Using Ecphoric Confidence Ratings to Discriminate Seen from Unseen Faces: The Effects of Retention Interval and Distinctiveness

James D. Sauer

University of Portsmouth

Nathan Weber & Neil Brewer

Flinders University

Author Note

This research was supported by a University of Portsmouth Department of Psychology Small Grant and an Australian Research Council Grant DP1093210. Thanks to Erika Näsholm for her assistance with data collection. Correspondence concerning this paper should be addressed to James Sauer, Department of Psychology, University of Portsmouth, King Henry Building, King Henry I street, Portsmouth, PO1 2DY. Email: James.Sauer@port.ac.uk. Ph: 44-23-9284 6330.

Abstract

Theories of confidence processing for recognition judgments suggest that confidence indexes the degree of match between a presented stimulus and an image in memory (ecphoric similarity). Recent research demonstrates that having participants rate their confidence that a face had been previously seen provided an equivalent or better index of the stimulus' status than eliciting a simple binary response (Sauer, Brewer, & Weber, 2008). Using a face recognition paradigm, we manipulated retention interval and stimulus distinctiveness to directly test the suggestion that confidence indexes ecphoric similarity, and probe boundary conditions for using confidence ratings to discriminate seen from unseen faces. Consistent with the proposed ecphoric basis for confidence ratings, mean confidence was higher for previously seen than unseen faces, and conditions conducive to the formation of strong memories improved confidence-based discrimination. In all conditions, after the application of a classification algorithm, confidence ratings provided a more sensitive index of face status (i.e., seen or unseen) than did binary responses.

Keywords: confidence, face recognition, memory, metamemory

Using Ecphoric Confidence Ratings to Discriminate Seen from Unseen Faces: The Effects of Retention Interval and Distinctiveness

The drastic consequences of eyewitness identification errors raise important questions such as is it possible to directly assess a witness' memory, while avoiding non-memorial influences that contribute to errors? Optimizing indices of stimulus-memory match, though central to eyewitness identification reliability, has not been a focus for memory researchers. Thus, the effectiveness of the memory probes used to test eyewitness recognition memory has not been systematically examined. Sauer, Brewer, and Weber (2008) suggested that confidence ratings may provide a relatively direct measure of stimulus-memory match (ecphoric similarity: Tulving, 1981). Across various paradigms and stimuli, results showed that having participants rate their confidence that a face had been previously seen (i.e., 'How confident are you that you have seen this stimulus before?') provided an equivalent or better index of the stimulus' status than eliciting a simple binary response (i.e., 'Have you seen this stimulus before?'). Here we tested whether confidence ratings index ecphory, and explored two boundary conditions for using confidence ratings to discriminate previously seen from unseen faces. We also investigated mechanisms underlying the diagnostic advantage associated with 'ecphoric confidence' ratings.

Ecphoric confidence ratings are distinct from both retrospective confidence judgments and typical confidence response scales (e.g., ratings from *sure old* to *sure new*). Retrospective confidence judgments reflect assessments of decision accuracy. Similarly, typical confidence response scales include an inherent old/new decision, with scale points representing gradations of confidence in that decision. Conversely, while ecphoric confidence ratings possibly involve an implicit decision, they do not demand an old/new decision, only an assessment of the match between a stimulus and item in memory. Avoiding explicit decisions

may attenuate non-diagnostic influences on decision criteria, offering a more sensitive index of ecphory, and resulting in improvements in discrimination. Alternatively, more fine-grained dependent variables may simply offer more sensitive measures of recognition. To investigate the effects of scale-grain-size on discrimination we calculated 'recognition ratings' (ranging from -100 [certain new] to 100 [certain old]; cf. Tenney, MacCoun, Spellman, & Hastie, 2007) by combining binary responses with retrospective confidence ratings, and compared classification performance using these ratings to performance using ecphoric confidence and binary responses.

Using a face recognition paradigm, we investigated participants' ability to use ecphoric confidence ratings to index recognition. First, we investigated how changes in retention interval and stimulus distinctiveness affected confidence ratings given to studied and unstudied stimuli and, consequently, the level of discrimination afforded by confidence ratings. Second, we examined how these manipulations affected classification performance using ecphoric confidence and binary judgments.

Recognition memory performance declines as retention interval increases (e.g., Ebbinghaus, 1964; Schacter, 1999). We examined how delay-related declines in memory quality affect ecphoric confidence. This is important for two reasons. First, if ecphoric confidence ratings are insensitive to memory quality, they will not provide a reliable index of ecphory/recognition. Studies examining retrospective confidence ratings consistently demonstrate that retrospective confidence judgments are less sensitive than memory performance itself to a variety of manipulations (e.g., Gigerenzer, Hoffrage, & Kleinboelting, 1991; Weber & Brewer, 2004). Although retrospective and ecphoric confidence are distinct, both presumably index memory and stimulus discriminability (see Macmillan & Creelman, 1991; Van Zandt, 2000; Wixted & Mickes, 2010). Importantly, this research is the first to test whether ecphoric confidence ratings track changes in memory quality. Second, increasing

delay reduces discriminability. If confidence indexes stimulus-memory match, increased delay should reduce a) confidence for old faces, b) the difference between confidence ratings for old and new faces and, consequently, c) discrimination using ecphoric confidence. Essentially, weaker memories provide an impoverished basis for comparisons supporting assessments of ecphoric similarity. Delay-induced reductions in ecphoric confidence ratings for old faces would support the theoretical claim that confidence indexes ecphory. However, low ecphoric confidence ratings may reflect newness or a paucity of evidence of oldness. If the latter influence is too extreme, the usefulness of ecphoric confidence in discriminating old from new faces will be undermined. To begin testing boundary conditions for the usefulness of confidence ratings, we compared delay-related declines in discriminability using ecphoric confidence ratings with those for a binary response comparison.

We also used distinctive and typical face stimuli. Increasing stimulus distinctiveness improves recognition memory performance (e.g., Light, Kayra-Stuart, & Hollander, 1979; Semmler & Brewer, 2006) and influences ecphoric confidence ratings (cf. Sauer, et al., 2008). Compared to typical stimuli, distinctive stimuli produce stronger and more readily accessible memory traces. This would lead to higher confidence ratings for old distinctive, compared to old typical, stimuli and lower confidence ratings for new distinctive, compared to new typical, stimuli (cf. Dodson & Schacter, 2002). Similarly, signal detection theory (Macmillan & Creelman, 1991) holds that signal strength distributions for old and new faces show less overlap for distinctive, than for typical, stimuli. We tested whether ecphoric confidence ratings tracked these changes in memory strength. While ecphoric confidence ratings may provide effective classification for distinctive stimuli (even after a delay; Shepherd, Gibling, & Ellis, 1991), combining reduced memory quality due to increased delay with reduced discriminability for typical stimuli may undermine discrimination using

ecphoric confidence ratings. We tested whether such effects on ecphoric confidence ratings exceeded those for a binary response comparison.

In sum, we addressed three questions. First, did variations in memory quality affect discrimination using ecphoric confidence ratings? We answered this question by examining the effects of our manipulations on a) mean ecphoric confidence ratings for studied (old) versus unstudied (new) faces, and b) measures of calibration and resolution. Given the proposed memorial basis for ecphoric confidence ratings, we expected delay-induced reductions in memory quality to reduce mean confidence for old faces and, consequently, discrimination. Based on previous research, we expected higher (lower) mean confidence for old (new) distinctive stimuli than for old (new) typical stimuli. Second, when memory quality was reduced, did classification performance using ecphoric confidence remain superior, or at least equivalent, to that for binary responses? We expected reduced memory quality to reduce discrimination using ecphoric confidence. However, how this would affect classification performance was unclear. We assessed classification performance using measures of discrimination (d') and criterion placement[1] (c). Finally, we compared classification performance using ecphoric confidence and recognition ratings to investigate the contribution of scale-grain-size to the benefits associated with use of ecphoric confidence ratings (cf. binary responses).

Method

Participants

Ninety-six (68 male) undergraduate students, aged 16 to 63 ($M = 26$, $SD = 11$), participated.

Design

We used a 3 (retention interval: immediate test, 1 week delay, 2 weeks delay) × 2 (response type: binary response, confidence) × 2 (distinctiveness: distinctive, typical) × 2

(face status: old, new) mixed design. Participants were randomly allocated to one of the six cells created by crossing retention interval with response type. Face status and distinctiveness were varied within-subjects. Participants viewed equal numbers of distinctive and typical faces, with half of each old at test.

Stimuli

We used the 96 color photographs of faces used by Sauer et al. (2008, Expt 1). Photographs showed male and female individuals of Caucasian descent, ranging in age from young- to elderly-adults. Photographs were obtained from databases at Flinders University and the University of Stirling, and the AR Face Database (Martinez & Benavente, 1998). Photographs displayed the head and neck of the individual. At study and test, photographs were displayed on 19 inch monitors at a size of 200 x 200 pixels, with a resolution of 1024 x 768 pixels.

Photographs were previously sorted according to distinctiveness. Semmler and Brewer (2006) had 34 participants rate faces on a 7-point distinctiveness scale (1 = *typical*; 7 = *distinctive*). Photographs were divided into three categories (distinctive, moderate and typical) according to mean distinctiveness ratings. We selected faces from only the distinctive and typical categories. Mean distinctiveness ratings for distinctive male and female faces were $M = 4.59$ ($SD = 0.53$) and $M = 4.28$ ($SD = 0.50$), respectively. For male and female typical faces the mean distinctive ratings were $M = 2.82$ ($SD = 0.25$) and $M = 2.73$ ($SD = 0.23$).

Procedure

Participants were tested individually. Computers presented instructions and stimuli, and recorded responses. Participants completed a study phase and 3 minute filler task then, immediately or following a delay, a test phase. The study phase presented a series of 48

photographs for 500 ms each, with an inter-stimulus interval of 500 ms. Participants were asked to attend closely to the photographs as they would be questioned on them later.

The test phase included 96 trials. Each trial required participants to respond to a single face. Binary response condition participants indicated whether or not each face was presented earlier by clicking a 'Yes' or 'No' button. After each response, participants rated their confidence in the accuracy of their decision (from 0% to 100% with decile response options). Participants in the confidence condition rated their confidence, from 0% to 100% (with decile response options) that each face was presented earlier. Participants were not given any other instructions (or verbal anchors) for interpreting the confidence scale.

Results

Effect sizes were measured using Cohen's *f*. Cut–off values for small, medium and large effects are 0.10, 0.25 and 0.40, respectively. Analyses comparing the immediate testing condition with each of the delayed conditions produced similar patterns of results[2] so we collapsed delay condition.

Ecphoric confidence ratings, calibration, and discrimination

We tested old-new confidence difference using a 2 (retention interval: immediate, delayed test) × 2 (distinctiveness: distinctive, typical) × 2 (face status: old, new) mixed ANOVA (see Tables 1 and 2 for descriptive and inferential statistics, respectively). The significant main effect of face status indicated higher mean confidence for old than new faces. The significant Face Status × Retention Interval interaction revealed that this difference decreased as retention interval increased. The significant Face Status × Distinctiveness interaction showed a stronger effect of face status for distinctive than for typical faces.

These effects must be interpreted alongside the small but significant Face Status × Distinctiveness × Retention Interval interaction. Simple effects analyses revealed a moderate

Face Status × Retention Interval interaction for distinctive faces, $F(1, 46) = 24.52$, $p = .00$, $f = 0.35$. As retention interval increased, the difference in mean confidence ratings for old and new distinctive faces decreased. As expected, after a delay, confidence ratings for old distinctive faces decreased. Further, confidence ratings for new distinctive faces increased. However, the interaction for typical faces was small and non-significant (after a Bonferroni-correction), $F(1, 46) = 4.52$, $p = .04$, $f = 0.12$. Contrary to expectations, increased delay did not affect confidence ratings for typical faces. Perhaps ecphoric confidence ratings are insensitive to changes in memory quality. Alternatively, they may offer a robust index of stimulus discriminability. Analyses of *d'*, reported below, suggest the latter.

We plotted calibration curves (e.g., Weber & Brewer, 2003) to further investigate the effects of varied retention interval and distinctiveness on the utility of confidence ratings in discriminating old from new faces (see Figure 1). Confidence data were collapsed from the eleven initial confidence categories to five weighted categories (i.e., 0-20, 30-40, 50-60, 70-80 and 90-100% confidence), providing a more stable representation of the relationship. The calibration functions reveal, in all conditions, a generally linear, positive relationship between the level of confidence expressed and the probability that a face had been seen before. For typical faces, this relationship is most evident in the upper half of the confidence scale (consistent with previous research demonstrating that individuals are better at discriminating degrees of 'oldness' than degrees of 'newness'; e.g., Weber & Brewer, 2004).

The adjusted normalized resolution index (*ANRI*; see Table 3) ranges from 0 (no discrimination) to 1 (perfect discrimination), measuring the extent to which confidence discriminated old from new faces. In all conditions, *ANRI* statistics were significantly greater than zero. Thus, confidence discriminated old from new faces. A 2 × 2 mixed ANOVA on the within-subjects *ANRI* statistics (Table 4) revealed results consistent with those for mean confidence. The significant main effect of retention interval showed decreased resolution

with increased delay. However, the Distinctiveness × Retention Interval interaction revealed that this reduction was significant for distinctive, $t(46) = 5.25$, $p = .00$, $f = 0.77$, but not typical faces, $t(46) = 1.20$, $p = .24$, $f = 0.18$. The significant main effect of distinctiveness indicated superior resolution for distinctive compared to typical stimuli.

These findings support previous research suggesting that confidence ratings index recognition (Mickes, Wixted, & Wais, 2007; Ratcliff & Starns, 2009; Sauer, et al., 2008). The *ANRI* statistics confirm that confidence discriminates studied from unstudied faces, but the level of discrimination varies according to memory quality.

Classification performance: Discrimination and bias

Next we investigated a) whether, after the application of a classification criterion, confidence ratings could be used to reliably separate studied from unstudied stimuli, b) how variations in retention interval and stimulus distinctiveness affected classification performance, and c) how these effects compared to effects on binary response classifications. As per Sauer et al. (2008, Expt 1), for each participant we determined the confidence criterion that maximized overall accuracy. The criterion was then applied to classify the confidence ratings in each condition as indicative of an old or new stimulus. Stimuli producing confidence ratings equaling or exceeding this criterion were classified as 'Old'; those producing ratings falling below the criterion were classified as 'New'. Thus, we were able to compute measures of discriminability (or *sensitivity)* (*d'*) and criterion placement (*c*) for each participant, and for each condition (see Table 5). Criteria were derived from participants' data, not designated by the experimenters. We sought to maximize the diagnostic value of participants' memorial information. Thus, separate criteria were calculated for distinctive ($M = 51.48$, $SD = 20.68$) and typical ($M = 54.79$, $SD = 21.83$) face trials for each participant[3]. Using the same process we classified faces based on recognition ratings (i.e., binary recognition plus retrospective confidence), with separate criteria for distinctive ($M = -15.58$,

SD = 48.97) and typical (*M* = -17.50, *SD* = 47.58) faces. Descriptive statistics are included in Table 5.

Changes in *d'* were assessed using a 2 (retention interval: immediate, delayed test) × 2 (distinctiveness: distinctive, typical) × 2 (response type: confidence, binary response) mixed ANOVA (Table 6), with distinctiveness as the within-subjects variable. Large main effects of retention interval and distinctiveness revealed greater discrimination in the immediate than the delayed test condition, and for distinctive compared to typical stimuli. Importantly, the moderate response type main effect indicated superior classification performance for the confidence condition, compared to the binary response condition. No interactions involving response type were significant. These findings have three implications. First, consistent with the assumption that confidence ratings and binary responses share an evidential basis, manipulations that affected binary responses also affected confidence-based classifications. Second, compared to binary responses, confidence ratings allowed consistently superior classification performance. Third, given no evidence of a significant effect of retention interval on confidence for typical old stimuli, and no significant interactions with response type on *d'*, we have no reason to doubt that the superiority of the confidence ratings (evidenced by the significant main effect) is robust to these factors that impair discriminability.

An identical 2 × 2 × 2 mixed ANOVA on *c* (Table 6) revealed a small but significant main effect of retention interval; after a delay, participants (and the classification algorithm) were less likely to classify a test stimulus as ‘Old’ (see Footnote 2). Together with the discriminability results, this finding precluded the possibility that improved performance using the confidence procedure, compared with the binary response condition, simply reflects a more/less conservative classification method. Effects of retention interval on classification criteria placement did not interact with response type.

To test if the improved classification demonstrated above reflected a more fine-grained dependent measure, we ran 2 (response type: binary, recognition rating) × 2 (retention interval: immediate, delayed test) × 2 (distinctiveness: distinctive, typical) mixed ANOVAs on *d'* and *c*, with retention interval as the between-subjects variable (Table 7). A significant main effect of response type on *d'* indicated superior discrimination using recognition ratings. Thus, a more fine-grained scale improved discrimination. The Response Type × Distinctiveness interaction revealed a moderate effect for typical faces, but only a small effect for distinctive faces. The small Response Type × Distinctiveness interaction on *c* indicated a more conservative criterion for binary response classifications (cf. classification based on recognition ratings), for typical, but not distinctive faces.

However, 2 (response type: confidence, recognition rating) × 2 (retention interval: immediate, delayed test) × 2 (distinctiveness: distinctive, typical) mixed ANOVAs, with distinctiveness as the within-subjects variable, on *d'* and *c* (Table 8) revealed a significant main effect of response type on *d'* - indicating superior discrimination using ecphoric confidence ratings - with no significant interactions involving response type. No significant effects on *c* were observed. Thus, the superiority of ecphoric confidence ratings over recognition responses persists for fine-grained recognition responses.

Summary

First, ecphoric confidence ratings were significantly higher for old than new faces. This difference was greater in conditions where the evidential basis for confidence was stronger. These findings suggest confidence indexes ecphory. However, confidence for old typical faces showed no decline after a delay. Second, calibration curves and *ANRI* statistics demonstrate that confidence discriminated previously seen from unseen faces in all conditions. Finally, analyses of classification performance indicated superior discrimination for the confidence procedure in all comparisons, with no evidence of a difference in response

bias (or in the effects of our manipulations on response bias) between the confidence and binary response conditions, or between ecphoric confidence and recognition ratings.

Discussion

Our findings generally support a memorial basis for ecphoric confidence. Across conditions, mean confidence was higher for old, than for new faces. Further, conditions producing stronger memories generally led to increased confidence for old faces, and increased resolution. Confidence ratings for old typical faces showed no effect of delayed testing. Thus, ecphoric confidence may sometimes be insensitive to changes in memory quality. However, analyses of *d'* demonstrated that ecphoric confidence ratings were more robust than binary responses to factors that impaired discriminability. This advantage was not reduced by increasing delay.

Using more fine-grained measures improves discrimination. However, classification performance using ecphoric confidence ratings exceeded performance using recognition ratings (potentially because recognition ratings incorporate retrospective confidence which is vulnerable to non-diagnostic influences). Thus, the superiority of the confidence procedure over binary responses cannot be accounted for by the grain size of the response scale. This advantage suggests a more sensitive index of ecphory.

Encouragingly, although retention interval and distinctiveness affected ecphoric confidence and resolution, calibration curves and *ANRI* statistics indicated a) monotonic, positive relationships between ecphoric confidence and the probability that a face had been previously viewed, and b) that ecphoric confidence discriminated seen from unseen faces, in all conditions. Further, after applying classification criteria, discriminability using the confidence procedure exceeded that for binary responses in all conditions. This result was not attributable to differences in response bias. These findings have significant theoretical and, potentially, practical implications.

Manipulating memory strength exerted similar effects on performance for both classification procedures. This supports previous research demonstrating that ecphoric confidence ratings a) index the evidential basis for recognition decisions, and b) can reliably discriminate between complex stimuli participants do and do not recognize (Koriat, 1993; Mickes, et al., 2007; Sauer, et al., 2008). The consistently superior classification performance using the confidence procedure (and the improvements associated with using recognition ratings) indicates that participants providing binary responses are not making optimal use of the evidence available to them. Thus, procedures that do not require overt decisions may avoid errors encountered when participants control the placement of their decision criteria.

Sub-optimal placement of decision criteria can have serious consequences in applied settings. This is particularly pertinent for the eyewitness identification task, which presents numerous barriers to optimum criterion placement. When viewing a lineup, witnesses often assume they are *expected* to pick someone (see Wells & Olson, 2003, for a review). Combined with stimulus ambiguity, this perceived pressure to pick may lead witnesses to lower their decision criteria, increasing the risk of false identification (Wells, 1993). Alternatively, witnesses aware of the potential consequences of false identifications may set overly conservative criteria, and fail to identify a culprit who is present in the lineup. When testing witness memory, probes should allow access to the information that best discriminates studied from unstudied stimuli. The present results demonstrate that binary recognition decisions are not the best test for this purpose. Procedures capable of ameliorating the effects of criterion placement and/or providing a more sensitive index of recognition would be of considerable practical value. Sauer et al. (2008) demonstrated that ecphoric confidence ratings discriminate target from foil stimuli in lineup tasks. However, the treatment of this type of evidence by the courts requires further investigation.

In sum, ecphoric confidence ratings discriminate studied from unstudied faces, and can be used to reliably classify faces as previously studied or unstudied, even when memory quality is reduced. The similar effects of our memory manipulations on confidence-based and binary response classification suggest that confidence accesses the evidential basis for recognition memory decisions. Finally, the improved performance associated with the confidence procedure, when compared to the binary response group, suggests that the confidence procedure may attenuate non-memorial influences on recognition memory decisions and allow more direct access to the evidence upon which recognition decisions are based.

References

Dodson, C. S., & Schacter, D. L. (2002). When false recognition meets metacognition: The distinctiveness heuristic. *Journal of Memory and Language, 46*, 782-803. DOI: 10.1006/jmla.2001.2822.

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology*. New York: Dover. (Original work published 1895).

Gigerenzer, G., Hoffrage, U., & Kleinboelting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*, 506-528.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100*, 609-639.

Light, L. L., Kayra-Stuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 212-228.

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

Martinez, A. M., & Benavente, R. (1998). The AR Face Database. Barcelona, Spain: Computer Vision Center, Universitat Autonoma de Barcelona.

Mickes, L., Wixted, J. T., & Wais, P. E. (2007). A direct test of the unequal-variance signal detection model of recognition memory. *Psychonomic Bulletin & Review, 14*, 858-865.

Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological Review, 116*, 59-83. DOI: 10.1037/a0014086.

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

Schacter, D. L. (1999). The seven sins of memory. *American Psychologist, 54*, 182-203.

Semmler, C., & Brewer, N. (2006). Postidentification feedback effects on face recognition confidence: Evidence for metacognitive influences. *Applied Cognitive Psychology, 20*, 895-916.

Shepherd, J. W., Gibling, F., & Ellis, H. D. (1991). The effects of distinctiveness, presentation time and delay in face recognition. *European Journal of Cognitive Psychology, 3*, 137-145.

Tenney, E. R., MacCoun, R. J., Spellman, B. A., & Hastie, R. (2007). Calibration trumps confidence as a basis for witness credibility. *Psychological Science, 18*, 46-50.

Tulving, E. (1981). Similarity relations in recognition. *Journal of Verbal Learning & Verbal Behavior, 20*, 479-496.

Van Zandt, T. (2000). ROC curves and confidence judgments in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 582-600.

Weber, N., & Brewer, N. (2003). The effect of judgment type and confidence scale on confidence-accuracy calibration in face recognition. *Journal of Applied Psychology, 88*, 490-499.

Weber, N., & Brewer, N. (2004). Confidence-accuracy calibration in absolute and relative face recognition judgements. *Journal of Experimental Psychology: Applied, 10*, 156-172.

Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist, 48*, 553-571.

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology, 54*, 274-295.

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review, 117*, 1025-1054. DOI: 10.1037/a0020874.

Footnotes

[1] Here *c* has an atypical interpretation. As confidence condition participants did not make old/new responses, *c* does not index participants' criterion placement. Rather, it indexes the placement of the optimal criterion identified by the classification algorithm.

[2] The effect of retention interval on *c* was significant when immediate testing was compared to the short, but not long, delay condition. No other differences were significant.

[3] The only difference resulting from the use of single criterion was the emergence of a small ($f = 0.22$) main effect of response type on *c*. Optimal criteria for confidence-based classifications were more conservative than participants' binary response criteria.

Table 1

Mean, SD, and 95% Confidence Interval (CI) Statistics for Confidence According to Distinctiveness, Face Type and Retention Interval for the Confidence Response Condition

	Retention Interval	
Distinctiveness & Face Type	Immediate Test	Delayed Test
Distinctive Faces		
Old		
M	.56	.46
SD	.11	.14
95% *CI*	.50 - .62	.41 - .51
New		
M	.23	.30
SD	.09	.11
95% *CI*	.18 - .28	.26 - .34
Overall		
M	.40	.38
SD	.08	.12
95% *CI*	.36 - .44	.34 - .42
Typical Faces		
Old		
M	.46	.42
SD	.11	.12
95% *CI*	.40 - .51	.38 - .46
New		
M	.33	.35

SD	.11	.11
95% *CI*	.27 - .39	.31 - .39
Overall		
M	.40	.39
SD	.10	.11
95% *CI*	.34 - .45	.35 - .43
Overall		
Old		
M	.51	.44
SD	.10	.12
95% *CI*	.46 - .56	.40 - .48
New		
M	.28	.33
SD	.09	.10
95% *CI*	.18 - .28	.26 - .34
Overall		
M	.40	.38
SD	.08	.10
95% *CI*	.35 - .44	.35 - .42

Table 2

Mixed ANOVA on Ecphoric Confidence Ratings

Effect	*df*	*F*	*f*	*p*
	Between-subjects			
Retention Interval (R)	1	0.18	0.05	.68
R error	46	(382.38)		
	Within-subjects			
Face status (F)	1	175.54	0.71	.00
F × R	1	19.62	0.24	.00
F error	46	(71.04)		
Distinctiveness (D)	1	0.02	0.01	.90
D × R	1	0.22	0.02	.64
D error	46			
F × D	1	79.01	0.32	.00
F × D × R	1	11.07	0.12	00
F × D error	46	(31.49)		

Note: Values in parentheses represent mean-square errors.

Table 3

Mean, SD, and 95% CI ANRI statistics for Distinctive and Typical Faces for Confidence Group Participants in the Immediate and Delayed Testing Conditions

Distinctiveness & Retention Interval	*ANRI*		
	M	*SD*	*95% CI*
Distinctive			
Immediate Test	.29	.16	.20 - .38
Delayed Test	.10	.09	.07 - .13
Overall	.16	.15	.12 - .21
Typical			
Immediate Test	.07	.07	.03 - .11
Delayed Test	.05	.06	.03 - .07
Overall	.06	.06	.04 - .07
Overall			
Immediate Test	.16	.08	.12 - .20
Delayed Test	.06	.05	.04 - .07
Overall	.09	.08	.07 - .11

Table 4

Mixed ANOVA on ANRI statistics

Effect	*df*	*F*	*f*	*p*
	Between-subjects			
Retention Interval (R)	1	31.29	0.57	.00
R error	46	(0.01)		
	Within-subjects			
Distinctiveness (D)	1	37.94	0.69	.00
D × R	1	14.47	0.42	.00
D error	46	(0.01)		

Note: Values in parentheses represent mean-square errors.

Table 5

Mean, SD, and 95% CI Statistics for d' and c According to Retention Interval and Distinctiveness for Classifications for Classifications Based on Ecphoric Confidence Ratings, Binary Responses, and Recognition Ratings

	Response type								
	Ecphoric confidence			Recognition			Binary		
Distinctiveness	*M*	*SD*	95% *CI*	*M*	*SD*	95% *CI*	*M*	*SD*	95% *CI*
	Immediate								
Distinctive									
d'	1.52	0.49	1.26 - 1.78	1.39	0.46	1.15 - 1.63	1.31	0.46	1.07 - 1.56
c	.31	.57	.01 - .61	.21	.22	.09 - .32	.21	.19	.11 - .31
Typical									
d'	0.93	0.42	0.71 - 1.16	0.85	0.33	0.68 - 1.03	0.69	0.45	0.45 - 0.93
c	.44	.79	-.16 - .73	.09	.52	-.19 - .36	.16	.53	-.12 - .45
Overall									
d'	1.13	0.52	0.85 - 1.41	1.10	.26	0.96 - 1.24	0.98	0.31	0.82 - 1.15
c	.62	.43	.39 - .85	.14	.31	-.03 - .31	.18	.29	.02 - .34

					Delay				
Distinctive									
d'	0.96	0.35	0.84 - 1.09	0.67	0.49	0.49 - 0.85	0.52	0.59	0.31 - 0.73
c	.52	.62	.29 - .74	.19	.65	-.04 - .43	.31	.40	.17 - .46
Typical									
d'	0.68	0.28	0.58 - 0.79	0.59	0.36	0.46 - 0.72	0.28	0.40	0.13 - 0.42
c	.52	.77	.24 - .80	.26	.78	-.02 - .54	.45	.46	.28 - .62
Overall									
d'	0.76	0.32	0.64 - 0.87	0.62	0.32	0.51 - 0.74	0.39	0.39	0.25 - 0.53
c	.49	.74	.22 - .76	.23	.69	-.02 - .47	.37	.37	.24 - .51
					Overall				
Distinctive									
d'	1.15	0.47	1.01 - 1.29	0.91	0.58	0.74 - 1.08	0.78	0.66	0.59 - 0.98
c	.45	.61	.27 - .63	.20	.54	.04 - .36	.28	.35	.18 - .38
Typical									
d'	0.77	0.35	0.66 – 0.87	0.68	0.37	0.57 - 0.79	0.42	0.46	0.28 – 0.55

c	.44	.79	.21 - .67	.20	.70	-.01 - .41	.35	.50	.21 - .50
Overall									
d'	0.88	0.43	0.75 - 1.01	0.78	0.38	0.67 - 0.89	0.59	0.46	0.45 - 0.72
c	.53	.66	.34 - .73	.20	.59	.03 - .37	.31	.35	.21 - .41

Table 6

Mixed ANOVAs on d' and c for Classifications Based on Ecphoric Confidence Ratings and Binary Responses

Measure and effect	*df*	*F*	*f*	*p*
d'				
	Between-subjects			
Response type (Resp)[a]	1	20.65	0.36	.00
Retention Interval (R)	1	49.67	0.56	.00
Resp × R	1	1.99	0.11	.16
Error	92	(0.22)		
	Within-subjects			
Distinctiveness (D)	1	50.85	0.48	.00
D × R	1	7.91	0.19	.01
D × Resp	1	0.00	0.00	.98
D × R × Resp	1	0.08	0.02	.78
D error	92	(0.16)		
c				
	Between-subjects			
Resp	1	1.56	0.10	.22
R	1	4.37	0.17	.04
Resp × R	1	0.02	0.01	.89
Error	92	(0.42)		

		Within-subjects		
D	1	0.04	0.01	.84
D × R	1	0.49	0.04	.49
D × Resp	1	0.11	0.02	.75
D × R × Resp	1	0.27	0.03	.60
D error	92	(0.25)		

Note: Values in parentheses represent mean-square errors.

[a] Participants either provided binary responses or confidence ratings.

Table 7

Mixed ANOVAs on d' and c for Classifications Based on Binary Responses and Recognition Ratings

Measure and effect	*df*	*F*	*f*	*p*
d'				
	Between-subjects			
Retention Interval (R)	1	33.31	0.58	.00
Error	46	(0.38)		
	Within-subjects			
Response type (Resp)	1	20.98	0.18	.00
Resp × R	1	2.25	0.06	.14
Resp error	46	(0.06)		
Distinctiveness (D)	1	16.24	0.39	.00
D × R	1	5.18	0.22	.03
D × Resp	1	6.01	0.07	.02
D × R × Resp	1	0.67	0.02	.42
D error	46	(0.36)		
c				
	Between-subjects			
R	1	1.15	0.12	.29
Error	46	(0.70)		

		Within-subjects		
Resp	1	1.79	0.09	.19
Resp × R	1	0.60	0.05	.44
Resp error	46	(0.23)		
D	1	0.01	0.01	.93
D × R	1	1.71	0.08	.20
D × Resp	1	4.65	0.03	.04
D × R × Resp	1	0.01	0.01	.98
D error	46	(0.01)		

Note: Values in parentheses represent mean-square errors.

Table 8

Mixed ANOVAs on d' and c for Classifications Based on Ecphoric Confidence Ratings and Recognition Ratings

Measure and effect	*df*	*F*	*f*	*p*
d'				
	Between-subjects			
Response type (Resp)	1	5.87	0.18	.02
Retention Interval (R)	1	52.53	0.54	.00
Resp × R	1	0.49	0.05	.48
Error	92	(0.16)		
	Within-subjects			
Distinctiveness (D)	1	39.33	0.45	.00
D × R	1	10.51	0.23	.00
D × Resp	1	1.17	0.08	.28
D × R × Resp	1	0.41	0.05	.53
D error	92	(0.15)		
c				
	Between-subjects			
Resp	1	3.27	0.16	.07
R	1	1.52	0.11	.22
Resp × R	1	0.33	0.05	.57
Error	92	(0.65)		

		Within-subjects		
D	1	0.06	0.01	.81
D × R	1	0.48	0.04	.49
D × Resp	1	0.01	0.01	.91
D × R × Resp	1	0.27	0.03	.61
D error	92	(0.25)		

Note: Values in parentheses represent mean-square errors.

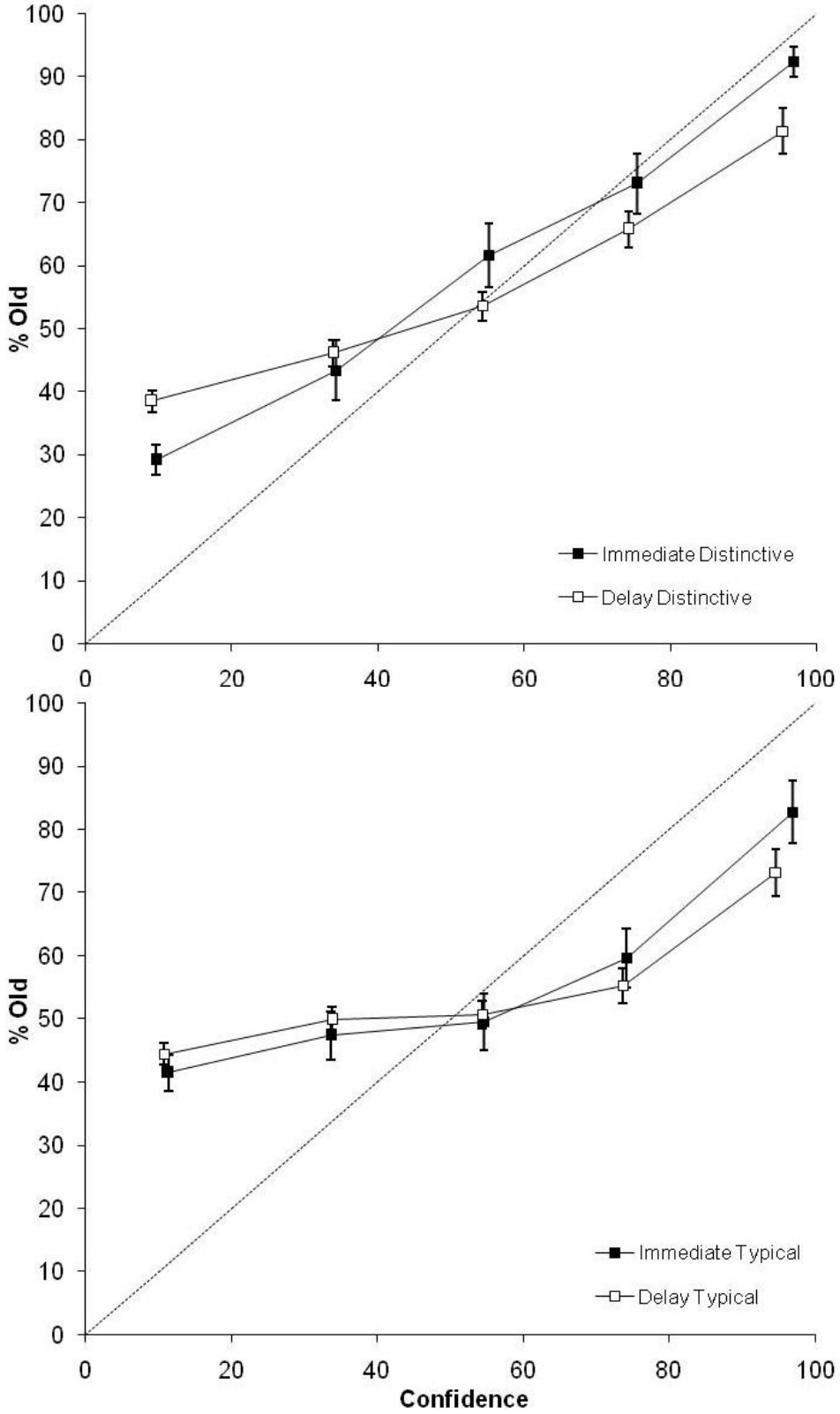

Figure 1. Calibration curves for distinctive (upper panel) and typical (lower panel) face trials, for confidence group participants in the immediate and delayed testing conditions. Error bars represent standard errors.