Supporting Information

Wilmer et al. 10.1073/pnas.0913053107

SI Text S1: Twin Model-fitting Analyses

We used the statistical modeling software package Mx (http://www.vcu.edu/mx) to fit standard maximum likelihood-based models of genetic and environmental influence to raw MZ and DZ twin data. We first fit a full ACE model to the data, controlling for age and gender. The ACE model divides familial resemblance (A + C) into additive genetic effects (A) and shared environmental effects (C) and attributes the remaining variation (E) to some combination of measurement error and unique environmental effects (1). Because the ACE model estimated the influence of shared environment (C) on face recognition performance as 0% (95% CI: 0–18%), we fit a second "AE" model that dropped C. The AE model fit our data as well as the ACE model, as indicated by a zero χ^2 statistic, and did so more parsimoniously, with one fewer parameter. Akaike's Information Criterion, an indication of model performance that rewards goodness of fit, while penalizing additional parameters, decreased for the AE model (from 0 to −2), indicating that it outperformed the ACE model. According to the AE model, additive genetic factors account for 68% of total variation in face recognition performance (95% CI: 59–74%), or 100% of familial resemblance in face recognition performance (95% CI: 87–100%). An "ADE" model that included both additive (A) and nonadditive (D) genetic effects did not fit our data better than the AE model [$\chi^2(1) = 2.21$, $P = 0.14$] despite the added (D) parameter. These modeling analyses confirmed that familial resemblance in face recognition can be attributed to genetic effects.

SI Text S2: Reliability Analyses

Nonfamilial variation in face recognition performance (E in our ACE model) accounted for 32% of CFMT variation (95% CI: 26–41%). Given reliability data, E can, in theory, be parsed into measurement error (1.0 − reliability) and nonfamilial environmental influence (reliability − MZ correlation). We have computed several reliability statistics.

We computed several measures of internal consistency reliability, including Cronbach's α, Guttman's λ2, McDonald's ωt, and greatest lower bound (2), separately on both our twin and nontwin data. These measures ranged between 0.89 and 0.90. Test-retest reliability was 0.70 in 389 nontwin participants who completed CFMT twice with a mean delay of 6 months (95% CI: 0.64–0.74). Scores rose moderately from test to retest (mean = 76.9%, SD = 12.9% to mean = 83.2%, SD = 12.9%), as did Cronbach's α (0.86 to 0.90), and there were no outliers (all Cook's D's <0.25). Alternate forms reliability was 0.76 in 42 nontwin participants who completed CFMT after its alternate form with a mean delay of 2 months (95% CI: 0.59–0.86). Scores were moderately higher for the alternate form than for the original CFMT (mean = 78.6%, SD = 16.3% and mean = 72.0%, SD = 16.4%, respectively), Cronbach's α's were comparable (0.86 and 0.88), and there were no outliers (all Cook's D's <0.50).

CFMT's high internal consistency reliability suggests little measurement error. Test-retest and alternate-forms reliability are lower, perhaps partially attributable to the so-called "state" factors, such as wakefulness, testing conditions, and mood, that exhibit transient fluctuations. Such fluctuations could differentially affect twins in a pair and do not contribute to stable individual differences. However, learning effects can also attenuate test-retest reliability, and nonperfect equivalence between forms can attenuate alternate-forms reliability.

In sum, the reliability estimates vary substantially (0.70–0.90), but the 0.20 difference between the MZ correlation and our highest reliability estimate is consistent with a modest but nontrivial nonfamilial environmental contribution to CFMT performance.

SI Text S3: Control Task Analyses

Our two nonface memory control tests account for a far smaller portion of the variance in CFMT performance (7.4% for both VPAM and AAM combined in our large nontwin cohort) than the variance in CFMT performance shared between our MZ twins (48.8% for our MZ twins). Therefore, even if the entire 7.4% overlap between CFMT and control tasks contributed to CFMT's MZ correlation, subtracting this variance out would still leave 44.7% of CFMT's variance shared between MZ twins [(48.8 − 7.4%)/(100 − 7.4%) = 44.7%], corresponding to an MZ correlation of 0.67. This "worst case" correlation decreases only slightly when differential reliability between tasks is taken into account (0.63) and/or when correlations between tasks are determined from the small twin cohort that completed all three tasks ($n = 120$) (0.66 using raw twin data, 0.61 using reliability-corrected twin data) rather than the large web-recruited sample ($n > 1{,}500$).

SI Text S4: Correlations Between CFMT and Memory of Famous Faces

We administered both CFMT and a 20-item Famous Faces Recall test (FFR) to 780 nontwin web-recruited participants (Cronbach's α for FFR = 0.87). The robust association between these measures [$r(778) = 0.51$, $P < 0.0001$] links CFMT to the long-term face memory processes necessary to recognize celebrities in FFR. As a further (rough) probe of memory for celebrity faces, we asked a subset of twin participants simply to rate on a five-point Likert scale their endorsement of the statement "I can recognize famous celebrities in photos or on TV," as well as a control statement, "I can remember a seven-digit telephone number long enough to write it down." CFMT performance correlated significantly with the former [$r(188) = 0.37$, 95% CI: 0.24–0.49] but not at all with the latter [$r(188) = 0.01$, 95% CI: −0.13 to 0.15]; a linear fit finds an average 1.1 SD higher CFMT performance for those who report "always" recognizing celebrities compared with those who report "not usually" recognizing celebrities.

1. Plomin R, DeFries JC, McClearn GE, McGuffin P (2008) *Behavioral Genetics* (Worth, New York), 5th Ed.

2. Sijtsma K (2009) Reliability beyond theory and into practice. *Psychometrika* 74: 169–173.