Men are from Earth, Women are from Venus: Gender Differences in Predicting Actor and Partner Effects with Machine Learning in Initial Romantic Attraction

Bachelorarbeit von

Kimberly Klugescheid

Angefertigt unter den Bedingungen der Prüfungsordnung für den Bachelor-Studiengang Psychologie an der Georg-August-Universität Göttingen, Fakultät für Biologie und Psychologie, vom 8. März 2023

Erstprüfer: Dr. rer. nat. Tobias Kordsmeyer-Storp

Zweitprüferin: Dr. rer. nat. Judith Schmitz

Betreuerin: Laura Botzet, M.Sc.

Betreuer: Eric Grunenberg, M.Sc.

Kimberly Klugescheid

Matrikelnummer 16504970

kimberly.klugescheid@outlook.de

Raiffeisenstraße 3

34399 Wesertal

Abstract

The current study aimed to predict romantic desire of actor and partner in a speed-dating paradigm. It used a similar methodological approach as Joel et al. (2017) did and also the same data while extending it by researching gender differences, using behavioral predictor variables in addition to self-reported ones and validating the results through nested cross-validation. The aim was to investigate if and how the amount of explained variance changed through these extensions and identify high-impact predictors. Data was collected from 187 university students who reported their personality characteristics, demographic variables, and descripted themselves and their ideal partner. They then participated in one of seven speed-dating sessions and reported their romantic desire for each date, which serves as the outcome variable. Additionally, independent raters rated their flirting behavior. The data was analyzed with random forest models and nested cross-validations. In a descriptive comparison, the additional behavioral variables increased the mean explained variance only slightly. For this sample, more variance could be explained in male actor and partner desire than their female equivalents. Then, there was a significant variation in explained variance scores in the results of the nested cross-validation, which suggests that this methodological extension was effective in preventing an over- or underestimation of the effect. Moreover, feature importance scores for the predictors were extracted, showing that male and female romantic desire correlate in some cases with the same but often with different predictors. Overall, the current study enhances our insight into the factors driving romantic attraction.

Keywords: romantic relationships, machine learning, random forests, gender differences, flirting, nested cross-validation

German Abstract

Ziel der aktuellen Studie war es, das romantische Begehren von Akteur und Partner in einem Speed-Dating-Paradigma vorherzusagen. Dafür wurde ein ähnlicher methodischer Ansatz wie bei Joel et al. (2017) verwendet, sowie dieselben Daten. Als Erweiterungen wurden geschlechtsspezifische Unterschiede untersucht, zusätzlich zu den selbstberichteten Variablen behaviorale Variablen verwendet, und die Ergebnisse durch verschachtelte Kreuzvalidierung validiert. Ziel war es, zu untersuchen, ob und wie sich die Varianzaufklärung durch diese Erweiterungen verändert, und relevante Prädiktoren zu identifizieren. Es wurden Daten von 187 Universitätsstudierenden erhoben, die ihre Persönlichkeitsmerkmale und demografischen Daten angaben und sich selbst und ihren idealen Partner beschrieben. Anschließend nahmen sie an einer von sieben Speed-Dating-Sitzungen teil und berichteten über ihr romantisches Begehren für jedes Date, was als Kriteriumsvariable dient. Zusätzlich bewerteten unabhängige Beurteilende ihr Flirtverhalten. Die Daten wurden mit Random-Forest-Modellen und verschachtelten Kreuzvalidierungen analysiert. In einem deskriptiven Vergleich erhöhten die zusätzlichen Verhaltensvariablen die mittlere Varianzaufklärung nur geringfügig. Bei dieser Stichprobe gab es eine höhere Varianzaufklärung für männliches Akteur- und Partner-Begehren als für die weiblichen Pendants. Allerdings gab es in den Ergebnissen der verschachtelten Kreuzvalidierung signifikante Variationen in der Varianzaufklärung, was darauf schließen lässt, dass diese methodische Erweiterung eine Über- oder Unterschätzung des Effekts wirksam verhindert hat. Darüber hinaus wurden für die Prädiktoren Werte für ihre jeweilige Merkmalsbedeutung extrahiert, die zeigen, dass romantisches Begehren bei Männern und Frauen in einigen Fällen mit denselben, oft aber mit unterschiedlichen Prädiktoren korreliert. Insgesamt verbessert die aktuelle Studie unseren Einblick in die Faktoren, die romantische Anziehung beeinflussen.

Schlagworte: romantische Beziehungen, maschinelles Lernen, Random Forest, Geschlechterunterschiede, Flirten, verschachtelte Kreuzvalidierung

Men are from Earth, Women are from Venus: Gender Differences in Predicting Actor and Partner Effects with Machine Learning in Initial Romantic Attraction

"Men are from Earth, Women are from Venus." – the assumption that men and women are fundamentally different in a various range of aspects is widely popular. One aspect, where these two planets collide and gender differences should become visible, is the world of dating. However, what are those differences? And how do context and other factors influence these differences? The investigation of gender differences in romantic relationship research is a promising topic as it advances our understanding of dyadic relationships while exploring gender-specific tendencies that influence initial impressions, mate selection, and shape the early stages of romantic relationships (Fisman et al., 2006; Todd et al., 2007, Back et al., 2011b). Moreover, it can help dispel misconceptions about gender roles through evidence-based insights that may reshape societal norms and perceptions. Lastly, it provides an evaluation of different approaches in initial romantic contact that may support relationship-seeking individuals in their pursuits (Place et al., 2009).

Machine Learning

While a lot of research in this field has been accomplished, the use of artificial intelligence (AI) has been rather rare. However, AI methods hold promise for data analysis and predictive modeling in personality psychology as they offer new opportunities for understanding complex phenomena and making informed decisions (Stachl et al., 2020). Machine learning (ML) is a subset of AI that is used to create algorithms that can access data and use it to learn independently. The model is trained on the training set and learns during that process underlying patterns in the data. This then allows the algorithm to predict the criterium for a new, entirely unknown dataset, often called test set. ML has several advantages and presents itself as a promising methodological extension to the more “classical” statistical tools used in personality psychology such as linear regressions. The major advantages of ML algorithms are their ability to capture complex non-linear relationships, which also enables a better understanding of how different variables interact (Strobl et al., 2008). Furthermore, ML can handle a great number of predictors by implicitly reducing feature sets while also reducing the risk of overfitting, which is an overgeneralization based on the data it was trained on (Strobl et al., 2009). In comparison to classical regression models, ML algorithms usually also exhibit a higher level of robustness to certain violations of assumptions such as independence or linearity (Breiman, 2001a). This makes them appliable to a greater range of questions. They also provide an advantage in personality psychology as they can handle high-dimensional data and personality traits are often multidimensional and interconnected (Vélez, 2021). Finally, ML models regularly show a higher accuracy in their predictions than linear regression models (Caruana & Niculescu-Mizil, 2006). It is important to note that ML is not a replacement for more common statistical tools used in psychology but rather a complementary approach and suitable alternative for specific research questions. For example, Jacobucci & Grimm (2020) showed in a study with simulated data that ML algorithms only outperform linear regression when the reliability of the predictors is high, if not, the performance scores are similar.

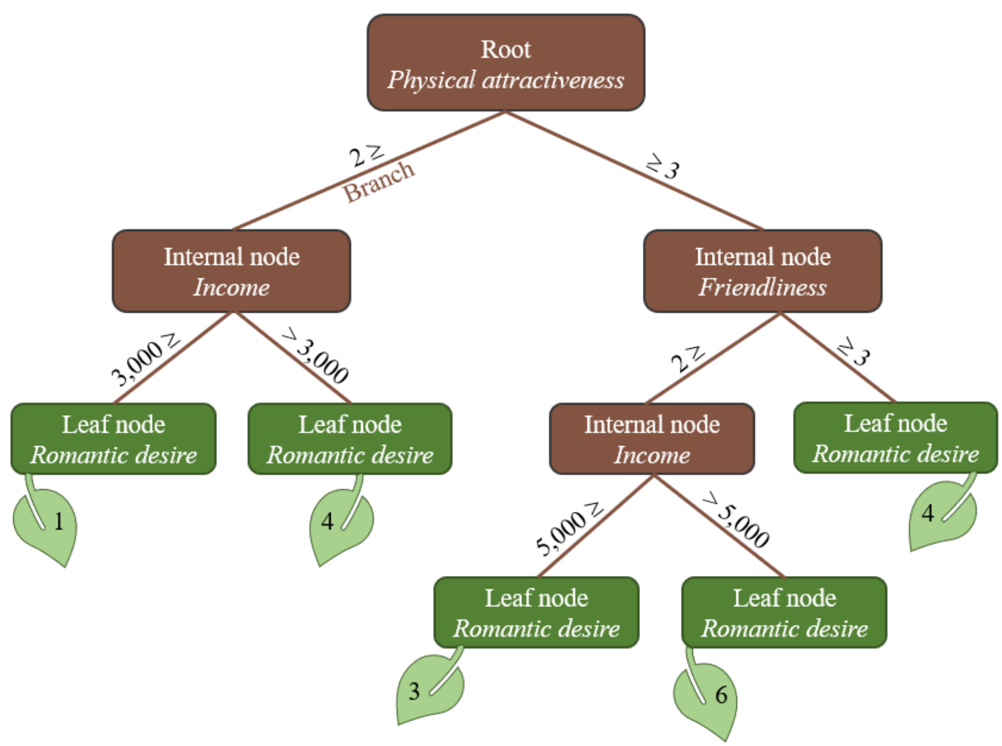
Random Forest

A common supervised ML algorithm is *random forest* (Breiman, 2001b). In the context of ML, supervised refers to algorithms that are trained on data providing predictor and outcome variables, while an unsupervised algorithm is trained without providing the outcome variable. Random forest is a so-called ensemble learning tool as it combines multiple algorithms to enhance its performance. Matching its name, a random forest combines the results from multiple *decision trees*. A decision tree is a predictive model which has a hierarchical, tree-like structure composed of nodes and branches (Song & Lu, 2015). Decision trees operate by selecting the variable that allows the best split of the dataset in subsets. This process is recursively repeated with each subset until a stopping criterion is met. This could be maximum depth, for example, no more than five parent nodes, or a minimum number of samples per subset. The “best split” can be defined differently based on the aim of the study. One common best-split criterium for classification is Gini impurity which measures the probability of wrongly classifying a randomly chosen element in the dataset if a certain feature is used as the predictor. Gini impurity is calculated for each possible split predictor and the predictor with the lowest Gini impurity is chosen. For regression, the entropy which deals with information gain can be useful to decide which feature should be used to split the data at the current node. Splitting the dataset into subsets is useful for several reasons. For once, with each split, the subsets within themselves become more homogenous which allows the decision tree to capture more distinct patterns and relationships. Then, through Gini impurity or information gain, the splits are hierarchically based on which features best discriminate between the single elements, enabling the decision tree to focus on the most informative variables and disregard less relevant ones (Quinlan, 1986). After the decision tree is constructed, it consists of internal nodes that represent the predicting features and leaf nodes that represent the predicted class in classificational problems or the regression value in regressive problems. To predict the criterium for new data, the decision tree algorithm follows a hierarchical top-down approach. The path starts at the root, deciding at each internal node (feature) which path to further follow. The prediction is finally determined by the reached leaf node. To illustrate it, one can imagine that we try to predict romantic desire with a range of variables, to limit the example we assume that we only have three predictors: physical attractiveness, income, and friendliness. At the root, the first node, the decision tree will select which of these three predictors allows the highest increase in information gain. In this example, physical attractiveness, a 4-point scale, could allow the greatest information gain, so the data is divided into two subsets, one with participants that have a “1” or “2” on the physical attractiveness scale, the other one with “3” and “4”. This principle continues recursively to select the next predictor that splits the new subset and so on. Finally leading to the leaves that then deliver in the end a predicted value for the outcome variable, which could be romantic desire in this case. If we then present a new person with values of the three measured variables to the decision tree, the decision tree will first assess their value in physical attractiveness, based on this follow down a specific branch to the next predictor, assess the person’s value in it, follow down the next branch and then arrive at a leaf that represents the predicted value in romantic desire for this specific person. Refer to Figure 1 for a visualization of this example. Because of the easily accessible decision structure, the decision tree algorithm leaps interpretable models which provide insight into the feature importance and therefore simultaneously allow feature selection. However, a shortcoming of this algorithm is overfitting, which negatively influences its predictive performance when tested on the test set. This is due to its high sensitivity to small changes in the training set which can possibly lead to an entirely different tree structure, mostly too complex with a high depth and a structure that memorizes the training set rather than finding an underlying general pattern (Jacobucci, 2018). The ensemble learning algorithm random forest addresses the limitations of the decision tree and improves the predictive accuracy by constructing an ensemble of multiple decision trees. Through bootstrap aggregation, the training set is randomly sampled with replacement which leaps multiple subsets of the original data with a similar sample size. Each bootstrapped subset is used to train a different decision tree.

Additionally, random forests use a mechanism called feature randomization, which essentially means that not each decision tree is receiving the same variables for their prediction but rather a randomly selected subset. This mechanism aims to enhance variation among the singular decision trees and reduces the reliance on single features. The decision trees are then grown recursively as described above. Finally, the random forest makes its prediction by aggregating the individual predictions of the decision trees. In the case of regression, the arithmetic means or medians of the decision trees predicted values are taken as the overall prediction. With this, random forests reduce overfitting and can capture patterns underlying the data while also profiting from all advantages that decision trees offer (Breiman, 2001b).

Figure 1

Decision tree – structure and example



Note. The example above represents only a simplified structure of a regressional decision tree.

Cross-Validation

To strengthen the generalizability of random forests, but also other algorithms, even further, *cross-validation* can be used (Little et al., 2017). This technique resamples the dataset to simulate how the model would perform on independent data. To accomplish this, the dataset is split into a user-defined number of *n* so-called folds. The selected model is then trained on *n-1* subsets and evaluated on the remaining subset, the test set. This process is repeated until each subset has been the test set once. Finally, the mean of all extracted performance scores allows for a more realistic insight into the actual model performance. Cross-validations follow different algorithms that are more or less fit for different problems. One worth pointing out in the context of the current study is *nested cross-validation*. It combines two separate cross-validations by implementing on each training set from the “outer” cross-validation a completely new “inner” cross-validation (Cawley & Talbot, 2010). The outer cross-validation serves the same purpose as described above (i.e., model evaluation), the inner one allows hyperparameter selection. If cross-validation with hyperparameter tuning is not implemented, the user must select the hyperparameters manually, however, there is often no mathematical solution to choosing the right parameters. In the example of random forests, hyperparameters that could be tuned are the number of decision trees, the maximum depth of a single decision tree, or the minimum sample size to split an internal node again. The inner cross-validation takes each training set from the outer cross-validation and separates it into *m* folds, following the same splitting principle of the outer cross-validation. In each inner fold, all combinations of the hyperparameters are tested. The parameter set that performed on average the best in the inner folds is returned to the outer fold, which then trains and tests the model with it. Nested cross-validation allows researchers to obtain more reliable and well-optimized models. In psychological research, this can be useful, especially in two aspects. First, psychological research often uses small sample sizes due to the challenges of participant recruitment and data collection (Koul et al., 2018). Nested cross-validation minimizes this shortcoming by evaluating the model on multiple train-test splits, which provides a more reliable estimate of the performance, even for small sample sizes. Second, the hyperparameter tuning on independent data enhances the generalizability to new individuals. This is important in psychological research as it aims to make inferences about broader populations (De Rooij & Weeda, 2020).

Permutation Feature Importance

In the specific case of studying romantic relationships, ML can be used to predict initial romantic attraction by integrating a wide range of predictors, while also giving insight into the relevance of each single predictor. Unlike the common assumption that ML models lack transparency in how they come to their decisions, it is possible to extract relevant predictors from the model to gain information about the variables influencing the criterium by utilizing feature importance scores. This allows a psychological interpretation of the model's prediction while using the advantages of AI. Several techniques allow the extraction of feature importance, providing metrics of the single predictor's relevance in the prediction of the criterium, one of them being *permutation feature importance*. It provides an importance score by permuting each predictor and calculating the decrease in model performance compared to the non-permuted version's performance score. In the case that the model performance decreases drastically due to the permutation of one predictor, it can be assumed that the model relied heavily on that predictor (Altmann et al., 2010). In psychological research, this can be useful when a wide range of predictors is entered into a model and it is yet unclear, which of them influences the criterium.

Romantic Relationship Research

In romantic relationship research, ML algorithms can be used to predict dyadic effects such as actor and partner effects and extract the importance of the variables influencing the prediction via permutation feature importance. Actor and partner effect are components of the *Social Relations Model* which helps understand the sources of variance in interpersonal behavior (Back & Kenny, 2010). In this context, the actor effect measures how strongly the individual's characteristics influence their own behavior in the social situation. Applied to initial romantic attraction, actor effects could refer to how much the individual romantically generally desires their partners. Through the examination of different interactions, researchers can estimate consistent factors or patterns that influence romantic desire, independent from the influence of a specific interaction partner. Partner effects, on the other hand, measure the response that the partner elicits in the actor. In the context of initial romantic attraction, it could be how much the individual is generally desired by their partners. The comparison of responses elicited in different partners by the same actor can give insight into what individuals are more likely to feel attracted to. Researching actor and partner effects as well as variables that influence them contributes to the understanding of how romantic connections are formed (Back et al., 2011a; Asendorpf et al., 2011; Joel et al., 2017).

Original Study

The study that also presents the methodological foundation of this paper used ML algorithms, specifically random forests, to predict romantic desire in a zero-acquaintance speed-dating study design, inter alia, the actor and partner effects. By using a range of self-reported predictors, Joel et al. (2017) were able to explain 4%–18% of the variance in the actor effect and 7%–27% of thevariance in the partner effect. Through permutation feature importance they were also able to extract which features are most relevant in the models and hence, important predictors for romantic desire. For the actor effect, they identified the actor's ideal partner's level of warmth as well as their own pickiness as important predictors. Their interpretation was that actors who desired more warmth from their partner or reported to be less picky when it comes to romantic partners felt overall more romantic desire for their speed-dating partners. For the partner effect, self-reports of attractiveness and mate value were found to be most relevant. Here their interpretation was that people who reported themselves to be physically attractive and also considered themselves to have a high mate value were more romantically desired by their dates.

Flirting

While the original study has used mainly self-reported predictors to investigate romantic desire in initial romantic attraction, it is also worth considering behavioral variables. As Furr (2009) mentioned in his paper: "What we think, feel, desire and value may have their most important effects when we actually act upon those thoughts, feelings, desires and values.", this ultimately led him to a call for focusing more on behavioral variables in personality psychology research. For example, in the case of interpersonal effects, Sprecher and Duck (1994) found the quality of communication to be predictive of attraction and the desire to see the partner again. They also found gender differences, namely that the quality of communication with the partner was more important for women's attraction towards men than for men's attraction towards women, and more important for friendship attraction than for romantic attraction. Another relevant behavioral predictor of romantic desire is flirting, a social behavior that expresses romantic or sexual interest in someone else via verbal and non-verbal cues. It can be assumed that gender differences can be found for flirting as a predictor of the actor effect as several studies reported that women's motives for flirting are more often rooted in fun and relational reasons whereas men's motives rather root in sexual interest (Frisby et al., 2010; Henningsen, 2004). At the same time, men also attribute sexual interest to female flirting while women attribute fun and relational reasons to male flirting (Frisby et al., 2010; Johnson et al., 1991). This misconception should logically lead to a discrepancy in the partner effect: While men should report a higher romantic desire for women who flirt with them, as they view it as sexual interest, women should not base their romantic interest on their partner's flirting behavior. However, Back et al. (2011a) reported different results: for both genders flirting positively correlated with romantic desire in the partner effect but not in the actor effect. This means that there was no relation between flirting with someone and choosing that person, but there was a relation between being flirted with by a partner and choosing that partner. Hence, the study results suggest that flirting as a behavior does not directly express mating interest but is interpreted as such, by both genders. This is especially interesting under the consideration that flirting is usually a reciprocal behavior, meaning that flirtatious behavior and being flirted with strongly correlate for both men and women (Grammer et al., 1998). The actual mate selection, on the other hand, does not seem to be reciprocal (Back et al., 2011a).

To gain more insight into flirting as a predictor of romantic desire, it can be valuable to differentiate between flirting styles. The interpersonal circumplex (Wiggins et al., 1988) offers a differentiation of interpersonal behaviors along the axes of agency and communion, therefore, flirting can be described as agentic or communal flirting. While an agentic flirting style is characterized by more dominant, assertive, and leading verbal and non-verbal actions, a communal flirting style is rather polite, benevolent, warm, and friendly (Dufner & Krause, 2023). Frisby et al. (2010) found that dominant flirting in men was not perceived as attractive nor effective by women which could imply that the effect on women's romantic desire for men is also dependent on the flirting style. Moreover, Kennair et al. (2022) evaluated flirting styles by distinguishing between gender and the mating context. They observed that a friendlier flirting style, comparable to communal flirting, is for both men and women more effective in a long-term mating context in comparison to a short-term context. However, in a short-term context, there were differences. While for women a dominant, agentic flirting style with physical contact was more effective, men's flirting style was more effective if it was less dominant but rather friendly. Furthermore, since men and women reported different motives for displaying flirting behavior (Henningsen, 2004), it is also possible that on the actor side, they use different flirting styles based on their motives.

Gender Differences

Partner Effect. While it is interesting to find general predictors of romantic desire in the speed-dating paradigm, gender differences in these predictors also present a promising field of research. This research dates to historical times: already in folktales people wrote about gender differences in mate choice, namely that men's romantic desire would be more based on the physical attractiveness of their partner compared to female's romantic desire (Gottschall, 2005). In the context of dyadic effects, this would signify that for the partner effect, physical attractiveness is a better predictor for men's desire than for women's desire. And indeed, Fisman et al. (2006) used a speed-dating study design similar to Joel et al. (2017) while emphasizing gender differences and found that men base their romantic desire more on physical attractiveness than women. Men's intelligence turned out to be a better predictor for women's partner desire. Men, on the other hand, did not report a higher romantic desire for partners with a higher intelligence level, on the contrary, they reported less romantic desire for women whose intelligence level exceeded their own. However, while other studies such as Luo and Zhang (2009) also found that men desire physically attractive women more, they discovered the same effect for women, finding no gender difference. Instead, Luo and Zhang (2009) identified in their college-sample speed-dating study predictors such as the partner's age, weight, political views, neuroticism, extraversion, conscientiousness, agreeableness, anxiousness, and self-worth as predictors for women's romantic desire. Men's romantic desire, on the other hand, seemed mainly led by women's physical attractiveness and interest in sports activities. Nevertheless, these variables also highly correlated with women's romantic desire for men and hence explained variance in the partner effect but do not present a gender difference in the importance of physical attractiveness. Asendorpf et al. (2011) confirmed this in a more heterogenous community sample, by also finding the physical attractiveness of the speed-dating partner as one of the main predictors of romantic desire for both men and women. Additionally, for women's romantic desire, they found the partner's sociosexuality, openness to experience, shyness, education, and income to be relevant. However, an earlier study from Eastwick and Finkel (2008) did not find gender differences in the association between a partner's income and the actor's romantic desire for them. It should be noted that this does not reflect the importance the actor implicitly gives to their partner's income: here women clearly based their romantic desire more on income than men did, while men reported finding physical attractiveness more relevant to their romantic desire (Eastwick & Finkel, 2008). Interestingly, many studies found that reported mate preferences do not predict actual mate choice or at least present little relevance for the evaluation of a potential mating partner. Thus, while some gender differences have been found in mate preferences, they did not become visible in actual mate selection (Li et al., 2013; Kurzban & Weeden, 2007; Eastwick et al., 2014, Todd et al., 2007). The meta-analysis of Eastwick et al. (2014) also found physical attractiveness to be a strong predictor of romantic desire for a partner, income also proved to be relevant but with a smaller effect size. However, no gender differences were found. Another meta-analysis by Feingold (1992) reported that women had a higher romantic desire for men with more ambitiousness, intelligence, “character”, status, and a higher socioeconomic status. However, they did not find gender differences in personality-related variables. Another important factor when considering gender differences in initial romantic attraction is the prospective timeframe or seriousness of the relationship. Regan (1998) primarily observed gender differences in short-term mating preferences, where women demonstrated higher romantic desire for older, more interpersonally responsive partners, compared to men, who applied less stringent selection criteria on their potential short-term mates. However, Sprecher and Duck (1994) observed gender differences also in long-term mating preferences with women preferring higher earning potential and men preferring youth and physical attractiveness.

Actor Effect. But what factors determine within the actor how much the actor romantically desires their partner? And are there gender differences too? Luo and Zhang (2009) found that men compared to women demonstrated a generally higher romantic desire for potential short-term mating partners when they reported higher interest in social or fun activities, had less politically conservative views, and a lower physical attractiveness. In contrast, women’s romantic desire in short-term mating correlated positively with having a lower age, higher body weight, lower interest in social activities and art, higher level of extraversion and openness to experiences as well as a more positive affect. Statistical significance was found for age, conservatism, positive affect, and fun as well as habitual activity, making them gender-dependent predictors for actor romantic desire. Moreover, Asendorpf et al. (2011) found gender differences in income and years of education: men reported overall less romantic desire when having a higher income or more years of education, but women did not show such links.

According to Fisman et al. (2006), the actor effect is also moderated by the group size of the potential partners. They found that while male selectivity is invariant to the change of group size, female selectivity is increasing with an increase in group size.

Current Study

To further investigate these differences in flirting styles and their relation to the actor and partner effect, the current study will add the variables “flirting”, “agentic flirting” and “communal flirting” to the predictor variables. As this study is based on Joel et al. (2017) and uses the same data, it will extend their research by adding previously named behavioral variables. Moreover, I will offer a more thorough investigation of possible gender differences in actor and partner effects by further analyzing which predictors are relevant for which gender. Lastly, I will evaluate the results of the current study and strengthen their generalizability by utilizing nested cross-validation, which also presents a methodologically new way to reassess the results already found by Joel et al. (2017). Consequently, the following three research questions emerge:

I. How accurately can actor and partner effects in initial romantic attraction be predicted
   (a) using only self-reported predictors compared to
   (b) a combination of self-reported and behavioral cues?

II. Do gender differences exist in the model performance of a random forest when predicting actor and partner effects?

III. Which predictors are for men and women the most relevant in actor and partner effect and are there differences?

Methods

Sample

Like Joel et al. (2017), this study will use the sample of the *Northwestern Speed-Dating Study* (Finkel et al., 2007), named “Sample B” in Joel et al. (2017). Participants were 187 heterosexual undergraduate students frequenting an American university in 2005 and participating in one of seven on-campus speed-dating events. The sample consists of 93 women and 94 men with a mean age of 19.6 years (*SD* = 1.2). Based on provided data and items from Joel et al. (2017), as well as the original study (Finkel et al., 2007), it cannot be determined whether the studies assessed biological sex or gender. Since I assume that participants are more likely to answer with their gender to the item “Are you male or female” and I also expect a high correlation between biological sex and gender, I will use the term “gender”. The previous studies do not state specific exclusion criteria. I excluded one male participant whose pre-event questionnaire data was not available. With this, 186 participants remained.

Procedure

After being recruited via flyers or emails, interested candidates signed up for the speed-dating event by filling out a 30-minute pre-event questionnaire which assessed a variety of psychological constructs as well as demographical variables. For this they were paid $5, given that they show up at the booked event. After about two weeks, the study team invited the participants to one of several speed-dating events, in which they met about twelve other participants of the opposite gender. With each of them, they had a four-minute-long speed date, which was recorded by a video camera, capturing audio and video. Directly after each speed date, the participants completed a questionnaire to report the experience they had just had with their date and whether they would like to meet them again. After filling out the questionnaire, one person went on to the next table where their new date was sitting. Between the sessions, it was varied whether the man or the woman changed tables. Participants received after one day an email showing them their matches and the matches' contact information. In the following 30 days, participants were sent ten follow-up questionnaires to report how their relationships with their matches developed. For each completed follow-up questionnaire, they received $3, if they completed at least nine of them, they received a bonus of $10.

Variables

Predictors

In the observational Northwestern Speed-Dating Study (Finkel et al., 2007) various variables were measured, from which Joel et al. (2017) used a total of 112 scales for their sample B. For more information on the exact scales and items that were used, see Database S1 in the Supplemental Material. The pre-event questionnaire included demographic variables such as gender ("Are you male or female?") and age. Moreover, it measured on a 7-point Likert scale the participant's desire for a long- or short-term relationship (e.g., "I desire a short-term romantic relationship", two items), in a check-all-that-apply format their speed-dating reasons, and in an open format the participant's expected selectivity and desirability.

A big part of the questionnaire was dedicated to self-reported personality characteristics, which were to be answered on a 7-point Likert scale, except sociosexuality, which also had an open-format item asking for a specific number. The used scales included the following constructs: attachment anxiety (e.g., "I need a lot of reassurance that I am loved by romantic partners.", six items), attachment avoidance (e.g., "I feel comfortable opening up to romantic partners." [reversed], six items), sociosexuality (e.g., "With how many different partners have you had sex within the last year?", six items), mating strategies (e.g., "I prefer a long-term relationship with one partner.", two items), mate value (e.g., "I am a desirable dating partner.", three items), pickiness ("I am very picky about my choice of romantic partners.", one item), relationship initiation (e.g., “I am very picky about my choice of partners.”, two items), ambivalent sexism (e.g., “A good woman should be set on a pedestal by her man.”, two items), gender identity (e.g., “I feel good about the gender group I belong to.”, two items), sexual orientation (“I am exclusively attracted to members of the opposite sex.”, one item), love styles (e.g., “The best kind of love grows out of a long friendship.”, two items), adversarial sexual beliefs (e.g., “In dating relationships, women are largely out to take advantage of men.”, two items), traditionalism (e.g., “I want a traditional family.”, two items), need to belong (“I have a strong need to belong.”, one item), sex drive (e.g., “I have a strong sex drive.”, two items), emotional intelligence (e.g., “I am a self-motivating person.”, four items), political affiliation (e.g., “I endorse many aspects of liberal political ideology.” [reversed], two items), desperation (e.g., “In a romantic domain, I feel desperate these days.”, two items), relationship theories (e.g., “Relationships often fail because people do not try hard enough.”, six items), self-clarity (e.g., “My beliefs about myself often conflict with one another.” [reversed], three items), goals in romantic relationships (e.g., “I strive to enhance bonding and intimacy in my romantic relationships.”, four items), Northwestern belonging (“I feel like I belong at Northwestern.”, one item), dispositional forgiveness (e.g., “I have a tendency to harbor grudges.” [reversed], three items), growth (e.g., “A successful relationship evolves through hard work and resolution of incompatibilities.”, three items), loneliness (e.g., “I frequently feel left out”, three items), self-esteem (e.g., “I have high self-esteem.”, three items), self-control (e.g., “I have less self-discipline than most people.” [reversed], three items), pragmatic love style (e.g., “I try to plan my life carefully before choosing a lover.”, three items), subjective well-being (e.g., “I am satisfied with my life.”, two items), self-construal (e.g., “I will sacrifice my self-interest for the benefit of the group I am in.”, four items), rejection sensitivity (e.g., “I generally expect that people will accept me." [reversed], two items), affect (e.g., "I feel anxious these days.", four items), capitalization ("I usually respond enthusiastically when loved-ones tell me about something good that has happened to them.", one item), narcissism (e.g., "I am an extraordinary person.", eight items), importance of partner status ("I sometimes fantasize about being in a relationship with someone who is socially powerful and wealthy.", one item), regulatory focus (e.g., "I often do well at different things that I try.", twelve items), love as basis for marriage (e.g., "Even if a person had all the qualities I desired in a romantic partner, I would never marry him/her if I were not in love with him/her.", two items), romantic beliefs (e.g., "I believe that to be truly in love is to be in love forever.", three items), romantic expression ("In general, I am comfortable expressing my romantic interest to someone I am interested in.", one item), social phobia (e.g., "When mixing socially I am uncomfortable.", three items), optimism (e.g., "I am always optimistic about my future.", two items) and the *Big Five* personality traits: neuroticism (e.g., "I seldom feel blue." [reversed], four items), extraversion (e.g., "I am the life of the party.", four items), conscientiousness (e.g., "I like order.", four items), agreeableness (e.g., "I feel others' emotions.", four items) and openness to experience (e.g., "I have a vivid imagination.", four items).

Furthermore, participants described their actual selves', their ideal selves', their ought selves', their ideal partners', and their ideal speed dates' characteristics on a 9-point Likert scale. Aggregated variables included physical attractiveness (two items), good earning prospects (two items), vitality (two items), warmth (three items), agency (three items), and intelligence (two items). Then, variables from the post-date questionnaires were included, which were also reported on a 9-point Likert scale. These include predicted chemistry (e.g., "You and he/she seem to have a lot in common.", three items), reciprocity (e.g., "He/she seemed to really like you.", two items), anxiety ("You worried that he/she didn't like you as much as you liked him/her.", one item), avoidance ("You felt comfortable opening up to him/her.", one item). Lastly, the attractiveness of each participant was rated on a 9-point Likert scale, and the current study expands the predictors by including the rated behavioral variables "flirting", "agentic flirting" and "communal flirting" which were rated on 7-point Likert scales.

Dependent Measure

The dependent variable[1], romantic desire, was aggregated and calculated for the actor and partner effect separately. The aggregation method is similar to the one used in the original study by Joel et al. (2017) and is based on their social-relations-model analysis that confirmed that the dependent measure comprises actor and partner variance. For the actor effect, I calculated for each participant the mean romantic desire that they reported to have for their dates. The actor effect grand mean of reported romantic desire was then subtracted from each of the participants' mean. This yielded the dependent variable for the actor effect. For the partner effect, I calculated for each participant the mean of romantic desire their dates reported to have for them. Then, the partner effect grand mean of reported romantic desire was also subtracted for each of the participants' mean, which yielded the dependent variable for the partner effect.

Rating

For each speed-dating session, two independent raters rated the variables "flirting", "agentic flirting" and "communal flirting" for participants of the opposite gender by watching the video recording of the speed date, with the screen cut in such a way that only the participant who was to be rated, was visible. Before the rating, all raters participated in mandatory training in which they learned about common judgment errors and practiced rating speed date videos.

---

[1] While in psychological research, the term "dependent variable" is commonly used in experimental study designs, in the ML context the notation is not as strict and an outcome variable in a correlational design can also be called dependent variable as the prediction of the model is dependent on the underlying pattern that the model has learned.

The predictor “flirting” deals with the flirting frequency and intensity (“How strongly is the actor flirting with their partner?”), “agentic flirting” relates to a dominant flirting style (“To what extent does the actor show a dominant flirting style, is bold and playfully provocative in his approaches?”, and “communal flirting” measures a friendlier flirting style (“To what extent does the actor show a warm-hearted flirting style, is courteous and friendly in his approaches?”). The raters were psychology students from different German universities, all a similar age to the participants’ age during the study. For some videos, raters reported sound problems which may have negatively influenced the rating accuracy. For each pair of raters and each flirting variable, interrater reliability was calculated with the *intraclass correlation coefficient* (ICC) that ranges from 0 to 1, with 1 signifying complete agreement among the raters. The mean ICC of all pairs of raters for general flirting is .542, for agentic flirting .520, and for communal flirting .504.

Analysis

To answer the first research question, I conducted four random forest models, varying between actor and partner effect, and a dataset of predictors with and without the behavioral flirting variables. Each of the random forest models was validated with nested cross-validation, using five outer folds and ten inner folds as suggested by Stachl et al. (2020). Tuned hyperparameters inside the inner folds included the number of decision tree estimators, the maximum depth of a single decision tree, the maximum number of features to consider when splitting at the parent node, the minimum number of samples required to split a parent node as well as the minimum number of samples required for leaf nodes. Inside the nested cross-validation, the random forest regressor was pipelined together with a scaler and an imputer. While random forests are relatively robust to the scale of features, scaling can still help equalize the importance of single variables by preventing features with larger scales from dominating the model. The imputer makes sure that no missing values occur by replacing them with the mean value of the predictor. Finally, the pipelining of this process prevents data leakage during the nested cross-validation. Data leakage happens when the model is not predicting on a completely unknown dataset. By applying the scaler and imputer only on the specific training set of each fold, instead of applying them beforehand on the complete dataset, data leakage is prevented, and the results become more accurate.

To answer the second research question, I evaluated random forest models that were trained and tested on only-male and only-female data. Moreover, to further analyze the contributions of the flirting variables to the model performance, I trained separate models for the dataset with and without flirting variables. Overall, this leaped eight random forest models which were again validated with nested cross-validation and also pipelined together with a scaler and an imputer.

To answer the third research question, I implemented permutation feature importance in the nested cross-validation that was already used for the second research question. This added function permuted each feature in each outer fold ten times and then extracted the means and standard deviations of the scores that resulted from the permutation, with a higher permutation feature score indicating that a predictor is more relevant to the model's prediction. Another function then evaluated, based on the mean permutation feature importance score, the top ten predictors with the highest scores in each fold. This resulted in a distribution of feature importance for each model. It should be noted that this distribution gives insight into the most important predictors but not into the effect direction in which they relate to the outcome. In this analysis, I used only the dataset that includes all three flirting variables because if the flirting variables were not relevant to the model prediction, they would be eliminated in the process of this analysis anyway.

Open science and sharing research code are essential practices that promote transparency, reproducibility, and collaboration within the scientific community. By openly sharing research findings and providing access to the underlying code used in data analysis, researchers enable others to validate and build upon their work. In this spirit of openness and collaboration, the current study also embraces the principles of open science and provides the research code[2] written for this project on the website of Open Science Framework.

Results

With vs. Without Flirting

As shown in Table 1, both random forests for the actor effect were not able to explain any variance in the actor's romantic desire by using the predictor variables, regardless of whether the behavioral variables were included or not. However, within the outer folds, performance scores varied a lot, ranging from about -5% to 6% in both models, indicating that the models were rarely able to accurately predict how much a person would generally desire their speed dates. Compared to the actor effect, the random forest models were able to explain descriptively more variance in the partner effect, on average about 12% more. This means that the model had a higher predictive power when predicting how much a person would be generally desired by their speed dates. Here, the model that included the three flirting variables as predictors was able to explain on average 3% more variance in romantic desire than the model that used the reduced dataset. However, both models showed high standard deviations, with performance scores

[2] The code can be accessed through the following view-only link: https://osf.io/9wejz/?view_only=e83c098ca6d74ab186fb1beea1ae6ce0

ranging from -16% to 24% in the reduced model and -11% to 25% in the model which included flirting variables.

Table 1

Results random forest models with and without flirting predictors

	R^2			
	Without flirting		With flirting	
	M	*SD*	*M*	*SD*
Actor effect	0.002	0.044	0.004	0.043
Partner effect	0.113	0.133	0.141	0.149

Male vs. Female

As visible in Table 2, random forest models were able to explain descriptively more variance for men than for women in models for actor effect and partner effect as well as in models excluding or including flirting variables as predictors. While the explained variance in the male actor effect was greater than the overall one where both genders were included, the one for women was even negative and hence, worse than the overall one. A negative R^2 value indicates that the prediction made by the model is worse than if the prediction was simply the overall estimated mean of the criterium and therefore, indicates poor model performance. For men, the explained variance ranged from -7% to 16% within the outer folds, for women it ranged from -29% to -1%. Compared to the actor effect, for the partner effect model performance scores were again higher. Even though, also for this effect, the explained variance was higher for men than for women, the female explained variance was not negative as it was for the actor effect. However, the scores in all models had large standard deviations: in the models for men, scores ranged from -7% to 28% of explained variance, and in the models for women from -9% to 17%.

Table 2

Results random forest models with and without flirting predictors separated by gender

	R^2			
	Male		Female	
	M	*SD*	*M*	*SD*
Actor effect, without flirting	0.036	0.069	-0.174	0.079
Actor effect, with flirting	0.058	0.078	-0.160	0.081
Partner effect, without flirting	0.082	0.124	0.022	0.091
Partner effect, with flirting	0.108	0.085	0.030	0.088

Feature Importance Analysis

Actor Effect

Figure 2 shows the distribution of how often each predictor made it into the top ten predictors during all folds, separated by gender and effect. In the actor effect, for women 18 predictors were most often, meaning at least two times, among the top ten predictors in the five outer folds: actual self-evaluation of physical attractiveness, intelligence, agency, as well as warmth, attachment anxiety, avoidance of goals in romantic relationships, emotional intelligence, communal flirting, general flirting, their ideal partner’s agency, their ideal partner’s physical attractivity, the general chemistry they subjectively reported to have had with their speed dates, loneliness, the feeling of belonging to their university, self-control, social phobia, love style storge, and objective attractivity. As shown in Figure 3, the variable importance scores ranged for all these predictors from 0.01 to 0.04, meaning that the permutation of one predictor decreased the percentage of explained variance by 1% to 4% compared to the baseline model without permutation of the predictor. Men’s romantic desire was best predicted by 11 predictors: self-evaluation of their warmth, their agreeableness, their ideal speed date’s warmth, the general reciprocity they subjectively reported to have received during the speed date, the affective component towards their gender identity, their ought self's warmth, their preventional regulatory focus, the importance they gave to their partner's status, their traditionalism, their expected selectivity and how comfortable they felt opening up to their date. The decrease in explained variance compared to the baseline model ranged for these predictors from 1% to 6% (see Figure 3).

When descriptively comparing the most relevant predictors for women's and men's actor effects, some differences become noticeable. For once, from the predictors in which participants described their actual selves, for women a wider range of predictors, including intelligence, attractiveness, and warmth, were relevant, while for men it was only warmth. However, warmth also had for men on average a higher permutation feature importance. Moreover, for women, psychosocial variables such as loneliness, social phobia, and emotional intelligence were relevant in their actor desire, while these variables did not seem to play an important role in men's actor desire. From the additional behavioral variables, general flirting as well as communal flirting seemed to be relevant for predicting women's actor desire. For men, on the other hand, communal and agentic flirting showed a certain relevance, but only in one fold and also only with relatively low permutation feature importance. Whereas some of the variance in men's actor desire could be explained from post-date variables that assessed how comfortable the man felt opening up to his date or in how far he believed that she liked him, such variables seemed to be less important for women in their romantic desire. Another descriptive difference can be found for objectively rated attractiveness. While this attractiveness rating predicted women's actor desire, such an effect was not found for men. At the same time, the attractiveness of their ideal partner predicted women's actor desire but not men's. Instead, men's actor desire could be predicted by predictors such as traditionalism or their own status, which did not turn out to be relevant predictors for women.

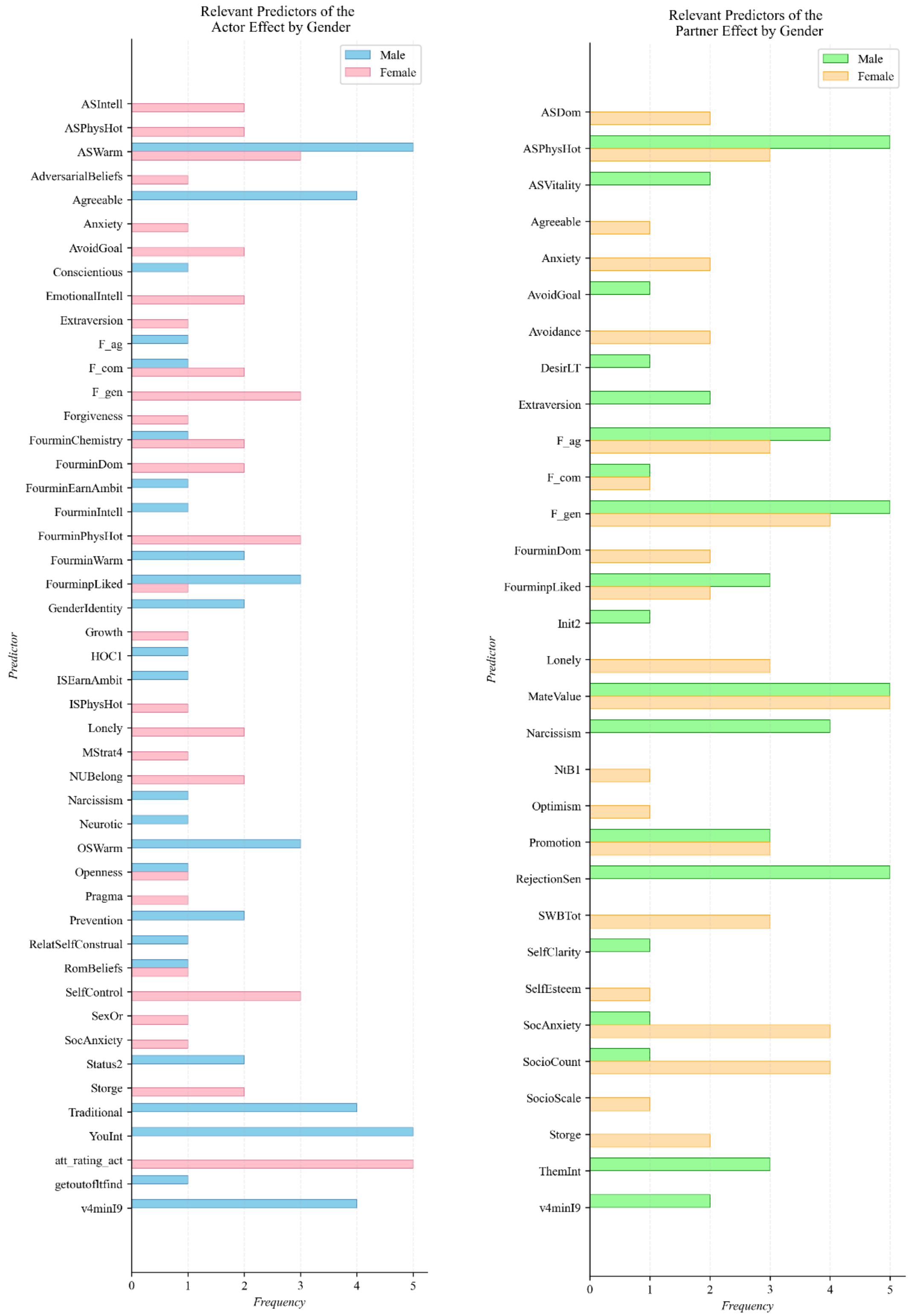
Partner Effect

The 15 most relevant predictors for women for their partner effect, so for how strongly men desired that woman on average, were the woman's actual-self evaluations of their agency and physical attractiveness, their level of anxious and avoidant attachment styles, their general and agentic flirting, their ideal speed date's level of agency, the general reciprocity they subjectively reported to have received during the speed date, their loneliness and social phobia, their subjective well-being, promotional regulatory focus, sociosexuality. mate value, and love style storge. The permutation feature importance score of these predictors was around 0.02, however, some predictors such as agentic flirting, ideal speed date's level of agency and mate value, had on average a higher impact with scores around 0.03 (see Figure 3), and mate value also being selected a top predictor in all five folds.

For the men's partner effect, or what variables within the man explained variance in the women's general romantic desire for him, 12 predictors were most relevant: their actual-self evaluations of their physical attractiveness and vitality, their extraversion, their agentic and general flirting, the general reciprocity they subjectively reported to have received during the speed date, their mate value, narcissism, promotional regulatory focus, rejection sensitivity, their expected desirability, and how comfortable they felt opening up to their speed date. Permutation feature importance scores ranged from 0.01 to 0.05, indicating that the permutation of some variables such as mate value and physical attractiveness lowered the explained variance by 4% to 5% (see Figure 3). In a descriptive comparison, it becomes noticeable that both men and women were influenced in their romantic desire for their partner by their partner's self-evaluated physical attractiveness but also mate value. But while the feature importance scores for women's partner desire for these two variables was about 0.02, it was on average 0.04 for men's partner desire, indicating that women relied more heavily on the variables when determining romantic desire for a partner, than men did. Then, while men's romantic desire for women could be predicted from variables such as the women's level of social phobia, subjective well-being, loneliness, agency, and sociosexuality, these variables seemed to have relatively low predictive power for women's romantic desire for men. Women's romantic desire, on the other hand, could be predicted by men's extraversion, rejection sensitivity, and narcissism. Moreover, the man's expected desirability and his assessment of how comfortable he felt opening up to his date could predict how much women desired him in general. It should be noted that the extracted permutation feature importance scores do not give insight into the direction of the effect and theoretically, the effect direction could even differ between the genders. Meaning, that if a predictor proved important for both genders, it could still be that the correlation between predictor and outcome is positive in one case and negative in the other.

Figure 2

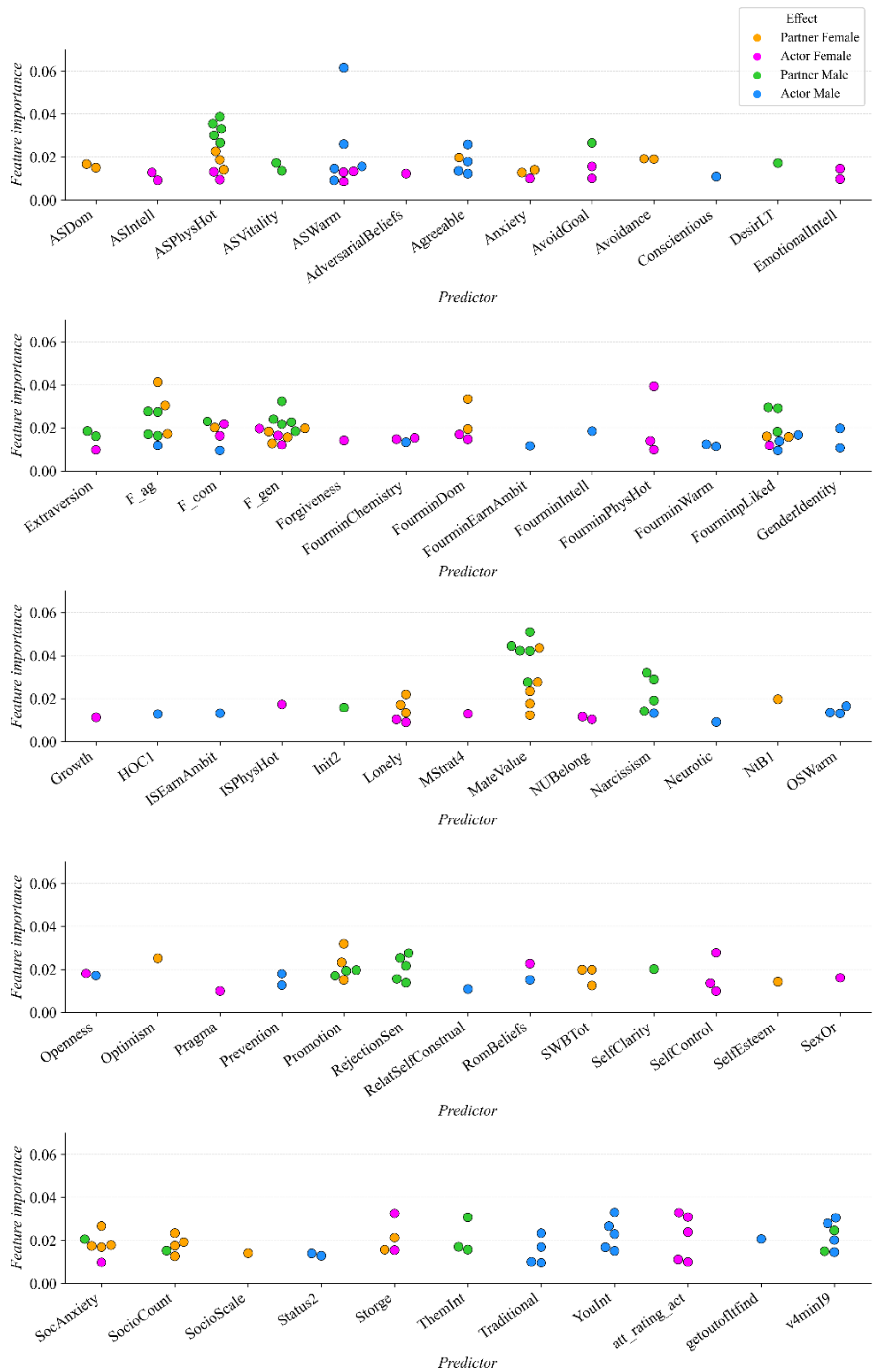
Distributions of the most relevant predictors in cross-validation separated by effect and gender



Note. Frequency describes in how many of the five folds the predictor was in the top ten most relevant predictors. Predictor names are abbreviated with their scale names in the figure. Abbreviations: actual self agency (ASDom), actual self intelligence (ASIntell), actual self physical attractiveness (ASPhysHot), actual self vitality (ASVitality), actual self warmth (ASWarm), adversarial sexual beliefs (AdversarialBeliefs), agreeableness (Agreeable), attachment anxiety (Anxiety), goal avoidance (AvoidGoal), attachment avoidance (Avoidance), conscientiousness (Conscientious), desire for long-term relationships (DesirLT), emotional intelligence (EmotinoalIntell), extraversion (Extraversion), agentic flirting (F_ag), communal flirting (F_com), general flirting (F_gen), dispositional forgiveness (Forgiveness), speed date prediction of chemistry (FourminChemistry), agency of ideal speed date (FourminDom), good earning prospects of ideal speed date (FourminEarnAmbit), intelligence of ideal speed date (FourminIntell), physical attractiveness of ideal speed date (FourminPhysHot), warmth of ideal speed date (FourminWarm), speed date prediction of reciprocity (FourminpLiked), gender group identity (GenderIdentity), growth (Growth), capitalization (HOC1), good earning prospects of ideal self (ISEarnAmbit), physical attractiveness of ideal self (ISPhysHot), romantic expression (Init2), loneliness (Lonely), mating strategy infidelity (MStrat4), mate value (MateValue), Northwestern belonging (NUBelong), narcissism (Narcissism), neuroticism (Neurotic), global need to belong (NtB1), warmth of ought self (OSWarm), openness to experience (Openness), optimism (Optimism), pragmatic love style (Pragma), preventional regulatory focus (Prevention), promotional regulatory focus (Promotion), rejection sensitivity (RejectionSen), interdependent self-construal (RelatSelfConstrual), romantic beliefs (RomBeliefs), subjective well-being (SWBTot), self-clarity (SelfClarity), self-control (SelfControl), self-esteem (SelfEsteem), sexual orientation (SexOr), social phobia (SocAnxiety), sociosexuality count (SocioCount), sociosexuality (SocioScale), importance of partner status (Status2), storge as love style (Storge), expected desirability (ThemInt), traditionalism (Traditional), expected selectivity (YouInt), objective attractivity rating (att_rating_act), speed-dating reasons long-term (getoutofltfind), speed date prediction of avoidance (v4minI9)

Figure 3

Permutation feature importance scores for the top ten predictors from all outer folds separated by effect and gender



Note. For an explanation of the predictor abbreviations see note of Figure 2.

Discussion

Flirting Variables and Cross-Validation

The first research question asked whether the addition of behavioral variables would improve the explained variance score in the prediction of actor and partner effect. Overall, the mean explained variance in the five outer folds of the cross-validation was descriptively greater when the behavioral variables were used as additional predictors. However, these results do not allow a generalization on other populations outside of this sample. Another aspect is that the absolute mean differences in R^2 scores for the actor effect are minimal. Compared to the partner effect, the mean explained variance increased more in the partner effect through the addition of the flirting variables. The small absolute difference in the actor effect resonates with Back et al. (2011a) who found that flirting correlated positively with romantic desire in the partner effect but not the actor effect. The study's descriptive results found for this specific sample therefore, support the results of Back et al. (2011a) that flirting does not necessarily express the actor's romantic interest in their partner, but still influences the partner's romantic interest in the actor.

In this analysis, more variance could be descriptively explained for the partner effect than for the actor effect. These results match with the original study from Joel et al. (2017) which also explained more variance in the partner effect than the actor effect but did not use cross-validation. However, the exact explained variance differs quite a bit from this study. For the actor effect, they reported about 5% to 9.5% of explained variance, depending on a more liberal or more stringent variable selection. In their supplemental material, they also published R^2 scores for random forest models not using a reduced set of predictors. Here, the explained variance for the actor effect was -5%. Since my study did not apply variable selection, the extracted mean score should be compared to Joel et al.'s (2017) score without variable selection.

In this case, it seems that no matter whether flirting variables were added or not, cross-validation descriptively increased the explained variance in the actor effect. This indicates that the single random forest model that Joel et al. (2017) used may have underestimated the explained variance as the cross-validated model in the current study delivers more robust results. However, since the variance of extracted R^2 scores in my study was also quite large, it is not surprising that Joel et al. (2017) found negative scores, as the scores in the cross-validation had a great range. The fact that the variance is so great makes it even more clear why a cross-validation is needed: Single scores from the random forest model can easily give an over- or underestimation of the model's performance, hence, researchers should opt for a solution that integrates or averages all these predictions – such as a cross-validation. For the partner effect, Joel et al. (2017) extracted R^2 scores ranging from about 0.18 to 0.26 based on how stringently they selected the variables. However, if no variable selection was applied, they were able to explain about 15% of the variance. All these scores are higher than the explained variance in the current study, no matter if the flirting variables were included or not. This indicates, that in this case, cross-validation prevented an overestimation of the explained variance.

In comparison to the original study, the results of the current study show that cross-validation is a valid and important addition to the analysis, providing more accurate results that find deeper underlying patterns and are also not as much influenced by extreme results through the process of averaging the single scores. Moreover, flirting variables do not seem to add significantly to the explained variance, as the generalizability of the small descriptive mean differences that were found is questionable.

Gender Differences and Cross-Validation

For the second research question, analysis aimed to determine whether gender differences existed in the model performance in actor and partner effect. Descriptively, it was found that more variance could be explained for male actor and partner effect than for female actor and partner effect, regardless of whether the flirting variables were included or not. However, these results can also not be generalized to the entire as their statistical significance was not tested. If these results were proven to be statistically significant, it would suggest that the models may not have been a good fit to predict female actor and partner desire. One reason for this could be that the chosen predictors were not the ones relevant for female effects. Another one could be that the female data was more heterogenous than the male data. This would have made it more challenging for random forest models to find underlying patterns, especially due to the limited sample size, and would also explain the lower scores. On the other hand, the limited sample size also may have contributed to more extreme results that differ from the true patterns that can be found in the population.

In the cross-validation, the extracted R^2 scores varied a lot. But taking the average scores, in this sample the male scores were for the actor effect better than the overall score for both genders in the actor effect, for females worse than the overall one. For the partner effect, both scores were worse than the overall one. In this analysis, the addition of the flirting variables did not change the results, the explained variance was in any case higher for male effects, indicating that the flirting variables did not have a substantial effect on only one gender.

Features

Actor Effect

During the feature importance analysis, it became noticeable that for men’s and women’s actor and partner effects, relevant predictors differed quite a lot. For the women’s actor effect, a wide range of predictors that were related to their evaluation of their actual-self proved to be important, such as intelligence, attractiveness, and warmth. For men only the latter one. Moreover, psychosocial variables like loneliness or social phobia were only relevant for women's romantic desire but not for men. The additional behavioral variables of flirting proved to be important for both. Interestingly, objectively rated attractiveness was only predictive for women's actor effect, while men's actor effect relied on variables such as traditionalism, or social status. In certain parts, these results resonate with Luo and Zhang (2009) as they also found that political conservatism and traditionalism predict men's actor effect. However, they also reported that physical attractiveness is a relevant predictor for men, but in this study, neither their self-reported nor their objective physical attractiveness influenced men's romantic desire for their partners. While Luo and Zhang (2009) found variables such as age and body weight predictive of women's actor effect, this information was not available in the current data set. Interestingly, they also found that the Big Five personality domains extraversion and openness to experiences predicted women's actor desire. Yet, in the current study, these predictors played a more marginal role, being both extracted only in one outer fold as one of the top ten predictors. While the current study confirms some of the results from Luo & Zhang (2009), it also found a lot of other variables to be relevant. While both studies used university samples, results may still differ due to differences in the methodology. One important factor could be the group size of each speed-dating event. In the current study, each participant had twelve speed dates, in Luo & Zhang (2009) it ranged from six to ten speed dates. But since the actor effect is moderated by the group size, with women's selectivity increasing with the group size while men's selectivity stays the same, results may differ (Fisman et al., 2006). Even though Joel et al. (2017) extracted feature importance scores, they cannot be compared meaningfully across models as they used variable selection which leaped different predictors in each model. Nonetheless, in their model with liberal variable selection, they extracted for example warmth, social phobia, and physical attractiveness as important predictors. Due to the differentiation between the genders, the current study offers the insight that social phobia and physical attractiveness are indeed relevant to the actor's romantic desire, however, only if the actor is a woman. At the same time, the smaller sample size in the current study that resulted from the gender-split datasets, increased variability and led to a higher chance of discovering random effects since variations can have a more substantial impact. This aspect limits the findings in comparison to Joel et al. (2017). Findings concerning the flirting variables somewhat contradict previous literature. Frisby et al. (2010) and Henningsen (2004) stated that women rather flirted for fun and relational reasons and men rather out of sexual interest. However, in this study, all three flirting variables were more predictive for female romantic desire than for male romantic desire.

Partner Effect

For the partner effect, both genders' romantic desire was predicted by their partner's self-evaluated physical attractiveness and their mate value. Additionally, men's romantic desire was related to the woman's loneliness, agency, and sociosexuality. Women's romantic desire was more based on the man's extraversion, rejection sensitivity, and narcissism. This confirms Asendorpf et al. (2011) and Luo and Zhang's (2009) results who found physical attractiveness to be highly predictive for the partner effect, in both men and women. At the same time, for women's partner desire, the man's physical attractiveness was in all five outer folds selected as a top ten predictor while for men's partner desire only three times. Even though random variations in the data should be expected due to the limited sample size, this could indicate that women base their romantic desire more on the physical attractiveness of their partner than men do. This would directly contradict studies such as Fisman et al. (2006) and Eastwick and Finkel (2008) who found an effect in the opposite direction. Whereas Feingold (1992) did not find any gender differences in personality-related predictors, the current study revealed that these variables in men are more predictive of women's partner desire for them than the other way around. When comparing the results to Joel et al.'s (2017) the same issue of variable selection persists as already named when discussing the actor effect. Yet, still it should be noted that with liberal variable selection, they also extracted mate value and self-rated physical attractiveness as some of the most important predictors. Concerning the additional flirting variables, results correspond to Back et al. (2011a) who also found flirting to be predictive of partner desire but did not find gender differences. At the same time, this effect was mainly found for agentic and general flirting, and not communal flirting. This indicates that a friendlier and warm flirting style may not create a romantic desire in the dating partner.

Limitations

Methodology

While machine learning methods, or in this specific case random forest models, offer a lot of advantages, as previously explained, they also have some shortcomings. Due to the small sample size of the current study, it is harder for the model to find deeper underlying patterns. Splitting the data into even smaller subsets in the nested cross-validation amplifies this issue. This is because it becomes more likely that the single subsets are more heterogenous in themselves but also differ strongly from the other subsets. This problem results in high variations of the performance metrics across the folds, which is also something noticeable in the current study since the R^2 scores had in many analyses very large standard deviations. Another factor is the measurement reliability of the predictors, as machine learning models' performance drastically decreases with unreliably measured predictors compared to linear regression models

(Jacobucci & Grimm, 2020). This is not unlikely something that lowered the random forests' model performances as at least the three flirting variables had quite low reliability in their measurement. To combat this, future studies could use larger sample sizes, even though this can be a challenging aspect of psychological studies. Moreover, the number of outer and inner folds could be adjusted or at least controlled in such a way, that the subsets in the single folds are more homogenous. Lastly, to see whether the performance is lower due to unreliably measured predictors, there could be a model performance comparison to a linear model such as regression.

In addition, the current study only analyzed descriptive differences. This prevents the results from being generalizable to the broader population, making them only relevant for this specific sample. Again, due to the small sample size, it would have been hard to find significant results as the power would likely not have been sufficient for the effect size that may would have been found. Nonetheless, future studies that use larger samples should test for statistical significance so more general conclusions can be drawn on the prediction of romantic desire.

Rating

Even though the three behavioral flirting variables turned out to be some of the best predictors for actor and partner effect, the differences in explained variance did not drastically increase through the addition of them nor do these descriptive differences indicate that they also exist in the whole population. It should therefore be noted that there are also some limitations to the rating of them. First, the ICC scores are quite low and can be considered with the guidelines from Koo and Li (2016) as "moderate". This indicates that there was to some extent agreement among the raters but not enough to reliably measure the variables. It can be assumed that this affected the quality of the predictive model. Second, some dates were not comprehensible due to either issues with the audio recording or the participants communicating in another language.

While these issues do not apply to all dates, the overall quality was influenced by them, impacting the results of the current study.

Sex, Gender, and Orientation

As stated in the sample description, the data does not allow clear differentiation of whether the item “Are you male or female?” measured “sex” or “gender” and hence, the decision was made to refer to it as gender in the current study, as I expected the participants to rather answer with their gender than sex. Nonetheless, since gender refers to socially constructed identities of men, women, or gender-diverse people, sex is based on biological and physiological differences, it should be expected that results would differ in a certain way, dependent on which variable is measured. Data collection of both variables in the current but also future studies would provide a more comprehensive picture of the population that is being studied, give more visibility and understanding to the experiences of non-binary and gender-nonconforming individuals, create the opportunity to research the relations between biological sex and gender identity and also enhance the quality and accuracy of the data. This could be easily put into practice by creating an item that clearly asks for biological sex “What is your biological sex?” and one that explicitly asks for gender.

For similar reasons for including gender and sex as separate variables, it is also beneficial to study non-heterosexual dyads in romantic relationship research. This could be done by creating speed-dating events specifically for non-heterosexual people. In the current study, only binary gender data and heterosexual pairs were considered, aligning with most already existing research. However, I encourage upcoming studies to explore gender distinctions more intricately, especially in contexts involving non-binary individuals and non-heterosexual dyads. With that, our understanding will encompass a broader range of gender identities and diverse relationship dynamics.

Effect Direction

Due to the exploratory nature of the current study and the used analysis methods, a wide range of important predictors for actor and partner effect was identified. However, permutation feature importance does not provide information on the direction of the relation between predictor and criterium. This means for example, that we know that extraversion predicts how much a person is desired by others, but we do not know if a higher or a lower level of extraversion leads to more desire. It also means, that if a predictor is deemed predictive for both genders, that this predictor does not necessarily influence the outcome variable in the same direction. If we knew, for example, that extraversion has a high permutation feature importance for both men and women in the actor effect, it could still be that high extraversion relates to more romantic desire in men, but in women, a low level of extraversion relates to more romantic desire, or vice versa. In future studies, it could hence be interesting to analyze the identified predictors with a different statistical tool that also allows to discover the direction of the relation.

Sample Population

The use of a *WEIRD* (Henrich et al., 2010), which stands for Western, educated, industrialized, rich, and democratic, sample likely introduced several aspects that limit the validity and generalizability of the results. The homogeneity of the sample may hinder the understanding of the influences of different cultures' romantic desire in dyadic effects. Therefore, the findings of this study may not hold true for non-WEIRD populations. Moreover, this sample is also characterized by their young age and the fact that it only includes university students, which makes it even more difficult to generalize the findings on other WEIRD populations. In future research, it could hence be interesting to test whether similar results can be obtained when the data is derived from more inclusive WEIRD populations but especially non-WEIRD populations.

Zero-Acquaintance

During the rating of the flirting, it became clear that quite many participants knew each other, which violates the assumption that the speed-dating paradigm provides zero-acquaintance data. It is not surprising that this particular sample was quite well-acquainted with each other as it was a sample from a university and singular sessions were divided by university year, raising the chances to have a date with a person one might already know from a common course. To battle this shortcoming, future research may more rigidly control this assumption, for instance by adding an item to the post-date questionnaire, so the data provides insight on zero-acquaintance and already acquainted participants can be analyzed separately.

Permutation Feature Importance

The current study used permutation feature importance to determine which predictors are the most important ones in predicting actor and partner effect. However, permutation feature importance has two shortcomings. First, it is dependent on the overall model performance, leaping lower feature importance scores for worse-performing models. This issue was combated by determining the top ten predictors within each outer fold, and not between each outer fold. Meaning, that the top ten predictors were always selected from each model separately. Otherwise, if one model in the outer folds performed drastically better than the other models, it would have been likely that mainly the predictors with the highest permutation feature importance scores from that model would have been selected. Still, an issue that persists is the accurate comparison of the permutation feature importance scores across the models in the outer folds. Because in this case the permutation feature importance scores can be over- or underestimated based on the model's performance. To solve this issue, a visualization of the scores relative to the model performance could be useful in the future.

Second, permutation importance does not take dependencies between predictors into account. This means if feature 1 is correlated with feature 2 and for example, feature 1 has only an increased feature importance when feature 2 is present, permutation feature importance fails to take this into account. Different variants of feature importance such as conditional permutation importance could help to combat this problem by permuting the feature 1 data conditioned on the other features. However, this would also drastically increase complexity and computational effort (Strobl et al., 2008).

Future Research

Compared to Joel et al. (2017) the current study did not use variable selection to reduce the number of predictors. Even though Joel et al. (2017) were able to explain substantially more variance in their analysis that included variable selection, their results are limited by the method they chose to select the variables as it increased the likelihood of overfitting the dataset by risking data leakage. Therefore, future research should select a different method of variable selection since irrelevant or noisy features can negatively impact the random forest's ability to capture meaningful patterns in the data. For this, the nested cross-validation could be useful as it allows to select the features on the training data and test selected predictors on the test data.

Moreover, the current study was not able to show significant differences in the prediction of romantic desire by including behavioral variables. Future research should hence be encouraged to use models with bigger samples that can obtain more generalizable and significant results. Furthermore, to expand the scope, a more diverse array of behavioral variables should be used. For instance, researchers could consider adding behavioral measures related to general communication styles, emotional expression, but also nonverbal behaviors such as eye contact, body language, or spatial proximity.

Another interesting aspect for future research may be the use of actual mate choice instead of mate preference as the outcome measure. While mate preferences give insight into individuals' stated desires, mate choice delves into real-life choices. The comparison of mate preference versus mate choice as outcome variables could be especially interesting since Regan (1998) already observed differences between short-term and long-term mating preferences and hence, actual mate choice should differ from mate preference (Asendorpf et al., 2011).

Implications and Conclusion

Conclusively, the current study addressed the topic of predicting romantic desire and identifying the most important predictors within this context. Due to its descriptive nature, the current study was unable to find results that can be generalized to the population outside of the sample at hand. Nonetheless, for this sample, it found differences in the explained variance of male and female romantic desire and was also able to marginally increase the explained variance by adding flirting variables. In addition, the use of cross-validation yielded results possessing greater reliability by minimizing the chance of over- or underestimating the descriptive effects. Overall, the current study signifies progress in extending psychological research methods and a deeper understanding of what makes us romantically desire others and others romantically desire us. Within that, a focus on behavioral variables revealed interesting aspects that have potential but demand future research. Finally, separated analysis by gender indicated that the aggregation of the gender variable may hinder the surfacing of gender-specific effects and information. Like two planets residing within the Milky Way, there are undoubtedly similarities between the genders. Nevertheless, it is crucial to also recognize inherent differences, because after all, men are from Earth, but women are from Venus.

References

Altmann, A., Tolosi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: A corrected feature importance measure. *Bioinformatics*, *26*(10), 1340–1347. https://doi.org/10.1093/bioinformatics/btq134

Asendorpf, J. B., Penke, L., & Back, M. D. (2011). From dating to mating and relating: Predictors of initial and long–term outcomes of speed–dating in a community sample. *European Journal of Personality*, *25*(1), 16–30. https://doi.org/10.1002/per.768

Back, M. D., & Kenny, D. A. (2010). The social relations model: How to understand dyadic processes. *Social and Personality Psychology Compass*, *4*(10), 855–870. https://doi.org/10.1111/j.1751-9004.2010.00303.x

Back, M. D., Penke, L., Schmukle, S. C., Sachse, K., Borkenau, P., & Asendorpf, J. B. (2011a). Why mate choices are not as reciprocal as we assume: The role of personality, flirting and physical attractiveness. *European Journal of Personality*, *25*(2), 120–132. https://doi.org/10.1002/per.806

Back, M. D., Penke, L., Schmukle, S. C., & Asendorpf, J. B. (2011b). Knowing your own mate value. *Psychological Science*, *22*(8), 984–989. https://doi.org/10.1177/0956797611414725

Breiman, L. (2001a). Statistical modeling: The two cultures. *Quality Engineering*, *48*(1), 81–82. http://biostat.jhsph.edu/~mmccall/articles/breiman_2001.pdf

Breiman, L. (2001b). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. In *ICML '06: Proceedings of the 23rd international conference on Machine* *learning*. Association for Computing Machinery. https://doi.org/10.1145/1143844.1143865

Cawley, G. C., & Talbot, N. L. C. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, *11*(70), 2079–2107. https://doi.org/10.5555/1756006.1859921

De Rooij, M., & Weeda, W. D. (2020). Cross-Validation: A method every psychologist should know. *Advances in Methods and Practices in Psychological Science*, *3*(2), 248–263. https://doi.org/10.1177/2515245919898466

Dufner, M., & Krause, S. (2023). On how to be liked in first encounters: The effects of agentic and communal behaviors on popularity and unique liking. *Psychological Science*. https://doi.org/10.1177/09567976221147258

Eastwick, P. W., & Finkel, E. J. (2008). Sex differences in mate preferences revisited: Do people know what they initially desire in a romantic partner? *Journal of Personality and Social Psychology*, *94*(2), 245–264. https://doi.org/10.1037/0022-3514.94.2.245

Eastwick, P. W., Luchies, L. B., Finkel, E. J., & Hunt, L. L. (2014). The predictive validity of ideal partner preferences: A review and meta-analysis. *Psychological Bulletin*, *140*(3), 623–665. https://doi.org/10.1037/a0032432

Feingold, A. (1992). Gender differences in mate selection preferences: A test of the parental investment model. *Psychological Bulletin*, *112*(1), 125–139. https://doi.org/10.1037/0033-2909.112.1.125

Finkel, E. J., Eastwick, A., & Matthews, A. J. (2007). Speed-dating as an invaluable tool for studying romantic attraction: A methodological primer. *Personal Relationships*, *14*(1), 149–166. https://doi.org/10.1111/j.1475-6811.2006.00146.x

Fisman, R., Iyengar, S. S., Kamenica, E., & Simonson, I. (2006). Gender differences in mate selection: Evidence from a speed dating experiment. *Quarterly Journal of Economics*, *121*(2), 673–697. https://doi.org/10.1162/qjec.2006.121.2.673

Frisby, B. N., Dillow, M. R., Gaughan, S., & Nordlund, J. R. (2010). Flirtatious communication: An experimental examination of perceptions of social-sexual communication motivated by evolutionary forces. *Sex Roles*, *64*(9–10), 682–694. https://doi.org/10.1007/s11199-010-9864-5

Furr, R. M. (2009). Personality psychology as a truly behavioural science. *European Journal of Personality*, *23*(5), 369–401. https://doi.org/10.1002/per.724

Gottschall, J. (2005). The heroine with a thousand faces: Universal trends in the characterization of female folk tale protagonists. *Evolutionary Psychology*, *3*(1). https://doi.org/10.1177/147470490500300108

Grammer, K., Kruck, K., & Magnusson, M. K. (1998). The courtship dance: Patterns of nonverbal synchronization in opposite-sex encounters. *Journal of Nonverbal Behavior*, *22*(1), 3–29. https://doi.org/10.1023/a:1022986608835

Henningsen, D. D. (2004). Flirting with meaning: An examination of miscommunication in flirting interactions. *Sex Roles*, *50*(7/8), 481–489. https://doi.org/10.1023/b:sers.0000023068.49352.4b

Henrich, J., Heine, S., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/s0140525x0999152x

Jacobucci, R. (2018). *Are decision trees stable enough for psychological research?* PsyArXiv. http://dx.doi.org/10.31234/osf.io/f2utw

Jacobucci, R., & Grimm, K. J. (2020). Machine learning and psychological research: The unexplored effect of measurement. *Perspectives on Psychological Science*, *15*(3), 809–816. https://doi.org/10.1177/1745691620902467

Joel, S., Eastwick, P. W., & Finkel, E. J. (2017). Is romantic desire predictable? Machine learning applied to initial romantic attraction. *Psychological Science*, *28*(10), 1478–1489. https://doi.org/10.1177/0956797617714580

Johnson, C. L., Stockdale, M. S., & Saal, F. E. (1991). Persistence of men's misperceptions of friendly cues across a variety of interpersonal encounters. *Psychology of Women Quarterly*, *15*(3), 463–475. https://doi.org/10.1111/j.1471-6402.1991.tb00421.x

Kennair, L. E. O., Wade, T. J., Tallaksen, M. Ø., Grøntvedt, T. V., Kessler, A. M., Burch, R., & Bendixen, M. (2022). Perceived effectiveness of flirtation tactics: The effects of sex, mating context and individual differences in US and Norwegian samples. *Evolutionary Psychology*, *20*(1). https://doi.org/10.1177/14747049221088011

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Koul, A., Becchio, C., & Cavallo, A. (2018). Cross-validation approaches for replicability in psychology. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01117

Kurzban, R., & Weeden, A. J. (2007). Do advertised preferences predict the behavior of speed daters? *Personal Relationships*, *14*(4), 623–632. https://doi.org/10.1111/j.1475-6811.2007.00175.x

Li, N. P., Yong, J. C., Tov, W., Sng, O., Fletcher, G. J. O., Valentine, K. A., Jiang, Y. F., & Balliet, D. (2013). Mate preferences do predict attraction and choices in the early stages of mate selection. *Journal of Personality and Social Psychology*, *105*(5), 757–776. https://doi.org/10.1037/a003377

Little, M. A., Varoquaux, G., Saeb, S., Lonini, L., Jayaraman, A., Mohr, D. C., & Kording, K. P. (2017). Using and understanding cross-validation strategies. Perspectives on Saeb et al. *GigaScience*, *6*(5). https://doi.org/10.1093/gigascience/gix020

Luo, S., & Zhang, G. (2009). What leads to romantic attraction: Similarity, reciprocity, security, or beauty? Evidence from a speed-dating study. *Journal of Personality*, *77*(4), 933–964. https://doi.org/10.1111/j.1467-6494.2009.00570.x

Place, S. S., Todd, P. M., Penke, L., & Asendorpf, J. B. (2009). The ability to judge the romantic interest of others. *Psychological Science*, *20*(1), 22–26. https://doi.org/10.1111/j.1467-9280.2008.02248.x

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*(1), 81–106. https://doi.org/10.1007/bf00116251

Regan, P. C. (1998). Minimum mate selection standards as a function of perceived mate value, relationship context, and gender. *Journal of Psychology & Human Sexuality*, *10*(1), 53–73. https://doi.org/10.1300/j056v10n01_04

Song, Y., & Lu, Y. (2015). Decision tree methods: Applications for classification and prediction. *Shanghai Arch Psychiatry*, *27*(2), 130–135. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4466856/

Sprecher, S., & Duck, S. W. (1994). Sweet talk: The importance of perceived communication for romantic and friendship attraction experienced during a get-acquainted date. *Personality and Social Psychology Bulletin*, *20*(4), 391–400. https://doi.org/10.1177/0146167294204006

Stachl, C., Pargent, F., Hilbert, S., Harari, G. M., Schoedel, R., Vaid, S. S., Gosling, S. D., & Bühner, M. (2020). Personality research and assessment in the era of machine learning. *European Journal of Personality*, *34*(5), 613–631. https://doi.org/10.1002/per.2257

Strobl, C., Boulesteix, A., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, *9*(1). https://doi.org/10.1186/1471-2105-9-307

Strobl, C., Malley, J. D., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, *14*(4), 323–348. https://doi.org/10.1037/a0016973

Todd, P. M., Penke, L., Fasolo, B., & Lenton, A. P. (2007). Different cognitive processes underlie human mate choices and mate preferences. *Proceedings of the National Academy of Sciences*, *104*(38), 15011–15016. https://doi.org/10.1073/pnas.0705290104

Vélez, J. I. (2021). Machine learning based psychology: Advocating for a data-driven approach. *International Journal of Psychological Research*, *14*(1). https://doi.org/10.21500/20112084.5365

Wiggins, J. S., Trapnell, P. D., & Phillips, N. E. (1988). Psychometric and geometric characteristics of the revised interpersonal adjective scales (IAS-R). *Multivariate Behavioral Research*, *23*(4), 517–530. https://doi.org/10.1207/s15327906mbr2304_8

Hiermit versichere ich, dass ich die Arbeit mit dem Titel „Men are from Earth, Women are from Venus: Gender Differences in Predicting Actor and Partner Effects with Machine Learning in Initial Romantic Attraction“ selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt habe.

Datum 17.08.2023…………… Unterschrift……………………………