

1-1-2008

Forgetting the Once-Seen Face: Estimating the Strength of an Eyewitness's Memory Representation

Kenneth A. Deffenbacher

University of Nebraska at Omaha, kdeffenbacher@unomaha.edu

Brian H. Bornstein

University of Nebraska at Lincoln, bbornstein2@unl.edu

E. Kiernan McGorty

University of Nebraska-Lincoln

Steven D. Penrod

John Jay College of Criminal Justice, New York

Follow this and additional works at: <https://digitalcommons.unomaha.edu/psychfacpub>



Part of the [Criminology and Criminal Justice Commons](#), and the [Psychology Commons](#)

Please take our feedback survey at: https://unomaha.az1.qualtrics.com/jfe/form/SV_8cchtFmpDyGfBLE

Recommended Citation

Deffenbacher, Kenneth A.; Bornstein, Brian H.; McGorty, E. Kiernan; and Penrod, Steven D., "Forgetting the Once-Seen Face: Estimating the Strength of an Eyewitness's Memory Representation" (2008). *Psychology Faculty Publications*. 3.

<https://digitalcommons.unomaha.edu/psychfacpub/3>

This Article is brought to you for free and open access by the Department of Psychology at DigitalCommons@UNO. It has been accepted for inclusion in Psychology Faculty Publications by an authorized administrator of DigitalCommons@UNO. For more information, please contact unodigitalcommons@unomaha.edu.

An earlier version of the meta-analysis included in this article and curve fits for two of the empirical forgetting functions included in Table 2 were presented at the "Off the Witness Stand: Using Psychology in the Practice of Justice" conference held March 1-3, 2007, in New York, NY.

Submitted October 9, 2007; revised February 20, 2008; accepted February 25, 2008.

Forgetting the Once-Seen Face: Estimating the Strength of an Eyewitness's Memory Representation

Kenneth A. Deffenbacher, Department of Psychology, University of Nebraska-Omaha

Brian H. Bornstein, Department of Psychology, University of Nebraska-Lincoln

E. Kiernan McGorty, Department of Psychology, University of Nebraska-Lincoln

Steven D. Penrod, Department of Psychology, John Jay College of Criminal Justice

Corresponding author – Kenneth Deffenbacher, Department of Psychology,
University of Nebraska-Omaha, NE 68182-0274; email kdeffenbacher@mail.unomaha.edu

Abstract

The fidelity of an eyewitness's memory representation is an issue of paramount forensic concern. Psychological science has been unable to offer more than vague generalities concerning the relation of retention interval to memory trace strength for the once-seen face. A meta-analysis of 53 facial memory studies produced a highly reliable association ($r = .18$, $d = 0.37$) between longer retention intervals and positive forgetting of once-seen faces, an effect equally strong for both face recognition and eyewitness identification studies. W. A. Wickelgren's (1974, 1975, 1977) theory of recognition memory provided statistically satisfactory fits to 11 different empirical forgetting functions. Applied to the results of field studies of eyewitness memory, the theory yields predictions relevant to fact finders' evaluations of eyewitness credibility. A plausible upper limit for witness initial memory strength corresponds to a probability of .67 of being correct on a fair six-person lineup. Furthermore, not only can the percentage of remaining memory strength be determined for any retention interval, but this strength estimate can be translated into an estimated probability of being correct on a fair lineup of a specified size.

Keywords: eyewitness memory strength, forgetting of faces, retention interval, single-trace fragility theory

Unless the state possesses incriminating physical evidence, eyewitness identification testimony is crucial whenever the prosecution attempts to prove that the defendant and the perpetrator are one and the same. The reliability of an identification is affected by two classes of variables, system variables and estimator variables (Wells, 1978). System variables are those under the control of the criminal justice system, instructions given to eyewitnesses before they consider a lineup or photospread or the method by which members of the lineup other than the suspect are chosen, for instance. Estimator variables are those beyond the control of the criminal justice system and whose effects can only be estimated. These factors include, among many other estimator variables, the duration of the eyewitness's exposure to the perpetrator, lighting conditions at the crime scene, and retention interval, the length of the interval between observation of the suspect and test of the eyewitness's memory.

Given the controllability of system variables, a considerable amount of research has been focused on them, given the greater promise that such research would lead to increases in the reliability of eyewitness memory testing procedures. Indeed, sufficient research progress on system variables had accumulated in the last 2 decades of the 20th century that the U.S. Department of Justice issued guidelines for the collection of eyewitness evidence based on these findings (Techni-

cal Working Group for Eyewitness Evidence, 1999). Police in a number of jurisdictions around the United States have already adopted these guidelines as standard practice.

Progress in producing forensically useful empirical generalizations has not been nearly as great in the case of estimator variables. Nevertheless, research on these variables may have the potential to produce not only greater understanding of situations in which eyewitnesses may experience perceptual or memorial problems but also to yield empirical generalizations that may assist the trier of fact (judge or juror) when he or she must assess the fidelity of an eyewitness's memory representation (cf. Wells, Memon, & Penrod, 2006).

In making this assessment, the key estimator variables are initial memory strength for the perpetrator's face and length of the retention interval. Many other estimator variables have their effect only as they affect initial memory strength. These variables include duration of exposure to the perpetrator, illumination conditions, presence or absence of other foci of attention (e.g., a weapon), eyewitness stress level, and whether the perpetrator is of a different race, among others. To make a proper assessment, the trier of fact would not only need to have an estimate of the witness's initial memory strength for the perpetrator and to know the length of the retention interval but also to understand the nature of the forgetting function for the human face. The forgetting function, of course, is

the curve that specifies the strength of the memory representation over the retention interval. That is, the forgetting function specifies how rapidly memory strength, plotted on the ordinate of the graph, decreases as a function of time, plotted on the abscissa. Knowing the rate of memory strength loss and retention interval length allows one to specify the proportion of original memory strength remaining. To specify in absolute terms how much memory strength remains, one must know the initial memory strength, the "starting point" on the ordinate of the forgetting function.

Typically, the length of the retention interval for an eyewitness can easily be established to a reasonable degree of precision from information provided in police reports. Until now, however, psychological science has not had a means to provide at least a ballpark estimate of initial memory strength for a witness. Furthermore, cognitive psychologists have not established whether the forgetting curve for the human face is even of the same form as Ebbinghaus (1913) had determined. For that matter, it has not always been abundantly clear whether there even is a statistically reliable association between retention interval length and facial recognition memory (Deffenbacher, 1986). For example, in the period between 1970 (approximately the beginning of modern research on eyewitness testimony) and 1985, studies testing the effect of retention interval length on memory for the human face included a substantial minority reporting a null effect. An initial meta-analysis of this literature by Deffenbacher (1986) included 15 studies reporting a null effect out of a total of 33 studies, even though overall he found a highly reliable effect ($p < .0001$) of memory test delay on face recognition memory: The average z was 1.46, yielding a meta-analytic Z of 8.38 and an equivalent correlation of .25 (as retention interval or delay increased, forgetting increased). Including retention interval as part of a much more comprehensive meta-analysis than Deffenbacher's, Shapiro and Penrod (1986) also documented statistically reliable effects of retention interval length on face recognition memory.

With results of these previous meta-analyses in hand, an immediate attempt to describe the forgetting function for once-seen faces might seem in order. However, there are good reasons to conduct a more up-to-date meta-analysis of face memory studies before searching for a suitable theoretical forgetting function. In the more than 2 decades since 1986, the published body of research findings concerning the effect of delay has increased by more than 60%. The number of null or negative ("negative" forgetting or reminiscence) results has also continued to increase.

A further concern is the proportion of studies that have been conducted in the context of the eyewitness identification paradigm rather than with the standard face recognition task in the tradition of cognitive psychology. The eyewitness identification paradigm usually exposes witnesses to a single target face (perpetrator) in a scripted scenario. Memory for this face is tested by embedding it in a 5- to 9-person live or photo lineup (target present) or by substituting someone else who is a match to the perpetrator's description (target absent). Witnesses are asked to identify the perpetrator or to indicate that he or she is not present. The recognition memory task, on the other hand, exposes observers to a relatively large number of target faces. A recognition memory test typically includes the target set plus an equal number of unfamiliar distracter faces. Observers are exposed to faces serially and are to respond "yes" or "no" as to whether a given face has been seen previously. It turns out that

the proportion of eyewitness identification studies has more than doubled, increasing from 27% of published studies concerned with the effect of delayed memory test in Deffenbacher's (1986) meta-analysis to 57% at present.

As a result, not only has there been a considerable increase in the proportion of studies with greater forensic applicability, but it is entirely possible that the effect size for retention interval could be different for eyewitness identification studies than for face recognition studies. Consider one possibility. Results of face recognition memory studies have typically been assumed to represent high estimates of the amount of facial memory obtaining in real-world eyewitness identification settings. If eyewitness identification studies did indeed produce lower estimates of initial memory strength than did face recognition memory studies, then there would be less room for the decline of any forgetting function to occur. Thus, retention interval effects might be less for studies in the eyewitness identification paradigm because of the greater probability of a restriction of range in possible loss of memory strength, as compared with face recognition studies. On the other hand, the direction of a difference in the effect size for retention interval could well be in the opposite direction. A number of published meta-analyses of the effects of other independent variables have yielded generally larger effects on memory for eyewitness identification studies than for laboratory face recognition studies. For instance, Deffenbacher, Bornstein, Penrod, and McGorty (2004) found a considerably larger negative effect of heightened stress on memory for witnesses in studies conducted in the more forensically relevant eyewitness identification tradition than for witnesses in face recognition studies.

For all these reasons, before attempting a theoretical description of the forgetting function for face memory, we deemed it advisable to conduct an up-to-date meta-analysis of the effects of retention interval on the strength of a witness's memory representation for the once-seen face. We next present the methodology followed and the results obtained from this meta-analysis.

Meta-Analysis of Retention Interval Effects

Method

Sample characteristics. Clark (2005), Deffenbacher *et al.* (2004), and Reisberg and Heuer (2007) have all agreed that the legal standards for proffered scientific testimony established in *Daubert v. Merrell Dow Pharmaceuticals* (1993) have strengthened the legal system's preference for meta-analytic conclusions based on a body of well-conceived, well-executed, and easily retrievable studies. Hence, we made the decision to include only published studies in our sample. A thorough search of relevant citation retrieval systems was made. These systems included PsycINFO, Medline, and Social SciSearch (the Social Science Citation Index). We also examined the citations in published research and in social science convention proceedings. The present study sample consists of 39 published articles, books, and book chapters. These sources, listed in Table 1, generated 53 independent tests ($N = 5,405$) of the hypothesis that longer retention intervals have a negative effect on memory strength for the once-seen face. Individual sample sizes ranged considerably, from a low of 8 to a high of 590 ($M = 101.98$). Retention intervals associated with these studies ranged from 1 s to 350 days.

Table 1. Effect Sizes for Proportion Correct Recognition Memory or Identification Accuracy

Study	<i>n</i>	RI	<i>z</i>	<i>r</i>
Scapinello & Yarmey (1970)	40	20 min	0.00	.00
Smith & Nielsen (1970)	144	10 s	3.30	.28
Wallace <i>et al.</i> (1970)				
Children	200	5 min	0.00	.00
Adults	200	5 min	0.00	.00
Goldstein & Chance (1971)	52	2 days	0.00	.00
Shepherd & Ellis (1973)	36	35 days	1.92	.32
Laughery <i>et al.</i> (1974)	292	1 week	0.00	.00
Chance <i>et al.</i> (1975)	144	2 days	0.00	.00
Egan <i>et al.</i> (1977)	86	54 days	1.65	.18
Davies <i>et al.</i> (1978)	40	19 days	1.96	.31
Walker-Smith (1978)	8	19 s 2.58	.91	
Yarmey (1979)	84	30 days	2.32	.25
Ellis <i>et al.</i> (1980)	48	1 week	2.58	.37
Courtois & Mueller (1981)	128	28 days	2.76	.24
Deffenbacher <i>et al.</i> (1981): Control: 2 min/2 wk	22	2 weeks	1.96	.42
Krouse (1981)	76	2.5 days	2.58	.30
Mauldin & Laughery (1981)	100	47.5 hr	0.00	.00
Barkowitz and Brigham (1982)	237	1 week	2.58	.17
Brigham <i>et al.</i> (1982)	88	22 hr	5.65	.60
Shepherd <i>et al.</i> (1982)				
Experiment 2	40	343 days	3.58	.57
Experiment 3	104	90 days	−0.72	−.07
Krafka & Penrod (1985)				
TP/no context	24	22 hr	−0.22	−.04
TP/context	20	22 hr	0.45	.10
TA/no context	21	22 hr	2.16	.47
TA/context	20	22 hr	1.41	.32
Cutler <i>et al.</i> (1986): Experiment 2	287	23 days	2.75	.16
Chance & Goldstein (1987)	59	5 days	0.00	.00
Cutler <i>et al.</i> (1987a)	165	1 week	−1.60	−.12
Cutler <i>et al.</i> (1987b)	290	12 days	−0.47	−.03
Peters (1988)	212	6 days	0.00	.00
Read <i>et al.</i> (1989)				
Early rehearsal	68	1 week	2.63	.32
Late rehearsal	68	1 week	−1.30	−.16
Ellis & Flin (1990)	153	1 week	1.96	.16
Podd (1990)	90	2 weeks		1.75 .18
Read <i>et al.</i> (1990)	90	100 min	0.00	.00
Goodman <i>et al.</i> (1991)	48	4.5 days	1.04	.15
Peters (1991)				
Experiment 1	71	26 days	0.00	.00
Experiment 3	64	13 days	0.00	.00
Shepherd <i>et al.</i> (1991)	96	1 month	1.96	.20
Wixted & Ebbesen (1991): Experiment 2	195	2 weeks	2.81	.20
Yarmey <i>et al.</i> (1996)				
TP:5 min/24 hr	69	24 hr	1.45	.17
TA: 5 min/24 hr	76	24 hr	0.82	.09
Sh/TP: 5 min/24 hr	69	24 hr	1.29	.16
Sh/TA: 5 min/24 hr	70	24 hr	3.05	.36
Peters (1997): Experiment 2	96	6 months	3.00	.31
MacLin <i>et al.</i> (2001)	64	30 min	1.34	.17
Memon <i>et al.</i> (2003)				
Older adults: TP	45	1 week	1.86	.28
Younger adults: TP	42	1 week	0.40	.06
Older adults: TA	42	1 week	3.75	.58
Younger adults: TA	42	1 week	−0.97	−.15
Yarmey (2004)	590	4 hr	0.00	.00
Brewer <i>et al.</i> (2006)				
TP	37	30 min	0.00	.00
TA	66	30 min	0.00	.00

RI = length of delay between shortest and longest retention intervals. TP = target present lineup; TA = target absent lineup; Sh = showup.

Statistical procedures. As we always compared the longest and shortest retention intervals in each study to determine effect size, we selected z scores for a difference between proportions as the primary dependent measure. For the studies in our sample, a z score for a difference between proportions was occasionally reported or, more often, could be calculated post hoc. In instances in which a test of the hypothesis was reported as not statistically significant, but no statistics were cited, we followed the conservative procedure of entering a z of zero (Rosenthal, 1995). Otherwise, we entered a z score associated with the p value of the effect size estimate, 1.65 for $p = .05$, one-tailed, for instance.

To test the statistical reliability of any estimate of typical effect size, we calculated a one-sample t test and an associated 95% confidence interval (Rosenthal & DiMatteo, 2002). Given that r and d are more frequently encountered measures of effect size, and in the case of r , may be a more generally useful measure, we have reported mean effect sizes in terms of r and d as well. In the case of r , we first converted the z -score measures of effect size for each study to r by dividing z by the square root of n , a conversion formula recommended by Rosenthal and DiMatteo (2002). Each of these biserial correlation coefficients between retention interval (short or long) and memory accuracy was then normalized by conversion to the equivalent Fisher's z' score before averaging. Values of Cohen's d equivalent to the mean effect size expressed in terms of r were obtained by use of the expression $d = 2r(1 - r^2)^{-1/2}$.

Results and Discussion

For each study, we subtracted the proportion correct associated with the longer retention interval from the proportion correct associated with the shorter retention interval. Thus, a positive result represented positive forgetting, a loss of memory. A negative result represented negative forgetting, or reminiscence. When we report effect size in terms of r , then a positive r means that longer retention intervals were associated with more forgetting.

The unweighted mean r was .18, significantly different from zero, $t(52) = 4.78$, $p < .005$, with the 95% confidence interval (CI) extending from .10 to .26. The mean effect size for r in this instance is equivalent to $d = 0.37$, a small to medium effect size (Cohen, 1988). We should note that when all possible pairwise comparisons of retention intervals in studies that had more than two retention intervals were treated as independent effect sizes, the sample size increased from 53 to 78, and the mean effect size was .17 ($d = 0.34$), remarkably close to the results we obtained when only the longest and shortest retention intervals were compared.

We next applied a test of homogeneity of variances across the sample of weighted effect sizes to determine whether the degree of variability exceeded that expected on the basis of sampling error alone. A chi-square value of 23.19 ($df = 52$, $p > .05$) indicated that the degree of variability did not exceed that expected on the basis of sampling error. Strictly speaking, then, no moderator analyses were required. However, given our prediction that studies conducted in the context of the eyewitness identification paradigm might well show more or even less of an effect of retention interval on memory strength as compared with face recognition studies, we nevertheless calculated mean effect sizes across 23 face recognition memory studies and 30 eyewitness identification studies. In the former

case, the mean r was .21, $t(22) = 3.18$, $p < .005$, 95% CI = .08–.34; in the latter case, it was .16, $t(29) = 3.58$, $p < .005$, 95% CI = .07–.25. However, the difference between these two correlations was not significant by a two-sample t test, $t(51) = .21$, $p > .05$. Hence, nature of the research paradigm was not a moderator of average effect size.

Post hoc, it was suggested to us that a particularly strong moderator of the effect size for delay might be the duration of delay itself. In the third column of Table 1 (RI), we have included the length of delay between the shortest and longest retention intervals for each study in our sample. Noting that the most commonly encountered delay for British police has been a month (Pike, Brace, & Kynan, 2002), we estimated the average r to be .27, $t(6) = 3.39$, $p < .01$, 95% CI = .08–.44, for the seven studies with delays of a month or more. For studies with lesser durations of delay, we estimated the average r to be .17, $t(45) = 3.99$, $p < .005$, 95% CI = .08–.25. The difference in magnitude of these two correlations suggests that duration of the memory test delay itself might moderate effect size. This conjecture cannot be supported, however, because the difference between effect sizes at shorter and longer durations of delay was not significant, $t(51) = .38$, $p > .05$. Even so, it is interesting to note that the upper bound of the confidence interval for the studies with a maximum delay of a month or more was .44, as compared with a comparable figure of .25 for studies with a maximum delay of less than a month.

Thus, despite 22 of the sample of 53 effect sizes being null or negative, we have found a statistically reliable effect size estimate for the effect of retention interval on proportion of correct recognition judgments, a measure of memory accuracy for the human face. Furthermore, our effect size estimate does not vary as a function of whether it is a product of studies done in the face recognition memory paradigm or of studies conducted in the eyewitness identification tradition. Hence, it is reasonable to conclude that increased delay of a test for recognition memory for the once-seen face portends decreased probability of correct recognition judgments. This decreased probability presumably reflects loss of underlying memory trace strength.

Our estimate of the effect size for retention interval on memory for faces is also likely an underestimate of the actual value. The 28% of studies reporting forgetting effects that were not statistically significant but for which no statistics were cited resulted in our entering a conservative value of $z = 0.00$ in each instance. Most likely a small but positive amount of forgetting was actually exhibited by participants in such studies.

Our meta-analyses put us in a better position to specify what happens over time to a person's memory representation for an unfamiliar face. At least now we can say with some assurance that memory strength will be weaker at longer retention intervals than at briefer ones. However, our meta-analyses do not permit us to specify the shape of the forgetting function and answers to related questions, such as whether the memory representation will ever be truly lost, much less when. To address these questions, we would need to be able to specify a theoretical forgetting function that would satisfactorily fit empirical forgetting functions, particularly for studies in which facial recognition memory was tested at three or more retention intervals. The latter requirement would enable us to assess fit to nonlinear functions.

Finding a Theoretical Forgetting Function for the Human Face

Criteria

As indicated earlier, the trier of fact has had no useful way to estimate the initial strength of an eyewitness's memory representation for the once-seen face. Clearly, for it to have forensic applicability, any candidate theoretical forgetting function must (a) be able to provide an estimate of initial memory strength; (b) be accurate at predicting where future points will fall as retention interval increases, a strong test of the theory (Wixted & Carpenter, 2007); and (c) be able to satisfactorily fit group forgetting data, the form in which empirical forgetting functions exist in studies of memory for the human face included in our meta-analyses.

If a theoretical forgetting function is found that meets these criteria, eyewitness memory researchers should finally have evidence bearing directly on their belief that the forgetting function for the once-seen face is Ebbinghausian in nature. That is, 93% of experts in the field of eyewitness testimony research, when surveyed most recently (Kassin, Tubb, Hosch, & Memon, 2001), agreed that there was a research basis for the notion that the rate of memory loss for an event is greatest right after an event and then levels off over time. A still large majority (83%) of these same experts agreed that this generalization was reliable enough for psychologists to present in courtroom testimony. There has been little direct evidence provided to date, however, that the faith of these experts is justified when it comes to specifying the forgetting function for the once-seen human face. Consider the critique provided by Elliott (1993):

The Ebbinghaus forgetting curve ... is another dubious metaphor for most eyewitness circumstances, both because the human face seems to have special properties as a stimulus, and because the retention intervals that are pertinent to identification scarcely ever include the very short ones where most forgetting presumably occurs. There is now a large enough number of results that are null or negative with respect to the Ebbinghaus hypothesis that their presence ought certainly to form part of any testimony that might be given: They should no longer be treated simply as error. (p. 429)

Selection of a Theory of Forgetting

The only theory meeting the first criterion for forensic applicability, provision of an estimate of initial memory strength at 0 s after stimulus cessation, is Wickelgren's (1972, 1974, 1975, 1977, 1979) single-trace fragility theory of recognition memory. Thus, Wickelgren's theory is the only one that we evaluate for its ability to meet the remaining two criteria. In its least complex version (Wickelgren, 1975, 1977), the form of the retention function is $m = Lt^{-D}e^{-It}$, where m represents memory strength at a given retention interval, t seconds after target stimulus exposure has ended; L is initial memory strength at 0 s after stimulus exposure ends; D is the rate parameter for a time-decay process, which is inversely proportional to the rate of memory consolidation; and I is the rate parameter for the loss of memory strength due to interference, which is directly proportional to the similarity of the target stimulus to subsequently encountered stimuli. Of course $e = 2.72$, the base of the natural or Napierian system of logarithms. It is important to note that Wickelgren (1974) proposed that at least for recognition mem-

ory, an interval-scale measurement of memory strength (d') is possible by making relatively weak, yet plausible assumptions concerning how statistical decision theory would translate strength into yes-no decisions. For all practical purposes, then, both m and L are measured in terms of d' units.

Wickelgren's (1972, 1974, 1975, 1977, 1979) theory is unique in that rather than distinguishing between short- and long-term traces, it posits a single memory trace and two mechanisms productive of forgetting. An interference-free, time-decay process produces rapid forgetting in the first seconds and minutes of the retention interval because initially trace fragility is very high. As the neurophysiological process of consolidation begins to decrease trace fragility, however, the rate of forgetting slows in a negatively accelerated fashion, and less is forgotten per unit time. Consolidation, showing a negatively accelerated increase over time, is assumed to continue to decrease trace fragility and its susceptibility to the time-decay process throughout the life of the memory trace. The negative power component of Wickelgren's forgetting function, t^{-D} , would appear to be a plausible model of the negatively accelerated loss of trace fragility over time and therefore the continually decreasing amount of trace strength lost per unit time.

As the contribution of the time-decay process to forgetting declines in power function fashion with increases in the retention interval, the second process, a storage interference process, operating in a negative exponential fashion (e^{-It}), would be expected to increase its influence on the rate of forgetting at longer retention intervals. This prediction might explain a result noted by Deffenbacher, Carr, and Leu (1981), who found that for recognition memory of both faces and words, the amount of forgetting due to retroactive interference with an item's trace in storage increased over a 2-week retention interval.

We should note that a simpler version of Wickelgren's (1972, 1974, 1975, 1977, 1979) theory has been proposed (e.g., Wixted & Carpenter, 2007). This version, in effect, contains only two free parameters, initial memory strength and the rate of forgetting due to time decay. The version we have selected (Wickelgren, 1974, 1975, 1977) contains a third parameter, rate of forgetting due to interference generated subsequent to encoding of the stimulus. It would be prudent to justify the necessity of the additional free parameter.

We have found it necessary to retain the interference parameter to secure an adequate fit to empirical forgetting functions that included retention intervals greater than 1–2 weeks in length. The two-parameter version provides about the same degree of fit as the three-parameter one for retention intervals up to this length. At longer intervals, however, face recognition memory appears to require a source of forgetting in addition to time decay. Values of the time-decay parameter sufficient for a good fit at shorter intervals were not sufficient to account for the considerable additional forgetting at longer intervals. Indeed, a plot of log memory strength (d') against log time reveals a downward inflection in empirical forgetting functions that occurs between an interval corresponding to about 1 week and ones corresponding to a month or more (Deffenbacher, 1986). Interestingly, Valentine, Pickering, and Darling (2003) found in their analysis of 314 lineups conducted by the London Metropolitan Police that the probability of identifying the suspect decreased drastically in the interval between 1 week and 1 month, declining from .66 to .34. Finally, face recognition in a forensic context often includes an institutional source of interference subsequent to encoding of the perpetrator's face, the

exposure of the witness to mugshots before a memory test by means of a live or photographic lineup (Deffenbacher, Bornstein, & Penrod, 2006). This sort of interference could be problematic if the later lineup were a target-absent one.

Previous Tests of the Theory

In the first decade after the introduction of the single-trace fragility theory of forgetting in the 1970s, a modest amount of empirical support was generated. For instance, Wickelgren (1972, 1974, 1975) found that the theory provided an excellent fit to forgetting functions obtained for episodic memory representations for frequently encountered words and for pictures of commonly encountered objects. In the three publications just cited, Wickelgren reported a dozen experiments resulting in 35 separate r^2 statistics, averaged across 3–10 research participants in each instance. The median r^2 was .89, the proportion of empirical forgetting function variance accounted for by single-trace fragility theory. All these experiments used yes-no recognition memory tasks, with memory for verbal and pictorial materials being tested under a variety of conditions and measured at retention intervals up to 2 years in length.

To the best of our knowledge, there have been only two previously published attempts to fit any theory of forgetting to face recognition memory forgetting functions. Deffenbacher (1986) not only conducted a meta-analysis of memory for the once-seen face as a function of retention interval but also conducted a preliminary test of the ability of Wickelgren's (1972, 1974, 1975, 1977, 1979) single-trace fragility (power-exponential) theory to fit empirical forgetting functions for face recog-

nition memory. He found that Wickelgren's power-exponential theory provided relatively good fits to the functions from five different studies. Wixted and Ebbesen (1991, Experiment 2) showed that a simple power function was an excellent fit to their empirical forgetting function for face recognition memory tested at retention intervals ranging from 1 hr to 2 weeks in length.

Unfortunately, except for the single effort of Wixted and Ebbesen (1991, Experiment 2), neither Deffenbacher nor anyone else ever followed up these first curve-fitting forays with any further theory testing or development in regard to the forgetting of faces. Furthermore, neither Deffenbacher nor anyone else ever made a serious attempt to determine to what extent either Wickelgren's (1972, 1974, 1975, 1977, 1979) power-exponential theory or any other theory of forgetting might have forensic application. In the next section, we attempt to remedy the first of these two deficiencies. We remedy the second deficiency in a subsequent section.

New Tests of the Theory

Table 2 illustrates the results of our fitting Wickelgren's (1972, 1974, 1975, 1977, 1979) power-exponential theory to 11 empirical forgetting functions obtained from the face recognition memory and eyewitness identification literatures. These 11 data sets were obtained from studies that included at least three retention intervals, that obtained a significant effect for retention interval (positive forgetting), and for which sufficient information was provided to calculate d' values as a measure of memory strength at each of the tested retention intervals.

Table 2. Fit of Single-Trace Fragility Theory to Empirical Forgetting Functions

Study	Observed and (predicted) d' memory strength after various delays				
	0 s	5 min	2 days	7 days	
Barkowitz & Brigham (1982)	(1.70)	1.47 (1.47)	1.19 (1.24)	1.14 (1.18)	
Chance & Goldstein (1987)	0 s	10 min	2 days	7 days	
White faces	(2.49)	2.12 (2.12)	1.96 (1.82)	1.61 (1.72)	
Japanese faces	(1.53)	1.30 (1.30)	1.02 (1.02)	0.88 (0.76)	
Courtois & Mueller (1981)	0 s	1 min	2 days	28 days	
	(3.26)	2.94 (2.94)	2.41 (2.39)	1.93 (1.95)	
Ellis & Flin (1990)	0 s	5 min	1 day	7 days	
7 years/2-s encoding time	(1.13)	0.98 (0.98)	0.74 (0.84)	0.70 (0.72)	
10 years/2-s encoding time	(1.78)	1.54 (1.54)	0.98 (1.20)	0.72 (0.62)	
10 years/6-s encoding time	(1.98)	1.72 (1.72)	1.72 (1.47)	1.23 (1.26)	
Shepherd & Ellis (1973)	0 s	3 min	6 days	35 days	
	(1.97)	1.73 (1.73)	1.24 (1.28)	0.78 (0.74)	
Shepherd, Ellis, & Davies (1982)	0 s	7 days	30 days	90 days	350 days
	(2.78)	1.92 (1.92)	1.62 (1.64)	1.47 (1.17)	0.00 (0.29)
Wixted & Ebbesen (1991, Experiment 2)	0 s	1 hr	1 day	1 week	2 weeks
	(2.47)	2.01 (2.01)	1.75 (1.83)	1.46 (1.57)	1.41 (1.37)
Yarmey (1979)	0 s	1 min	7 days	30 days	
	(3.39)	3.12 (3.12)	2.44 (2.30)	1.47 (1.50)	

Six of the data sets were obtained from studies published since Deffenbacher's (1986) preliminary test of Wickelgren's theory of forgetting.

We should note that these 11 functions were of necessity fitted by eye so as to minimize the sum of absolute deviations of predicted and observed values. Least-squares or maximum likelihood estimates of parameter values were not possible, given that each forgetting function contained only three or four retention intervals and that the observed values at each retention interval were group d' scores.

Fortunately, we were able to begin our curve-fitting exercise by taking advantage of parameter values required to fit Wickelgren's (1975) data for frequently encountered English words and Ryback, Weinert, and Fozard's (1970) data for recognition of pictures of common everyday objects. We discovered, however, that the value of the time-decay parameter needed to fit our data for unfamiliar faces was only one-tenth that required for the data by Wickelgren (1975) and Ryback *et al.* (1970). The same value of the time-decay parameter (.025) provided good fits for 10 of the 11 forgetting curves. A value of .02 improved the fit slightly for the remaining study (Yarmey, 1979). Values of the interference parameter that we used here were up to an order of magnitude smaller (6×10^{-8}) than that required to fit the data of Wickelgren (1975) and Ryback *et al.* (1970), 6×10^{-7} . The forgetting data from Barkowitz and Brigham (1982), Courtois and Mueller (1981), and Shepherd, Ellis, and Davies (1982) and Chance and Goldstein's (1987) data from Caucasians viewing Caucasian faces were fit with the 6×10^{-8} value of the interference parameter, and the data from the remaining seven studies were fit by values of the interference parameter that were up to 16 times greater.

The values provided in the 0-s column of Table 2 are estimates of L , the initial memory strength parameter. Given that all the data for memory measurement as a function of retention interval were group, rather than individual, in nature, and given the lack of any previously established estimates of initial memory strength for unfamiliar faces, we obtained initial strength estimates by substituting for the predicted value of d' in Wickelgren's (1975, 1977) equation the observed value of d' obtained from the first retention interval at which face memory was measured and then solving for L . It should therefore not be surprising that the predicted and observed values match perfectly at the first retention interval for each forgetting function. Clearly, any statistical assessment of the adequacy of fit includes only the degree of fit at retention intervals subsequent to the first. This approach also permits assessment of how well the theory predicts where future points will fall as retention interval increases beyond Time 0.

A statistical assessment of the fit of Wickelgren's (1972, 1974, 1975, 1977, 1979) power-exponential theory to the 11 empirical face forgetting functions was made by applying a chi-square goodness-of-fit test in each instance. In no instance was the chi-square test significant. Hence, in each case the null hypothesis that both observed and predicted values represent the same forgetting function could not be rejected. An omnibus chi-square test of the fit of retention interval data from all 11 functions (23 df) was also not significant. The quality of the curve fits by the power-exponential theory is especially encouraging when one notes that the observed values of d' are by necessity group scores rather than being based on individually computed scores such as Wickelgren obtained from continuous recognition memory experiments.

Thus, it can be said that Wickelgren's (1972, 1974, 1975, 1977, 1979) power-exponential theory has met all three criteria set out earlier for any theory of forgetting to have potential forensic applicability. The theory provides an estimate of initial memory strength, predicts accurately where future points will fall on the forgetting function as retention interval increases, and satisfactorily fits functions based on group data. Before considering the forensic applicability of Wickelgren's (1975, 1977) theoretical forgetting function, however, we should first note some additional aspects of its theoretical utility.

Further Theoretical Observations

Faces as a stimulus category. Finding a theory that fits face recognition memory forgetting functions has been rewarding in terms of a number of additional insights gained, insights accrued beyond the mere promise of having a more precise accounting of the loss in fidelity for the memory representation of the once-seen face. One particularly intriguing finding is that not only is virtually the same value of the time-decay parameter required to fit the power-exponential theory to each of the 11 empirical forgetting functions but it is an order of magnitude smaller than the value required by Wickelgren (1975) to fit forgetting functions for common English words and pictures of common objects. Apparently, there is a more rapid rate of decline in trace fragility (a more rapid rate of increase in trace consolidation) for unfamiliar faces than there is for episodic traces of familiar English words and pictures of familiar objects. Thus, even though face recognition memory forgetting functions may be fit by the same theory as forgetting functions for words and objects, episodic memory for unfamiliar faces may decline more slowly. As Deffenbacher (1986) noted, perhaps this phenomenon should not be all that surprising, given selection pressures in the evolutionary history of our species to promote efficient processing of faces. After all, faces constitute a very important category of stimuli, providing a very rich source of socially relevant information, continually requiring all of us to make fine discriminations among them, and needing a relatively large quantity of human cortex to be devoted to their processing (e.g., Sergent, Ohta, & MacDonald, 1992). Our finding that memory representations for the unfamiliar face may be consolidated more rapidly than for certain other visual stimulus classes reinforces the notion that faces may not be unique stimuli but they are at least somewhat special (Ellis & Young, 1989).

Cross-race effect

A second interesting byproduct of our theoretical search is revealed as a result of our fit of the power-exponential theory to two forgetting functions provided by the data of Chance and Goldstein (1987; see Table 2). Here we have additional illumination of mechanisms underlying the cross-race effect, a forensically relevant phenomenon whereby once-seen faces of another race or ethnic grouping are discriminated from one another less well and later recognized less well than are once-seen faces of the observer's own race (Meissner & Brigham, 2001). Chance and Goldstein's observers were Caucasians exposed to Caucasian and Japanese faces, blocked in counter-balanced order by race of face. The statistically reliable cross-race effect obtained here was clearly due to superior encoding of the Caucasian faces. This effect was documented by an initial memory strength superiority of approximately 1 d' unit for Caucasian faces when encoded by Caucasian observers

as compared with the initial memory strength engendered by their encoding of Japanese faces. It is also of interest that even though Japanese faces were consolidated at the same rate as Caucasian faces, requiring the same value of the time-decay parameter for a fit of theory to forgetting function, the forgetting function for the Japanese faces required a value of the interference parameter 10 times as great, 10^{-7} versus 10^{-8} . This finding provides support for the view that same-race faces are more easily discriminated from one another than are other-race faces (Meissner & Brigham, 2001), providing an opportunity for them to be encoded in a more discriminating fashion, yielding thereby at least some of the encoding advantage for same-race faces. Same-race faces would therefore likewise be expected to withstand better the ravages of interference generated by subsequent encounters with other faces during any particular face's retention interval. Indeed, Meissner and Brigham (2001) found that cross-race effects were greater at longer retention intervals.

Target face exposure. The effect of increases in target face exposure time is illuminated by our fit of Wickelgren's (1972, 1974, 1975, 1977, 1979) theory to two forgetting functions yielded by the work of Ellis and Flin (1990). Examining the face recognition memory functions generated by two different groups of 10-year-old Scottish schoolchildren (Table 2), one group having an encoding time of 2 s per face and the other having encoding time of 6 s, we observed that the extra 4 s of encoding time provided an initial memory strength advantage of about 0.20 of a d' unit. This initial advantage leveraged a memory strength advantage of 0.50 of a d' unit at a 7-day retention interval. A recent meta-analysis by Bornstein, Deffenbacher, McGorty, and Penrod (2007) has confirmed that longer exposures of target faces are associated with both higher hit rates ($r = .23$) and lower false alarm rates ($r = -.12$). Clearly, more research is needed in an effort to try to establish just how much additional initial memory strength can be purchased by n additional seconds of exposure time.

Age differences. Still another phenomenon of face recognition memory is the age effect associated with efficiency of face processing, such that older children show superior memory for faces. It would appear that the age effect associated with the greater recognition memory shown by Ellis and Flin's (1990) 10-year-olds with 2 s of encoding time per face as compared with 7-year-olds with the same 2 s of encoding time is strictly due to the former being able to encode unfamiliar faces more effectively in the time available, yielding an initial memory strength advantage of 0.65 of a d' unit (Table 2). Up until the middle teen years, face recognition memory shows continual improvements in the ability to discriminate same-race faces from one another and later to recognize them (Chance & Goldstein, 1984). Thus, 10-year-olds have had another 3 years of fine tuning of their brain's perceptual learning "machinery," permitting enhanced ability to respond to more subtle differences among faces they typically encounter.

Failures to find retention interval effects

Finally, power-exponential theory can help us to understand at least one potential contributor to the frequent failure to find statistically significant retention interval effects for face recognition memory. Clearly, one factor in producing findings of a lack of a statistically reliable effect of forgetting occurs primarily when just two retention intervals are measured

and both intervals are at points on the forgetting function between which little forgetting would be expected to occur. Although it is more informative to test retention at more than two intervals, if only two intervals are to be tested, investigators should at least ensure that one occurs within minutes after encoding, given the relatively rapid roll-off in memory strength in the first minutes after encoding. After all, memory strength for the once-seen face loses 15% of its strength in the first 10 min of the retention interval. Consider an illustration provided by the data from Wixted and Ebbesen (1991, Experiment 2, 11-s stimulus duration condition); see Table 2. Note that there was about a 12% loss of original memory strength ($d' = 2.47$) between Day 1 and Day 7. However, the actual amount of original memory strength lost since encoding was about 41%. If one were to have only measured face recognition memory at the Day 1 and Day 7 intervals, one would have underestimated the actual amount of forgetting by a factor of more than three to one. Hence, Wickelgren's (1972, 1974, 1975, 1977, 1979) power-exponential theory and its ability to provide an estimate of original memory strength permits us to see clearly that just where retention intervals fall on the theoretical forgetting function will affect the likelihood of finding statistically reliable amounts of forgetting.

Forensic Applicability of Power-Exponential Theory

As we indicated in the introduction to this article, to make a proper assessment of the strength of an eyewitness's current memory representation, a trier of fact needs to have an estimate of the witness's initial memory strength, to know the length of the retention interval, and to understand the nature of the forgetting function for the once-seen human face. As retention interval length can usually be specified with some precision, acquiring an estimate of the eyewitness's initial memory strength and a knowledge of the precise nature of the forgetting function represent the key forensic needs.

Let's consider the latter forensic need first. Psychologists are now able to provide a much greater degree of specific knowledge to the trier of fact as regards the nature of the forgetting function for human face recognition memory. It turns out that we can now offer the judge or juror an estimate of what proportion of memory strength, regardless of its initial value, remained at the time the eyewitness's memory for the perpetrator was tested. We can do this for three reasons. First, we have clearly demonstrated the ability of Wickelgren's (1972, 1974, 1975, 1977, 1979) power-exponential theory to fit forgetting functions that have included retention intervals ranging in length from 1 min to nearly 1 year. Second, we have demonstrated the remarkable constancy of the value of the time-decay parameter (.025) needed for a fit. Third, we have likewise demonstrated the relatively narrow range of values of the interference parameter needed for a fit, practically all falling within an order of magnitude of each other, 10^{-7} to 10^{-8} . For a conservative estimate of the proportion of remaining memory strength, we can simply plug into Wickelgren's (1975, 1977) equation given earlier in this article the values of the forensically relevant retention interval (in seconds), the value of the time-decay parameter at .025, and the value of the interference parameter at 10^{-8} ; if a case involves a cross-race identification, the interference parameter should be instead set at 10^{-7} . For a given retention interval, the resulting calculation yields the estimated proportion of initial memory strength remaining. Having an expert on eyewitness memory be able to testify to

this estimate should aid the trier of fact considerably in his or her task of assessing the fidelity of an eyewitness's memory representation. Knowing, for example, that at memory test, an eyewitness had only 50% of original memory strength remaining would represent a real improvement in specificity over what could be provided by an expert before the present. More valid assessments of eyewitness credibility can only increase the quality of justice rendered by a trier of fact.

Let's now return to the other key forensic need for triers of fact to be able to add precision to their assessment, the need for an estimate of initial memory strength for the perpetrator's face. Wickelgren's (1972, 1974, 1975, 1977, 1979) theory of forgetting is indeed unique among theories of forgetting in its specification of an estimate of initial memory strength. However, the forensic situation is neither a laboratory nor a field experiment and does not yield an estimate of initial memory strength.

There is, nevertheless, a way to yield a conservative estimate of initial memory strength for a typical eyewitness, conservative in the sense that the estimate would very likely represent an upper bound on initial memory strength for many forensic situations. The results of an interesting field experiment (Pigott, Brigham, & Bothwell, 1990) provide us with the opportunity. Pigott *et al.*'s (1990) participants were 47 Florida bank tellers, each of whom interacted with one of two men who attempted to cash a crudely altered U.S. Postal Service money order during a scripted 1.5-min encounter. More than 75% of these tellers had training in eyewitness techniques but were not made aware that their encounter with the perpetrator of attempted bank fraud was not genuine until after their recall and recognition had been measured 4 hr later. Averaged across two target-present and two target-absent lineups, their mean proportion correct was 0.55, which for a seven-alternative, forced-choice recognition memory task (six lineup members plus the alternative of rejecting the entire lineup) is equivalent to a d' score of 1.41 (Hacker & Ratcliff, 1979). It seems reasonable to account for the alternative of rejecting the lineup as an additional choice. After all, for a target-absent lineup, the correct choice is rejection of the lineup. If this had not been done in the present instance, the six-alternative, forced-choice d' would have been 1.30.

If we substitute 1.41 for the value of m in Wickelgren's (1975, 1977) equation and solve for L , we come up with a very plausible estimate of initial memory strength for the bank teller eyewitnesses of Pigott *et al.* (1990), $d' = 1.79$, equivalent to 67% correct on a seven-alternative, forced-choice task (Hacker & Ratcliff, 1979). After 4 hr, then, memory strength for the perpetrator of attempted bank fraud was just 79% of what it had been originally. Had the tellers not been tested until a week after the encounter with the perpetrator, memory strength would have been approximately $d' = 1.24$, equivalent to a predicted performance score of 49% correct—considerably lower, although still better than chance.

The d' value of 1.79 plausibly represents an upper limit for initial memory strength for eyewitnesses in many forensic situations, at least for those not having a highly distinctive or memorable perpetrator. That is, it is fair to say that the forensic scene for Pigott *et al.*'s (1990) bank tellers represents a close to optimal situation for an eyewitness. Consider that the perpetrator was in full view for 1.5 min, an amount of target exposure greater than the "critical value" of 1.0 min noted by Valentine, Pickering, and Darling (2003) in their massive study of lineups conducted by the London Metropolitan Police. The banks were well illuminated. The perpetrators were not dis-

guised. There was no alternative focus of attention present that might ordinarily have been expected to draw the teller's attention away from the perpetrator's face, such as a weapon; only the face of the money order proffered by the man trying to perpetrate bank fraud required some attention. Tellers should not have been operating under high stress levels, given the absence of any personal threat to them.

To the extent that one or more of the optimal conditions of the Pigott *et al.* (1990) study are not met in any given forensic situation, then, the predicted initial memory strength for an eyewitness should be less than the figure of 67% correct predicted for the Florida bank tellers in Pigott *et al.*'s field experiment. Until further research is conducted—testing three or more retention intervals and the effect of varying durations of target person exposure, target person distinctiveness, and illumination levels, for instance—conservative advice to a trier of fact would be that the typical eyewitness viewing a perpetrator's face that was not highly distinctive would be expected to have no more than a 50% chance of being correct in his or her lineup identification (six-person lineup) at a 1-week delay. Clearly the trier of fact would still need to consider other specific facts of the case to decide how much less than 50%, if any, the chance of a correct identification might be in these less than optimal witnessing conditions. Retention interval benchmarks other than 1 week can be readily calculated, of course, using Wickelgren's (1972, 1974, 1975, 1977, 1979) theory of forgetting and the data provided by Pigott *et al.* These calculations do assume that the lineup's construction and administration have been conducted fairly. However, a post hoc assessment of lineup fairness is relatively straightforward (Malpass, Tredoux, & McQuiston-Surrett, 2007).

To illustrate in particular the need for more research on the relationship of variations in target face distinctiveness to initial memory strength, consider the results of using power-exponential theory to predict initial memory strength for the eyewitnesses in the field studies of Read, Tollestrup, Hammersley, McFadzen, and Christenson (1990). Averaging across four different photo lineup conditions and 212 retail clerks, performance was at a level of 76% correct at a 48-hr retention interval, equivalent to a d' of 2.10. Using this value of memory strength to estimate initial memory strength, we find a d' value of 2.87 (91% correct), considerably higher than the 67% correct initial memory strength figure predicted for Pigott *et al.*'s (1990) bank tellers. Predicted performance level for Read *et al.*'s clerks would have been 73% correct at a memory test delay of 1 week, equivalent to a memory strength value of $d' = 1.98$. However, because only one perpetrator was used across the four different 48-hr retention interval conditions to which Read *et al.*'s clerks were exposed, it may well be that their higher performance level than that obtained by Pigott *et al.*'s (1990) eyewitnesses and those in other field experiments was due to the single perpetrator having a rather distinctive, and hence memorable, face.

Conclusions

Psychological science is now in a position to offer the trier of fact more than vague generalities regarding the relationship between retention interval and strength of the memory representation for the once-seen face. The results of our meta-analysis confirm that there is indeed a statistically reliable association between longer retention intervals and decreased face recognition memory, an association equally true of face recog-

nitition memory and eyewitness identification studies. That is, there is an increase in positive forgetting as the delay increases between encoding of a face and test of one's memory for it.

The present meta-analytic review of the literature also provides support for Wickelgren's (1972, 1974, 1975, 1977, 1979) theory of forgetting from recognition memory, using data from studies of memory for once-seen faces that used diverse methodologies and a wide range of retention intervals. Fitting Wickelgren's power-exponential theory to 11 different empirical forgetting functions providing group data resulted in statistically satisfactory fits, fits predicting where future points will fall on the function as retention interval increases. In addition, power-exponential theory provides an estimate of initial memory strength. This latter feature of the theory, particularly useful when applied to the results of field studies, permits calculation of not only an estimate of initial memory strength (d') but also calculation of a strength estimate at any given retention interval. Hence, not only can percentage of initial memory strength remaining be determined at any retention interval, but the strength estimate at a particular retention interval can also be readily translated into a probability of being correct on a fair lineup of a specified size. Of course, to be practically useful, these estimates would need to be calculated for and clearly explained to the trier of fact by an eyewitness memory expert.

When considering the applicability of our findings, at least two concerns might be raised. First, throughout this article, we have assumed that the amount forgotten is a function of the current strength of the memory representation. It could be objected that what is remembered is also very much determined by retrieval conditions, such as the type of memory test. Recognition memory tests are often more sensitive measures of memory than are recall tests for the same material, for instance. Furthermore, the encoding specificity principle proposes that the amount of forgetting at retrieval is a function of the degree of match between encoding context and retrieval context. Although we do not deny the validity of these objections, we do not see them as problematic for the applicability of Wickelgren's (1972, 1974, 1975, 1977, 1979) theory to eyewitness memory. For one thing, the retrieval tasks in the studies we have reviewed, eyewitness identification memory tests and laboratory face recognition tests, differ somewhat but are still essentially tests of recognition memory. One of our moderator analyses showed that type of retrieval task, recognition memory or eyewitness identification, was not a moderator of effect size for the correlation between retention interval length and memory strength. In addition, even though the match of encoding and retrieval context is typically greater for recognition memory tasks than for eyewitness identification tasks, this difference did not affect the strength of the relationship between retention interval and memory strength, either. Furthermore, Wickelgren's theory was equally effective at predicting memory strength at any given retention interval for both eyewitness identification and face recognition memory studies.

A second concern that might be raised relates to the fact that our curve-fitting exercise could only be applied to 11 forgetting functions from just eight published studies. The robustness of the data sets underlying these 11 functions clearly depends on the overall quality of the eight published papers. Our considered judgment is that there is little to be concerned about in this regard. Quality of fit of theory and data shows no obvious relationship to any perceived differences in quality of

publication source or any minor differences in quality of methodology and data analysis.

In any event, psychologists interested in the psychology of testimony now have much more abundant direct evidence bearing on their belief that the forgetting function for the once-seen face is Ebbinghausian in nature (cf. Kassin *et al.*, 2001): Rate of memory loss for an unfamiliar face is greatest right after the encounter and then levels off over time. Psychological science can now also provide to both these same psychologists and triers of fact rather more specific details concerning the decline and fall of a face's memory representation over time and succeeding facial encounters.

References

References marked with an asterisk () indicate studies included in the meta-analysis.*

- * Barkowitz, P., & Brigham, J. C. (1982). Recognition of faces: Own-race bias, incentive, and time delay. *Journal of Applied Social Psychology*, 12, 255-268.
- Bornstein, B. H., Deffenbacher, K. A., McGorty, E. K., & Penrod, S. D. (2007). *The effect of cognitive processing on facial identification accuracy: A meta-analysis*. Unpublished manuscript, University of Nebraska—Lincoln.
- * Brewer, N., Caon, A., Todd, C., & Weber, N. (2006). Eyewitness identification accuracy and response latency. *Law and Human Behavior*, 30, 31-50.
- * Brigham, J. C., Maass, A., Snyder, L. D., & Spaulding, K. (1982). Accuracy of eyewitness identifications in a field setting. *Journal of Personality and Social Psychology*, 42, 673-681.
- Chance, J. E., & Goldstein, A. G. (1984). Face-recognition memory: Implications for children's eyewitness testimony. *Journal of Social Issues*, 40, 69-85.
- * Chance, J. E., & Goldstein, A. G. (1987). Retention interval and face recognition: Response latency measures. *Bulletin of the Psychonomic Society*, 25, 415-418.
- * Chance, J. E., Goldstein, A. G., & McBride, L. (1975). Differential experience and memory for faces. *Journal of Social Psychology*, 97, 243-253.
- Clark, S. E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law and Human Behavior*, 29, 575-604.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- * Courtois, M. R., & Mueller, J. H. (1981). Target and distracter typicality in face recognition. *Journal of Applied Psychology*, 66, 639-645.
- * Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987a). The reliability of eyewitness identifications: The role of system and estimator variables. *Law and Human Behavior*, 11, 233-258.
- * Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987b). Improving the reliability of eyewitness identification: Putting context into context. *Journal of Applied Psychology*, 72, 629-637.
- * Cutler, B. L., Penrod, S. D., O'Rourke, T. E., & Martens, T. K. (1986). Unconfounding the effects of contextual cues on eyewitness identification. *Social Behaviour*, 1, 113-134.
- Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579. (1993).
- * Davies, G. M., Ellis, H. D., & Shepherd, J. W. (1978). Face identification: The influence of delay upon accuracy of Photofit construction. *Journal of Police Science and Administration*, 6, 35-42.

- Deffenbacher, K. A. (1986). On the memorability of the human face. In H. D. Ellis, M. A. Jeeves, F. Newcombe, & A. Young (Eds.), *Aspects of face processing* (pp. 61-70). Dordrecht, the Netherlands: Martinus Nijhoff.
- Deffenbacher, K. A., Bornstein, B. H., & Penrod, S. D. (2006). Mugshot exposure effects: Retroactive interference, mugshot commitment, source confusion, and unconscious transference. *Law and Human Behavior*, 30, 287-307.
- Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior*, 28, 687-706.
- * Deffenbacher, K. A., Carr, T. H., & Leu, J. R. (1981). Memory for words, pictures, and faces: Retroactive interference, forgetting, and reminiscence. *Journal of Experimental Psychology: Human Learning and Memory*, 7, 299-305.
- Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenius, Trans.). New York: Teachers College, Columbia University. (Original work published in 1885).
- * Egan, D., Pittner, M., & Goldstein, A. G. (1977). Eyewitness identification: Photographs vs. live models. *Law and Human Behavior*, 1, 199-206.
- Elliott, R. (1993). Expert testimony about eyewitness identification: A critique. *Law and Human Behavior*, 17, 423-437.
- * Ellis, H. D., & Flin, R. H. (1990). Encoding and storage effects in 7-year-olds' and 10-year-olds' memory for faces. *British Journal of Developmental Psychology*, 8, 77-92.
- * Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1980). The deterioration of verbal descriptions of faces over different delay intervals. *Journal of Police Science and Administration*, 8, 101-106.
- Ellis, H. D., & Young, A. W. (1989). Are faces special? In A. W. Young & H. D. Ellis (Eds.), *Handbook of research on face processing* (pp. 1-26). Amsterdam: North-Holland.
- * Goldstein, A. G., & Chance, J. (1971). Visual recognition memory for complex configurations. *Perception & Psychophysics*, 9, 237-241.
- * Goodman, G. S., Hirschman, J. E., Hepps, D., & Rudy, L. (1991). Children's memory for stressful events. *Merrill-Palmer Quarterly*, 37, 109-158.
- Hacker, M. J., & Ratcliff, R. (1979). A revised table of d' for M-alternative forced choice. *Perception & Psychophysics*, 26, 168-170.
- Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the "general acceptance" of eyewitness testimony research: A new survey of the experts. *American Psychologist*, 56, 405-416.
- * Krafka, C., & Penrod, S. D. (1985). Reinstatement of context in a field experiment on eyewitness identification. *Journal of Personality and Social Psychology*, 49, 58-69.
- * Krouse, F. L. (1981). Effects of pose, pose change, and delay on face recognition performance. *Journal of Applied Psychology*, 66, 651-654.
- * Laughery, K. R., Fessler, P. K., Lenorovitz, D. R., & Yoblick, D. A. (1974). Time delay and similarity effects in facial recognition. *Journal of Applied Psychology*, 59, 490-496.
- * MacLin, O., MacLin, K. M., & Malpass, R. S. (2001). Race, arousal, attention, and delay: An examination of factors moderating face recognition. *Psychology, Public Policy, & Law*, 7, 134-152.
- Malpass, R. S., Tredoux, C. G., & McQuiston-Surrett, D. (2007). Lineup construction and lineup fairness. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology, Vol. 2: Memory for people* (pp. 155-178). Mahwah, NJ: Erlbaum.
- * Mauldin, M. A., & Laughery, K. R. (1981). Composite production effects of subsequent facial recognition. *Journal of Applied Psychology*, 66, 351-357.
- Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the role of the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, & Law*, 7, 3-35.
- * Memon, A., Bartlett, J., Rose, R., & Gray, C. (2003). The aging eyewitness: Effects of age on face, delay, and source-memory ability. *Journal of Gerontology: Psychological Sciences and Social Sciences*, 58B, 338-345.
- * Peters, D. P. (1988). Eyewitness memory and arousal in a natural setting. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues, Vol. 1: Memory in everyday life* (pp. 89-94). Chichester, England: Wiley.
- * Peters, D. P. (1991). The influence of stress and arousal on the child witness. In J. Doris (Ed.), *The suggestibility of children's recollections* (pp. 60-76). Washington, DC: American Psychological Association.
- * Peters, D. P. (1997). Stress, arousal, and children's eyewitness testimony. In N. L. Stein, P. A. Ornstein, B. Tversky, & C. J. Brainerd (Eds.), *Memory for everyday and emotional events* (pp. 351-370). Hillsdale, NJ: Erlbaum.
- Pigott, M. A., Brigham, J. C., & Bothwell, R. K. (1990). A field study on the relationship between quality of eyewitnesses' descriptions and identification accuracy. *Journal of Police Science and Administration*, 17, 84-88.
- Pike, G., Brace, N., & Kynan, S. (2002). *The visual identification of suspects: Procedures and practice* (Briefing note 2/02 Policing and Reducing Crime Unit, Home Office Research, Development and Statistics Directorate) [online]. Retrieved December 3, 2007, from <http://www.homeoffice.gov.uk/rds/prgbrief-pubs1.html>
- * Podd, J. (1990). The effects of memory load and delay on facial recognition. *Applied Cognitive Psychology*, 4, 47-59.
- * Read, J. D., Hammersley, R., Cross-Calvert, S., & McFadzen, E. (1989). Rehearsal of faces and details in action events. *Applied Cognitive Psychology*, 3, 295-311.
- * Read, J. D., Tollestrup, P., Hammersley, R., McFadzen, E., & Christensen, A. (1990). The unconscious transference effect: Are innocent bystanders ever misidentified? *Applied Cognitive Psychology*, 4, 3-31.
- Reisberg, D., & Heuer, F. (2007). The influence of emotion on memory for forensic settings. In M. P. Toglia, J. D. Read, D. F. Ross, & R. C. L. Lindsay (Eds.), *The handbook of eyewitness psychology, Vol. 1: Memory for events* (pp. 81-116). Mahwah, NJ: Erlbaum.
- Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin*, 118, 183-192.
- Rosenthal, R., & DiMatteo, M. R. (2002). Meta-analysis. In J. Wixted (Ed.), *Stevens' handbook of experimental psychology* (Vol. 4, pp. 391-428). New York: Wiley.
- Ryback, R. S., Weinert, J., & Fozard, J. L. (1970). Disruption of short-term memory in man following consumption of ethanol. *Psychonomic Science*, 20, 353-354.
- * Scapinello, K. F., & Yarmey, A. D. (1970). The role of familiarity and orientation in immediate and delayed recognition of pictorial stimuli. *Psychonomic Science*, 21, 329-331.

- Sergent, J., Ohta, S., & MacDonald, B. (1992). Functional neuroanatomy of face and object processing. *Brain*, 115, 15-36.
- Shapiro, P. N., & Penrod, S. D. (1986). A meta-analytic analysis of facial identification studies. *Psychological Bulletin*, 100, 139-156.
- * Shepherd, J. W., & Ellis, H. D. (1973). The effect of attractiveness on recognition memory for faces. *American Journal of Psychology*, 86, 627-633.
- * Shepherd, J. W., Ellis, H. D., & Davies, G. M. (1982). *Identification evidence: A psychological evaluation*. Aberdeen, Scotland: Aberdeen University Press.
- * Shepherd, J. W., Gibling, F., & Ellis, H. D. (1991). The effects of distinctiveness, presentation time, and delay on face recognition. *European Journal of Cognitive Psychology*, 3, 137-145.
- * Smith, E. E., & Nielsen, G. D. (1970). Representations and retrieval processes in short-term memory: Recognition and recall of faces. *Journal of Experimental Psychology*, 85, 397-405.
- Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement*. Washington, DC: U.S. Department of Justice, Office of Justice Programs.
- Valentine, T., Pickering, A., & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real lineups. *Applied Cognitive Psychology*, 17, 969-993.
- * Walker-Smith, G. J. (1978). The effects of delay and exposure duration in a face recognition task. *Perception & Psychophysics*, 24, 63-70.
- * Wallace, G., Colheart, M., & Forster, K. I. (1970). Reminiscence in recognition memory for faces. *Psychonomic Science*, 18, 335-336.
- Wells, G. L. (1978). Applied eyewitness testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology*, 36, 1546-1557.
- Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest*, 7, 45-75.
- Wickelgren, W. A. (1972). Trace resistance and the decay of long-term memory. *Journal of Mathematical Psychology*, 9, 418-455.
- Wickelgren, W. A. (1974). Single-trace fragility theory of memory dynamics. *Memory & Cognition*, 2, 775-780.
- Wickelgren, W. A. (1975). Alcoholic intoxication and memory storage dynamics. *Memory & Cognition*, 3, 385-389.
- Wickelgren, W. A. (1977). *Learning and memory*. Englewood Cliffs, NJ: Prentice-Hall.
- Wickelgren, W. A. (1979). Chunking and consolidation: A theoretical synthesis of semantic networks, configuring in conditioning, S-R versus cognitive learning, normal forgetting, the amnesic syndrome, and the hippocampal arousal system. *Psychological Review*, 86, 44-60.
- Wixted, J. T., & Carpenter, S. K. (2007). The Wickelgren power law and the Ebbinghaus savings function. *Psychological Science*, 18, 133-134.
- * Wixted, J. T., & Ebbesen, E. B. (1991). On the form of forgetting. *Psychological Science*, 2, 409-415.
- * Yarmey, A. D. (1979). The effects of attractiveness, feature saliency and liking on memory for faces. In M. Cook & G. Wilson (Eds.), *Love and attraction* (pp. 51-53). Oxford, England: Pergamon Press.
- * Yarmey, A. D. (2004). Eyewitness recall and photo identification: A field experiment. *Psychology, Crime & Law*, 10, 53-68.
- * Yarmey, A. D., Yarmey, M. J., & Yarmey, A. L. (1996). Accuracy of eyewitness identifications in showups and lineups. *Law and Human Behavior*, 20, 459-477.