nature portfolio

https://doi.org/10.1038/s41562-021-01197-3

Supplementary information

Physiological synchrony is associated with attraction in a blind date setting

In the format provided by the authors and unedited

Supplementary Information for

Physiological synchrony predicts attraction in a blind date setting

E. Prochazkova[1,2], E. E. Sjak-Shie[1,2], F. Behrens[1,2], D. Lindh[3,4,5] & M. E. Kret[1,2*]

Correspondence to: Dr. M. E. Kret
Email: m.e.kret@fsw.leidenuniv.nl.

This PDF file includes:

- Figure S1 to S3
- Tables S1 to S20
- Quantification of physiological synchrony

Other supplementary materials for this manuscript include the following:

- Movies S1

Supplementary Figures

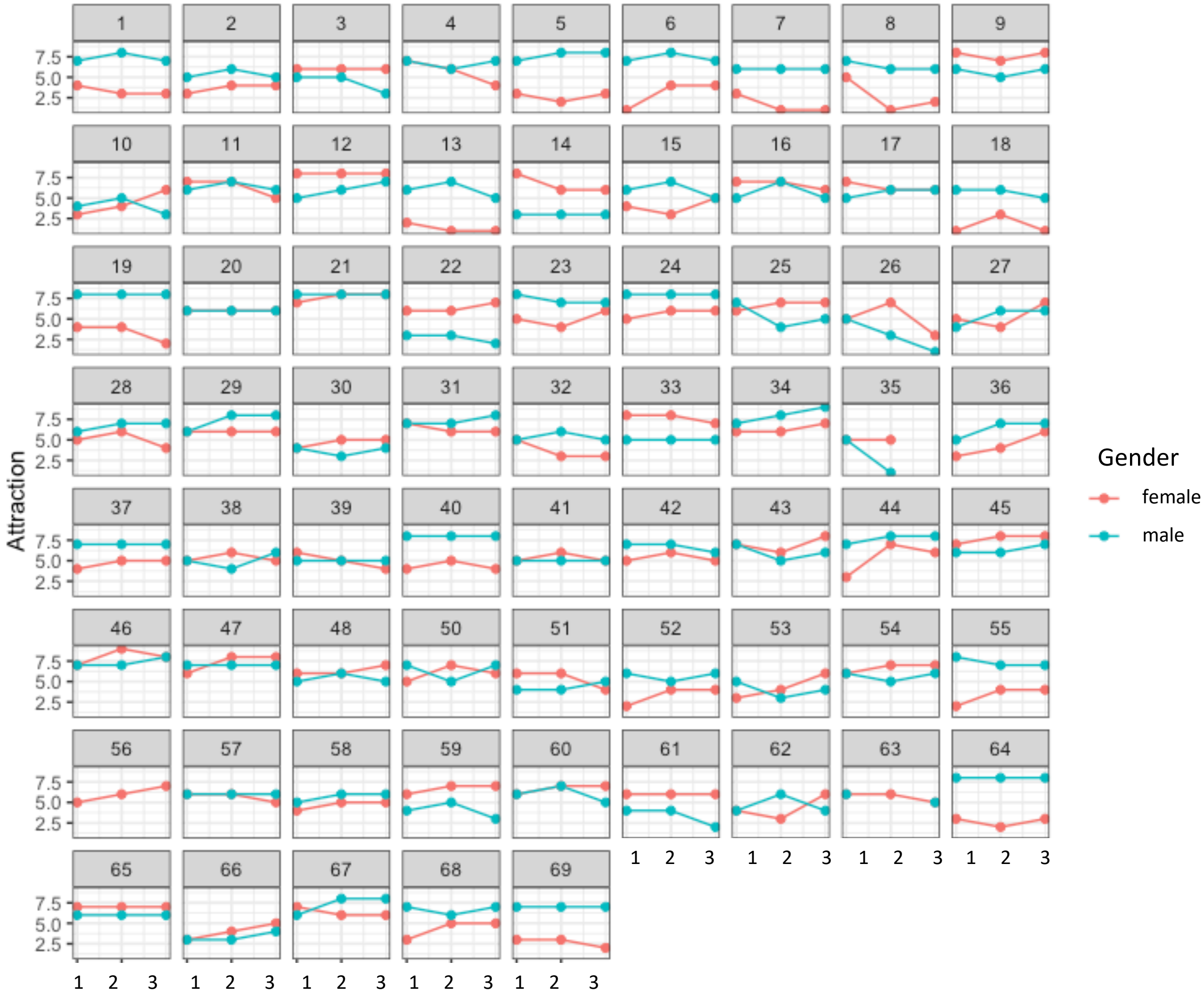

Figure S1. These line graphs provide an example of how attraction changed over time. Time: 1 = first impression, 2 = second interaction, 3 = third interaction. The rating scale was 0 – 9.

Spearman's rank order correlations assumptions

A monotonic relationship is a statistical definition which describes a scenario in which the size of one variable increases as the other variables also increases. A visual inspection suggested a monotonic relationship in all variables; aside from hand gestures, which initially looked non-monotonic. We then examined the data closer (see heatmap bellow) and saw that there have not been many hand gestures exchanged. Specifically, we only found 16 data points for males' hand gestures and 18 data points for females' hand gestures. At the same time, the males' and females' data showed a tight relationship where if one person in the couple displayed hand gestures, the other partner was very likely to show hand gesture as well (heatmap: Figure S2). Specifically, the top row of the heatmap represents females' gestures and the lower row represents male's hand gestures. The bars from top-down show that when females' made hand gestures, males were very likely to displayed hand movement as well (red color = no gestures, colder colors = more time gesturing). Thus, the heatmap bellow shows that participants tend to mimic each other.
Thus, the assumptions for Spearman's rank order correlations were met.

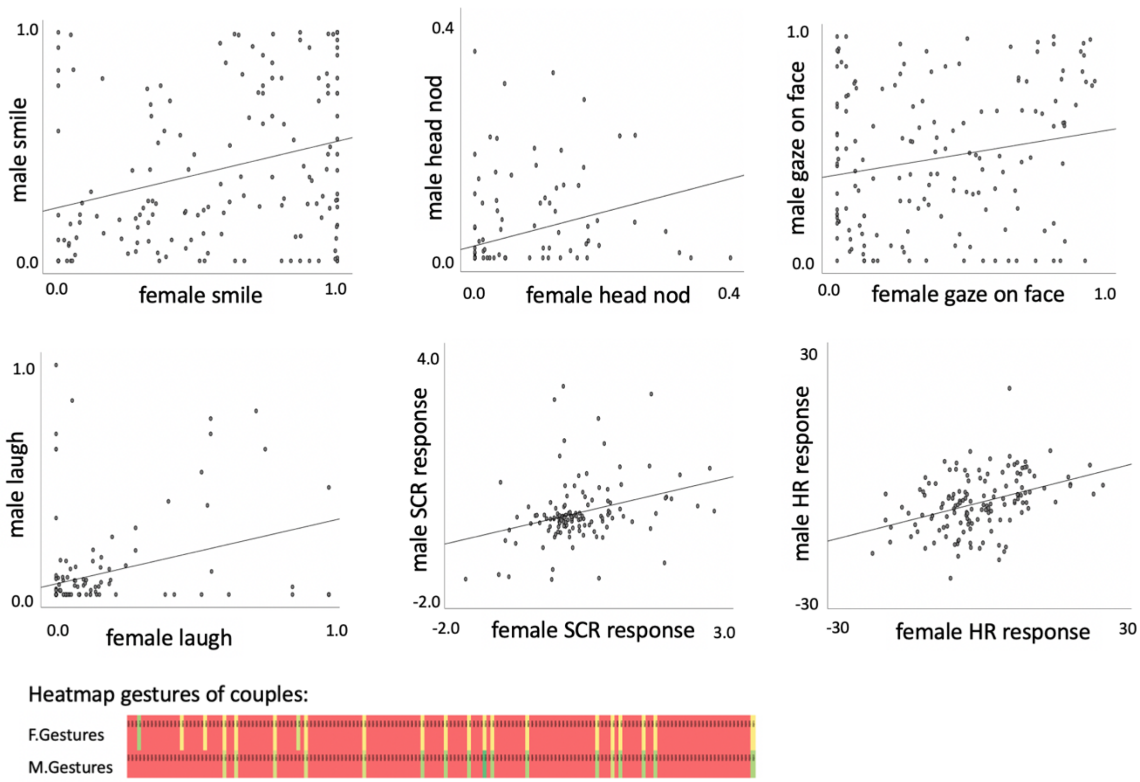

Figure S2. Scatter plots between males' and females' behavioral expressions and physiology.

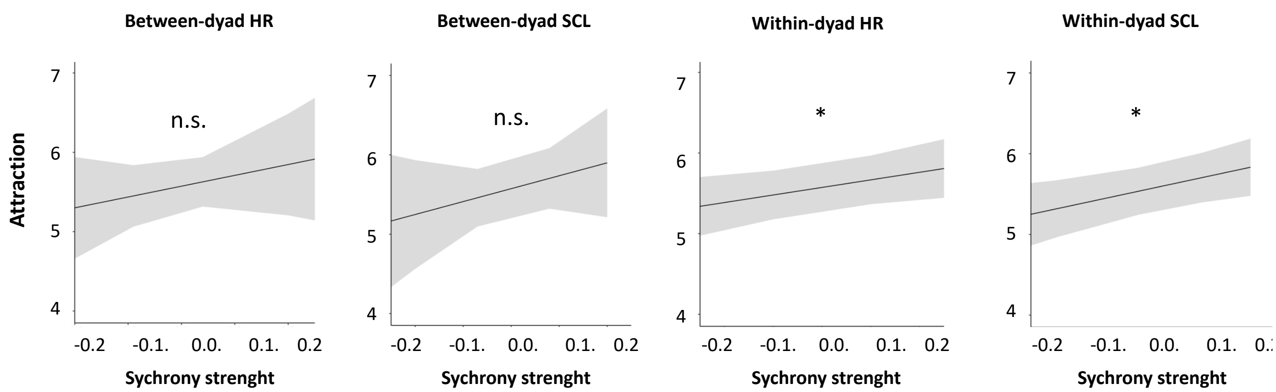

Figure S3: The line graphs represent slopes extracted from our Multilevel linear mixed model predicting attraction based on synchrony measures reflecting between-dyad variations and within-dyad variation. The shaded areas represent 95% confidence intervals.

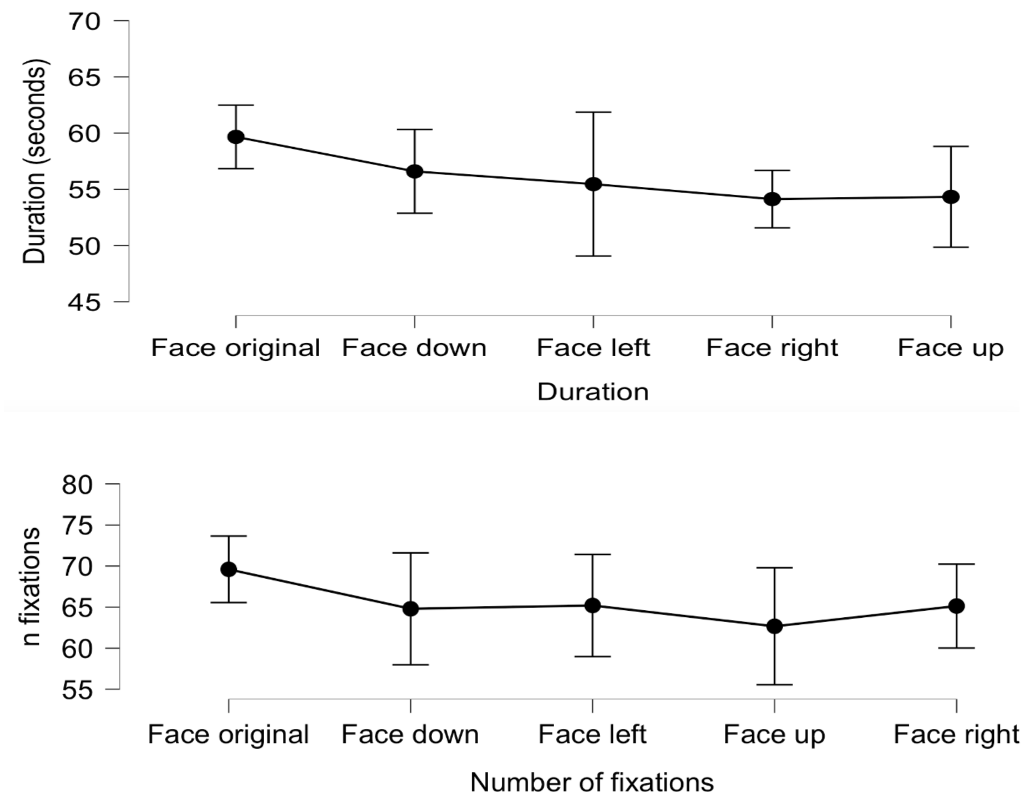

Figure S4. The effect of face AOI position on (top) fixation duration (bottom) number of fixations. Original face represents the face mask that corresponds to participants' actual face position, (ii) face down: face mask shifted down by 26px, (iii) face up: face mask shifted up by 26px and (iv) face left: face mask shifted left by 26px t, (v) face right: face mask shifted right by 26px. Error bars represent 95% confidence intervals.

Supplementary Tables

Table S1 shows synchrony associations (focusing on the circles in Figure 1) within real dating partners compared to randomly matched pairs. Significant evidence was found for seven types of synchrony in: smiles, laughs, head nods, hand gestures, face-to-face gaze, heart rate, and skin conductance. There is no significant difference in touching face, body gaze and eye contact fixations between true couples and randomly matched couples.

Table S1. Correlation comparisons between true couple and randomly matched couples

	True Male	Random Male	Fisher's Z	p
Female's Face	0.22	-0.50	**2.44**	**0.01**
Female's Body	0.29	0.15	1.25	0.21
Female's Laugh	0.51	-0.03	**5.24**	**0.00**
Female's Smile	0.31	0.09	**2.09**	**0.03**
Female's Hand gestures	0.87	-0.04	**12.11**	**0.00**
Female's Head nod	0.67	-0.10	**8.12**	**0.00**
Female's Touch Face	0.29	0.10	1.76	0.07
Female's Skin conductance	0.32	0.09	**2.16**	**0.03**
Female's Heart rate	0.39	0.16	**2.23**	**0.02**

The study design (Table S2): When we first designed this experiment, it was important that we introduced verbal and nonverbal conditions. This is because if we would have only included verbal interaction, it would be difficult to determine whether physiological synchrony influences attraction because it reflects the quality of the verbal interaction, or if physiological synchrony predicts attraction even when people do not communicate verbally. On the other hand, if we would include only nonverbal interactions (e.g., direct eye gazing), one may argue that this type of nonverbal interaction rarely happens in real life (questioning the ecological validity of this study). Thus, the key reason why we chose to include verbal/nonverbal conditions was to control for verbal and nonverbal influences on attraction. At the same time, we hoped that these conditions would not have a significant influence on the target variable. This was confirmed in bellow full model Supplementary Table S2, which summarizes results of the Multilevel linear mixed models where we investigated whether our experimental design impacted on participant's attraction ratings (0-9). The multilevel model had following structure: three time points (Level 1) nested in participants (Level 2) nested in Dyads (Level 3). We included factors of gender, time (first impression, first interaction, second interaction), the type of interaction (verbal/nonverbal), the order of interaction (first verbal/nonverbal) and two-way and three-way interactions between: gender * interaction type, gender * verbal interaction first.

Table S2. The Summary of Multilevel linear mixed models predicting attraction ratings (0-9) based on gender, time, the type of interaction and the order of interaction

	Attraction			
Predictors	*Estimates*	*CI*	*t-value*	*p*
(Intercept)	4.77	4.16 – 5.38	15.29	**<0.001**
Time	0.18	0.03 – 0.32	2.32	**0.020**
Gender (0 = female, 1 = male)	1.12	0.28 – 1.96	2.60	**0.009**
Verbal interaction first (0 = no, 1 = yes)	0.44	-0.41 – 1.28	1.02	0.309
Interaction type (0 = nonverbal/ 1 = verbal)	0.04	-0.36 – 0.43	0.18	0.855
Gender * Verbal interaction first	-0.38	-1.57 – 0.81	-0.62	0.533
Gender * Interaction type	-0.30	-0.86 – 0.26	-1.04	0.300
Interaction type * Verbal interaction first	-0.33	-0.94 – 0.27	-1.08	0.282
Gender * Verbal interaction first * Interaction type	0.41	-0.39 – 1.21	1.00	0.320
Random Effects				
σ^2	0.74			
$\tau_{00\ ID\_num}$	0.02			
$\tau_{00\ ID\_num.1}$	2.10			
$\tau_{00\ Dyad}$	0.00			
ICC	0.05			
N_{Dyad}	54			
N_{ID_num}	108			
Observations	324			

Table S3 summarizes results of the Multilevel linear mixed models where we investigated how different types of interpersonal synchronies impact on participant's attraction ratings (0-9). The multilevel model had following structure: three time points (Level 1) nested in participants (Level 2) nested in dyad (Level 3). We included all 7 synchrony predictors including synchrony in (i) smiles, (ii) laughs, (iii) head nods, (iv) hand gestures, (v) face-to-face, (vi) heart rate, and (vii) skin conductance. The full model further included factors of gender, time (first impression, first interaction, second interaction), the type of interaction (first impression, verbal, nonverbal), the order of interaction (verbal/nonverbal first) and two-way interactions between the type of interaction * and the type of synchrony (smiles, laughs, head nods, hand gestures, face-to-face, heart rate, and skin conductance) and the type of interaction*interaction order. The final model was selected with a backward stepwise selection of fixed effects.

Table S3. The summary of the full multilevel linear mixed model predicting attraction based on synchrony measures, gender, time, the type of interaction, the order of interaction and interactions between the type of interaction * synchrony

	Attraction			
Predictors	*Estimates*	*CI*	*t-value*	*p*
(Intercept)	4.56	3.99 – 5.13	15.67	**<0.001**
Time	0.38	0.15 – 0.60	3.28	**0.001**
Gender (0 = female, 1 = male)	0.89	0.32 – 1.46	3.08	**0.002**
Verbal interaction first (0 = no, 1 = yes)	0.46	-0.14 – 1.06	1.49	0.136
Interaction type (0 = nonverbal/ 1 = verbal)	0.46	-1.95 – 2.86	0.37	0.710
SCL synchrony	1.38	0.24 – 2.51	2.38	**0.017**
HR synchrony	0.98	0.12 – 1.83	2.22	**0.026**
Face-to-face contact	-0.26	-1.03 – 0.51	-0.67	0.502
Smile mimicry	0.18	-0.32 – 0.68	0.70	0.484
Laugh mimicry	0.09	-0.96 – 1.14	0.17	0.866
Nodding mimicry	-2.04	-7.29 – 3.21	-0.76	0.446
Gesture mimicry	-10.13	-30.74 – 10.47	-0.96	0.335
Verbal interaction first * Interaction type	-0.51	-1.06 – 0.05	-1.80	0.072
Interaction type * SCL synchrony	0.54	-1.48 – 2.55	0.52	0.601
Interaction type * HR synchrony	0.49	-1.75 – 2.74	0.43	0.667
Interaction type * Face-to-face contact	0.16	-0.64 – 0.96	0.40	0.689
Interaction type * Smile mimicry	0.17	-1.06 – 1.41	0.28	0.783
Interaction type * Laugh mimicry	-2.29	-5.69 – 1.11	-1.32	0.187
Interaction type * Nodding mimicry	32.95	-207.67 – 273.57	0.27	0.788
Random Effects				
σ^2	0.70			
$\tau_{00\ ID\_num}$	0.01			
$\tau_{00\ ID\_num.1}$	2.03			
$\tau_{00\ Dyad}$	0.00			
N_{Dyad}	54			
N_{ID_num}	108			
Observations	324			

Note: Time had three time points: first impression, first interaction, second interaction

Table S4. The summary of the final multilevel linear mixed model predicting attraction based on synchrony measures

	Attraction			
Predictors	*Estimates*	*CI*	*t-value*	*p*
(Intercept)	4.93	4.50 – 5.36	22.49	**<0.001**
Time	0.20	0.07 – 0.32	3.07	**0.003**
Gender (0 = female, 1 = male)	0.90	0.30 – 1.49	3.00	**0.004**
SCL synchrony	1.33	0.35 – 2.31	2.67	**0.008**
HR synchrony	0.98	0.16 – 1.80	2.34	**0.020**
Random Effects				
σ^2	0.71			
$\tau_{00\ ID\_num}$	0.02			
$\tau_{00\ ID\_num.1}$	2.08			
$\tau_{00\ Dyad}$	0.00			
ICC	0.05			
N_{Dyad}	54			
N_{ID_num}	108			
Observations	324			

Note: Time had three time points: first impression, first interaction, second interaction

Control analysis – Does within or between dyad synchrony predict attraction?

In the previously described model, the variables for heart rate and skin conductance level synchrony included within- and between-dyad level variation in synchrony. It is therefore unclear whether couples that are highly attracted to each other synchronize more than those who are not (i.e., between-dyad effect), or whether changes in physiological synchrony over time predict attraction changes (i.e., within-dyad effect). To disentangle the two types of variations, we computed two variables: (1) Between-dyad SCL synchrony: the averaged synchrony level across time points per dyad, and (2) Within-dyad SCL synchrony: the deviation in synchrony level (per time point) from the dyad's averaged synchrony level (within-dyad centering). Both variables were included in a Multilevel linear mixed model with a three-level structure (three-time points (Level 1), nested in participants (Level 2) and dyad (Level 3). We also included a random intercept effect (across participants) and a random slope for Time, but not allowing a correlation between both random effects. The slope for time indicated the evolution of attraction over time.

Table S5. The summary of the final multilevel linear mixed model predicting attraction based on synchrony measures reflecting between-dyad variations (dyad's overall level of synchrony), and within-dyad variation (changes in synchrony level over time within each dyad)

	Attraction			
Predictors	*Estimates*	*CI*	*t-value*	*p*
(Intercept)	4.93	4.51 – 5.36	22.32	**<0.001**
Gender	0.90	0.32 – 1.47	2.98	**0.002**
Time	0.19	0.07 – 0.32	3.01	**0.002**
Within-dyad SCL synchrony	1.30	0.28 – 2.31	2.47	**0.012**
Between-dyad SCL synchrony	1.63	-1.46 – 4.72	1.01	0.302
Within-dyad HR synchrony	0.94	0.10 – 1.78	2.16	**0.028**
Between-dyad HR synchrony	1.36	-1.48 – 4.21	0.92	0.347
Random Effects				
σ^2	0.71			
$\tau_{00\ ID\_num}$	2.08			
$\tau_{00\ ID\_num.1}$	0.02			
ICC	0.74			
N_{ID_num}	108			
Observations	324			

Note: Time had three time points: first impression, first interaction, second interaction.

Cross-validation analysis

Moreover, to evaluate robustness of our data, we considered using additional cross-validation analysis. Specifically, we additionally applied clustered bootstrapping to estimate the confidence intervals of the coefficients. Thus, in addition to the parametric approach of determining statistical significance of fixed effects with conditional F-Tests and marginal significance of fixed-effect coefficients conditional t-tests, their respective non-parametric confidence intervals were calculated by fitting generalized linear model with the cluster bootstrap function (clusbootglm) in R. The Table 1 bellow shows that after 1000 permutations, the coefficient of the HR and SC synchrony were the only coefficient for which the confidence interval of the bootstrap analysis did not include zero, indicating its stability**.** This bootstrap analysis supported that synchrony effects, for HR and SC synchrony were both robust.

Table S6. Clustered bootstrapping to estimate the confidence intervals of the coefficients

Attraction	Nonparametric CI	
	2.5%	97.5%
(Intercept)	4.5	5.4
Gender	0.3	1.4
Time	0.07	0.31
Within-dyad SCL synchrony	**0.08**	**2.49**
Between-dyad SCL synchrony	-1.28	4.48
Within-dyad HR synchrony	**0.09**	**1.78**
Between-dyad HR synchrony	-1.904	4.83

Control analysis – Does increase in physiological arousal predict attraction?

In the current study we observed that synchrony in skin conductance level and heart rate could predict attraction. One possible confound is that it is not the synchrony on the dyadic level, but the arousal responses of the two individuals that drive these findings. For example, skin conductance levels might rise if a participant feels attracted to his/her partner. Consequently, the responses of the two participants would highly correlate reflecting the individuals' attraction rather than an interpersonal process. To test this, we conducted a control analysis where attraction was regressed against the participants' baseline corrected heart rate and skin conductance levels per each interaction (first impression, verbal, nonverbal). For the physiological measures, we first computed the mean skin conductance and heart rate change from the baseline and then standardized the responses per participant. Consistent with the model of the main analysis, we included gender and time as a control variable including individual as a random intercept effect. The model summary is shown in Table S6 which shows that attraction could not be predicted by the arousal responses of the two individuals.

Table S7. The summary of multilevel linear mixed model with the heart rate (hr) and skin conductance level (scl) predicting participants' attraction ratings

	Attraction			
Predictors	*Estimates*	*CI*	*t-value*	*p*
Intercept	4.98	4.54 – 5.42	22.21	**<0.001**
Gender	0.92	0.32 – 1.51	3.03	**0.003**
Time	0.13	-0.01 – 0.26	1.83	0.067
SC response	0.01	-0.01 – 0.02	0.28	0.779
HR response	0.03	-0.06 – 0.12	0.60	0.544
Random Effects				
σ^2	0.74			
$\tau_{00\ \mathrm{ID\_num}}$	0.03			
$\tau_{00\ \mathrm{ID\_num.1}}$	2.11			
$\tau_{00\ \mathrm{Dyad}}$	0.00			
N_{Dyad}	54			
$N_{\mathrm{ID\_num}}$	108			
Observations	324			

Note SC = Skin conductance, HR = Heart rate

Control analysis – is attraction a valid outcome variable?

One may wonder whether we really measured attraction in this study or possibly something else. To control for this possibility, throughout the experiment we also collected other ratings including trust, liking, feeling of connection and click. We also asked whether participants felt awkward or anxious. Theses scores were then compared with attraction ratings **(Part 1)** and participants choice to go for another date (yes/not) with the partner **(Part 2)**.

Table S8. The descriptive statistics of participants' ratings

	N	Minimum	Maximum	Mean	Std. Dev.
			Self-ratings		
Valence	362	2	9	6.01	1.632
Arousal	362	1	9	5.94	1.757
Self-confident	397	1	9	5.65	1.467
Awkward	395	1	9	4.80	2.197
Shy	398	1	9	4.34	2.003
			Partner ratings		
Trustworthy	408	1	9	6.87	1.438
Intelligent	409	2	9	6.59	1.318
Funny	409	1	9	5.96	1.611
Attractive	408	1	9	5.57	1.711
			Self - Partner ratings		
Similar personality	408	1	8	4.86	1.712
Connection	411	1	8	4.48	1.838
Click	404	1	9	4.38	1.871
Sex. Attraction	410	1	9	3.83	1.945

The scale for all ratings ranged between 0 – 9, Descriptive statistics are based on 138 participants (N = 69 dyads) rating their partner three times (after first impression, verbal and nonverbal interaction). Valence = higher number represents positive valence. Arousal = higher number represents more arousal levels.

Control analysis – is attraction a valid outcome variable? (Part 1)

To identify the common dimensions of ratings, we took all thirteen ratings and submitted them to a principal component analysis (PCA), using the Oblimin with Kaiser Normalization rotation method). The first principal component (PC) accounted for 37.7% of the variance and the second PC accounted for 17.2% of the variance of the mean trait judgments. All positive judgments (e.g., attractive, funny, similar in personality, feeling of click, connection) had positive loadings, and all negative feelings (e.g., awkward, shy, low self-confidence) had negative loadings on the first PC (Table S4), suggesting that it can be interpreted as valence evaluation.

Table S9. The principal component analysis: loadings of participants' ratings

	Component 1 (Positive)	Component 2 (Negative)
Click	**.876**	-.035
Connection	**.861**	-.003
Sexual Attraction	**.826**	.070
Attractive	**.775**	.089
Funny	**.720**	.094
Similar personality	**.699**	.011
Intelligent	.562	.002
Trustworthy	.514	-.104
Valence	.483	-.285
Arousal	.235	.169
Awkward	-.049	**.880**
Shy	.183	**.851**
Self-confident	.092	**-.751**

Note. Self-confidence is negatively loaded to feelings of awkwardness and shyness because more confident people were less awkward and shy, they felt. The PCA was based on N = 344 valid cases.

Control analysis – is attraction valid outcome variable? (Part 2)

Multilevel binary logistic regression investigates how different types of ratings predict participants' choice to go for another date (yes/ no, coded 1 and 0 respectively). We included all 13 ratings as predictors. The results showed that the model was highly predictive of participants' choice to date their partner again (Overall percentage reached 80.4% accuracy). Among all the ratings only positive affect and attraction predicted participants decisions significantly, whereas attraction ratings explained the most variance in participants' binary decision to date their partner (yes/no).

Table S10. The Multilevel binary logistic regression: ratings predict participants' choice to go for another date

	Date partner (yes/no)			
Predictors	*Estimates*	*CI*	*t-value*	*p*
Intercept	-1.701	-6.54– -3.14	-.696	.488
Valence	.432	0.04 – 0.82	2.188	.031
Arousal	-.111	-0.43 – 0.21	-.672	.503
Shy	-.268	-0.75 – 0.22	-1.091	.278
Awkward	.154	-0.26 – 0.57	.734	.464
Self-confident	-.398	-0.88 – 0.85	-1.637	.105
Attractive	.623	0.13 – 1.11	2.534	**.013**
Funny	.039	-0.39 – 1.14	.179	.858
Intelligent	-.785	-1.46 – -0.11	-2.307	.023
Trustworthy	.245	-0.38 – 0.86	.779	.438
Similar personality	-.162	-0.60 – 0.28	-.721	.472
Connection	.375	-0.12 – 0.87	1.506	.135
Sex. Attraction	.276	-0.24 – 0.79	1.050	.296
Click	-.045	-0.59 – 0.50	-.161	.873
Random Effects				
$\tau_{00\ ID\_num}$	0.10			
N_{Dyad}	68			
Observations	324			

Table S11. Participants' Demographics and Other Descriptive Statistics

	Female			Male			
Characteristics	N	M	SD	N	M	SD	t
Age	69.00	23.45	4.18	69.00	25.71	4.64	-2.96 **
Weight	68.00	65.84	10.34	67.00	78.06	8.70	-7.42 ***
Height	69.00	171.90	6.98	68.00	182.51	6.44	-9.24 ***
Number of alcohol drinks	68.00	0.01	0.55	68.00	0.58	0.55	-3.13 **
How much commitment (0-9)	64.00	5.92	1.64	65.00	5.72	1.43	0.73
Time single in months	62.00	38.69	65.31	59.00	37.03	65.45	0.14
	Average percent						
Education	Female (N = 69)			Male (N = 69)			
VMBO	0.02			0.00			
HAVO	0.04			0.02			
VWO	0.02			0.15			
MBO	0.13			0.08			
HBO	0.30			0.38			
WO	0.47			0.38			

Note: *** $p < .001$, VMBO: the lowest completed high-school level, WO: the highest level (scientific education, Bachelor or Master degree). How much commitment is on 0-9 scale.

Table S12. Comparisons (t-test) variables by sex: Liebowitz Social Anxiety Scale (LSAS), Positive and Negative Affect Schedule (PANAS) and Sexual Desire Inventory (SDI). The SDI is comprised of 11 items about various sexual behaviours, on a 5-point Likert scale. The total score on the SDI is the sum of all 11 items, with higher scores reflecting a higher sexual desire. The LSAS is comprised of two subscales: performance and social interaction. The 24 questions ultimately lead to six subscale scores: total fear, fear of social interaction, fear of performance, total avoidance, avoidance of social interaction and avoidance of performance. The statements had to be answered on a 4-point scale (0 = not at all, 4 = totally). The PANAS: consists of two 10-item mood scales, measuring positive affect (PA) and negative affect (NA). Participants are asked to rate their experience with a certain emotion on a 5-point scale (1 = very slightly or not at all, 5 = very much).

Table S12. Comparisons (t-test) variables by sex

	Female		Male			
Variable	M	SD	M	SD	t	p
SDI	4.25	1.47	4.68	0.93	-1.692	0.094
LSAS Fear	0.46	0.49	0.45	0.64	0.099	0.921
LSAS avoidance	0.77	0.67	0.81	0.69	-0.271	0.787
PANAS negative	1.69	0.45	1.72	.40	-.280	0.780
PANAS positive	3.20	0.56	3.41	.53	-1.853	0.067
Trust baseline	8.10	0.94	7.75	1.02	2.084	**0.039**
Trust overall	6.85	1.56	6.89	1.29	.186	0.817

Note: Trust baseline measures how trustworthy a potential partner should be, trust overall measures average trust across three interaction periods.

Table S13: Comparison between real world and manual mapping fix numbers

Face fixations	*Real world mapping:*	*Manual mapping:*
First impression (fix number)	*4 (out of 30) 13%*	*4 (out of 30) 13%*
Nonverbal interaction	*101 (out of 1200) 84%*	*100 (out of 1200) 83%*
Eye fixations	*Real world mapping:*	*Manual mapping:*
First impression (fix number)	*0 (out of 30) 0%*	*0 (out of 30) 0%*
Nonverbal interaction	*47 (out of 1200) 3,9%*	*45 (out of 1200) 3,8%*

Control: Time-lagged mimicry vs. direct mimicry

For comparison reasons, we also quantified 'direct mimicry 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). The proportion of mimicry was calculated for each condition (the first impression, verbal and nonverbal interaction) resulting in N dyads * 3 results * for mimicry in smiles, laughs, head nods, hand gestures, face-to-face fixations. Not surprisingly, lagged classification resulted in more data being classified as mimicry compared to direct mimicry measures (see Supplementary Table S14). Subsequently, we run our main analyses twice, once with direct and once with lagged mimicry values. This multilevel linear mixed model investigated how different types of interpersonal synchronies impact on participant's attraction ratings (0-9). The multilevel models had following structure: three time points (Level 1) nested in participants (Level 2) nested in Dyad (Level 3). We included all 7 synchrony predictors including synchrony in (i) smiles, (ii) laughs, (iii) head nods, (iv) hand gestures, (v) face-to-face, (vi) heart rate, and (vii) skin conductance. The model further included factors of gender, time (first impression, first interaction, second interaction)**,** the type of interaction (first impression, verbal, nonverbal), the order of interaction (verbal/nonverbal first). Gender (0 = female, 1 = male), Interaction type (0 = nonverbal/ 1 = verbal), Verbal interaction first (0 = no, 1 = yes) and two-way interactions between verbal first*interaction type and all synchrony variables*interaction type.

Comparison between the two approaches of quantifying synchrony depicted in the **Supplementary Table S3** and **Table S15** showed that the "direct mimicry" and "lagged mimicry" measures yield highly comparable results. This made us confident that both lagged and direct synchrony provide very similar information.

Table S14. Descriptives: comparison between the direct synchrony values vs. lagged synchrony values

		Direct mimicry		Lagged mimicry	
	N	Mean	Std.	Mean	Std.
Face-to-face contact	324	0,0132	0,0634	0,0984	0,1765
Face-to-face contact	324	0,1930	0,2426	0,3030	0,2914
Head-to-head gaze	324	0,2393	0,2556	0,3642	0,2985
Smile mimicry	324	0,3014	0,3123	0,3508	0,3285
Laugh mimicry	324	0,0355	0,1144	0,0447	0,1319
Nodding mimicry	324	0,0033	0,0111	0,0105	0,0267
Gesture's mimicry	324	0,0016	0,0051	0,0020	0,0059
Touching face mimicry	324	0,0027	0,0198	0,0047	0,0325
Average	-	**0,0987**	0,1281	**0,1473**	0,1615

Direct mimicry = Original synchrony values where person A directly does the same expression as person B
Lagged mimicry = turn-taking component of synchrony with 5 sec. lag

Table 15. The summary of multilevel linear mixed model predicting attraction with <u>direct mimicry</u>

	Attraction			
Predictors	*Estimates*	*CI*	*t-value*	*p*
(Intercept)	4.56	3.99 – 5.13	15.67	**<0.001**
Time	0.38	0.15 – 0.60	3.28	**0.001**
Gender (0 = female, 1 = male)	0.89	0.32 – 1.46	3.08	**0.002**
Verbal interaction first (0 = no, 1 = yes)	0.46	-0.14 – 1.06	1.49	0.136
Interaction type (0 = nonverbal/ 1 = verbal)	0.46	-1.95 – 2.86	0.37	0.710
SCL synchrony	1.38	0.24 – 2.51	2.38	**0.017**
HR synchrony	0.98	0.12 – 1.83	2.22	**0.026**
Face-to-face contact	-0.26	-1.03 – 0.51	-0.67	0.502
Smile mimicry	0.18	-0.32 – 0.68	0.70	0.484
Laugh mimicry	0.09	-0.96 – 1.14	0.17	0.866
Nodding mimicry	-2.04	-7.29 – 3.21	-0.76	0.446
Gestures mimicry	-10.13	-30.74 – 10.47	-0.96	0.335
Verbal interaction first * Interaction type	-0.51	-1.06 – 0.05	-1.80	0.072
Interaction type * SCL synchrony	0.54	-1.48 – 2.55	0.52	0.601
Interaction type * HR synchrony	0.49	-1.75 – 2.74	0.43	0.667
Interaction type * Face-to-face contact	0.16	-0.64 – 0.96	0.40	0.689
Interaction type * Smile mimicry	0.17	-1.06 – 1.41	0.28	0.783
Interaction type * Laugh mimicry	-2.29	-5.69 – 1.11	-1.32	0.187
Interaction type * Nodding mimicry	32.95	-207.67 – 273.57	0.27	0.788
Random Effects				
σ^2	0.69			
$\tau_{00\ \mathrm{ID\_num}}$	0.02			
$\tau_{00\ \mathrm{ID\_num.1}}$	2.04			
$\tau_{00\ \mathrm{Dyad}}$	0.00			
N_{Dyad}	54			
$N_{\mathrm{ID\_num}}$	108			
Observations	324			

Note: Time had three time points: first impression, first interaction, second interaction.

Quantifying behavioral mimicry

Mimicry is defined broadly as 'doing what others are doing'. This includes mimicry of mannerisms, gestures, postures, and gaze movements. While some studies are very loose on their definition of mimicry; for instance, mimicry might be defined as any movement following the other person's movement (Fujiwara & Daibo, 2016; Tschacher, Rees, & Ramseyer, 2014). We adopt a stricter definition of mimicry where behavioral mimicry occurs when two or more people engage in the same behavior within a short window of time, typically no longer than five seconds (for review see, Chartrand and Lakin, 2013). To quantified mimicry for each dyad and interaction, motor movements (smiling, laughing, head nod, hand gestures, face touching) were coded by four independent raters (two raters for males and two for females). Eye fixations falling on pre-defined areas of interests were mapped manually or with assisted mapping with Tobii Pro Lab analyzer. Both emotional expression or eye fixation were classified per tenths of seconds as binary variable (1 for occurrence, 0 for no occurrence). We then classified as mimicry when a person A directly repeats the same expression as person B within five seconds window. Specifically, if subject A showed a behavior and subject B showed the same behavior within that time point or within a duration of (5 seconds = 50 windows) after the time point, an instance of lagged mimicry was noted. We did the same for the person B to A. "Mutual mimicry" was then calculated by summing the number of noted mimicries from A to B and B to A and divided by twice the total number of timepoints. This variable accounted for proportion of time both participants displayed mimicry (smiling, laughing, head nod, hand gestures, face touching) and gaze fixations (looking at partners' face and body). The proportion of mimicry was calculated for each condition (the first impression, verbal and nonverbal interaction) resulting in N dyads * 3 results * for mimicry in smiles, laughs, head nods, hand gestures, face-to-face fixations.

Quantification of physiological synchrony

A variety of methods have been proposed in previous literature to quantify synchrony including correlations, regressions, structural equation models and recurrence quantification analyses. These approaches differ in their assumptions, their operationalization of synchrony, and the type of synchrony they measure (for reviews, see Gates & Liu, 2016; McAssey et al., 2013; Schoenherr et al., 2018; Thorson, West, & Mendes, 2017). In the current article, we focus on continuous time series measures in dyads. For this type of data, it is important that the method captures responses that happen "in sync" but also responses that occur with a small-time delay (e.g., one individual responds to another or at a different pace). Furthermore, the method needs to allow for changes in the level of synchrony as it will vary depending on the events happening in a conversation with moments of stronger and weaker synchrony.

Two methods that take non-stationarity into account are lagged windowed cross-correlation (Boker, Xu, Rotondo, & King, 2002) and recurrence quantification analysis (Gates & Liu, 2016). Cross-RQA, much like crosscorrelation, can quantify the strength of relation between one person's behavior at a given time point with the other person's behavior at an earlier time point. At its simplest, RQA arrives at a "recurrence rate," or the overall probability that states will reoccur. While both analyses allow to measure the magnitude of synchrony, the recurrence quantification analysis (RQA) has been predominantly applied in the psychological sciences on time series containing nominal data (e.g., observational coding, Gates & Liu, 2016). As physiology was measured on continuous scale, for this reason we decided to use lagged windowed cross-correlation (Boker, Xu, Rotondo, & King, 2002). Additionally, we feel that windowed cross-correlation is more intuitive to interpret. Consequently, we decided to apply this method which provides measures of the strength of synchrony. The objective of the lagged windows-cross correlations analysis (Boker et al., 2002) is to calculate the strength of association between two time series while taking into account the non-stationarity of the signals and the lag between responses, that is, to consider the dynamics of a dyadic interaction. Specifically, the time series are segmented into smaller intervals, calculating the cross-correlation for each segment. This allows the means and variances to differ between segments accounting for non-stationarity. This is important as the level of synchrony may change during the experiment, sometimes having moments of strong synchronization while during other times responding less strong to one another. Additionally, as the strength of association between two time points may differ depending on how far apart they are from each other, the segments are moved along the time series by an increment such that two adjacent segments overlap. Hence, segmenting the time series into smaller intervals and partially overlapping these intervals while moving along the time series provides a better estimate of the local strength of association between the physiological signals of two participants.

Besides the dynamics in the strength of synchronization during the course of the experiment, participants differ in how fast one might respond to a certain event or the other person. In other words, participants might not always be perfectly "in sync" whereby one participant might sometimes respond to the other person or vice versa introducing a delay between the responses of two individuals. To account for this, for each segment, the signals of the two participants are lagged in relation to one another. Specifically, the signal of participant 1 is kept constant while the signal of participant 2 is shifted more and more by a specified lag increment until a maximum lag is reached. Next, the same procedure is performed the other way around with participant 2 being kept constant. The maximum lag determines what is still considered synchrony. For example, if the maximum lag is four seconds, responses from two participants that are four seconds apart from each other are still considered synchronized. On the other hand, if one participant reacts to a certain event and the other participant shows a response 5 seconds later, it is not considered a response to the same event anymore and therefore does not count as synchrony. Based on this approach, there are four parameters that need to be determined: (1) the length of each segment, referred to the window size w_{max}; (2) the increment with which the segments are moved along the time series, the window increment w_{inc}; (3) the maximum with which two segments can be lagged from one another, the maximum lag τ_{max}; and (4) the increment with which two segments are lagged from each other, the lag increment τ_{inc}. We determined the parameters following an extensive process by comparing previous studies using similar statistical methods, by 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, & Ferrer, 2013), 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 2 seconds (40 samples), the maximum lag was 4 seconds (80 samples) and the lag increment was 100ms (2 samples).

Calculating the cross correlations of each lag for each window segment generates a result matrix with each row representing one window segment and each column indicating a lag. The middle column represents the cross-correlation with a lag of zero, while the first and last column contain the cross-correlations for the maximum lag of participant 1 and 2. Hence, the number of columns in the result matrix is $(2^* \tau_{max} / \tau_{inc}) + 1$. The number of rows is given by $(N - w_{max} - \tau_{max})/ w_{inc}$, with N being the number of observations in the whole time series. Based on this result matrix, a so-called peak picking algorithm is applied. For each segment (i.e., each row in the matrix), the maximum cross-correlation across the lags is detected closest to the zero-lag (i.e., across all columns in a given row). If that maximum correlation is preceded and followed by smaller correlations, it is marked as a peak. For example, if participant 2 synchronizes with participant 1 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). If the signals are lagged further away from each other, the cross-correlation decreases again. If, however, a peak cannot be detected, the algorithm assigns a missing value for that segment. This might be the case, for example, if people do not respond to an event or to each other (e.g., both participants wait and do nothing). The peak picking algorithm outputs a matrix with two columns, containing the value of the maximum cross-correlation (the peak) and the corresponding lag at which the peak cross-correlation is detected. The output has the same number of rows as the result matrix as it searches for a peak cross-correlation for each window segment.

Both the windowed cross-correlations and the peak picking algorithm were 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 54 dyads * 6 result and peak picking matrices. Finally, the mean cross-correlations of all window segments were calculated for both physiological measures for each condition per dyad.

Control: Quantification of the impact of systematic error on the eye-trackers ability to discriminate between the face and body AOI hits

1. Quantify the impact of systematic error in the gaze position signals.

A recent publication by Niehorster et al. (2020) implicated that Tobii glasses' gaze quality measures changes with the target location. In subsequent post-hoc tests, we therefore decided to validate whether the accuracy of the eye-tracker is sufficient for our main analyses including our smallest AOI: Face. We did this by assessing the impact that deviations reported by Niehorster et al. (2020) would have on our findings, given our existing data. Specifically, we were interested to what extent deviations in gaze-position accuracy are likely to influence our main result where we show that face-to-face gaze synchrony had no significant impact on attraction.

We then explain the following steps:

I. Using the deviations reported by Niehorster et al. (2020), we tested the effect of the glasses' inaccuracy on the face AOI hits. To do this, first, we assessed whether partners were mostly viewed in the center position or also the other positions (in terms of the various errors). Retrospective visual inspection of all videos confirmed that participants were sitting approximately in the center of partner's visual field.

II. Next, we measured the length of the table width on the screen, in px, and divide that by the actual width of the table, to get the pixel/cm ratio. In our case, the width of the table was 1000 mm, which corresponded to 550 px. The 550-pixel width of the table was calculated from a snapshot image - for example, see Figure 4. in the manuscript. Thus, 1000 mm/550 px corresponded to 1.81 mm/px.

III. We then calculate deviations in cm from the degrees mentioned by Niehorster et al. (2020) paper. When compared to the grid reported by Niehorster et al. (2020), the AOI for face mostly fallen between the center and top quadrant depicted in the Niehorster et al. (2020) diagram below. Thus, we expected that the deviations for face correspond to **max 2.7° error**.

IV. We then quantified deviations for face, whereas the distance to the other person's face was 1000 mm. (The table had a square shape 1000 x 1000 mm). Deviation (face) = 1000*tan (2.7°) = 47 mm. To Pixels: Deviation (face) = 47 mm * 0.55 mm/px = 26 px (The 0.55 being derived from the 1 px/1.8 mm the authors mention above).

V. Shift AOIs for face: up, down, left, right by 26 pixels found above.

VI. Rerun analysis for 15 interaction segments.

VII. Compare the difference between the existing AOIs and the shifted AOIs.

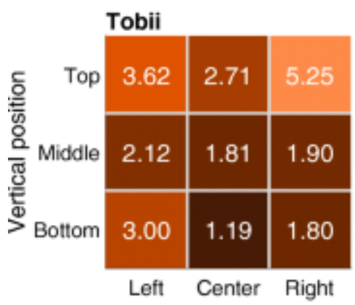

Fig. 6 a) Accuracy deviation adapted from Niehorster et al. (2020).

Knowing this, we reasoned that if an error of 26 pixels could affect our results, moving the face mask around should have a significant effect on fixation duration and the number of fixations when running the same AOI analysis again. Since we want to evaluate the absence of an effect, we ran a Bayesian repeated measures ANOVA (using JASP) to evaluate evidence for or against there being a difference between conditions. Figure 7 shows that by shifting the mask, the number of fixations and the duration of the fixations slightly decreased. However, we found that the bayes factor for the null-model (i.e. there is no difference when moving the face mask) for fixation duration was $BF_{01} = 3.729$, meaning that the null-hypothesis is 3.7 times more likely given our data and indicating strong evidence against our manipulation having any effect. Similarly, the bayes factor for the null-hypothesis in regards to the number of fixations was $BF_{01} = 6.673$. See below for plots, relevant descriptive statistics and model comparisons.

To conclude, given the size of our face AOIs, and the distance between them, we posit that even with the inaccuracy of the tracker, the discrimination stength between the face, body AOI hits was acceptable for our purposes. In the manuscript itself we highlight that because of the anticipated technical limitation of eye tracker, eyes-to-eye fixations were omitted from the main analysis. In the discussion we touch upon technical limitations of current set-up, including the deviations.

Table S16. Descriptives for fixation time

				95% Credible Interval	
Duration	**Mean**	**SD**	**N**	**Lower**	**Upper**
Face down	56.600	43.585	15	32.463	80.737
Face left	55.467	43.184	15	31.552	79.381
Face original	59.667	45.017	15	34.737	84.596
Face right	54.133	41.811	15	30.979	77.287
Face up	54.333	42.885	15	30.585	78.082

Table S17. Descriptives for number of fixations

				95% Credible Interval	
Number of fixations	**Mean**	**SD**	**N**	**Lower**	**Upper**
Face down	64.800	53.159	15	35.361	94.239
Face left	65.200	53.065	15	35.814	94.586
Face original	69.600	56.804	15	38.143	101.057
Face right	65.133	52.784	15	35.903	94.364
Face up	62.667	47.737	15	36.231	89.102

Table S18. Model Comparison fixation duration

Models	P(M)	P(M|data)	BF_M	BF_{01}	error %
Null model (incl. subject)	0.500	0.789	3.729	1.000	
Duration	0.500	0.211	0.268	3.729	0.644

Table S19. Model Comparison for number of fixations

Models	**P(M)**	**P(M|data)**	**BF_M**	**BF_{01}**	**error %**
Null model (incl. subject)	0.500	0.870	6.673	1.000	
Number of fixations	0.500	0.130	0.150	6.673	0.991

Differences in gaze behavior (face gaze) and expressions (smiles) between the three events

The events all took place in similar settings – all Dutch events surrounding the topic of art, film and music and participants could walk in freely. We tried to keep the all relevant variables as consistent as possible across events. The seating and lighting conditions across testing sites were very similar as we brought our own equipment including the table, chair and lights. As for ambient noise, dyads were always situated in a separate room with minimal ambient noise. In terms of behavior, we conducted a post-hoc test to see if there was significant difference in gaze behavior (face gaze) and expressions (smiles) between the three events, the results of One-Way ANOVA showed that there were not find significant differences in gaze behavior (face gaze) and expressions (smiles) between the three events.

Table S20. One-way ANOVA mean duration of smiles and mean gaze face fixation across events

	Event	Mean	Std. Deviation	F	Sig.
Face	1.00	0.33	0.29	1181,00	.311
	2.00	0.49	0.31		
	3.00	0.35	0.36		
Smile	1.00	0.74	0.32	1359,00	.261
	2.00	0.63	0.30		
	3.00	0.63	0.37		
	3.00	0.01	0.03		