Supplement

S1. Oversampling Procedure

Prior to recruitment, 163 children were screened for BI using the 30-item Behavioral Inhibition Questionnaire (Bishop et al., 2003). We used cutoffs established in previous research on extreme temperament in children between the ages of 4 and 15 (Broeren & Muris, 2010) to define BI as a BIQ total score ≥ 119 or BIQ social novelty score ≥ 60. Scores below the BI cutoff were designated as not Behaviorally Inhibited (BN). Of the 163 potential participants, 39 children were above the BI cutoff. From this sample, 21 dyads were recruited and completed the Free Play episode. In the current sample, 15 children were above the BI cutoff. Although we recruited children via oversampling for the larger study, we use continuous BIQ score in the current analyses and use the cutoffs only for visualization purposes.

S2. Creation of Videos for MET Coding

We conducted manual gaze correction using the plug-in from Pupil Player v.0.9.12 (Pupil Labs) based on our validation procedure (Pérez-Edgar et al., 2019). Two trained RAs independently conducted a manual gaze correction for each participant's video by determining whether the red circle created in Pupil Player aligned with where the child was looking during the validation procedure (Pérez-Edgar et al., 2019). The master coder compared the manual gaze corrections from the two coders. If they were within 0.03 of each other, the master coder's coordinates were chosen. If there was a discrepancy greater than 0.03, the participant was excluded (N = 3). We exported the eye-tracking videos with the selected manual gaze corrections. We then used Final Cut Pro (Apple) to sync the eye-tracking videos with their respective room videos. The integrated recordings were inspected to ensure that the two recordings were not out of sync for more than three frames. These integrated videos were then

exported for MET and affect coding with a resolution of 1920x1080 pixels at 30 frames-per-second.

S3. Assessing the Effect of Dyadic Behavioral Inhibition

An interaction between each dyad partner's level of behavioral inhibition could influence the course of the social interaction. Thus, we tested for this emergent effect by including a self behavioral inhibition x peer behavioral inhibition interaction in Models 2 and 3.

For Model 2 assessing likelihood of gazing at the peer vs anywhere else in the room, we did not find that including a self behavioral inhibition x peer behavioral inhibition interaction changed our central finding that children were more likely to look at the peer than anywhere else in the room when self-expressing positive affect (b = 0.70, p < .001). Additionally, we did not find that self behavioral inhibition (b = -0.30, p = .433), peer behavioral inhibition (b = -0.31, p = .419) or the self behavioral inhibition x peer behavioral inhibition interaction (b = -0.00, p = .993) was related to gaze. This model was not a superior fit to the model presented in the manuscript.

For Model 3 assessing likelihood of self-expressing positive vs neutral affect, we did not find that including a self behavioral inhibition x peer behavioral interaction changed the overall meaning of our central finding that children at higher levels of BI were less likely to self-express positive affect in the presence of peer expressed positive affect (b = -0.37, p = .057). We also continued to see a significant AOI x peer-expressed affect interaction (b = -1.16, p = .024). We additionally saw a main effect of peer behavioral inhibition such that when children interacted with a peer with a higher behavioral inhibition score, children were less likely to self-express positive affect (b = -1.03, p = .031). We did not find that self behavioral inhibition (b = -0.78, p = .099) or the self behavioral inhibition x peer behavioral inhibition interaction (b = -0.56, p = .128) was related to self-expressed affect. We advise caution in interpreting these results as the

model including the self behavioral inhibition x peer behavioral inhibition was a significantly worse fit than the model presented in the manuscript ($X^2 = 9.46$, $p = .009$).

Table S1: Model comparisons for model 1

	Model Formula	AIC	BIC	logLik	deviance
Model 4	count ~ ptss + ages + aoi2	847.53	861.90	-417.77	835.53
Model 2	count ~ ptss + ages + aoi2 + (1 | id)	849.53	866.29	-417.77	835.53
Model 3	count ~ ptss + ages + aoi2 + (1 | peerGroup)	849.53	866.29	-417.77	835.53
Model 1	count ~ ptss + ages + aoi2 + (1 | peerGroup/id)	851.53	870.69	-417.77	835.53

Note: Model 4 was the best fit and is presented in full in the manuscript. All other model details are available in the data analysis code (Pérez-Edgar et al., 2019).

Table S2: Model comparisons for model 2

	Model Formula	AIC	BIC	logLik	deviance
Model 3	aoi2 ~ ptss + ages + biqs + affects2 + affectp2 + (1 | peerGroup/id)	3145.6	3196.7	-1564.8	3129.6
Model 8	aoi2 ~ ptss + ages + onsets + biqs + affects2 + affectp2 + (onsets || peerGroup/id)	3098.7	3168.9	-1538.3	3076.7
Model 7	aoi2 ~ ptss + ages + onsets + biqs + affects2 + affectp2 + (onsets | peerGroup/id)	3085.3	3168.4	-1529.7	3059.3
Model 2	aoi2 ~ ptss + ages + biqs * affects2 + biqs * affectp2 + affects2 * affectp2 + (1 | peerGroup/id)	3148.7	3219.0	-1563.3	3126.7
Model 1	aoi2 ~ ptss + ages + biqs * affects2 * affectp2 + (1 | peerGroup/id)	3150.7	3227.3	-1563.3	3126.7
Model 6	aoi2 ~ ptss + ages + biqs + affects2 + affectp2 + (affects2 + affectp2 || id)	3150.9	3234.0	-1562.5	3124.9
Model 4	aoi2 ~ ptss + ages + biqs + affects2 + affectp2 + (affects2 + affectp2 | peerGroup/id)	3160.4	3275.4	-1562.2	3124.4
Model 5	aoi2 ~ ptss + ages + biqs + affects2 + affectp2 + (affects2 + affectp2 || peerGroup/id)	3164.5	3292.2	-1562.2	3124.5

Note: Model 3 was the best fitting model however Model 8 is the best fitting model including time. Model 8 is presented in full in the manuscript. All other model details are available in the data analysis code (Pérez-Edgar et al., 2019).

Table S3: Model comparisons for model 3

	Model Formula	AIC	BIC	logLik	deviance
Model 3	affects2 ~ ptss + ages + biqs * aoi2 + biqs * affectp2 + aoi2 * affectp2 + (1 | id)	1532.9	1596.8	-756.46	1512.9
Model 7	affects2 ~ ptss + ages + onsets + biqs * aoi2 + biqs * affectp2 + aoi2 * affectp2 + (onsets || id)	1483.8	1560.5	-729.92	1459.8
Model 6	affects2 ~ ptss + ages + onsets + biqs * aoi2 + biqs * affectp2 + aoi2 * affectp2 + (onsets | id)	1480.3	1563.4	-727.15	1454.3
Model 2	affects2 ~ ptss + ages + biqs * aoi2 * affectp2 + (1 | id)	1534.8	1605.1	-756.42	1512.8
Model 1	affects2 ~ ptss + ages + biqs * aoi2 * affectp2 + (1 | peerGroup/id)	1536.8	1613.5	-756.42	1512.8
Model 4	affects2 ~ ptss + ages + biqs * aoi2 + biqs * affectp2 + aoi2 * affectp2 + (aoi2 + affectp2 | id)	1537.6	1633.5	-753.81	1507.6
Model 5	affects2 ~ ptss + ages + biqs * aoi2 + biqs * affectp2 + aoi2 * affectp2 + (aoi2 + affectp2 || id)	1540.7	1642.9	-754.34	1508.7

Note: Model 3 was the best fitting model however Model 7 is the best fitting model including time. Model 7 is presented in full in the manuscript. All other model details are available in the data analysis code (Pérez-Edgar et al., 2019).

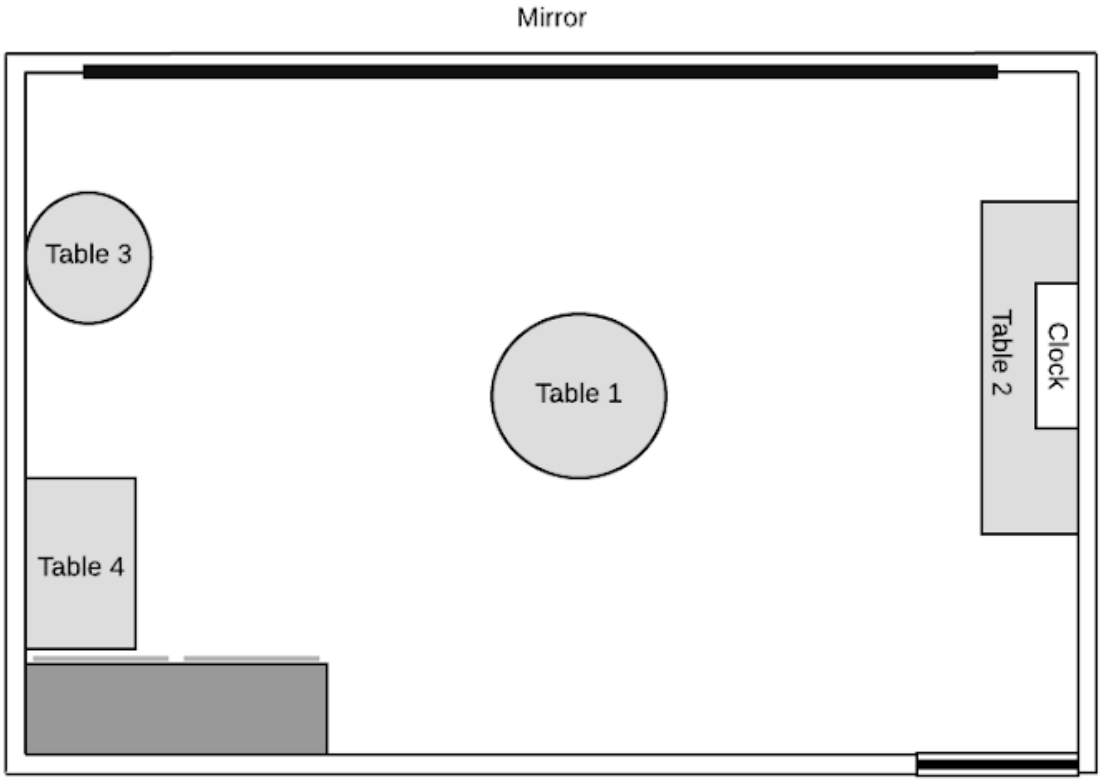

Figure S1. Free play room setup. Table 1: Activity books and drawing materials; Table 2: Jenga; Table 3: Candyland, 2 boxes of 48-piece puzzles; Table 4: 2 dolls, 3 toy dinosaurs.

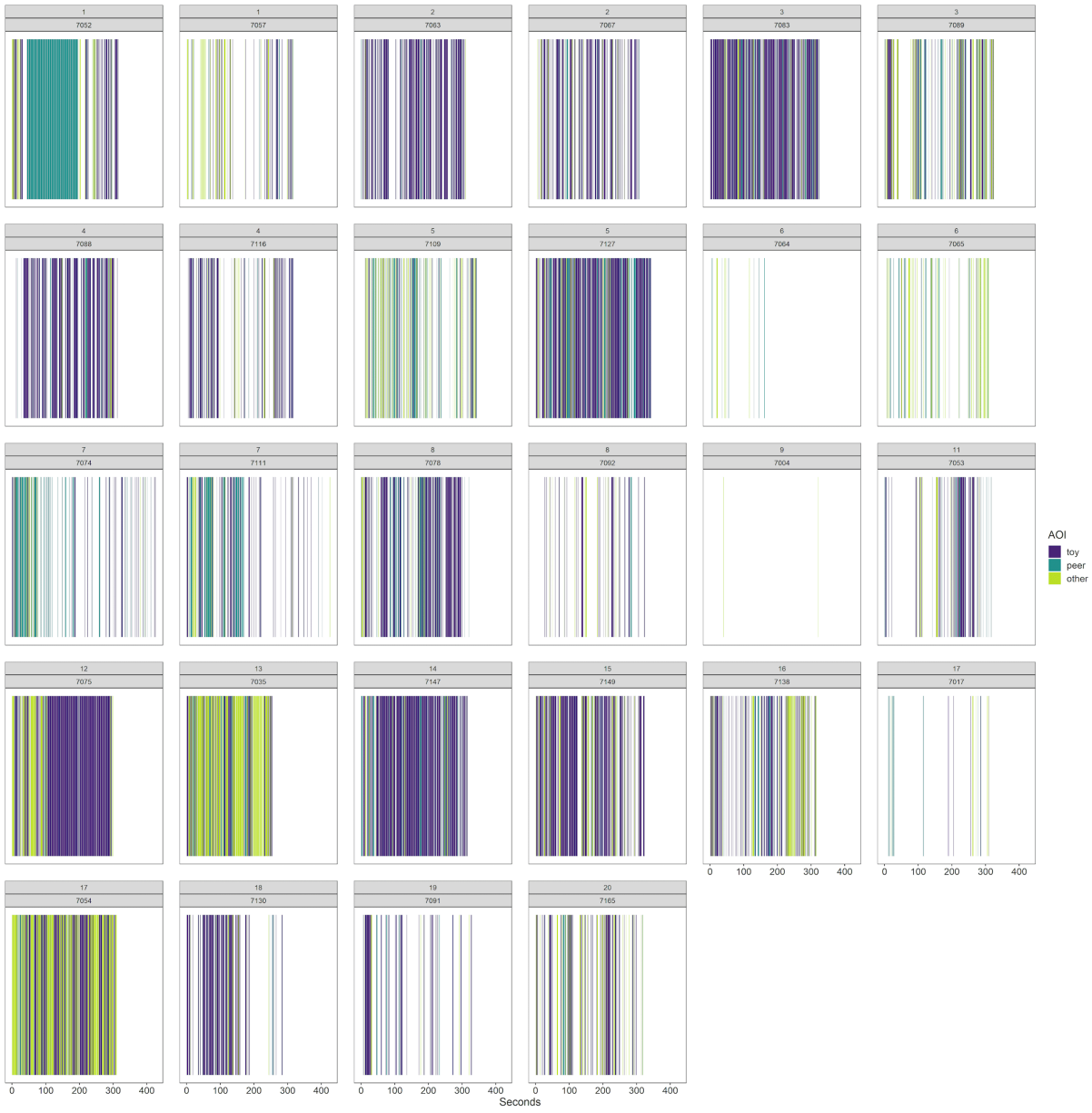

Figure S2. Dwell to AOIs for participants over the course of the free play episode. Upper number represents dyad and lower number represents individual participant.

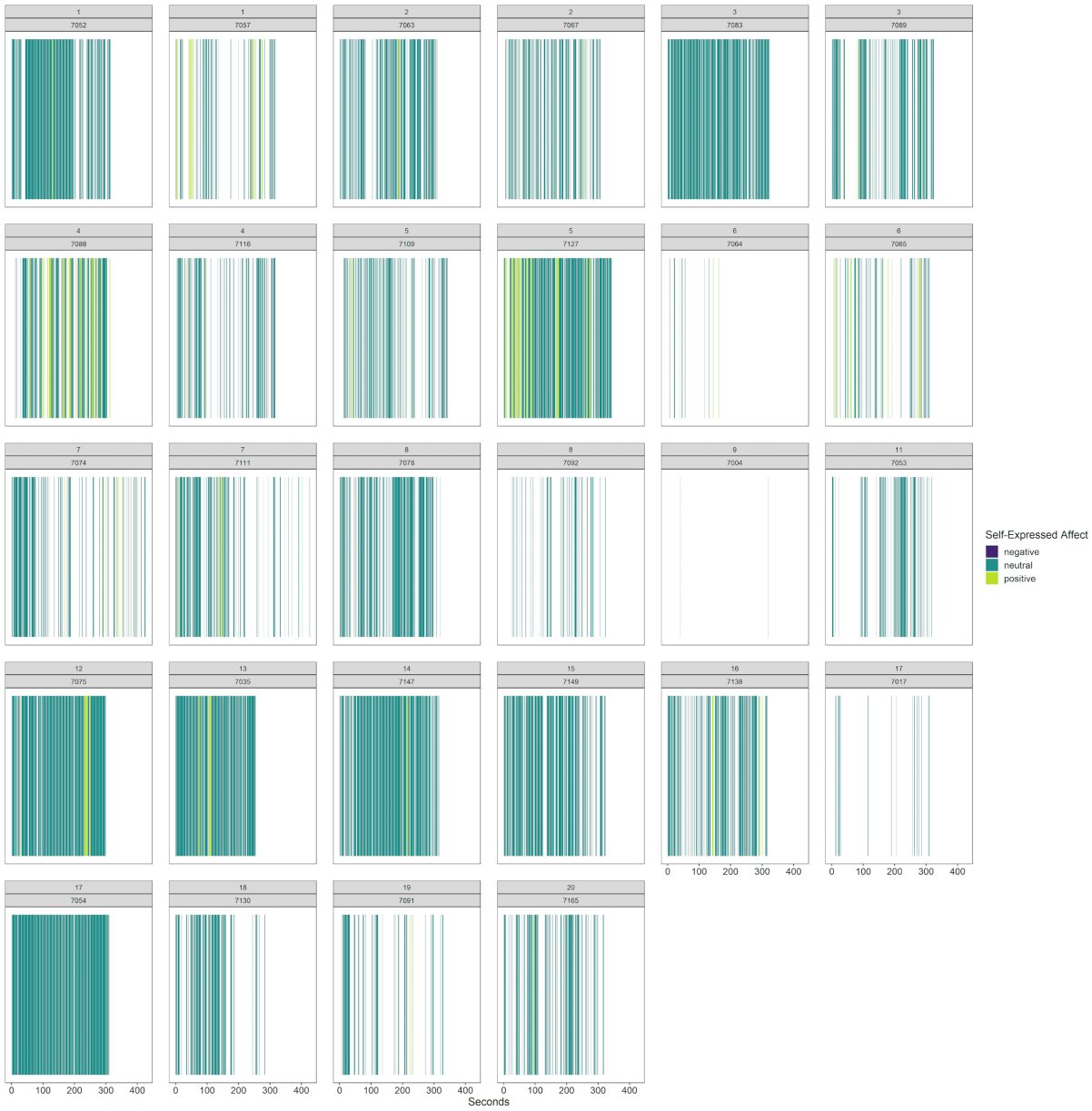

Figure S3. Self-expressed affect for participants over the course of the free play episode.

Upper number represents dyad and lower number represents individual participant.

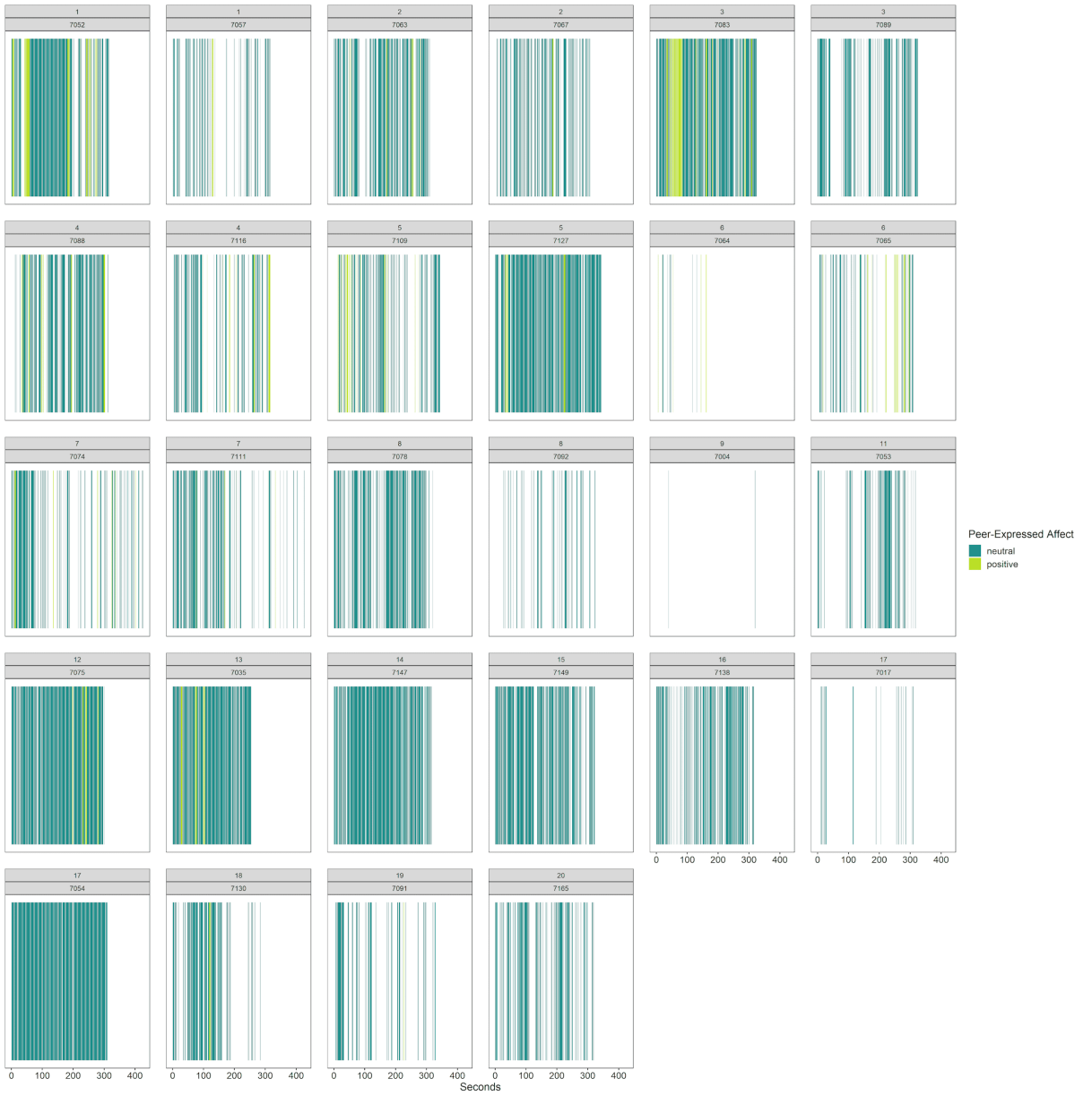

Figure S4. Peer-expressed affect for participants over the course of the free play episode.

Upper number represents dyad and lower number represents individual participant.

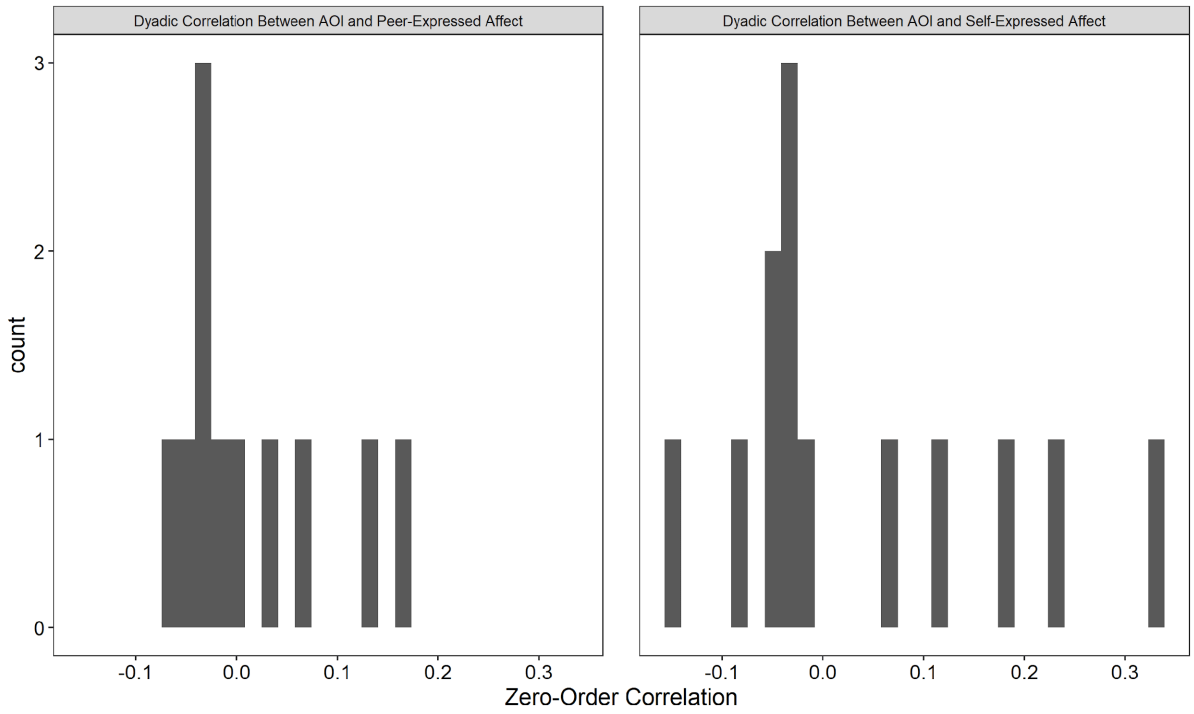

Figure S5. Variability in within dyad correlations between gaze and affect.

References

Bishop, G., Spence, S. H., & Mcdonald, C. (2003). Can parents and teachers provide a reliable and valid report of behavioral inhibition? *Child Development*, *74*(6), 1899–1917. https://doi.org/10.1046/j.1467-8624.2003.00645.x

Broeren, S., & Muris, P. (2010). A psychometric evaluation of the behavioral inhibition questionnaire in a non-clinical sample of dutch children and adolescents. *Child Psychiatry & Human Development*, *41*(2), 214–229. https://doi.org/10.1007/s10578-009-0162-9

Pérez-Edgar, K., Fu, X., MacNeill, L., & Gunther, K. (2019). *iTRAC Study: Mobile eye-tracking as a tool for studying socioemotional development*. https://doi.org/10.17605/OSF.IO/QAWC4