature recognition on confidence and accuracy.

## Models and Predictions

We presented naturalistic pictures for varying brief durations to our participants who then received a recognition test in which studied pictures were randomly intermingled with an equal number of distracters. Participants indicated whether each picture was old or new, and rated their confidence on a 4-point scale. As described later, in one experiment we manipulated and in another we observed which test responses were feature-based and which were familiarity-based. This allowed us to separately measure confidence and accuracy as functions of exposure duration for feature-based and familiarity-based responses.

### *Unidimensional and Multidimensional Models*

Figure 1 shows a unidimensional model (left panel) and a multidimensional model (right panel) as they apply to our experiments. By the unidimensional model, both exposure duration (d) and whether or not the picture contains a remembered feature (r) are assumed to influence a single underlying dimension of memory, called strength (S). Strength, in turn, determines both confidence and accuracy in that both are assumed to be monotonic functions, $m_P$, and $m_S$, of S. The key prediction of the unidimensional model is this: *any two combinations of d and r (e.g., a short-duration feature picture and a longer-duration* *familiarity picture) that produce equal levels of accuracy must, because they imply equal strength, produce equal confidence as well.*

***Figure 1 here***

By a multidimensional model confidence and accuracy are not based on identical underlying information. The right panel of Figure 1 shows one simple multidimensional model adapted from Busey et al. (2000). This model includes two underlying dimensions. As in the unidimensional model, duration (d) and the presence of a memorable feature (r) jointly determine the value of S. However, the presence of a memorable feature increases the value of a second dimension called certainty (denoted as T). We conceptualize certainty as being based on internal assumptions regarding factors relevant to accuracy. For instance, if a participant assumes that remembering a specific feature is associated with greater accuracy, then feature recollection increases T. Although a participant's assumptions about factors relevant to accuracy can influence the participant's confidence, they do not in and of themselves influence accuracy. Here accuracy is a monotonic function of S, but confidence is a function of both S and T. As a result, combinations of d and r that lead to equal accuracy will not necessarily lead to equal confidence.

To summarize, unidimensional models predict that any combination of d and r which produce equal accuracy will also produce equal confidence. The simple multidimensional model we described predicts that d-r combinations that produce equal accuracy will not produce equal confidence; rather, confidence will be higher for feature-based compared to familiarity-based recognition responses because of the increase in certainty that results from remembering a feature.

***Specific Predictions***

The predictions of the unidimensional and multidimensional models for our experiments are shown in the left and right panels of Figure 2, respectively. These predictions were generated using simple and specific if somewhat arbitrary choices for the functions shown in Figure 2 relating S to d and r, and relating confidence and accuracy to S and T; for additional discussion see Busey et al. (2000), pp. 29-32.

The top and middle panels show accuracy and confidence, respectively, as functions of exposure duration. As with all other graphs in this article, up-facing arrows indicate feature-based responses, and down-facing arrows indicate familiarity-based responses. As expected by any reasonable model, both accuracy and confidence increase with d. In keeping with previous findings, we expect that for all durations accuracy will be greater for feature-based than for familiarity-based responses. The accuracy advantage for feature-based responses implies generally higher strength relative to familiarity-based responses. Because confidence is a function of strength this leads to the prediction that for all durations, confidence, like accuracy, will be higher for feature-based than for familiarity-based responses.

***Figure 2 here***

The bottom panels show the critical predictions of the two models. Here data from the top and middle panels are combined to produce confidence-accuracy scatterplots showing accuracy as a function of confidence for each of the 12 response type-duration conditions. It is evident that the unidimensional model predicts a perfect rank-order correlation over the 12 conditions; that is, the curves for the feature-based and familiarity-based responses overlap perfectly across their shared range because all points across the two response types that produce identical accuracy must also produce identical confidence. Bamber (1979) details the logic of these *state-trace plots* that are shown in the two bottom panels of Figure 2.

The prediction of the multidimensional model is that the curves for feature-based and familiarity-based responses will be separated. This is because accuracy is a function only of strength, while certainty contributes to confidence and is a function of whether or not the stimulus contains a memorable feature. Consequently, conditions that produce equal accuracy across the response types will not produce equal confidence because for any accuracy level certainty will increase confidence for feature-based responses compared to familiarity-based responses.

## Manipulating the Basis of Recognition Responses

To test between the critical predictions of the models it was necessary to separate familiarity-based responses from feature-based ones. One way to do this is to have participants indicate the basis of each of their responses during the test. However, this entails treating response type as a dependent variable. This is difficult to reconcile with the models in Figure 1 where both duration and the presence of a feature are treated as independent variables that contribute to memory strength. For this reason it was desirable to directly manipulate whether responses were feature-based or familiarity-based, so that we could ensure an equal number of each response type, employ appropriate counterbalancing, and provide a good match between our method and the models that motivated it.

Accordingly we created a stimulus set comprising equal numbers of *feature pictures* and *familiarity pictures*. In a pilot experiment 108 participants were each presented with 204 study pictures. In a subsequent recognition test these were randomly intermingled with an additional 204 distracters and participants indicated whether each “old”/“new” response was based on memory for a specific feature in the picture or instead on a feeling of familiarity. Across participants all 408 pictures were used equally often as study pictures and distracters. Some pictures received primarily feature responses and others tended to receive familiarity responses. From the complete set of 408 stimuli used in the pilot experiment we assigned those 102 pictures with the greatest proportion of familiarity responses to be familiarity pictures, and those 102 with the greatest proportion of feature responses to be feature pictures. More specifically, we examined the rate at which participants reported utilizing a feature for recognition of a stimulus, referred to herein as the *naming rate*. We took as *feature pictures* those 102 stimuli which had a naming rate in the top half at both the shortest (17 ms) and longest (533 ms) exposure durations. *Familiarity pictures* were those with a naming rate in the bottom half for both shortest and longest durations[1]. The finding that this procedure categorized exactly half the stimulus set as either feature-type or familiarity-type is a coincidence, though a convenient one. Figure 3 shows examples of feature (left panels) and familiarity (right panels) pictures; not surprisingly, feature pictures but not familiarity pictures tended to contain obvious memorable features.

***Figure 3 here***

In Experiment 1 our analysis of feature-based and familiarity-based recognition performance involved comparing these two picture sets. To confirm that our picture sets were correlated with participants' phenomenal experiences during the test, Experiment 2 participants indicated whether each of their responses was based on familiarity or memory for a specific feature, and we performed two tests between the models: one using self-report data to separate feature-based from familiarity-based responses, and the other by comparing between the picture sets.

## Experiment 1

In Experiment 1 we used a picture-recognition paradigm. The two independent variables were study-picture exposure duration and picture type (familiarity/feature). Participants indicated whether each test picture was old or new and rated their confidence on a 4-point scale.

### Method

#### *Participants*

One hundred, twenty University of Washington undergraduate students participated for credit in their various Psychology classes. They were run in 12 groups of 6-16 participants per group.

#### *Stimulus and Apparatus*

The stimuli were 192 naturalistic color photographs depicting landscapes, cityscapes, and seascapes. Of these, 96 had been demonstrated in the pilot experiment to usually be recognized on the basis of familiarity (familiarity pictures) and the other 96 were usually identified on the basis of feature recognition (feature pictures). Pictures were presented at low contrast to avoid ceiling effects. Pictures were presented on a white wall in the front of the testing room using an LCD projector. A Windows-based computer using the MATLAB Psychophysics toolbox (Brainard, 1997) controlled stimulus

[1] The **average (SD)**, lowest, and highest naming rates were **0.83 (0.17)**, 0.64, and 1, respectively, for feature pictures and **0.40 (0.22)**, 0.10, and 0.61 for familiarity pictures. Note that naming rates are averages across the shortest and longest exposure durations; for all pictures naming rates were lower for brief durations than for longer durations.

presentation and timing. Data were collected using 8 numeric keypads that allowed running of up to 8 participants at once.

### *Design and Procedure*

Participants received eight blocks of trials, each consisting of a study phase followed by a test phase. In the study phase six familiarity pictures and six feature pictures, combined with the 6 durations, were presented in random order. Exposure durations for feature pictures were: 17, 33, 67, 133, 267, or 533 ms. For familiarity pictures the 17 ms duration was omitted and a 1067 exposure duration was added to promote similar accuracy ranges across the conditions. At test the 12 study pictures were randomly intermingled with 6 new familiarity pictures and six new feature pictures. Using the keyboards participants first indicated whether each picture was old or new, and then indicated their confidence on a four-point scale where 0 indicated guessing and 3 indicated that they felt sure. Each test picture remained until all participants made both responses.

Across the 12 separate groups of participants each picture was presented equally often at each relevant exposure duration and was used equally often as a target and a distracter.

## Results and Discussion

The mean (standard deviations in parentheses) false alarm rates were .036 (0.019) and .067 (0.019) for the feature and familiarity conditions respectively. Figure 4 shows our main results. Accuracy and confidence have been corrected for false alarms[2]. Note that because the groups differed in size we used groups, rather than participants, as the unit of analysis.

Both accuracy (top panel) and confidence (middle panel) increase with exposure duration and are greater for feature compared to familiarity pictures. As previously discussed these results do not obviously distinguish between a unidimensional and a multidimensional model. Indeed, because the graphs fail to show a classic dissociation between confidence and accuracy, one might be tempted to conclude from them that both confidence and accuracy are influenced identically by our independent variables. Such a conclusion would be incorrect: the bottom panel shows the confidence-accuracy scatterplot for each of our picture-duration combinations, and as can be seen, the functions for the two picture types are separated. This disconfirms the unidimensional model which predicts that conditions which produce equal accuracy must also produce equal confidence. Instead, the results show that confidence and accuracy result from different sources of information.

***Figure 4 here***

[2] The corrected accuracy measure is $(H_{ij}-FA_j)/(1-FA_j)$ where $H_{ij}$ and $FA_j$ are hit-and false alarm rates corresponding to study duration i and j indexes feature picture/familiarity pic. The corrected confidence measure is $C(H_{ij})-C(FA_j)$ where $C(H_{ij})$ and $C(FA_j)$ indicate mean confidence associated with hits and false alarms in each condition.

Note in the bottom panel of Figure 4 that for any arbitrary accuracy level confidence is higher for feature than for familiarity pictures. This suggests that remembering a feature produces an increase in confidence. This is precisely the effect that is predicted by the multidimensional model of the sort shown in Figure 1 where feature recollection increases certainty. Because certainty influences confidence but not accuracy, feature recollection leads to increased confidence.

In summary, Experiment 1 indicates that confidence and accuracy arise from different underlying sources of information for feature-based and familiarity-based memories. While it is true that in general feature-based responses are more accurate than familiarity-based ones, they are also accompanied by higher confidence such that when accuracy is held constant, feature-based pictures are accorded higher confidence than familiarity-based ones.

# Experiment 2

In Experiment 1 we treated response basis as an independent variable where the levels were defined by our two picture sets. For our conclusions to be compelling it is important to show that these picture sets reasonably map onto familiarity-based and feature-based responses. In Experiment 2 we replicated our finding that functions for feature-based and familiarity-based responses are separated on a confidence-accuracy scatterplot. The method was the same as in Experiment 1 except that participants indicated whether each recognition response was based on memory for a specific feature or instead on familiarity. We then separately compared the two response types when they were defined by self report on the one hand or by our picture sets on the other hand.

## Method

### *Participants*

Ninety-five University of Washington undergraduates participated in return for credit in their Psychology classes. They were run in 12 groups of 6-11 participants per group.

### *Stimulus and Apparatus*

The stimuli and apparatus were the same as in Experiment 1.

### *Design and Procedure*

The design and procedure were identical to those of Experiment 1 except that participants made three responses to each test picture. As in Experiment 1, they indicated whether the picture was old or new, and then rated their confidence on a four-point scale. Then in addition, they indicated whether their response was based on memory for a specific feature in the picture, or instead on familiarity. The specific instructions presented to the participants were:

*"Here we will ask you if your answer was based on a* S*PECIFIC* F*EATURE or* F*AMILIARITY. By this, I mean the following. Based on past research, we know that a person's decision about whether they've seen a picture or not can either be based on Feature or Familiarity. For instance, if you remember that a picture was presented before, it could be because you remember some specific feature in the picture. For example in the picture shown on the screen, you may recognize the image because you remember the yellow road sign right in the middle. On the other hand, it's possible that you know you haven't seen an image before, because you know that you would have remembered a feature. Using the same image as example, you may know that this image hasn't been shown before because you would have remembered the street sign in the middle. So whether you believe that you saw it before or you didn't, your Yes or No answer was based on a specific feature. On the other hand, your answers can sometimes be based on familiarity. You may know that you saw a picture before even if you don't remember a specific feature, simply because the picture seems familiar to you."*

Each test picture remained on the screen until each participant made all three of their responses.

## Results and Discussion

Feature and familiarity false alarm rates (with standard deviations in parentheses) respectively were .031 (0.019) and .052 (0.016) when defined by the picture sets. When defined by self report the false alarm rates (and standard deviations) were by coincidence also .031 (0.015) and .052 (0.021). Our main results are presented in Figure 5. For the left panels, response basis is defined by our picture sets and so these graphs represent a direct replication of Experiment 1. For the right panels feature-based and familiarity-based responses are defined by participants' self report[3]. As before the top and middle panels show accuracy and confidence, respectively, as functions of duration, while the bottom panels show the confidence-accuracy scatterplots.

***Figure 5 here***

There are two noteworthy findings. First, the patterns of results are essentially identical to those of Experiment 1: the scatterplots show separate functions for the two response types such that for any given accuracy level, feature-based responses yielded higher confidence than familiarity-based responses. Second, this is true regardless of whether response type is defined by our picture sets or instead by self report, which indicates that our picture sets effectively manipulated the underlying basis of recognition responses. From a practical perspective it is useful for researchers to know that one may manipulate the basis of recognition responses by creating appropriate picture sets. From a theoretical perspective it is useful to note that different pictures are reliably associated with different subsequent recollective experiences.

[3] The proportion of feature pictures that elicited feature responses was .63, and the proportion of familiarity pictures that elicited familiarity responses was .52.

Experiments 1 and 2 disconfirm unidimensional models which propose that confidence and accuracy arise from identical underlying information for both feature-based and familiarity-based memories. The data imply a multidimensional model in which feature recognition produces increased confidence.

## General Discussion

### Summary of Results

We used a within-subjects design and showed that responses based on memory for features in pictures tend to be more accurate than those based on the pictures' general familiarity. However, we also showed that this moderate accuracy gain is accompanied by a confidence difference that cannot be accounted for by a unidimensional model. We discuss the implications below.

### Implications

#### *Separate information underlies confidence and accuracy*

Our results show that confidence and accuracy are not based on the same underlying information in the cases of feature-based and familiarity-based memories, but rather that a multidimensional model is required to account for the data. Many specific multidimensional models are possible but for the present data, a multidimensional model with two underlying dimensions is sufficient. The first dimension (Strength) determines accuracy for both response types. The second dimension (Certainty) increases when a feature is encoded (and subsequently recognized), and contributes to confidence but not to accuracy. The result is that for any given level of accuracy confidence is higher for feature-based compared to familiarity-based responses.

Signal detection theory provides a popular basis for psychological theories describing confidence-accuracy relationships. By a simple signal detection account, accuracy is determined by the average difference in memory strength between the target and distracter distributions and confidence is determined by the distance between a single response criterion and the target distribution. Our results show that this single-parameter model is insufficient to account for the effects of feature recognition. A signal-detection account for our data must assume that the effects of feature recognition are twofold: in each of our experiments there is only a single distracter distribution, so average memory strength must be higher for feature pictures than for familiarity pictures (resulting in higher accuracy) and, second, the response criterion must be substantially more liberal for feature pictures than for familiarity pictures to account for the higher confidence that accompanies feature-based responses when accuracy is held constant. Note that the latter is the same as saying that feature recognition produces a confidence increase that is independent from memory strength, so this theory is a specific version of the multidimensional model we proposed in the introduction. Additionally, the results show that familiarity-based memories are not just “weak” feature-based memories. Some authors have proposed that familiarity is simply the reflection of a weak memory trace (see, e.g., Dunn, 2004). By this view memories vary across a continuum of memory strengths. Memories that fall below a criterion strength are so poorly remembered that they lack details. It is this lack of detail that causes the memory to be interpreted as resulting from familiarity. By this proposal there is no fundamental difference between feature-based and familiarity-based memories other than their underlying strength. Here familiarity responses should be accompanied by low accuracy and confidence, and feature responses should be associated by high accuracy and confidence. These predictions are disconfirmed by our experiments.

Finally, our data clearly show that people are more confident in their feature-based than in their familiarity-based decisions even when accuracy is equal, and that a multidimensional model is needed to explain the effects of feature recognition on accuracy and confidence. One possible explanation is that feature recognition boosts confidence. It is alternatively possible that failure to remember a feature deflates confidence in a way that is unrelated to accuracy. Two considerations support this possibility. First, false alarm confidence was higher in both experiments for familiarity-based than for feature-based responses. The opposite would be expected if confidence increased whenever someone thought they recognized a feature. Second, familiarity has often been proposed to be largely implicit, with little conscious access to the processes that produce it (e.g., Tulving, 1985). This lack of explicit remembering might produce doubt regarding one’s accuracy. Either case is consistent with the signal-detection explanation presented earlier in this section.

***Relation to dual-process theory***

The feature-familiarity distinction that motivates our research is similar but not identical to the common distinction between “remembering” and “knowing” that arises from dual-process theory (Tulving, 1985). We have conceived of the feature-familiarity distinction as an independent variable determined by the presence or absence of at least one memorable feature, whereas “remember”/”know” is a dependent variable intended to measure underlying recollective states. However, like feature-based responses, “remember” recognition responses are associated with both higher confidence and accuracy than “know” or familiarity-based responses (Yonelinas, 2002). Palmer, Brewer, McKinnon, and Weber (2010) recently had participants view a video showing five adults under conditions of full or divided attention; they then attempted to identify them from a photo lineup. Participants made “remember”/ “know” responses and also rated their confidence on an 11-point scale. “Remember” responses were associated with both higher confidence and accuracy than were “know” responses. However, taking “remember”/ “know” into account did not predict accuracy above confidence alone. This may imply that, as proposed by Donaldson (1996), Wixted (2007), and Dunn (2008), “remember” and “know” responses can be accommodated by a unidimensional model. However, those authors used a between-subjects design where stimuli were pictures of faces. Several authors have demonstrated that confidence-accuracy relations are often different depending on whether a within-subjects or between-subjects design is used (e.g., Perfect, Watson, & Wagstaff, 1993), and there are many demonstrations of face-picture differences in recognition. Results may be different in a between-subjects design using a state-trace methodology with naturalistic scenes as stimuli. If our feature-familiarity distinction maps directly onto the "remember"/ "know" distinction then the implication is that, given equal confidence, "know" responses should on average be more accurate than "remember" responses. It is interesting to note that this pattern was reported by Perfect, Williams, and Anderton-Brown (1995), who found that word-recognition accuracy was higher for "Know" than for "Recollect" responses at each of the three levels of confidence that they measured. Thus there have been conflicting depictions of the confidence-accuracy relationship for "Remember"/"Know" responses; additional research is needed to clarify this issue.

### Final Comments

We showed that our results generalize across two different ways of defining feature-based and familiarity-based responses: our patterns of results were identical when we compared between picture sets and when we compared across response types defined by self report. Although there is danger in applying results obtained using within-subjects designs to the courtroom we have provided evidence that large confidence differences may accompany equally accurate memory reports depending on their phenomenal basis. This raises the possibility that accurate familiarity-based reports may be discounted by jurors, since jurors tend to use confidence as a way to assess accuracy (e.g., Cutler, Penrod, & Stuve, 1998). Finally, it remains unclear whether our results generalize to other stimuli, such as faces. Many authors have proposed that, unlike scenes, faces are typically recognized on the basis of their overall configuration rather than on features (e.g., Tanaka & Farah, 1993). This leads to the possibility that feature responses may be associated with reduced accuracy for faces. This potential difference indicates the need to replicate our results using faces as stimuli; this effort is currently under way.

## References


Bamber, D. (1979). State-trace analysis: A method of testing simple theories of causation. *Journal of Mathematical Psychology, 19*(2), 137-181. doi:10.1016/0022-2496(79)90016-6

Bothwell, R. K., Deffenbacher, K. A., & Brigham, J. C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology, 72*(4), 691-695. doi:10.1037/0021-9010.72.4.691

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433-436.

Brewer, N., & Wells, G.L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, functional size, and target-absent base rates. *Journal of Experimental Psychology: Applied, 12(1),* 11-30. Doi: 10.1037/1076-898X.12.1.11

Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence–accuracy relation in recognition memory. *Psychonomic Bulletin & Review, 7*(1), 26-48. Retrieved from www.csa.com

Cutler, B. L., Penrod, S. D., & Stuve, T. E. (1988). Juror decision making in eyewitness identification cases. *Law and Human Behavior, 12*(1), 41-55. doi:10.1007/BF01064273

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24*(4), 523-533. Retrieved from www.csa.com

Dunn, J. C. (2004). Remember-know: A matter of confidence. *Psychological Review, 111*(2), 524-542. doi:10.1037/0033-295X.111.2.524

Dunn, J. C. (2008). The dimensionality of the remember-know task: A state-trace analysis. *Psychological Review, 115*(2), 426-446. doi:10.1037/0033-295X.115.2.426

Loftus, G. R., & Bell, S. M. (1975). Two types of information in picture memory. *Journal of Experimental Psychology: Human Learning and Memory, 1*(2), 103-113. doi:10.1037/0278-7393.1.2.103

Loftus, G. R., & Kallman, H. J. (1979). Encoding and use of detail information in picture recognition. *Journal of Experimental Psychology: Human Learning and Memory, 5*(3), 197-211. doi:10.1037/0278-7393.5.3.197

Palmer, M. A., Brewer, N., McKinnon, A. C., & Weber, N. (2010). Phenomenological reports diagnose accuracy of eyewitness identification decisions*.* *Acta Psychologica, 133*(2), 137-145. doi:10.1016/j.actpsy.2009.11.002

Perfect, T.J., Watson, E.L., Wagstaff, G.F. (1993). Accuracy of confidence ratings associated with general knowledge and eyewitness memory. *Journal of Applied Psychology*, 78, 144-147.

Perfect, T.J., Williams, R.B., & Anderton-Brown, C. (1995). Age differences in reported recollective experience are due to encoding effects, not response bias. *Memory*, 3(2), 169-186.

Sauer, J. D., Brewer, N., & Weber, N. (2008). Multiple confidence estimates as indices of eyewitness memory. *Journal of Experimental Psychology: General, 137*(3), 528-547. doi:10.1037/a0012712

Sauer, J. D., Brewer, N., Zweck, T.& Weber, N. (2009). The effect of retention interval on the confidence-accuracy relationship for eyewitness identification. *Law and Human Behavior,* doi: 10.1007/s10979-009-9192-x

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies*. Psychological Bulletin, 118*(3), 315-327. doi:10.1037/0033-2909.118.3.315

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 46A*(2), 225-245. Retrieved from www.csa.com

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne, 26*(1), 1-12. doi:10.1037/h0080017

Wixted, J.T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114*, 152-176. doi: 10.1037/0033-295X.114.1.152

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*(3), 441-517. doi:10.1006/jmla.2002.2864

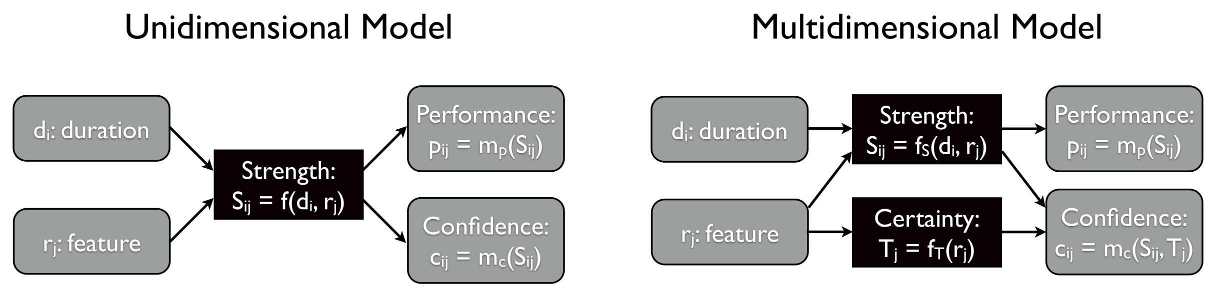


Figure 1. Unidimensional models and multidimensional models applied to the kinds of experiments under consideration. Here "feature" is depicted as an independent variable, i.e., a particular type of picture. As described by Bamber (1979) the same model also applies when "feature" is viewed as a dependent variable.

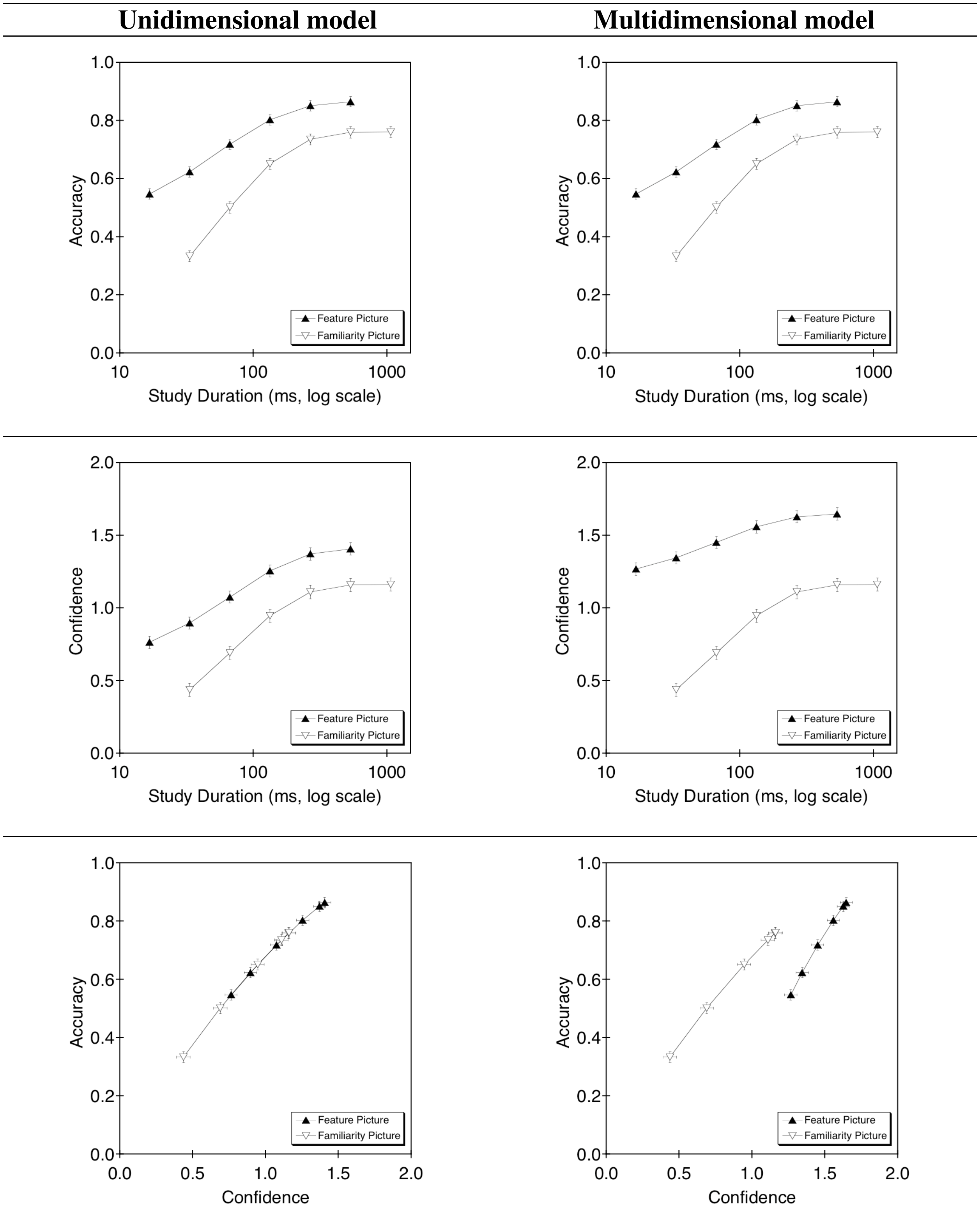


Figure 2. Predictions of a unidimensional model (left panels) and a specific multidimensional model (right panels).

| Feature Pictures | Familiarity Pictures |
|---|---|

Figure 3. Examples of feature pictures (left panels) and familiarity pictures (right panels).

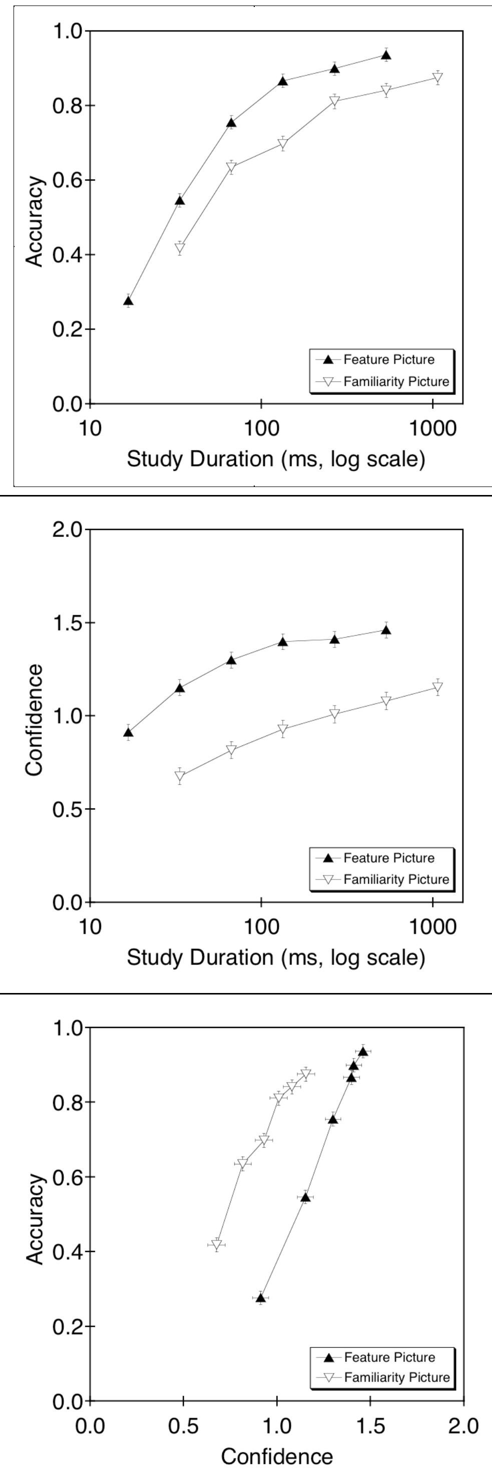


Figure 4. Experiment-1 results. The error bars show the standard errors. Note that feature/familiarity is experimenter-defined.

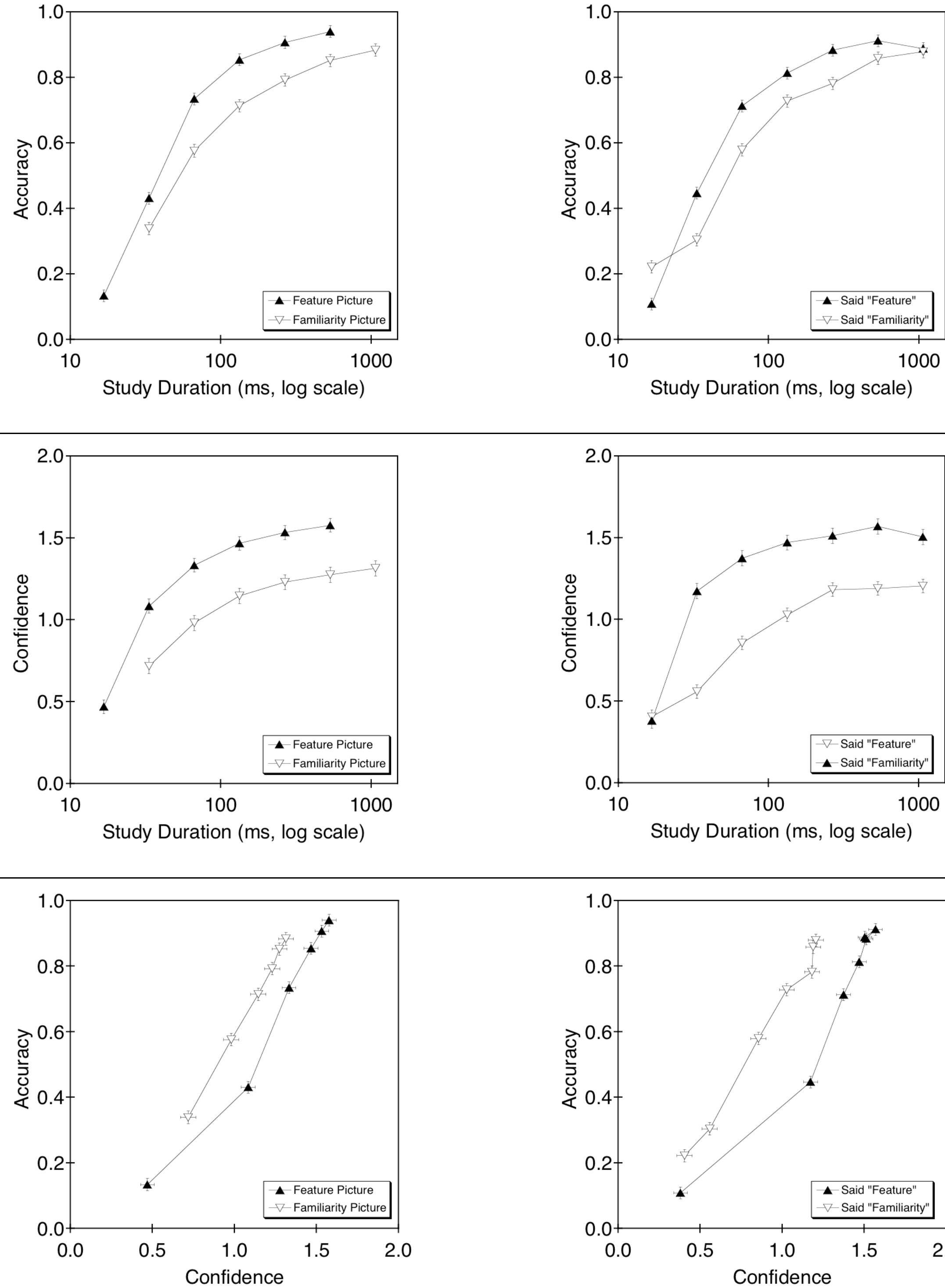


Figure 5. Experiment-2 results. The error bars show the standard errors. Feature/familiarity are experimenter-defined, thereby replicating Experiment 1 (left panels) or participant-defined (right panels).