The Effect of Personality on Romantic Preference: A Computational Linguistic Approach

Caleb Marshall

Dr. Erin M. Buchanan

Dr. Melissa Fallone

Missouri State UniversityAbstract

This study utilized Latent Semantic Analysis, to evaluate the relationship between personality and participants’ written personal account of an ideal romantic partner. From participants’ writing samples, thematic cosines (a measure of relatedness) were between each male and female participant in both writing conditions. Participants completed the Big Five Personality Questionnaire Short Form (Morizet 2014), a personality inventory which evaluated participants’ Openness, Extraversion, Agreeableness, Conscientiousness and Emotional Stability. For this study, we utilized five Multilevel Models to determine if similarity among two participants’ individual personality measures predicted similar romantic cosines. With the exception of Emotional Stability, models with random slopes and personality predictors were significantly better than intercept-only models. Of the predictor models, Extraversion, Agreeableness and Conscientiousness were statistically significant (using α = .05). This suggests that similar personality scores are predicative of similar romantic cosines. It stands to reason, then, that persons with similar Extraversion, Agreeableness and Conscientiousness traits may often desire similar romantic situations and partners as presented in the writing prompt.

The Interaction of Personality, Interest and Romantic Preference

Sexual and romantic desirability are vital in forming a basic unit of human culture, the mated pair. Through natural selection, general preference for certain traits, such as intelligence and physique, lead to our evolution as a species. However, romantic preference, which as an individual’s abstract set of sexual and socially favorable traits in a mate, is rarely aligned with concrete mating choices. Still, romantic preference defines many cultural phenomena. From beauty standards to courtship, sexual interaction and long-term relationships, our aggregate sexual preferences shape our environment in innumerable ways.

As a measure of desirability, romantic preference predictably varies from person-to-person. However, in a meta-analysis of 28 independent samples, Feingold observed unique differences between American men and women on responses to mate selection questionnaires. For example, the literature suggested that men value attractiveness more than women on questionnaires, but that attractiveness and mate selection were equally high for both sexes in concrete dating scenarios.

A seminal experiment on mate preference was conducted by David Buss in the late 1980’s. In a massive cross-sectional study, Buss examined sex differences in mate preference across 37 samples from 33 distinct cultural paradigms. (1989) This incredible work found consistent sex differences in mate preference across all cultures. Examples include: higher preference among women for fiscally stable partners; and higher preference among men for younger female partners.

Buss also carefully checked census data from each country to determine how mate preference influenced eventual mate choices. For example, across every culture, a significant age-gap of approximately 3-years was found between older men and younger women. (Buss, 1989) Yet, age differences are easily-measured, external variables. Moreover, as stated by Buss, age differences were the most statistically reliable findings in his study, which suggests that perhaps this result is more either a stronger effect between men and women or is more easily measured in his samples.

Often, however, the connection between mate preference and mate choice is convoluted. Castro, Hattori and Lopez (2012) studied differences between mate preference and mate choice using a survey based in-part on Buss’s original work (1989). In general, Castro’s results agreed with Buss’s on specific sex differences in mate preference, such as males valuing physical attractiveness more than women. However, Castro et al. found participant’s preferences in non-physical traits (i.e. humor, intelligence, etc.) did not necessarily correlate with concrete perceptions of current or recent mates. This suggests that intersex romantic preferences differ, but do not necessarily predict an individual’s perception of a real-world romantic partners. (Castro et al., 2012)

Castro’s findings highlight the difference between our abstract romantic preferences and our concrete sexual selective process. This implies that people often choose mates which do not fit their preferences. While interesting, this necessitates research into the intricacies of romantic preference and its role in evolutionary psychology and human cognition. Of course, while an individual’s romantic preferences may fail to predict their mate choices, certain social phenomena can be explained as a function of observed gender differences in romantic preference.

As an example, Feingold (1990) conducted a meta-analysis of several empirical methodologies used to study romantic preference and mate choice. He also compared this meta-data with linguistic analyses of advertisements and billboards targeted towards men or women specifically. Interestingly, he noted that advertisements targeting men focus on attractive female partners more than advertisements for women, a conclusion that mirrored Buss’s findings (1989) and meta-data collected from survey-based research in romantic preference. That advertisements dovetail with observed research shows the direct applicability of empirical research in romantic preference. It also reveals the influence of romantic preference in shaping our understanding of desirability across two distinct genders.

Of course, that romantic preference influences society suggests it also motivates individuals and influences their actions. Botwin, Buss and Shackelford found that individuals from both sexes prefer romantic partners whose personality traits mirror their own. Not only this, long-term partners were likely to exhibit similar personality traits, showing a distinct connection between personality preferences in romantic partners and successful long-term romantic relationships. Even more, among all participants, Botwin found that certain personality traits were unappealing. These included disagreeableness, emotional instability and non-equal Intellect-Openness scores between partners. This result was even more distinct when the personality differences were observed in existing long-term relationships .

Together, Botwin’s et al. (1997) and Feingold’s (1990) results suggest a conversation between romantic preference and mate choice, both on an individual or societal scale and across physical and personality traits. In observed concrete mate choices, similar personality scores are strong indicators of relational satisfaction. Yet, personality is a factor which Castro (2012) suggests plays a lesser role in abstract romantic preference, especially among males. Back et al. (not on reference list) also observed that, in short-term sociosexual interactions (i.e. speed dating scenarios), personality factors not only predicted desirability but also significantly correlated with participants’ ability to predict their desirability among fellow participants.

This initial distinction between romantic preference and sexual selection is too general to account for the diversification observed in human sexual selection. For example, Schwarz and Hassbrauck (2012) examined gender and age differences in a rich cross-sectional sample of heterosexual participants. Their research corroborated Castro’s and Feingold’s observation that men emphasize physical attraction. Beyond this, Schwarz and Hassbrauck showed that, in envisioning potential marriage partners, women were far more demanding than men, especially when it came to fiscal solvency. Even more fascinating, Schwarz and Hassbrauck found identical gender differences across the lifespan, which suggests that gender differences in romantic preference and sexual selection are largely fixed, irrespective of age.

The restrictive nature of female romantic preference encourages the selection of specific traits in offspring, such as intelligence, dominance, social extraversion and industriousness. (Schwarz & Hassbrauck, 2012) However, male emphasis on physical appearance, health and fitness more readily dovetails with bearing and raising offspring. (Castro et al., 2012) But what, from an evolutionary perspective, is the function of these gender differences in romantic preference?

One theory is Runaway Sexual Selection (Miller, 2000), which theorizes that within a population, common selection of a trait may lead to homogenous occurrence of that trait. As an example, take intelligence: if all women in a generation are predisposed to sexually select intelligent partners, they will likely produce intelligent offspring. These offspring will likely select for intelligence if they are female, or obtain satisfactory and healthy mates if they are male. Regardless, the male population is seen to adjust to the more stringent preferences of the female population. As can be seen from this example, the biological drive for Runaway Sexual Selection is genetic, and depends on the co-occurrence of male traits and female preferences in each generation. (Haufe, 2008) With the assumption of co-occurrence, Runaway Sexual Selection explains the unilateral appearance of certain physical and cognitive traits across both sexes in a population.

An interesting contradiction of Runaway Sexual Selection is the comparatively rapid development of the human brain in size and neural density (Haufe, 2008; Miller, 2000). Brain development across human ancestry demonstrates the rigorous and directional selection of intelligence and cerebral growth. Targeted selection is difficult to understand in the pseudo-random mating environment-model presented by Runaway Selection. However, Haufe concludes that Runaway Selection can at least co-direct human evolution alongside other known and unknown properties. More so, such directional evolution seems feasible considering the narrow selection bias of males and consistent, if precise, preferences of females in selecting long-term mating partners .

This study examined the interaction among gender, personality and common interests by examining participants’ written romantic preferences and personal hobbies. We hypothesized that, like previous non-linguistic research, similarity in participants’ personality scores would predict similar romantic preference as demonstrated by a written prompt. We also examined the effect of within-gender and between-gender comparison and participants’ personal interests, again collected through a writing prompt.

Quantitative data were derived from participants’ writing samples using Latent Semantic Analysis (LSA), a technique from Linear Algebra which generates a vector space, where each vector represents a direct connection between individual terms or documents. Cosines between these connections quantify thematic similarities between two participants’ written responses, and greater cosines indicate similarity in themes among two writing samples. (Landauer et al., 1998) In this sense, one can think of LSA as a quasi-correlational method of quantifying linguistic data for inferential statistical analysis.

Latent Semantic Analysis transforms word co-occurrence and frequency into quasi-correlational data, which can be used to measure similarities in thematics and contextual semantics across distinct writing samples. LSA is an application of Singular Values Decomposition, a technique from Linear Algebra which factorizes a real or complex matrix into three component matrices. Geometrically, these component matrices correspond to a rotation, scaling and final rotation which allows an inner product (in this case, cosine) to be calculated between any members of the vector space generated by the original linguistic corpora. We utilized LSA as a method of data-transformation for our participants’ written responses, with thematic cosines being the unique score connecting each participant to one another.

**Method**

**Participants**

A sample of undergraduate students (N=105) was recruited from a large Midwestern college. All participants were enrolled in an introductory psychology course and received two research-participation credits for completing the study. To measure gender effects accurately, relatively even samples of male (N=53) and female (N=52) participants were recruited. The average age of the participant was under 21 years of age, and the majority were white. Sample collection occurred over a two-month period from October through early-December.

**Materials and Procedure**

All participants received online survey materials through Qualtrics, an internet survey platform. Initially, participants were informed of the potential risks, incentives and requirements of the study. After reporting demographic information, participants completed the Big Five Personality Trait Short Questionnaire (Morizot, 2014), which assessed Openness, Extraversion, Agreeableness, Conscientiousness and Emotional Stability. Finally, in random order, participants responded to a pair of writing prompts. One concerned their interests and hobbies (“Describe your interests and/or hobbies”), while the other asked them to describe their ideal romantic partner (“Describe an ideal date with your perfect romantic partner”). The order of prompts was counterbalanced, and responses had to exceed a minimum of 2200 characters. This was to ensure enough information density in the writing samples to guarantee usable latent semantic data.

For this specific study, we did not utilize the interests-and-hobbies written data. In the future, we may analyze whether similarities in writing on other dimensions (i.e. interests, personal statements, etc.) moderates the relationships of personality and romantic writing. However, in this study, we only tested the relationship between similarity across each personality measure with romantic writing.

**Results**

Data analysis was conducted in two major steps: Latent Semantic Analysis and a Multilevel Model (MLM) examining the influence of individual participants’ personality differences on romantic writing similarity as measured by thematic cosines .

**Latent Semantic Analysis**

Raw written data were marked with a participant number, gender and prompt number. Latent Semantic Analysis (LSA), which was conducted in R using the <lsa> and <SnowballC> packages, was performed on each participant’s response to the romantic writing prompt. Initially, LSA encodes the word frequency and co-occurrence of each participant’s written response in a text-frequency matrix. Through a serious of algebraic manipulations, this text matrix was scaled and normalized, which allowed for the calculation of thematic cosines between each participant’s written response. These cosines, which act as a measure of thematic relatedness, were then encoded as a cosine matrix and exported. Full copies of all datum and R-scripts may be obtained from the principle author upon request. For insight into the algebraic niceties which underpin Latent Semantic Analysis, see Axler’s (2010) *Linear Algebra Done Right*, chapter seven.

**Data Reshaping and Screening**

Following Latent Semantic Analysis, a new data set was constructed which contained each male-to-female pair’s romantic cosines as well as a difference score for each personality measure. Difference scores were calculated by subtracting our male participant’s score from our female participant’s score across each personality variable. Following this, we took an absolute value in order to normalize the order effects of subtraction on our personality measure.

Before conducting our Multilevel Model, Mahalanobis outlier analysis was conducted on the cosine scores and personality responses. Out of 5485 participant comparisons, only one participant-pair fell outside the Mahal cutoff score, and was excluded. Data were then screened for accuracy, additivity, normality, linearity and heteroscedasticity. The data were slightly right-skewed and had issues with linearity in the extremities of the data. However, this is to be expected with highly correlated data, and further justifies the use of a Multilevel Model.

**Analysis**

Following outlier analysis, descriptive statistics were calculated for romantic cosines and personality measures across both males and females. The average romantic cosine (*M* = 0.19, *SD* = 0.17) was relatively small and showed a comparatively large standard deviation. Personality scores ranged from 10-50 on an interval scale. Although we utilized a difference score in our Multilevel Model; however, for convenience, table 1 shows personality means, standard deviations and Cohen’s *d* across both males and females.

In our analysis, each personality variable was analyzed in a separate Multilevel Model. We chose this design to streamline our analysis while still controlling for the correlated error introduced by our thematic cosine measure. As is standard, within each MLM, we compared an intercept-only model, a random-intercepts model without predictors, and a random-intercept model with personality differences as a predictor. Except for the MLM examining Openness, the random-intercepts model with predictors was the best fit for our data in each MLM. Individual model’s degrees of freedom, intercepts, as well as significance among all models can be found in Table 2. (Note that the intercept-only model and random-intercept model is identical in each MLM, and hence is only listed once.)

Within our statistically reliable random-intercepts models, we found that differences in Extraversion (*b* = -0.002, *t* = -3.16, *p* = 0.002), Agreeableness (*b* = -0.003, *t* = -3.92, *p <* .0001) and Conscientiousness (*b* = -0.002, *t* = -4.37, *p* < .0001) were all significant predictors of similarities in thematic cosines across romantic writing. With negative slopes, this suggests that smaller differences in personality predicted larger thematic cosines. Difference in Emotional Stability (*b* < .001, *t* = 0.47, *p* = 0.64) was not a significant predictor similarity in thematic cosines. For convenience, see Table 3 for predictors, intercepts, significance levels and standard errors of each predictor.

**Discussion**

These results do not provide many answers, since few of the predictors in the final model were significant independent variables. However, the fact that higher scores in Openness did predict similarity in romantic cosines is consistent with current literature in Psychology, and demonstrates that symmetries in specific personality dimensions does imply similar romantic preferences across all participants.

It is also interesting and exciting to see methods from Quantitative Linguistics return similar results as more traditional analyses in Mate Selection and Evolutionary Psychology. While Latent Semantic Analysis is by no means the simplest or most obvious approach to quantifying romantic preference, in many ways, it reflects our day-to-day interactions in conversations, message boards, social media and romantic relationships.

Finally, as technology progresses, we better realize the importance of language in understanding cognition. As an interface with society and an organizing structure for the mind, language provides a direct means of analyzing cognitive phenomena. Also, as activity on social media has grown exponentially, so has the potential linguistic data for analysis on a global scale. Is it feasible that an ideal romantic partner could be identified based simply on profile pages or diaries on Facebook and Blogger? Of course, not yet. But if we intend to understand ourselves in the modern world, it is best to hedge our bets and investigate all methods of self-inquiry available to us.

Unfortunately, written data are inherently noisy. Latent Semantic Analysis cleverly side-steps much of this noise by using word frequency and co-occurrence to convert actual writing into quasi-correlational data. Yet, this introduces the entirely different problem of correlated error, which we addressed using a Multilevel Model. This issue is best seen in the numbers: 105 participants were responsible for generating over 5,400 data points. Once can control for correlated error. But it is difficult to run large Multilevel Models with enough uniquely different data points to present candid, believable effects without running out of processing capabilities.

As an added limitation, the reshaping of each cosine matrix is a time-consuming process and is prone to errors. Because of time constraints, we did not take the usual time creating these matrices and checking them by hand for accuracy. This will, of course, be corrected in future research on this hypothesis. All that hand-waving aside, we are confident in the statistical validity of our methods and findings, and will be de-bugging our datasets to present refined, complete testing of our experimental structure and hypothesis testing.

Soon, we will be reworking our data sets and running larger Multilevel Models to examine the effects of each personality trait on romantic cosines. Also, we would like to examine the interaction between the type of comparison (female-male, etc.) and personal interest cosines with romantic cosines to determine if factors beyond personality hypothesized to influence romantic preference present similar effects in romantic cosines. These analyses require significant computational power and time, and we will be investigating novel methods of addressing these issues in the future to better understand the interaction of personality, personal interests and gender effects on romantic preferences in writing. However, for now, we refrain from coming to any full conclusions on the relationship.

References

Botwin, M.D., Buss, D.M., & Shackelford, T.K. (1993). Personality and mate preference: Five factors in mate selection and marital satisfaction. *Journal of Personality*, *65*(1), 107-136. doi: 10.1111/j.1467-6494.1997.tb00531.x

Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences*, *12*, 1-49. doi:10.1017/S0140525X00023992

Castro, F.P., Hattori, W.T., & Lopes, F.A. (2012) Relationship maintenance or preference satisfaction? Male and female strategies in romantic partner choice. *Journal of Social, Evolutionary, and Cultural Psychology*, *6*(2), 217-226. doi: 10.1037/0099213

Feingold, A. (1990). Gender differences in effects of physical attractiveness on romantic attraction: A comparison across five research paradigms. *Journal of Personality and Social Psychology*, *59*(5), 981-993. doi: [10.1037/0022-3514.59.5.981](http://psycnet.apa.org/doi/10.1037/0022-3514.59.5.981)

Haufe, C. (2008). Sexual selection and mate choice in evolutionary psychology. *Biological Philosophy*, *23*, 115-128. doi: 10.1007/s10539-007-9071-0

Kenrick, D. T., Sadalla, E. K., Groth, G., & Trost, M. R. (1990). Evolution, traits, and the stages of human courtship: Qualifying the parental investment model. *Journal of Personality*, *58*, 97-116. doi:10.1111/j.1467-6494.1990.tb00909.x

Landauer, T. K., Folt, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2), 259–284. doi: 10.1080/01638539809545028

Miller G.F. (2000). *The mating mind: how sexual choice shaped the evolution of human nature*. New York City, New York: Doubleday.

Morizot, J. (2014). Construct validity of adolescents’ self-reported big five personality traits: Importance of conceptual breadth and initial validation of a short measure. *Assessment*, *21*(5), 580-606. doi: 10.1177/1073191114524015

Schwarz, S., & Hassebrauck, M. (2012). Sex and age difference in mate-selection preferences. *Human Nature*, *23*, 447-466. doi: 10.1007/s12110-012-9152-x

Table 1

Table 2

*Multilevel Model of participants’ romantic cosines with individual model’s Akaike Information Criterion and significance values among different models.*



Table 3

*Individual predictors included in the third and final random-intercepts model.*