All Codes are included in the **Analysis Code** section bellow

**Analysis description:**

**Analysis 1.** At the final ratings, we asked participants whether they thought their partner wanted to date them or not (yes =1/no=2). We subtracted these answers from partners’ actual response (partner really wants to date: yes =1/no=2). This resulted in either accurate (0), false negative (1) or false positive (-1) answers. We then binarized the accuracy variable (correct/incorrect (pooling 1 and -1)) to test for significance above chance level with Chi-square test (alpha = 0.05).

**Analysis 2.** Apart from categorical answers (date or not), we asked participants to rate how attractive is their partner and how much do they think partner likes them? We asked these questions to investigate whether participants’ impression of being liked relates to their attraction towards their partner. We used multilevel linear mixed model with 3 level structure: dyad (Level 3), participant (Level 2) and time (Level 1). In this model, Participants’ impression of being liked was used as the target variable predicted by participant’s attraction towards their partner. We further included gender and the interaction between gender and attraction.

We split the data by gender and averaged participant’s ratings across all three interactions. Then, two simple linear regressions were conducted: In the first linear regression, we used partner’s perceived attraction scores (How attractive is your partner? Scale: 0-9) as a predictor of participant’s impression of being liked (How much partner likes me? Scale: 0-9). In the second linear regression, we used the partner’s perceived attraction score as predictors of partner’s actual liking score.

**Analysis 3.** We tested whether females’ and males’ differ in frequency of their naturally occurring expressions, eye fixations, and physiological responses. To do so, for each interaction type (first impression, verbal interaction and nonverbal interaction) and each participant we calculated the proportion of time (min = 0, max = 1) that participants were (i) smiling, (ii) laughing, (iii) head shaking, (iv) making hand gestures, (v) touching their face or fixating on partners’ (vi) body, (vii) eyes, (viii) face, (ix) head. Physiological levels: (x) skin conductance and (xi) heart rate responses were baseline corrected (30 seconds prior to every interaction) and then z-scored. This resulted in eleven averaged values for each subject and interaction. We used 3 x 2 Multivariate Generalized Linear Mixed model to test for gender differences using the within subject factor interaction type (first impression, verbal and nonverbal interaction), gender (male, female) as between subject factor. To control for multiple comparisons we employed a false discovery rate (FDR) in all following models56. To check whether females look longer at the background than males do, we conducted a Generalized Linear Mixed Model. In this model, the data were nested in each subject and the individual intercept was random. The average time (in seconds) looking at the background was used as a dependent variable and gender, interaction type (first impression, verbal, nonverbal), gender \* interaction type were used as fixed effects (see Supplementary Fig.1).

**Analysis 4.** Apart from physiological arousal, we investigated whether males and females differ in their cognitive arousal by conducting Multivariate Generalized Linear Mixed model testing for gender differences on mood grids: (i) arousal (ii) valence, self-ratings reporting the level of (ii) shyness and (v) self-confidence.

**Analysis 5.** Furthermore, we tested whether males and females behave in a specific way when they feel attracted to their partner. To account for within subject and dyad dependencies, we conducted a series of mixed effects models with following structure: three time points (Level 1) nested in participants (Level 2), nested in dyads (Level 3). In each model, expression, gaze fixations frequencies and baseline corrected physiological responses were used as predictors of participant’s attraction scores (scale 0 – 9). Gender and genders \* expression/fixation interaction were used as additional predictors of attraction. Due to the multicollinear nature of the data, we carried out a model for each expression, gaze fixations frequencies and physiological response independently (11 mixed effects models). The mixed effects models were conducted such that the intercept terms were allowed to vary across dyads and participants, we further used an AR1 covariance matrix to account for time dependencies. We defined significance using an FDR < 0.05.

**Analysis 6.** We used machine-learning techniques to detect a specific behavioural and physiological pattern that would predict participant’s attraction level. Using a 10-fold cross-validated lasso regression model (alpha = 0.5) implemented in Sci-kit learn in python 3.63, we aimed to predict attraction directly following either the verbal or non-verbal interaction, separately for men and women. In our model, we included all the predictors including participants’ frequency of smiling, laughing, hand gestures, fixations duration on partner’s eyes, face, head, body and physiological responses (skin conductance, heart rate) as well as 2-way interactions between all those features (91 predictors in total) to predict males’ and females’ self-reported attraction scores. All predictors were z-scored. To evaluate the performance of our models we performed a permutation test with 3000 permutations, shuffling the attraction levels across participants and testing that our observed non-shuffled R2 was larger than 95% of the randomly shuffled R2.

**Analysis 7.** We ran a correlation between all measures. This resulted in a large correlation table showing associations between male’s and female’s expressions eye fixations and physiological measures as well as associations between female’s-female’s, male’s-male’s showing how nonverbal behaviors and physiological responses relate to each other within participants. Then in control analysis, each female was paired with a random male. To test for significance, we directly contrasted the (FDR corrected) correlations coefficients between true couples and randomly matched couples with cocor package in R studio 57 using gender an independent group, two sided test with alpha set to 0.05.

**Quantifying expressive mimicry and eye fixation synchrony.** We quantified mimicry for each dyad and interaction by calculating the proportion of time both participants’ directly reciprocated expressions (smiling, laughing, head nod, hand gestures, face touching) and gaze fixations (looking at partners’ head, eyes, face, body).

**Quantifying synchrony.** We quantified synchrony with lagged windowed-cross correlational analyses4. This method has the advantage of differentiating the degree of synchronization by quantifying it on a continuous (correlation) scale. This is important as the level of synchrony may fluctuate during the experiment. We determined the parameters following an extensive process by comparing previous studies using similar statistical methods, looking at what is physiologically plausible given the time course of the physiological signals and by employing a data-driven bottom-up approach where we investigated how changing the parameters affected the outcomes using a different dataset. As expected, the absolute values of the synchrony measures varied depending on the parameters, but as supported by McAssey, Helm, Hsieh, Sbarra, and Ferrer58, the relative results were not affected (e.g. a dyadic manifesting relatively high synchrony showed such tendency for the different parameters). Based on these three factors, we set the parameters as follows: the window size was 8 seconds (160 samples), the window increment was 3 seconds (40 samples), the maximum lag was 4 seconds (80 samples) and the lag increment was 100ms (2 samples). Then the peak picking algorithm was applied4 (Boker et al., 2002). This algorithm allows detecting the maximum cross-correlation across the lags for each time segment. For example, if a participant synchronizes with her partner with a lag of one second, the cross-correlations will become higher the closer the segments from the two participants are shifted towards the point where they are one second apart from each other. When the two signals are lagged by exactly one second the cross-correlation is highest (the peak). Both the windowed cross-correlations and the peak picking algorithm are conducted 6 times per dyad, once for the heart rate responses and once for the skin conductance responses for each condition (the first impression, verbal and nonverbal interaction) resulting in N dyads \* 6 result and peak picking matrices. Finally, the mean and standard deviation of the peak cross-correlations of all window segments and the mean of the absolute values of corresponding time lags are calculated for both physiological measures for each condition per dyad.

**Analysis 8.** We test whether attraction can be predicted by synchrony. In this model, we used synchrony in expressions (smiling, laughing, head nod, hand gestures, face touching) and gaze fixations (looking at partners’ body, head, eyes, face) and physiology (skin conductance, heart rate) as predictors of participant’s attraction. In addition, gender, interaction type (verbal, nonverbal), the order of interaction (verbal/nonverbal first) were used as additional predictors in the model. To allow for differences between dyads, the intercept terms were allowed to vary across dyads and we included a first-order autoregressive AR(1) residuals structure to account for time dependencies. The final model was selected with a backward stepwise selection of fixed effects in a generalized linear mixed-effects model. This method first tests interaction terms, and then drops interactions one by one to test for main effects. Main effects that are part of interaction terms were retained, regardless of their significance as main effects (full model summarized in Supplementary Table 8).