Journal of Experimental Psychology: Applied
2000, Vol. 6, No. 4, 336–348

1076-898X/00/$5.00 DOI: 10.1037//1076-898X.6.4.336

Illusory Correlations in Graphological Inference

Roy N. King and Derek J. Koehler
University of Waterloo

The authors investigate the illusory correlation phenomenon as a possible contributor to the persistence of graphology's use to predict personality. Participants unfamiliar with graphology inspected handwriting samples paired with fabricated personality profiles. In Experiment 1, handwriting samples and personality profiles were randomly paired. In Experiment 2, discernible correlations near unity were set between targeted handwriting-feature–personality-trait pairs in a congruent or incongruent direction with graphologists' claims. In both experiments, participants' judgments of the correlation between designated handwriting-feature–personality-trait pairs agreed with graphologists' claims, even after controlling for their actual statistical association. Semantic association between words used to describe handwriting features and personality traits was the source of biases in perceived correlation. Results may partially account for continued use of graphology despite overwhelming evidence against its predictive validity.

"Beware of a man whose writing sways like a reed in the wind." Like Confucius, the graphologist makes inferences about personality by examining aspects of handwriting. In the past several decades, organizations around the world have begun to use the graphologist's assessment as a decision aid in personnel selection.

The use of graphology in personnel selection is most prevalent in Europe, particularly France, where estimates for the percentage of organizations using the technique range from 38% (Shackleton & Newell, 1994) to 93% (Bruchon-Schweitzer & Ferrieux, 1991). In the United States, estimates for the number of organizations using graphology rose from 500 in 1970 (Mickels, 1970) to 3,000 in 1977 (Hager, 1977), and more recent reports suggest that graphology is quietly gaining acceptance in corporate America (McCarthy, 1988). Although it is difficult to accurately assess how many organizations are using graphology, it does appear that hiring decisions regarding a large number of job applicants around the world are determined, at least in part, by inferences made on the basis of their handwriting.

For the practice of graphology to persist, a perception among its users that the method bears some utility or predictive validity must exist. Indeed, many human-resource practitioners give positive testimony to the predictive power of graphology and continue to procure the services of graphologists (Hooper & Stanford, 1992; McCarthy, 1988). Although the idea of diagnosing personality from handwriting may bear some intuitive appeal, evidence for its validity is weak.

Despite some early support from the scientific community (Allport & Vernon, 1933; Downey, 1923), the results of recent research testing the validity of using handwriting for predicting personality traits have been consistently negative. For example, Furnham and Gunter (1987) investigated the "trait" method of graphology, which predicts specific personality traits from individual features of handwriting. Participants completed the Eysenck Personality Questionnaire (EPQ; Eysenck & Eysenck, 1975) and copied a passage of text in their own handwriting. The writing samples were coded on 13 handwriting-feature dimensions (e.g., size, slant) that graphologists report to be diagnostic of personality traits. Only chance-level correlations were observed between writing features and EPQ scores on the Extroversion, Neuroticism, Psychoticism, and Lie scales. As another example, Bayne and O'Neill (1988) asked graphologists to estimate people's Myers–Briggs type (Extrovert–Introvert, Sensing–Intuition, Thinking–Feeling, Judging–Perceiving) from handwriting samples. Though highly confident in their judgments, none of the graphologists' appraisals accurately predicted the Myers–Briggs profile of the writers.

In a meta-analysis of over 200 studies assessing the validity of graphological inferences, Dean (1992) found only a small effect size for inferring personality from handwriting and noted that the inclusion of studies with methodological shortcomings may have inflated the effect-size estimate. The liberal estimated effect size of $r = .12$ for inferring personality from neutral-content scripts (i.e., scripts with fixed content not under control of the writer) is not nearly large enough to be of any practical value and would certainly be too small to be perceptible to the human judge (Jennings, Amabile, & Ross, 1982). Thus, even a small, real effect—for which the evidence is mixed at best—cannot account for the magnitude of handwriting-feature–personality-trait relationships reported by graphologists or their clients.

Gender (Furnham, 1988), socioeconomic status (Hines, 1988), and degree of literacy (Osborne, 1929), all predictable from handwriting, may in turn predict some personality traits. Thus, any weak ability of graphology to predict personality may be merely based on gender or socioeconomic status information assessed from handwriting. Graphological accuracy attributable to these

Roy N. King and Derek J. Koehler, Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada.

This research was supported by Natural Sciences and Engineering Research Council of Canada Grant OGP 0183792. We are grateful to Emily Marks and Jen Pipe for their assistance with data collection.

Correspondence concerning this article should be addressed to Derek J. Koehler, Department of Psychology, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. Electronic mail may be sent to dkoehler@watarts.uwaterloo.ca.

variables is of dubious worth because simpler, more reliable methods for assessing them are available.

The results of research investigating the validity of graphology for predicting job performance has been similarly negative (Ben-Shakhar, Bar-Hillel, Bilu, Ben-Abba, & Flug, 1986; Rafaeli & Klimoski, 1983). Graphological assessments for personnel selection focus on desired traits such as determination, sales drive, and honesty. Given its apparent lack of validity for predicting personality, it would be somewhat surprising if graphology proved to be a valid predictor of job performance. Indeed, in a meta-analytic review of 17 studies and using stricter inclusion criteria for a study's methodological soundness than did Dean (1992), Neter and Ben-Shakhar (1989) found that graphologists performed no better than did nongraphologists in predicting job performance. When handwriting samples were autobiographical, the two groups achieved modest accuracy in prediction. When the content of the scripts was neutral (i.e., identical for all writers), neither group was able to draw valid inferences about job performance. Thus, belief in the validity of graphology, as it is currently used to predict job performance, lacks empirical support.

As a necessary (but not sufficient) condition for valid inference, the reliability of predictions based on graphology must first be established (Goldberg, 1986). However, reliability of graphological prediction has its own precondition: The predictors—handwriting features—must first be reliably encoded. This precondition appears to be met; the mean agreement between different judges measuring objective handwriting features, such as slant or slope, is $r = .85$, and the mean agreement about subjective handwriting features, like rhythm, is still respectable at $r = .60$ (Dean, 1992). Agreement about what these features signify is somewhat less impressive. In studies reviewed by Dean (1992), the mean agreement of interpretations (i.e., inferences) made by graphologists is $r = .42$. Interestingly, even lay judges exhibit some agreement in their naive interpretations, with a reliability ($r = .30$) only slightly lower than that of the graphologists.

From a social psychological perspective, the agreement between judges is itself intriguing because such agreement constitutes shared but apparently invalid beliefs about the relationship between personality and handwriting. Although the origin of such beliefs among graphologists may lie in their training, such an explanation cannot account for the agreement between naive judges, who have no formal training in graphological inference. In research by Vine (1974), naive judges made invalid predictions about personality from handwriting, yet the agreement between them was quite high. More recently, James and Loewenthal (1991) reported a consistent, invalid belief that naive participants formed in making judgments about depression from handwriting. Though not predictive of the criterion, naive judges consistently perceived untidy handwriting as diagnostic of depression. In a similar vein, Loewenthal (1975) found that participants knew how to alter their handwriting to convey to naive judges false impressions of personality dimensions such as methodicalness and originality.

In each of these studies, participants were compelled to form conjectural hypotheses about personality from handwriting on the basis of their intuitions because they were not exposed to information about the writer's personality. The practice of graphology, in contrast, is said to be based not on "armchair" speculation but on the observation of empirical associations between handwriting and personality, rigorous study of which dates back to the pioneering 19th century work of the French graphologist Michon. The extent to which practicing graphologists genuinely believe these purported relationships to be predictive is an open question. However, it is clear from previous research that, once established, belief in erroneous theories can endure even in the face of disconfirmatory evidence (Ross, Lepper, & Hubbard, 1975), particularly when those theories are intuitively appealing (Chapman & Chapman, 1967) or are consistent with one's general beliefs and attitudes (Lord, Ross, & Lepper, 1979).

As Ben-Shakhar et al. (1986) have pointed out, graphology "seems to have the right kind of properties for reflecting personality" (p. 176; see also Bar-Hillel & Ben-Shakhar, 1986). Intuitively, because both personality and handwriting differ from person to person, one might be expected to offer insight into the other. Although a similar argument could be made replacing "handwriting" in the previous sentence with "birthdates," "palm lines," or "bumps on the skull," graphology—unlike astrology, palmistry, or phrenology—provides a sample of expressive behavior from which to infer personality (Bar-Hillel & Ben-Shakhar, 1986; Ben-Shakhar, 1989). Graphologists view handwriting as a category of nonverbal behavior that illuminates underlying mental characteristics of the person producing the writing, potentially including characteristics that the writer would prefer not to disclose or perhaps is not even conscious of possessing. This perspective was compelling enough for a pair of eminent psychologists, Gordon Allport and Phillip E. Vernon, to provide graphology an early endorsement as a scientific discipline in their 1933 book *Studies in Expressive Movement.*

The nature of handwriting itself conveys the impression of having potential diagnostic value. That is, handwriting bears the richness of features that would be necessary to reflect the many facets of personality (Ben-Shakhar et al., 1986). Indeed, most graphological services claim to use at least 100 features of handwriting, and one major company, DataGraph, claims to use over 400 features of handwriting to draw inferences about personality.

The intuitive appeal of graphology appears to extend to the specific relationships that are claimed to exist between handwriting and personality. Indeed, with features as rich as the language used to describe personality, possible relationships between handwriting features and personality traits can be easily hypothesized on the basis of intuition alone. Although graphologists are often reluctant to explicate the specific features of handwriting that they consider diagnostic of personality, certain purported relationships that have been made explicit convey the appearance of face validity and seem almost metaphorical in nature. For example, size of handwriting is believed to be diagnostic with respect to the personality dimension of modesty–egotism, with small handwriting implying modesty and large handwriting implying egotism. In many examples like this, the empirical relationships between handwriting features and personality traits identified by graphologists closely parallel semantic associations between words used to describe handwriting features (e.g., *regular* rhythm) and personality traits (e.g., *reliable*).

Research by Chapman and Chapman (1967, 1971) suggests that where semantic relationships such as these exist, the intuitive statistician may infer nonexistent or illusory correlations in the direction dictated by semantic association. Chapman and Chapman (1967) believed that such an effect might account for the persistent use of a popular but invalid diagnostic tool used by clinical

psychologists, called the Draw-A-Person (DAP) Test (Machover, 1949). The DAP is a projective test in which patients are asked to draw a person, and from those drawings, clinicians make inferences about their underlying psychopathology. At the time Chapman and Chapman (1967) wrote their article, there was extensive empirical evidence documenting that the DAP had no predictive validity with respect to clinical diagnosis.

Chapman and Chapman (1967) presented naive judges (i.e., those unfamiliar with the DAP as a clinical tool) with a set of DAP drawings, along with contrived symptom statements describing the patient who provided the drawing. Though symptom statements were uncorrelated with features of the drawing, naive participants perceived illusory correlations between the same semantically related pairs of drawing features and clinical symptoms that clinicians believed to be related. For example, like clinicians, naive participants perceived drawing a big head as correlated with concerns about intelligence and elaboration of the eyes as correlated with paranoia.

Other research also suggests that intuitive assessments of statistical relatedness tend to be influenced by semantic association. Kahneman and Tversky (1973) demonstrated that the judged likelihood of an event is often assessed by how closely the event resembles a prototype. In a related vein, Shweder and D'Andrade (1980) have argued that perceived trait relationships that constitute lay theories of personality are determined by the conceptual similarity of traits rather than by their statistical associations. Similarity-based judgments have been observed in areas as diverse as clinical diagnosis (Chapman & Chapman, 1967, 1969; Dowling & Graham, 1976), personality-trait inference and social judgment (Bar-Hillel & Neter, 1993; Koehler, Brenner, Liberman, & Tversky, 1996; Shweder & D'Andrade, 1980), and predictions of organizational traits (Camerer, 1988). Consequently, people sometimes report seeing systematic statistical relationships where none exist (Chapman & Chapman, 1967, 1969; Gilovich, Vallone, & Tversky, 1986; Redelmeier & Tversky, 1996).

Beyond the observation that judgments of statistical association appear to be influenced by semantic association, Chapman and Chapman (1967) did not offer a detailed process account of the illusory correlation phenomenon, which is not surprising because the effect is likely overdetermined. It is possible, however, to identify from previous research a number of potential underlying mechanisms that could contribute to the effect, which can be viewed as one example of the ubiquitous "confirmation bias" (see Nickerson, 1998). As suggested by Chapman and Chapman, when the participant is inspecting the evidence in a search for systematic relationships, semantic association is likely to guide the formulation of hypotheses about what goes with what, producing a kind of expectation. Other potential relationships may not be considered and hence not detected even if they are consistent with the observed evidence (e.g., Beyth-Marom & Fischhoff, 1983; Doherty, Mynatt, Tweney, & Schiavo, 1979; Wason, 1960). When the evidence is inspected in light of semantically determined hypotheses, ambiguous aspects of the evidence (e.g., drawing features) may be interpreted in a manner consistent with the hypothesized relationship (e.g., Frank & Gilovich, 1988; Hastorf & Cantril, 1954; Lord et al., 1979). Cases or instances consistent with the hypothesized relationship, perhaps due to their salience, may be given greater weight in the final assessment of the relationship than disconfirming instances (e.g., Kuhn, Amsel, & O'Loughlin, 1988; Smedslund, 1963). Even in the absence of a biased assessment process, confirming instances may simply be better remembered, with semantic association enhancing both encoding and retrieval of cases consistent with the (semantically based) hypothesized relationship (e.g., Berndsen, van der Pligt, Spears, & McGarty, 1996; Greenwald, Pratkanis, Leippe, & Baumgardner, 1986; Rothbart, Evans, & Fulero, 1979).

Despite its lack of validity, the use of graphology persists. This persistence is likely driven by a host of judgmental biases that result in overestimation of the validity of graphology (for a complete review, see Dean, Kelly, Saklofske, & Furnham, 1992). In this article, we isolate one possible source. Using a paradigm similar to that of Chapman and Chapman (1967), we investigate the phenomenon of illusory correlation as a contributor to the appeal and persistence of graphology's use as a predictor of personality. Our results suggest that the apparent validity of graphology may indeed arise from illusory correlations between semantically related handwriting features and personality traits. Because our experimental methodology differs in some critical respects from that of Chapman and Chapman, as explained in the General Discussion section, the results also attest to the generalizability of the illusory correlation phenomenon.

Overview of Experiments 1 and 2

Participants unfamiliar with the practice of graphology inspected a set of 40 cases, where each case consisted of a handwriting sample and a corresponding personality profile said to belong to the person who gave the writing sample. The participants' task was to inspect the casebook and then judge the relatedness (i.e., degree of empirical association, or correlation) between features of handwriting (e.g., size) and personality dimensions (e.g., modesty–egotism) that comprised the profiles in the casebook.

In Experiment 1, handwriting samples were randomly paired with personality profiles, resulting in negligible correlations between handwriting features and personality traits. Under the null hypothesis, participants' relatedness judgments should be based exclusively on the statistical association between features and traits. The semantic-association hypothesis, in contrast, implies that the relatedness judgments will be biased in the direction of the semantic association between words describing handwriting features and personality traits. Specifically, we expected this phenomenon to extend to the particular handwriting-feature–personality-trait pairs that graphologists claim to be related.

The semantic-association hypothesis was further tested in Experiment 2, in which large, discernible correlations between targeted feature–trait pairs were built into the casebook data set. In one condition, discernible correlations were congruent with graphologists' claims (e.g., large handwriting was correlated $r = .98$ with egotism). In the second condition, the correlations were of the same magnitude, only this time in a direction incongruent with graphologists' claims (e.g., large handwriting was correlated $r = .98$ with modesty rather than egotism). Under the null hypothesis, relatedness judgments should correspond to the data; thus, the perceived association should be of equal magnitude but fall in the opposite direction in the congruent and incongruent conditions. The semantic-association hypothesis predicts that the strength of the perceived association will be greater when the correlation is

congruent rather than incongruent with graphologists' claims, on the assumption that such claims are in fact derived from semantic association.

Feature–Trait Relationships

Through the results of a graphology bookstore survey, we identified and consulted two books (Amend & Ruiz, 1980; Roman, 1996) that bookstore patrons reported to be most valuable in learning about graphology. Examination of these books helped us to identify handwriting features and personality traits reported to be correlated by graphologists. We chose to examine six specific purported relationships that, between them, covered a broad range of handwriting features and personality traits (see Table 1). It is worth noting that some schools of graphology focus on a holistic analysis of a script's "gestalt" and, as such, are less committed to the existence of systematic relationships between specific handwriting features and personality traits, though of course some empirical regularities must exist for any graphological technique to have validity.

For each handwriting-dimension–personality-dimension pair listed in Table 1, the left pole of the handwriting dimension is thought by graphologists to be associated with the left pole of the personality dimension, and the right pole of the handwriting dimension is thought to be associated with the right pole of the personality dimension. For example, in the first case, small handwriting is believed to be associated with modesty, whereas large handwriting is thought to be associated with egotism. The terms used to describe the poles of handwriting (which we refer to as "features") were taken directly from the graphology handbooks, with the exception of compact–expansive and ascending–descending, which we chose as the most appropriate terms to describe these features of handwriting. The expansiveness of a script is one of the handwriting factors examined by Allport and Vernon (1933, p. 110); Roman (1996) also makes reference to the script's "expansion" (p. 125). Roman (1996, pp. 296–297) uses "uphill–downhill" or "rising–falling" instead of our terms, ascending–descending, but we suspect our results would not be much different had we used her terms instead. For reasons elaborated below, our studies also included a seventh personality dimension (cooperative–competitive) that graphologists do not generally associate (and that we suspected would not be strongly associated semantically) with any of the handwriting features we investigated.

Collection of Handwriting Samples

Handwriting samples were collected from University of Waterloo students, who were asked to copy a 143-word cooking recipe in their usual cursive writing. In accordance with the practice of graphology, the sample script contained numbers as well as letters, and all handwriting samples were written on a sheet of unlined paper using a ballpoint pen. Samples in which words were misspelled, omitted, or crossed out or that exceeded one page were excluded. Samples that were printed rather than written in longhand were also excluded. The final collection of 40 handwriting samples, identical in content and differing only in terms of writing style, were photocopied and compiled in a bound casebook, which participants would inspect during the main experiments.

Rating of Handwriting Features

For purposes described later, it was necessary to obtain reliable ratings of each of the 40 handwriting samples in terms of their values on the six handwriting dimensions. Thus, the 40 handwriting samples were independently rated on each of the six handwriting dimensions by 10 judges (Roy N. King, Derek J. Koehler, and 8 department colleagues). Each judge was given the same illustrative example sheet of handwriting features that participants would use in the main experiments and was asked to rate each of the samples on the six dimensions using 7-point scales, labeled at the endpoints with the appropriate feature names (e.g., 1 = *small* and 7 = *large*, for size of handwriting). No further instruction or assistance was given. Mean interrater reliability exceeded $r(10) = .89$ for each of the handwriting dimensions (see Table 2). Average

Table 1
Targeted Pairs of Handwriting and Personality Dimensions, Reported as Related by Graphologists, Along With Mean Semantic Association Ratings and Mean Relatedness Judgments From Experiment 1

Handwriting dimension	Personality dimension	Semantic association	Judged relatedness
Size (small–large)	Modest–egotistical	0.93	1.30
Speed (slow–fast)	Cautious–impulsive	1.95	1.65
Rhythm (regular–irregular)	Reliable–unreliable	1.14	1.28
Shape (angular–rounded)	Analytical–intuitive	0.15	0.37
Spacing (compact–expansive)	Introverted–extroverted	1.01	0.85
Slope (ascending–descending)	Optimistic–pessimistic	0.77	0.76

Note. Data are coded so that positive values indicate response direction consistent with claims of graphologists. The leftmost pole of each handwriting dimension is thought to be indicative of the leftmost pole of the personality dimension paired with it, and the rightmost pole of each handwriting dimension is thought to be indicative of the rightmost pole of the personality dimension paired with it. Cooperative–competitive was added so that the otherwise equal number of personality and handwriting dimensions would not be interpreted as implying a one-to-one mapping.

Table 2
Reliability of Handwriting-Feature Encoding by 10 Raters

Handwriting dimension	*R*(10)
Size (small–large)	.97
Speed (slow–fast)	.89
Rhythm (regular–irregular)	.89
Shape (angular–rounded)	.93
Spacing (compact–expansive)	.92
Slope (ascending–descending)	.97

ratings across judges were computed to provide a single rating for the six handwriting dimensions in each of the 40 samples.

Semantic-Association Task

To assess the degree of semantic association between dimensions of personality and handwriting features, a separate group of participants (N = 80 introductory psychology students drawn from the same population used in the main experiments) made semantic-association judgments between the personality-trait terms and the words used to describe the handwriting features in our main experiments. Participants were told nothing of the main experiments, nor were they told that any of the words were intended to describe features of handwriting; they were simply informed that our interest was in "the perceived similarity in meaning between words used to describe people and words used to describe objects." In this manner, semantic-association ratings were elicited for each of the 168 possible combinations of handwriting features and personality traits.

The required initial set of 168 feature–trait semantic-association judgments was distributed equally among four questionnaire forms, to which participants were randomly assigned. On each form, 42 pairwise judgments were made between seven personality traits with six features of handwriting, subject to the constraint that only a single pole from each dimension appeared on a particular form. To achieve this, we broke the seven personality dimensions into their 14 poles (which we refer to as "traits") and randomly assigned them to Trait List 1 or Trait List 2, under the constraint that no two poles of a personality dimension could appear on the same trait list. An identical procedure was used for the six handwriting features, dividing their 12 poles into Feature List A and Feature List B, under the constraint that no two poles of a handwriting dimension could appear on the same feature list. Four questionnaire forms were created by combining the trait and feature lists (1A, 1B, 2A, 2B). Items on the questionnaire were arranged so that on each page, participants made judgments about the semantic association of a single personality-trait term with each of six words used to describe handwriting features.

To obtain direct measures of semantic association and to make the semantic-association task as distinct as possible from the relatedness task of the main experiments, we used an elicitation technique different from that of the main experiments. In the main experiments, participants were asked to make relatedness judgments between handwriting features, such as fast handwriting, and dimensions of personality, such as cautious–impulsive. Thus, one of the 84 relatedness judgments concerned the standing of a person with fast handwriting on the cautious–impulsive dimension, with larger numbers indicating greater impulsivity. For the semantic-association task, in contrast, rather than making association judgments about the standing of fast on the cautious–impulsive dimension, participants were asked to judge directly the association either between fast and cautious or fast and impulsive. Responses to these two scales could then be pooled to derive a semantic-association value for fast with cautious–impulsive. Semantic-association judgments were made on 7-point scales ranging from −3 (*opposite in meaning*) through 0 (*unrelated in meaning*) to +3 (*similar in meaning*).

For example, the average semantic-association judgment for the pair fast and cautious was −2.05, indicating that, to this degree, they were perceived as having opposing meanings. The average judgment for the pair fast and impulsive was 1.75, indicating that, to this degree, they were perceived as similar in meaning. By negating the fast and cautious mean association judgments and aggregating them with the fast and impulsive association judgments, a grand mean of 1.90 is derived for the relationship of fast with cautious–impulsive, indicating that fast was perceived as more semantically related to impulsive than to cautious. Using this process, we reduced the original data set of 168 average feature–trait semantic-association judgments to a set of 84 handwriting-feature–personality-dimension association scores. In doing so, the semantic-association judgments were rendered comparable with the relatedness judgments made in the main experiments.

Experiment 1

Method

Participants. The 78 participants were enrolled in introductory psychology and received credit for their 1-hr participation. Before the experiment, participants completed a questionnaire in which they were asked if they were familiar with handwriting analysis, and, if so, to report any handwriting-feature–personality-trait pairs they had learned or believed to be related. No participant reported familiarity with any of the pairs about which they would be asked to make relatedness judgments. Hence, all participants were regarded as essentially naive to graphology's claims regarding the feature–trait relationships of interest in our study.

Materials. For each of the 40 handwriting samples, an accompanying personality profile was fabricated using the six personality dimensions given in Table 1, plus a seventh personality dimension, cooperative–competitive. The seventh dimension was added so that the otherwise equal number of personality and handwriting dimensions would not be interpreted by participants as implying a one-to-one mapping of handwriting features to personality traits.

Each of the seven personality-dimension scores that comprised a writer's profile were generated randomly, under the constraint that intuitively related personality dimensions (e.g., cautious–impulsive with reliable–unreliable) would be moderately intercorrelated in the expected direction, with magnitudes in the range of r = .1 to r = .4. This procedure was used to limit the co-occurrence within a profile of intuitively incompatible traits (e.g., impulsive and reliable) that might lead participants to question the authenticity of the profiles.

Writer's scores on each personality dimension were converted to percentiles, representing a writer's score on a personality scale, relative to other writers. Scores were rounded to the nearest 10 and presented on an 11-point scale. For example, in providing the writer's relative standing on the modesty–egotism dimension, participants were presented with the writer's percentile score on a scale ranging from 0 (*most EGOTISTICAL [least modest]*) through 50 (*average modesty–egotism*) to 100 (*most MODEST [least egotistical]*) in increments of 10 percentile points. The writer's

supposed percentile score was hand-circled on the scale. Following the random generation procedure, three salient profiles that included three or more dimensions at percentiles of 0 or 100 were adjusted slightly to make them less distinctive.

The 40 resulting personality profiles were randomly paired with the 40 handwriting samples and presented alongside one another in a casebook, such that, on opening the first page of the casebook, a handwriting sample stamped "Case 1" would appear on the left page and a personality profile, consisting of seven personality dimension scores, said to belong to the writer (also stamped "Case 1"), would appear on the right page. By randomly pairing handwriting samples with personality profiles, any correlation between handwriting features and personality dimensions would be purely incidental. Indeed, the average correlation between handwriting features (taken from the average responses of our raters) and personality scores was zero, with a standard deviation of .15. None of the incidental correlation magnitudes exceeded $r = .40$. In addition, imperfect reliability in perceptual encoding of the handwriting features attenuates, to some extent, the apparent correlation between features and personality dimensions perceived by an individual judge. (This point is considered in more detail later.)

A second form of the casebook (Form B) was created using reversed profile scores from the first casebook (Form A). Applying to all Form A profile scores the formula $100 - n$, where n represents the percentile scores of Form A, the resulting scores yielded complementary personality profiles for the Form B casebook. By pooling judgments across the counterbalanced Forms A and B, any effects of within-form incidental correlations between personality scores and handwriting features on the relatedness judgments should average to zero under the null hypothesis that participants are influenced only by statistical—and not semantic—association.

Procedure. Participants were randomly assigned to one of the two forms of the casebook and were tested in groups. After completing the preexperimental questionnaire measuring familiarity with graphology, handwriting analysis was described to participants as a practice "used to assess people's personality traits by examining specific features of their handwriting." Participants were provided with a list of "some handwriting features of interest to graphologists," along with short, illustrative examples of each handwriting feature targeted in this study.[1] In most cases, these examples were adapted from those found in the Amend and Ruiz (1980) graphology handbook. Participants were warned that "some or all of these features of handwriting may be entirely unpredictive of personality characteristics, but these features are among the ones that interest graphologists."

Next, participants were presented with the seven dimensions of personality of interest in the study, on a separate sheet of paper that they could refer back to throughout the task. Brief definitions of each personality-dimension pole were included as follows: modest (aware of one's own limitations; humble) versus egotistical (vain, conceited); cautious (prone to act only after reflection) versus impulsive (prone to act without reflection); reliable (responsible, dependable) versus unreliable (irresponsible, undependable); analytical (inquiring, logical, problem-solving) versus intuitive (irrational; acts on hunches); introverted (prefers to be alone) versus extroverted (prefers to be with others); pessimistic (tends to take the least hopeful view of things) versus optimistic (tends to take the most hopeful view of things); and cooperative (prefers to work with others toward common end) versus competitive (prefers to compete against others). Again, participants were cautioned that these personality dimensions "may or may not be predictable from a person's handwriting."

Before reading the printed instructions outlining their task, participants were given a verbal overview by the experimenter. In the printed instructions, participants were led to believe that in previous research, a group of 40 people had provided us with a sample of their handwriting and had completed a personality inventory made up of a number of standardized tests, from which the personality profiles had been derived. Participants were referred to the casebook in which we had compiled the 40 handwriting samples and personality profiles of the writers. The concept of percentile scores was described to participants using examples from the casebook. Participants were then instructed as follows:

> The casebook includes the handwriting sample and corresponding personality profiles for each of the forty people from the group we studied. *We would like you to examine this casebook and, on the basis of your inspection, evaluate whether particular handwriting features appear to be indicative of certain personality features.* For example, consider the size of a person's handwriting (large vs. small). Do people with large handwriting differ in certain personality traits from people with small handwriting?
>
> You will be asked to make your judgments exclusively on the basis of your inspection of the casebook, and *not* on your own previous experience or opinions. Further instructions on exactly how to make your judgments will be provided after you have inspected the casebook. *You will have 15 minutes to inspect the casebook.*
>
> As you read through the casebook, you should keep your task in mind, because once you have completed your inspection you will not be allowed to re-inspect the casebook when making your judgments about the relationship between handwriting features and personality traits. Because we are interested in your general impressions, note taking or writing in the casebook or on any other materials is not permitted. While inspecting the casebook, however, you should feel free to turn back to earlier pages to review previous cases. In short, you should use the inspection period to form ideas about "what goes with what," that is, about which, (if any) handwriting features are associated with which personality traits.

Participants were also given a preview of the judgments they would be asked to make by referring them to the judgment task binder where they would make their responses following their inspection of the casebook. They were not, however, allowed to record any judgments during the inspection period.

We settled on a 15-min inspection period on the basis of the results of informal pilot testing. Participants given 15 min to inspect the casebook did not complain about not having enough time, and we feared that a longer inspection period might produce boredom or fatigue. Chapman and Chapman (1967) reported that the illusory correlation effect they obtained in their initial experiment, in which the patient drawings were presented one at a time for 30 s each, was not reduced at all in a second experiment in which the inspection process was carried out once per day over 3 consecutive days, suggesting that limited inspection time does not exaggerate the size of the illusory correlation effect. A subsequent experiment (Chapman & Chapman, 1967, Experiment 5) in which participants could view each drawing for as long as they wished also produced no attenuation of the illusory correlation effect.

While inspecting the casebook, participants were warned when 10, 5, and 1 min remained in the inspection period. At the end of the inspection period, the casebooks were collected from the participants. The judgment task booklets were then redistributed.

Although we are interested in the perceived correlation between handwriting dimensions (e.g., size) and each personality dimension, eliciting such judgments directly is complicated because it would require the judge to indicate the direction as well as the magnitude of the perceived statistical association. To simplify the task, we elicited relatedness judgments separately for each pole or handwriting feature (e.g., small and large). That is, judgments were decomposed and conditioned on handwriting features. An alternative approach would be to condition on personality traits rather than handwriting features. We chose to condition on handwriting features because the resulting judgment scale is anchored on personality traits, which

[1] A copy of the list is available upon request from Derek J. Koehler.

is more intuitively compatible with the scales presenting personality profile information in the original casebook.

As an example (one presented to participants as part of the judgment instructions), consider the assessment of a writer's standing on the cooperative–competitive dimension given that the writer has small handwriting. This item would be headed, "COOPERATIVE–COMPETITIVE: relative to other writers in this sample, people with *small* handwriting tend to be. . .." Participants would make such an assessment on a 9-point scale centered at zero (*average cooperativeness–competitiveness*). Scale values increased from 1 to 4 in both directions from the center scale value of 0, with the far-left value on the scale labeled *most COMPETITIVE* (*least cooperative*) and the far-right value labeled *most COOPERATIVE* (*least competitive*). This scale was designed to maximize compatibility with the scale on which writers' percentile scores were presented in the casebook. Presented with this sample judgment scale, participants were instructed:

> This question asks whether the casebook samples of small handwriting are associated with the personality dimension of cooperative–competitive. If, upon reviewing the casebook, you felt that people with small handwriting were no more cooperative/competitive than other writers in the casebook, you would circle the number "0," labeled "average cooperativeness/average competitiveness." If you decided that people with small handwriting were the *most cooperative* writers in the casebook, you would circle the "4" labeled "most COOPERATIVE (least competitive)" at the end of the arrow pointing right. Conversely, if you decided that the people with small writing were the *most competitive* writers in the casebook, you would circle the number "4" labeled "most COMPETITIVE (least cooperative)" at the end of the arrow pointing left.

On each page in the judgment task booklet, participants made relatedness judgments about a single handwriting feature's relationship with each of the seven dimensions of personality. For example, on the page for which the scale given above would appear, the phrase "small handwriting" was printed across the top of the page. The page would include seven judgment scales, one for each of the seven personality dimensions. Participants were given unlimited time to complete the judgment task.

Results

For each relatedness judgment, there are two possible predictors: the actual correlation in the casebook, which in this study is just an incidental result of the random pairing procedure, and the semantic association between the trait and feature, as rated by our separate group of participants given the semantic-association judgment task.

We used hierarchical regression, regressing statistical association first and semantic association second as predictors of the 84 average handwriting-feature–personality-dimension relatedness judgments, separately for each form. As can be seen in Table 3, incidental correlations in the data set accounted for nonsignificant, near-zero variance in relatedness judgments for each form. Semantic association accounted for an additional 39% of the variance in Form A judgments ($p < .01$) and an additional 44% of the variance in Form B judgments ($p < .01$). Under the null hypothesis, which assumes that people are attuned only to statistical association, the semantic-association measure should not account for any variance.

The bivariate correlation between the 84 average judgments on the two forms provides a simple analysis of participants' response tendencies. Because the Form B incidental feature–trait correlations actually present in the casebook were reversed relative to those of Form A, data-consistent judgments would imply a negative correlation between judgments from the two forms. However, because the variance in the incidental correlations was small ($SD = .145$), data-consistent responding might similarly produce limited variability. This would include the case of responding with a judgment of 0 (i.e., unrelated) to most of the 84 judgments, a response pattern arguably consistent with the casebook data, and one which would produce an inter-form correlation near zero. The null hypothesis that judgments are based exclusively on statistical association, then, implies that the correlation between judgments from the two forms should be negative or zero. Instead, the correlation is positive: $r = .83$. This result implies a systematic bias in the judgments, unrelated to the actual correlations in the casebook, which depends only on the feature–trait pair being evaluated.

The high correlation in judgments between the two forms permitted aggregation across forms for further analyses of feature–trait relatedness judgments. In this process, any variance in judgments attributable to the actual statistical association in the casebooks is essentially eliminated. Indeed, a strict interpretation of the null hypothesis would lead to a prediction that there should be no (explainable) variance in the resulting set of mean judgments, which instead should all be at or near zero. In fact, semantic association accounted for 47% of the (substantial) variance in the 84 mean relatedness judgments when pooled across forms.

More specific tests concern the six targeted handwriting-feature–personality-trait relationships that graphologists report to be correlated. We expected that participants would perceive illusory correlations for the targeted handwriting-feature–personality-dimension pairs because of their semantic association. That is, we expected both the semantic-association ratings and the relatedness judgments to fall in a direction consistent with graphologists' claims. Table 1 shows the relevant means.

Table 3
Summary of Hierarchical Regression Analysis of Relatedness Judgments in Experiment 1

Dependent variable	Predictor	B	SE_B	β	R^2	ΔR^2
Form A ($n = 39$)						
Relatedness judgment	Statistical association	0.22	0.54	.05	.002	
	Semantic association	0.75	0.10	.63**	.396	.394**
Form B ($n = 39$)						
Relatedness judgment	Statistical association	−0.93	0.70	−.14	.021	
	Semantic association	1.04	0.13	.67**	.458	.437**

** $p < .01$.

To derive a single value representing the perceived relationship between a dimension of handwriting and a dimension of personality in Table 1, two transformations were necessary. First, average values for handwriting-feature–personality-dimension pairs were coded as positive when the direction of responding was consistent with the claims of graphologists and coded as negative when the direction of responding was inconsistent with those claims. For example, the average feature–trait relatedness judgment of −1.40 for the judgment pair fast with cautious–impulsive indicated that fast was perceived as related to the right pole (impulsive) of the personality dimension cautious–impulsive. Because this was consistent with graphologists' claims, it was coded +1.40. Likewise, the average feature–trait relatedness judgment of 1.89 to the judgment slow with cautious–impulsive indicated that slow was perceived as related to the left pole (cautious) of the personality dimension cautious–impulsive, again consistent with the claims of graphology. Next, these data were aggregated to produce a single value representing the perceived relationship between a personality dimension and its associated handwriting dimension. For example, by averaging the mean values 1.40 and 1.89 given in Table 1, the resulting mean of 1.65 represents the slow–fast with cautious–impulsive relationship. A similar procedure was used to derive a single feature–trait measure of semantic association.

Table 1 shows the mean relatedness judgments and semantic-association ratings for the six targeted handwriting-dimension–personality-dimension pairs. In all six cases, participants' relatedness judgments fell in a direction consistent with graphologists' claims, despite the lack of any statistical association. Semantic association between the words describing the handwriting features and the personality traits also fell in a direction consistent with graphologists' claims. The rank-order and zero-order correlations between semantic-association and relatedness judgments in the table are $r = .83$ and $r = .91$, respectively. Semantic association, then, appears to be responsible for the direction and magnitude of relatedness judgments regarding the targeted pairs because no actual correlations held between handwriting and personality across the data sets.

The average magnitude of semantic-association ratings for feature–trait pairs that graphologists claim to be correlated ($M = .99$) was significantly greater ($p < .01$) than for those of the remaining feature–trait pairs ($M = .34$). In other words, as expected, the semantic association of the targeted feature–trait pairs reported by graphologists is significantly greater than the chance-level semantic association found in the set of all possible pairwise comparisons among the features and traits we examined.

Discussion

In this experiment, participants inspected a data set in which, as appears to be the case in practice, negligible correlations existed between handwriting features and personality dimensions. They then made judgments about their statistical relatedness. True, incidental statistical association accounted for near-zero variance in the relatedness judgments, whereas semantic association, as expected, was predictive of these judgments. Though entirely uncorrelated when pooled across counterbalanced forms of casebooks, mean judged relatedness between handwriting features and personality traits fell in a direction consistent with their semantic association. This perception of uncorrelated variables as empirically related constitutes what Chapman and Chapman (1967) referred to as illusory correlation.

Of the 84 judgments made by participants in this study, we were particularly interested in the targeted feature–trait pairs that graphologists (Amend & Ruiz, 1980) claim to be related. Naive judges, who expressed no prior intuitions about these relationships, "discovered" the same relationships as those identified by graphologists (see Table 1), despite the absence of any reliable statistical association in the casebooks. For example, like graphologists, naive judges consistently reported a relationship between ascending handwriting and optimism even though these dimensions were uncorrelated across the data sets. Semantic association between trait terms (e.g., *optimistic*) and words used to describe handwriting (e.g., *ascending*) was identified as the likely source of the illusory correlations experienced by our participants.

The negligible feature–trait correlations in the casebooks of Experiment 1 are consistent with what research evidence suggests to be their actual magnitude. Although, in this sense, random pairing rendered the contrived relationships ecologically valid, the lack of any feature–trait relationships that might be discernible to the human judge introduces the possibility of an alternative explanation of the results. It could be argued that a participant, expecting to discern correlations in the casebook, might have responded in a direction consistent with the semantic-association hypothesis despite not actually having perceived any correlations in the casebook.

This alternative explanation for the results of Experiment 1 relies rather heavily on the assumption that participants intentionally violated the task instructions to (a) respond with a 0 (i.e., unrelated) judgment if they did not perceive a relationship between a handwriting feature and personality dimension and (b) base their judgments solely on the data presented in the casebook rather than on opinion or intuition. Despite such instructions, however, it may conceivably be difficult in practice for a participant who genuinely detected no relationships to indicate this perception by giving a response of zero to each and every one of the 84 relatedness judgments. Put differently, some participants may have inferred from the nature of the task and the number of judgments they were asked to make that at least some predictive relationships must have existed in the data set, even if they failed to perceive them. Thus, feeling compelled to report perceiving at least some correlations, such a participant might have provided non-zero responses, even though in truth he or she perceived no relationships in the data.

Although such a response tendency would misrepresent the perceptions of an individual participant, it would not present a serious methodological problem, provided that judgments deviating from zero on the response scale were made on an arbitrary basis. Such a response pattern would simply introduce random error, the expected value of which would be consistent with the null hypothesis. However, if participants who perceived no relationships in the data chose to deviate from zero in a systematic manner on the basis of semantic association, their data would be apparently consistent with the semantic-association hypothesis.

Experiment 2 was designed to address the possibility that such demand characteristics might account for the results of Experiment 1. Given that the source of the possible demand characteristics is the absence of discernible relationships in the casebooks, in Experiment 2, large, discernible relationships were set between targeted handwriting-feature–personality-dimension pairs.

Experiment 2

Method

Participants. The 58 participants were enrolled in an introductory psychology course and received credit for their 1-hr participation. As in the first experiment, participants completed a questionnaire concerning their familiarity with graphology. None of the participants reported familiarity with the feature–trait pairs of interest in the study.

Procedure. The design and procedure of Experiment 2 were identical to those of Experiment 1, with the exception of the construction of the personality profiles. Using our raters' average feature ratings for the 40 handwriting samples, profiles were constructed and paired with samples such that correlations near unity were set for four of the six handwriting-dimension–personality-dimension pairs claimed to be related by graphologists (for a total of eight targeted handwriting-feature–personality-dimension pairs). Personality profile scores were derived under the constraint that the four targeted feature–trait correlations had to exceed $r = .95$, while still preserving the intuitive direction and approximate magnitude of the intercorrelations among personality dimensions found in the profiles of Experiment 1. The direction of the correlations were set such that two of the four relationships in the casebook data set would be congruent with graphologists' claims (and, hence, semantic association), and the other two would be incongruent with graphologists' claims. These four targeted handwriting-dimension–personality-dimension pairs and their near-unity correlations are listed in Table 4. As in Experiment 1, imperfect reliability in perceptual encoding of the handwriting features attenuates, to some extent, the magnitude of the perceived correlation between features and personality dimensions for an individual judge. Resulting incidental correlations between the remaining targeted pairs (angular–rounded with analytical–intuitive, and regular–irregular with reliable–unreliable) are also listed in Table 4; these values are non-zero as a consequence of the personality dimension intercorrelations built into the profiles.

The same method used in Experiment 1 for counterbalancing forms of casebooks was also used in Experiment 2. In this experiment, counterbalancing also influences congruency with graphologists' claims, such that when the targeted relationships were congruent on one form, they would be incongruent on the opposite form. Correlations among the remaining handwriting-feature–personality-dimension pairs were purely incidental consequences of the score-construction procedure outlined above. Because of the increased constraints placed on the profile-construction procedure in Experiment 2, however, the typical magnitude of the incidental correlations was somewhat larger in Experiment 2 (mean $r = .232$) than in Experiment 1 (mean $r = .145$). Given at least some degree of unreliability of an individual's encoding of handwriting features, though, we doubt that these somewhat larger incidental correlations were substantially more discernible than those found in Experiment 1. In any case, they still fall well below the level generally required to be perceived by an individual judge (Jennings et al., 1982).

It is important to recall the alternative explanation for the results obtained in Experiment 1: Given no actual statistical associations, the participant perceives no systematic relationships but nonetheless feels compelled to report some non-zero correlations and does so on the basis of intuitive judgments driven by semantic association. In Experiment 2, the participant who is compelled to provide non-zero responses now has the opportunity to do so, provided by the actual large discernible statistical relationships present in the casebook. Under the null hypothesis that perceived correlation depends only on statistical—and not semantic—association, the strength of the perceived relationship should not depend on whether the feature–trait pair under evaluation is semantically incongruent (e.g., large handwriting being associated with modest in Form A) or semantically congruent (e.g., large handwriting being associated with egotistical in Form B). Thus, the null hypothesis predicts that relatedness judgments should be consistent with correlations in the casebook data set, equal in magnitude but opposite in sign for the two forms.

The semantic-association hypothesis, in contrast, implies that—holding actual statistical association constant—the perceived strength of semantically congruent relationships (e.g., large handwriting correlated with egotistical in Form B) will be enhanced relative to that of semantically incongruent relationships (e.g., large handwriting correlated with modest in Form A). Thus, we expect that the strength of the perceived relationship between semantically congruent pairs will be greater than that between semantically incongruent pairs, despite their equivalence in terms of statistical association.

Results and Discussion

Once again, for each relatedness judgment, there are two possible predictors: the actual correlation in the casebook and the semantic association between the trait and feature. Unlike Experiment 1, in which all feature–trait correlations were incidental and arguably indiscernible, in Experiment 2, the correlations for four targeted handwriting-dimension–personality-dimension relationships were set near unity. These statistical associations were expected to be discernible, and thus, statistical association was expected to account for greater variance in the relatedness judgments than it did in Experiment 1. Hierarchical regression was used in the same manner as in Experiment 1, entering handwriting-feature–personality-dimension statistical correlations first and semantic association second to fit the 84 average handwriting-feature–

Table 4
Correlations and Congruity of Targeted Relationships Built Into Casebook Data Sets of Experiment 2

Feature–trait relationship	Correlation magnitude	Casebook	
		Form A	Form B
Correlations set near unity			
Small–large with modest–egotistical	.98	Incongruent	Congruent
Compact–expansive with introverted–extroverted	.98	Incongruent	Congruent
Slow–fast with cautious–impulsive	.98	Congruent	Incongruent
Ascending–descending with optimistic–pessimistic	.95	Congruent	Incongruent
Correlations for remaining target pairs			
Angular–rounded with analytical–intuitive	.19	Incongruent	Congruent
Regular–irregular with reliable–unreliable	.29	Congruent	Incongruent

Table 5
Summary of Hierarchical Regression Analysis of Relatedness Judgments in Experiment 2

Dependent variable	Predictor	B	SE_B	β	R^2	ΔR^2
Form A (n = 29)						
Relatedness judgment	Statistical association	0.99	0.18	.52**	.265	
	Semantic association	0.57	0.10	.48**	.490	.225**
Form B (n = 29)						
Relatedness judgment	Statistical association	0.78	0.27	.31**	.095	
	Semantic association	1.02	0.12	.64**	.508	.413**

** $p < .01$.

personality-dimension judgments separately for each form (see Table 5). As expected, statistical association accounted for significant variance on the first step for both forms, though its influence was more pronounced for Form A than for Form B. As in Experiment 1, for both Form A and Form B, semantic association accounted for significant additional variance in relatedness judgments even after statistical association had been taken into account.

We hypothesized that data-consistent responding would be greater when statistical relationships in the casebook data set were congruent with semantic association (and graphologists' claims) than when they were incongruent with semantic association. To test this, for each of the eight targeted pairs, we coded data-consistent judgments as positive and data-inconsistent responses as negative and then compared the resulting mean judgments in the congruent and incongruent conditions. Means for each of the targeted pairs are presented in Table 6. For each pair, results are consistent with the semantic-association hypothesis: Data-consistent responding (denoted by positive values) was greater when statistical relationships were congruent with semantic association than when they were incongruent. As an omnibus test of this effect, each participant's mean judgment over the four targeted congruent pairs was computed and compared with his or her mean judgment over the four targeted incongruent pairs in a repeated-measures analysis of variance with form (A vs. B) as an additional between-subjects factor. This analysis revealed a statistically significant effect of the congruency manipulation, $F(1, 56) = 67.7$, $MSE = 6.46$, $p < .001$. There was also a statistically significant congruency by form interaction, $F(1, 56) = 10.1$, $MSE = 6.46$, $p < .01$, arising from judgments involving one particular feature–trait pair, which is discussed below.

The effect of semantic association on judgments of targeted pairs held even though participants were generally able to identify the direction of the actual statistical association for these pairs in the casebook. For example, given a casebook in which handwriting size was highly correlated with egotism–modesty in the semantically congruent direction, people's judgments were highly responsive to the data, with small handwriting judged as highly associated with modesty and large handwriting highly associated with egotism. In contrast, when given a casebook in which the statistical relationship in the data was semantically incongruent, people's judgments generally indicated the statistically correct (though semantically incongruent) direction of the relationship, but the magnitude of this perceived relationship was much attenuated in comparison to the semantically congruent condition. Thus, the same underlying mechanism responsible for the illusory correlation observed in Experiment 1, namely semantic association, appears to account for the results of Experiment 2.

One particular feature–trait pair stood out from the rest in this analysis, namely cautious–impulsive with speed (fast vs. slow). Unlike the other judgments, participants' reports of statistical association were in the same direction and nearly as strong in the incongruent condition as in the congruent condition. In other words, even though fast (slow) writing was a nearly perfect pre-

Table 6
Degree of Data-Consistent Responding (After Coding Data-Consistent Judgments as Positive and Data-Inconsistent Responses as Negative) for Each of the Eight Targeted Pairs in Experiment 2

Trait dimension	Handwriting feature	Congruent	Incongruent	Difference
Cautious–impulsive	Fast	2.14	−1.55	3.69**
Cautious–impulsive	Slow	1.79	−1.59	3.38**
Pessimistic–optimistic	Ascending	0.93	−0.17	1.10†
Pessimistic–optimistic	Descending	1.14	0.21	0.93
Introverted–extroverted	Compact	2.14	0.45	1.69**
Introverted–extroverted	Expansive	1.66	−0.21	1.86**
Modest–egotistical	Small	1.83	0.55	1.28†
Modest–egotistical	Large	2.03	0.45	1.59*

Note. Values under the congruent and incongruent columns are coded so that positive values indicate judgments consistent with data (which, in the congruent condition, also implies judgments consistent with semantic association), and negative values indicate judgments inconsistent with data.
† $p < .10$. * $p < .05$. ** $p < .01$, by two-tailed t test.

dictor of a cautious (impulsive) personality in the casebook data set, participants reported perceiving a relationship in the opposite (semantically congruent) direction that was almost as strong as that reported by participants in the semantically congruent condition. Two factors may account for this anomaly. First, of all the targeted pairs, the semantic association was greatest for cautious–impulsive with speed (fast vs. slow). Second, the speed dimension was the most ambiguous of the handwriting dimensions, indicated by its lower reliability of encoding by our raters (see Table 2). This ambiguity may have enhanced the illusory correlation by providing greater room for interpretation to the judge, who might have a tendency to confirm semantically guided expectations.

The observation that statistical association accounted for less variance in Form B relatedness judgments than in Form A can be attributed to the presence of the incongruent relationship between speed (fast vs. slow) and the cautious–impulsive personality dimension on Form B. For these pairs, participants reported positively correlated variables (e.g., fast with cautious) as negatively correlated, diminishing the overall variance accounted for by actual statistical association in Form B judgments. The finding that, with this one exception, participants were generally able to detect actual statistical relationships in the casebook suggests that there was nothing about our experimental task (e.g., limited inspection period, casebook layout, judgment format) that precluded identification of valid statistical associations. Despite this, semantic association systematically influenced participants' relatedness judgments.

General Discussion

In Experiment 1, negligible correlations existed between handwriting features and personality traits in the casebooks provided as data to our participants. Despite the lack of any statistical association in the casebooks, participants reported detection of systematic relationships between handwriting features and personality dimensions. Consistent with our predictions, these relatedness judgments were biased by the direction and strength of semantic association between words used to describe handwriting and personality traits. This effect was particularly pronounced for the six targeted feature–trait pairs we investigated that graphologists claim to be related. In Experiment 2, large, discernible correlations were built into the casebook data sets, such that each participant was presented with two relationships that were congruent with graphologists' claims (and semantic association) and two that were incongruent with graphologists' claims. Consistent with the semantic-association hypothesis, relatedness judgments were greater for semantically congruent pairs than for semantically incongruent pairs, despite the equivalence of their statistical association in the casebook. We suggest that such biased assessments of statistical association may help to support and maintain belief in the validity of graphology, both by graphologists and by their clients.

Like Chapman and Chapman (1967), we found that naive judges "discovered" relationships in the data consistent with those claimed to exist by experts. The obvious implication is that semantic association, which appears to drive illusory correlation in both domains, is the origin of experts' theories. In reference to the resemblance between diagnostic relationships reported by clinicians and illusory correlations perceived by naive judges, Chapman and Chapman suggested that "this striking similarity leads one to suspect that many clinical interpretations of the DAP also have their genesis in illusory correlation arising from associative connection" (p. 203). Semantic association may guide initial hypotheses regarding potential diagnostic relationships, which then receive the appearance of empirical support through biased assessments of correlation.

Although the research of Chapman and Chapman (1967) served as the inspiration for our investigation, the methodology of the present experiments does differ in some important ways from that of their original studies. First, we obtained judgments from participants regarding the magnitude as well as the direction of the perceived association between handwriting features and personality traits. Such judgments were requested for every possible handwriting-feature–personality-dimension combination. By contrast, Chapman and Chapman simply asked participants to name any drawing features they recalled having been associated with a designated patient symptom in an open-ended question format. Consequently, Chapman and Chapman had no measure of the strength of perceived association between DAP features and patient symptoms. Our experimental design allows a stronger test of the hypothesis that perceived statistical association is substantially influenced by semantic association, by assessing that relationship over a large set of handwriting-feature–personality-trait pairs varying widely in both semantic relatedness and actual statistical association, allowing more precise measurement of their relative contributions to perceived statistical association.

We also indicated to participants the specific handwriting features that would be the focus of subsequent judgments before presenting them with the casebook evidence, accompanied by a warning that "some or all of these features of handwriting may be entirely unpredictive of personality characteristics." By contrast, Chapman and Chapman (1967) investigated those drawing features nominated by their participants following the inspection period and requested such identification in a manner that arguably implied the existence of at least some systematic relationships in the data set. The fact that our results coincide so closely with those of Chapman and Chapman despite such methodological differences attests to the robustness and generalizability of the illusory correlation phenomenon. Taken together, the results indicate that people are subject to illusory correlations whether or not their attention is drawn in advance to critical features of the evidence and whether or not they are explicitly warned that there may be no systematic relationships in the data set.

The illusory correlation phenomenon may contribute to the persistent use of graphology despite substantial empirical evidence that casts doubt on its validity. If the graphologist or client has information regarding the writer's personality independent of what can be gleaned from his or her handwriting, they will tend to perceive systematic relationships between the writer's handwriting and personality even in the absence of any statistical association.

Of course, unlike in our experiment, graphologists and their clients do not generally have access to a precise measure of the writer's standing on each in a set of designated personality dimensions. That is, graphologists do not typically administer standardized personality inventories to compare their script analysis to known personality traits of the writer. In this respect, the task we

presented to our participants differs significantly from the everyday experience of the graphologist. It is the case, though, that graphologists generally prefer spontaneous scripts, often including autobiographical information that presumably would convey some information about the personality of the writer. Graphologists are also encouraged to use any other personal information regarding the writer that they can obtain in carrying out their analysis (Tripician, 2000). Furthermore, the client—if not the graphologist—is likely to have an independent source of information about the writer's personality.

Even in the absence of any independent information about the writer's personality, we suggest, the phenomenon of illusory correlation may still play a role in guiding and validating the graphologist's assessments by creating expectations regarding intercorrelations among handwriting features. For example, the cooccurrence of writing with regular rhythm and with precisely dotted *i*'s and crossed *t*'s might be overestimated by the graphologist because of the semantic association of the traits they are believed to diagnose, reliability and conscientiousness, respectively. Through this process, the graphologist would experience a false sense of convergent validity, indirectly guided by semantic association, even if he or she never had any access to personality information regarding the writer. An interesting consequence of this process is that the client tends to receive a coherent personality profile from the graphologist, which is likely to increase confidence in the graphologist's assessments (Kahneman & Tversky, 1973).

Because our investigation did not involve trained graphologists as participants, our results most clearly pertain to the question of why clients fail to detect the apparent lack of predictive validity on the part of the graphologists whose services they employ. As elaborated above, the illusory correlation phenomenon may very well contribute to the continued use of graphology, but we doubt that it is the only factor sustaining belief in its validity. Instead, multiple judgmental mechanisms are almost certainly involved (Ben-Shakhar et al., 1986; Dean et al., 1992). For example, in assessing judgmental accuracy, people often tend to overlook the importance of the judgment's diagnostic value (Beyth-Marom & Fischhoff, 1983; Doherty et al., 1979; Nickerson, 1998; Skov & Sherman, 1986). In the case of graphology, this amounts to failing to consider the extent to which a personality profile derived from graphological assessment fits people other than the writer. Barnum statements consist of sets of universal truths about personality that, while often describing an individual quite accurately, describe everybody else equally well. For example, the diagnosis that "at times you are extroverted, affable, sociable, while at other times you are wary, and reserved" (Forer, 1949) can be quite compelling to the assessee, at least until he or she considers how well the statement applies to others. McKelvie (1990) gave 108 students identical generic diagnoses of personality, which they believed had been ascertained from their handwriting. After reading their diagnoses, the students' belief in graphology was significantly strengthened.

In our research, we investigated just one of several judgmental mechanisms that may serve to maintain clients' confidence in the validity of graphology. Despite considerable—perhaps sufficient—research to discredit it, the continued belief in and use of graphology is worthy of further investigation. This suggests a reconsideration of the usefulness of graphology. Although it appears to be of dubious value when used for personality assessment, its use as a setting for the study of human judgment and decision making remains promising.

References

Allport, G. W., & Vernon, P. E. (1933). *Studies in expressive movement.* New York: MacMillan.

Amend, K., & Ruiz, M. S. (1980). *Handwriting analysis: The complete basic book.* North Hollywood, CA: Newcastle.

Bar-Hillel, M., & Ben-Shakhar, G. (1986). The a priori case against graphology. In B. Nevo (Ed.), *Scientific aspects of graphology* (pp. 263–279). Springfield, IL: Charles C Thomas.

Bar-Hillel, M., & Neter, E. (1993). How alike is it versus how likely is it: A disjunction fallacy in stereotype judgments. *Journal of Personality and Social Psychology, 65,* 1119–1131.

Bayne, R., & O'Neill, F. (1988). Handwriting and personality: A test of some expert graphologists' judgments. *Guidance and Assessment Review, 4,* 1–3.

Ben-Shakhar, G. (1989). Non-conventional methods in personnel selection. In P. Herriot (Ed.), *Assessment and selection in organizations* (pp. 469–485). Chichester, England: Wiley.

Ben-Shakhar, G., Bar-Hillel, M., Bilu, Y., Ben-Abba, E., & Flug, A. (1986). Can graphology predict occupational success? Two empirical studies and some methodological ruminations. *Journal of Applied Psychology, 71,* 645–653.

Berndsen, M., van der Pligt, J., Spears, R., & McGarty, C. (1996). Expectation-based and data-based illusory correlation: The effects of confirming versus disconfirming evidence. *European Journal of Social Psychology, 26,* 899–913.

Beyth-Marom, R., & Fischhoff, B. (1983). Diagnosticity and pseudodiagnosticity. *Journal of Personality and Social Psychology, 45,* 1185–1195.

Bruchon-Schweitzer, M., & Ferrieux, D. (1991). Une enquete sur le recrutement en France [A study of recruitment in France]. *European Review of Applied Psychology, 41,* 9–17.

Camerer, C. (1988). Illusory correlations in perceptions and predictions of organizational traits. *Journal of Behavioral Decision Making, 1,* 77–94.

Chapman, L. J., & Chapman, J. P. (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology, 72,* 193–204.

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology, 74,* 271–280.

Chapman, L. J., & Chapman, J. P. (1971, November 18–22). Test results are what you think they are. *Psychology Today,* 106–110.

Dean, G. A. (1992). The bottom line and effect size. In B. L. Beyerstein & D. F. Beyerstein (Eds.), *The write stuff: Evaluations of graphology, the study of handwriting analysis* (pp. 269–341). Buffalo, NY: Prometheus Books.

Dean, G. A., Kelly, I. W., Saklofske, D. H., & Furnham, A. (1992). Graphology and human judgment. In B. L. Beyerstein & D. F. Beyerstein (Eds.), *The write stuff: Evaluations of graphology, the study of handwriting analysis* (pp. 342–396). Buffalo, NY: Prometheus Books.

Doherty, M. E., Mynatt, C. R., Tweney, R. D., & Schiavo, M. D. (1979). Pseudodiagnosticity. *Acta Psychologica, 43,* 111–121.

Dowling, J. F., & Graham, J. R. (1976). Illusory correlations and the MMPI. *Journal of Personality Assessment, 40,* 531–538.

Downey, J. E. (1923). *Graphology and the psychology of handwriting.* New York: World Book Company.

Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire.* London: Hodder & Stoughton.

Forer, B. R. (1949). The fallacy of personal validation: A classroom demonstration of gullibility. *Journal of Abnormal and Social Psychology, 44,* 119–123.

Frank, M. G., & Gilovich, T. (1988). The dark side of self and social perception: Black uniforms and aggression in professional sports. *Journal of Personality and Social Psychology, 54,* 74–85.

Furnham, A. (1988). Write and wrong: The validity of graphological analysis. *The Skeptical Inquirer, 13,* 64–69.

Furnham, A., & Gunter, B. (1987). Graphology and personality: Another failure to validate graphological analysis. *Personality and Individual Differences, 8,* 433–435.

Gilovich, T., Vallone, R., & Tversky, A. (1986). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology, 17,* 295–314.

Goldberg, L. R. (1986). Some informal explorations and ruminations about graphology. In B. Nevo (Ed.), *Handbook of scientific aspects of graphology* (pp. 281–293). Springfield, IL: Charles C Thomas.

Greenwald, A. G., Pratkanis, A. R., Leippe, M. R., & Baumgardner, M. H. (1986). Under what conditions does theory obstruct research progress? *Psychological Review, 93,* 216–229.

Hager, D. (1977, July–August). Pen and inc. *North Liner,* 9.

Hastorf, A. H., & Cantril, H. (1954). They saw a game: A case study. *Journal of Abnormal and Social Psychology, 49,* 129–134.

Hines, T. (1988). *Pseudoscience and the paranormal: A critical examination of the evidence.* Buffalo, New York: Prometheus Books.

Hooper, M. A., & Stanford, K. S. (1992). A script for screening. *Security Management, 36,* 72–77.

James, M., & Loewenthal, K. (1991). Handwriting stereotypes for the judgment of depression. *Journal of Social Psychology, 131,* 749–750.

Jennings, D. L., Amabile, T. M., & Ross, L. (1982). Informal covariation assessment: Data-based versus theory-based judgments. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 211–238). Cambridge, England: Cambridge University Press.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237–251.

Koehler, D. J., Brenner, L. A., Liberman, V., & Tversky, A. (1996). Confidence and accuracy in trait inference: Judgment by similarity. *Acta Psychologica, 92,* 33–57.

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills.* New York: Academic Press.

Loewenthal, K. (1975). Handwriting and self-presentation. *Journal of Social Psychology, 96,* 267–270.

Lord, C. G., Ross, L., & Lepper, M. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37,* 2098–2109.

Machover, K. (1949). *Personality projection in the drawing of the human figure.* Springfield, IL: Charles C Thomas.

McCarthy, M. (1988, August 25). Handwriting analysis as personnel tool. *Wall Street Journal,* p. 17.

McKelvie, S. J. (1990). Student acceptance of a generalized personality description: Forer's graphologist revisited. *Journal of Social Behaviour and Personality, 5,* 91–95.

Mickels, J. R. (1970). Handwriting analysis in personnel selection. *The Office, 72,* 43–44.

Neter, E., & Ben-Shakhar, G. (1989). The predictive validity of graphological influences: A meta-analytic approach. *Personality and Individual Differences, 10,* 737–745.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2,* 175–220.

Osborne, A. S. (1929). *Questioned documents* (2nd ed.). Montclair, NJ: Patterson Smith.

Rafaeli, A., & Klimoski, R. J. (1983). Predicting sales success through handwriting analysis: An evaluation of the effects of training and handwriting sample content. *Journal of Applied Psychology, 68,* 212–217.

Redelmeier, D. A., & Tversky, A. (1996). On the belief that arthritis pain is related to the weather. *Proceedings of the National Academy of Sciences, 93,* 2895–2896.

Roman, K. G. (1996). *Handwriting: A key to personality.* Columbus, OH: SFM Press. (Original work published 1952)

Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology, 32,* 880–892.

Rothbart, M., Evans, M., & Fulero, S. (1979). Recall for confirming and disconfirming events: Memory processes and the maintenance of social stereotypes. *Journal of Experimental Social Psychology, 15,* 343–355.

Shackleton, V., & Newell, S. (1994). European management selection methods: A comparison of five countries. *International Journal of Selection and Assessment, 2,* 91–102.

Shweder, R. A., & D'Andrade, R. G. (1980). The systematic distortion hypothesis. In R. A. Shweder (Ed.), *New directions for methodology of social and behavioral science, 4: Fallible judgement in behavioral research* (pp. 37–58). San Francisco: Jossey-Bass.

Skov, R. B., & Sherman, S. J. (1986). Information gathering processes: Diagnosticity, hypothesis confirmation strategies, and perceived hypothesis confirmation. *Journal of Experimental Social Psychology, 22,* 93–121.

Smedslund, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology, 4,* 165–173.

Tripician, R. J. (2000). Confessions of a (former) graphologist. *Skeptical Inquirer, 24,* 44–47.

Vine, I. (1974). Stereotypes in the judgement of personality from handwriting. *British Journal of Social and Clinical Psychology, 13,* 61–64.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12,* 129–140.

Received June 4, 1999
Revision received May 30, 2000
Accepted June 2, 2000 ■