**Supplemental Material**
**for:**

# Examining Factors that Facilitate Young Children's Social Connections: An Experimental Approach

**This file includes:**

Supporting Text
- Procedural Flow
- Participant Demographics
- Condition Assignment Details
- Exclusions
- Drawing Task
- Analysis Strategy

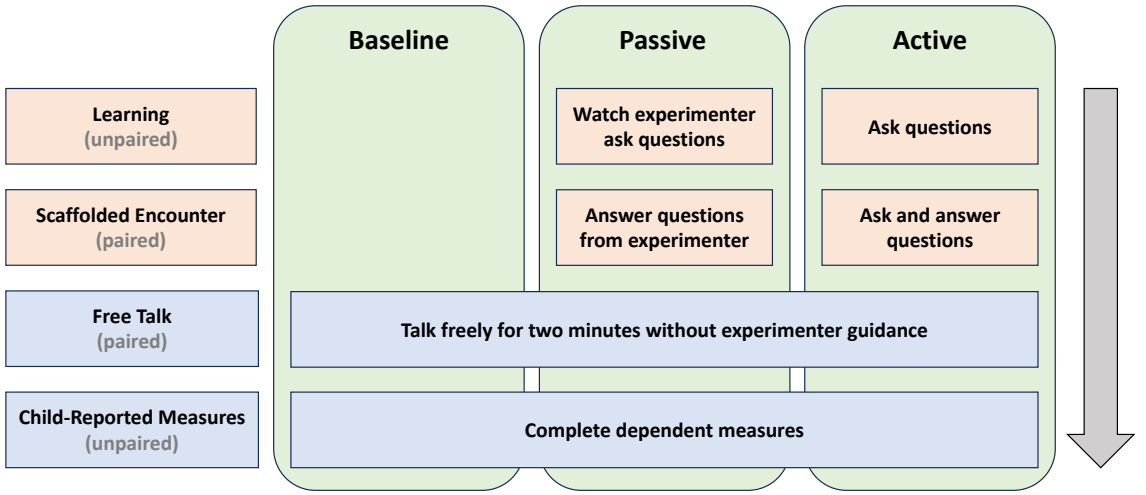
Supplemental Analyses
- Pair-Level Age Effects
- Pair-Level Reported Race/Ethnicity Effects
- Pair-Level Perceived Race/Ethnicity Effects
- Pair-Level Personality Effects
- Child-Level Personality Effects
- Pair Conversation Content
- Measure Correlations
- Experimenter Familiarity Effects
- Boy-Boy vs. Girl-Girl Pairs
- Partner Response Comparisons

Figure S1
Tables S1 to S6
Supplemental Material References

**Supporting Text**

**Procedural Flow.** The following figure (Figure S1) depicts the procedural flow of the *learning*, *scaffolded encounter*, *free talk*, and *child-reported measures* phases of the experiment by condition (*baseline*, *passive*, and *active*). Orange phases differ between conditions, and blue phases are identical across conditions. In paired phases, both children were located in the same virtual meeting room with both experimenters; in unpaired phases, each child was located in a separate virtual meeting room with one experimenter.

**Figure S1.**




**Participant Demographics.** Participant demographics (see Table S1) were collected via a demographic questionnaire (see OSF: https://osf.io/2zgy6/?view_only=28b09c680a5d465187177bdaf51b712e) that parents completed before their child participated in the study. The questionnaire included questions about the child's gender, race/ethnicity, experience with online social interactions, and personality.

**Table S1**

*Participant demographics by condition*

| | | Active ($N$ = 96) | Passive ($N$ = 96) | Baseline ($N$ = 96) |
|---|---|---|---|---|
| Participant Gender | Female | 46 | 47 | 49 |
| | Male | 50 | 48 | 47 |
| | Nonbinary | 0 | 1 | 0 |
| Participant Race/Ethnicity | American Indian or Alaskan Native | 1 | 0 | 1 |
| | Asian or Asian American | 7 | 9 | 8 |
| | Black or African American | 4 | 2 | 2 |
| | Hispanic or Latino | 5 | 6 | 6 |
| | White or Caucasian | 65 | 68 | 67 |
| | Multiracial | 14 | 9 | 10 |
| | Other | 0 | 0 | 2 |
| | Not Provided | 0 | 2 | 0 |

*Note*. Values are number of participants.

**Condition Assignment Details.** Pairs were assigned to one of three conditions: *baseline, active,* or *passive*. Condition assignment was completed via a randomized blocking design within each age and pair gender composition. For example, if one 5-year-old same-gender pair was randomly assigned to the *baseline* condition, then the next pair of the same type (i.e., 5-year-old same-gender pair) was assigned to the *active* condition, then the next pair of the same type was assigned to the *passive* condition, then the next pair of the same type was assigned to the *baseline* condition, and so on.

**Table S2**

*Count of pairs in each gender composition and reported race/ethnicity composition by condition*

| | | Condition | | |
|---|---|---|---|---|
| | | Active ($N$ = 48) | Passive ($N$ = 48) | Baseline ($N$ = 48) |
| Same-gender ($N$ = 72) | | | | |
| | Same-race/ethnicity | 13 | 13 | 13 |
| | Different-race/ethnicity | 11 | 11 | 11 |
| Different-gender ($N$ = 72) | | | | |
| | Same-race/ethnicity | 12 | 12 | 11 |
| | Different-race/ethnicity | 12 | 10 | 13 |
| | No response | 0 | 2 | 0 |

*Note.* Pairs sorted as "No response" contained at least one participant for whom parents chose not to provide racial or ethnic identity information.

After approximately two-thirds of the planned sample had been collected, we identified an imbalance emerging at random such that more same-gender pairs tended to be children of different races/ethnicities than of the same race/ethnicity, and different-gender pairs tended to be children of the same race/ethnicity than of different races/ethnicities. Such an imbalance could hinder our ability to test the effects of pair gender composition without race/ethnicity confounding these analyses. Therefore, we decided to switch our pairing system: Until this point, we allowed pairs to form naturally by allowing two families to sign up for the appointment slots of their choosing according to their availabilities. However, from this point forward, we first allowed one participant to sign up for each slot, then manually scheduled a second participant according to their gender and racial/ethnic identities. We continued to use the randomized blocking design to assign pairs to conditions until impossible to do so – that is, until cells were filled. Thus, as intended, within each condition and pair gender composition combination (e.g., same-gender pairs in the *active* condition), approximately half of the pairs were participants of the same race/ethnicity, and the other half of pairs were participants of different races/ethnicities

(see Table S2). We did not recruit participants of any particular race/ethnicity; however, given the larger proportion of White participants than participants from any other single racial or ethnic background (see Table S1), most pairs included one or more White participants. That is, White-White pairs were most common when race/ethnicity was the same, and White-non-White pairs were most common when race/ethnicity was different.

**Exclusions.** A total of 31 pairs were excluded from analyses due to: parent or sibling interference with the study procedures ($n$ = 27); experimental error ($n$ = 1); at least one member of the pair dropping out of the study during the *free talk* phase ($n$ = 1); or at least one member of the pair having already participated in the study previously ($n$ = 2).

Exclusions due to interference were determined based on transcript reviews conducted by two independent coders. Coder 1 was a research assistant unaware of condition assignment and hypotheses, and Coder 2 was one of the two first authors of this paper. Coder agreement was 91.00%, and disagreements were resolved by a third coder (whichever of the two first authors who was not Coder 2). Examples of exclusionary parent or sibling interference include: encouraging a child to speak during the *free talk* phase (e.g., “you sit here and talk to the kid, that's your job”); telling a child how to respond in the *scaffolded encounter* or *free talk* phases (e.g., “ask him if he has any brothers or sisters”); coaching a child on how to act or respond at any point prior to study completion (e.g., “remember, when you meet the other kid, you’re going to talk about all the questions you just learned”); or siblings disrupting the pairs’ interaction during *free talk* (e.g., a younger sibling inserting themself in the interaction and the participant repeatedly telling the sibling to leave). These interferences occurred despite parents receiving explicit instructions that family members should not interfere with the study session during the

*setup* phase. As evident in Table S3, most interference occurred in the *free talk* phase, and many more different-gender pairs than same-gender pairs were excluded due to interference.

**Table S3**

*Count of excluded pairs by phase(s) in which interference occurred, condition, and pair gender composition*

| Phase(s) at which interference occurred | | Condition | | |
|---|---|---|---|---|
| | | Active ($N$ = 8) | Passive ($N$ = 10) | Baseline ($N$ = 9) |
| Setup or learning ($N$ = 3) | Same-gender | 1 | 0 | 0 |
| | Different-gender | 1 | 0 | 1 |
| Scaffolded encounter ($N$ = 3) | Same-gender | 0 | 0 | NA |
| | Different-gender | 1 | 2 | |
| Free talk ($N$ = 18) | Same-gender | 1 | 3 | 1 |
| | Different-gender | 1 | 5 | 7 |
| Both scaffolded encounter and free talk ($N$ = 3) | Same-gender | 0 | 0 | 0 |
| | Different-gender | 3 | 0 | 0 |

**Drawing Task.** At the very end of the study session, participants completed an exploratory drawing measure. Each participant spent up to ten minutes drawing a picture of themself and their partner on a piece of paper using drawing supplies provided by the participant's parent. We planned for two coders unaware of condition assignment to rate the drawings for signs of connectedness (e.g., physical touch; positive facial affect) by viewing video recordings of the participants holding up their drawing to the camera. However, unexpectedly, only 64.6% of videos were sufficiently clear to identify the relevant details of the drawings (e.g., facial expressions) due to variability in participants' home environments (e.g., camera quality; lighting). Therefore, this measure was not included in analyses. Importantly, participants completed the drawing task after all other child-rated dependent measures, thus it could not have impacted their responses to other measures.

**Analysis Strategy.** All results reported in the main text were obtained using the following analysis strategy. First, to assess the effect of pair gender composition for pairs in the *baseline* condition specifically, we regressed each dependent variable on condition centered around *baseline* (*active* = 1, *passive* = 1, *baseline* = 0), pair gender composition (different gender = -0.5, same gender = 0.5), and the two-way interaction between those two variables.

Next, we regressed each dependent variable on the contrast between the *experimental* conditions (collapsed) and the *baseline* condition (Contrast 1; *baseline* = -2, *active* = 1, *passive* = 1), the contrast between the two *experimental* conditions (Contrast 2; *baseline* = 0, *active* = 1, *passive* = -1), pair gender composition, and the two-way interactions between pair gender composition and each of the two condition contrasts. Contrast 1 allowed us to evaluate whether the two *experimental* conditions (collapsed) differed from the *baseline* condition overall, and Contrast 2 allowed us to evaluate whether the two *experimental* conditions differed from one another. We also tested the interaction between Contrast 1 and pair gender composition to examine whether the *experimental* condition vs. *baseline* condition effect differed based on pair gender composition, and we examined the interaction between Contrast 2 and pair gender composition to examine whether the *active* and *passive experimental* condition effects differed based on pair gender composition.

To gain further insights into interaction effects for dependent variables with significant condition by pair gender composition interactions, we evaluated the simple effects of pair gender composition for each condition by centering condition around *active* (*active* = 0, *passive* = 1, *baseline* = 1), *passive* (*active* = 1, *passive* = 0, *baseline* = 1), and *baseline* (*active* = 1, *passive* = 1, *baseline* = 0), respectively, and estimating regression models using each condition variable in

turn. This allowed us to examine whether the effect of pair gender composition (i.e., gender in-group favoritism) was present in each condition individually.

Analysis scripts are provided on OSF: https://osf.io/2zgy6/?view_only=28b09c680a5d465187177bdaf51b712e

**Supplemental Analyses**

Below we report exploratory analyses incorporating the following supplemental demographic variables into the above-described analysis strategy: pair age, pair reported race/ethnicity composition, pair perceived race/ethnicity composition, and pair personality composition.

As described in the main text, video recordings were not available for eight pairs due to parents opting out of video recording. Thus, *free talk* results described in this section are based on the 136 pairs for whom video recordings were available.

**Pair-Level Age Effects.** First, to assess the effect of pair age at *baseline*, we regressed the dependent variable on an independent variable for condition centering around *baseline* (*active* = 1, *passive* = 1, *baseline* = 0), pair age (as calculated by averaging the two participants' ages in days within each pair), and the two-way interaction between those two variables. There was a significant effect of pair age on coder-rated connection and number of words spoken during *free talk*, such that older pairs in the *baseline* condition appeared more connected and talked more during the *free talk* phase than younger ones (coder-rated connection: $b = .17$, $F(1, 132) = 3.99$, $p = .047$; word count: $b = 13.55$, $F(1, 132) = 6.13$, $p = .015$). No other simple effects of pair age in the *baseline* condition were significant (all $p$s > .10).

Next, to evaluate whether the condition manipulations impacted younger versus older participant pairs differently, we tested for interactions between pair age and condition for each

dependent variable. We regressed the dependent variable on C1, C2, pair age, and the two-way interactions between pair age and each of the condition contrasts (C1 and C2). There were no significant interactions between pair age and either of the condition contrasts (all *p*s > .09), suggesting that there were no differences in condition effects based on participant age.

**Table S4**

*Count of pairs in each supplemental analysis variable by condition*

| | | Condition | | |
|---|---|---|---|---|
| | | Active ($N$ = 48) | Passive ($N$ = 48) | Baseline ($N$ = 48) |
| Pair Age | | | | |
| | 5-year-olds | 8 | 8 | 8 |
| | 6-year-olds | 8 | 8 | 8 |
| | 7-year-olds | 8 | 8 | 8 |
| | 8-year-olds | 8 | 8 | 8 |
| | 9-year-olds | 8 | 8 | 8 |
| | 10-year-olds | 8 | 8 | 8 |
| Reported race/ethnicity composition | | | | |
| | Same-race/ethnicity | 25 | 25 | 24 |
| | Different-race/ethnicity | 23 | 21 | 24 |
| | No response | 0 | 2 | 0 |
| Perceived race composition | | | | |
| | Perceived same-race | 32 | 31 | 28 |
| | Perceived different-race | 14 | 14 | 17 |
| | No video | 2 | 3 | 3 |
| Personality composition | | | | |
| | Mixed | 25 | 19 | 26 |
| | Extraverted | 11 | 13 | 13 |
| | Introverted | 11 | 15 | 8 |
| | No response | 1 | 1 | 1 |

*Note.* Pair age is reported categorically here, but it was treated as a continuous variable in analyses. Pairs sorted as "No response" contained at least one participant for whom parents chose not to provide racial or ethnic identity information. Pairs sorted as "No video" contained at least one participant whose parent or guardian declined to allow video recording of the study session.

**Pair-Level Reported Race/Ethnicity Effects.** First, to evaluate the effects of children interacting across versus within racial/ethnic group boundaries at *baseline*, we regressed the dependent variable on an independent variable for condition centering around *baseline*, pair reported race/ethnicity composition (different-race/ethnicity = -0.5, same-race/ethnicity = 0.5), and the two-way interaction between those two variables. Reported race/ethnicity composition of each pair was determined based on parent-reported demographics: if both children in a pair were identified as the same exact race/ethnicity (e.g., "Asian" and "Asian"), the pair was coded as "same-race/ethnicity"; if children in a pair were identified as different races/ethnicities (e.g., "White" and "Black-White multiracial"), the pair was coded as "different-race/ethnicity." Two parents did not report their children's racial/ethnic identities in the demographic questionnaire, so analyses of reported race/ethnic composition effects are based on 136 pairs (for the free talk measures) or 142 pairs (for other dependent variables) based on the availability of video recordings and racial/ethnic demographics. Of the 142 pairs for whom racial/ethnic information was provided, $N = 74$ pairs were "same-race/ethnicity" and $N = 68$ pairs were "different-race/ethnicity" (see Table S3).

None of the simple effects of pair reported race/ethnicity composition in the *baseline* condition were significant (all $p$s > .27). Next, to assess whether effects of condition differed between children interacting across versus within racial/ethnic group boundaries, we tested for the interactions between pair race/ethnicity composition and condition for each dependent variable. We regressed the dependent variable on C1, C2, reported race/ethnicity composition, and the two-way interactions between reported race/ethnicity composition and each of the condition contrasts.There were no significant interactions between reported race/ethnicity

composition status and either of the condition contrasts (all *p*s > .09), suggesting that condition effects did not vary between same-race/ethnicity and different-race/ethnicity pairs.

**Pair-Level Perceived Race/Ethnicity Effects.** Many people's racial and ethnic identities do not align with how others perceive their race/ethnicity based on appearance, and perceptions may be particularly influential during initial interactions in which people's true racial and ethnic identities are unknown. Indeed, research with adults supports this possibility; for example, Latinx people who are perceived as White tend to experience discrimination differently than Latinx people who are not perceived as White (Rodriguez, 2020). Therefore, we reasoned that children who hold different racial identities (based on parent reports) could nevertheless perceive one another as racial or ethnic ingroup members; at the same time, it is also possible that children who hold the same racial/ethnic identities sometimes perceive one another as belonging to different groups. Thus, in addition to evaluating effects of condition based on *parent-reported* race/ethnicity, we also examined effects of condition based on *perceived* race/ethnicity.

We created a "perceived race/ethnicity composition" variable based on ratings from adult coders. All pairs were rated by one coder who identified as Asian, and another coder who identified as White (the two monoracial identities most frequently reported by parents in the present sample). Coders watched video recordings of each pair and rated whether children appeared to belong to the same or different racial/ethnic groups. Two coders rated each pair, and disagreements were resolved via a third coder. Of the 136 pairs for whom videos were available, $N = 91$ pairs were rated as "perceived same-race/ethnicity" and $N = 45$ pairs were rated as "perceived different-race/ethnicity" (see Table S3).

First, to evaluate the effects of perceived race/ethnicity composition at *baseline*, we regressed the dependent variable on an independent variable for condition centering around

*baseline*, pair perceived race/ethnicity composition (different-race/ethnicity = -0.5, same-race/ethnicity = 0.5), and the two-way interaction between those two variables. There was a significant effect of pair perceived race/ethnicity composition on coder-rated connection and number of words spoken during free talk, such that perceived same-race/ethnicity pairs in the *baseline* condition appeared more connected and talked more during the *free talk* phase than perceived different-race/ethnicity pairs (coder-rated connection: $b = .77$, $F(1, 132) = 6.46$, $p = .012$; word count: $b = 43.70$, $F(1, 132) = 4.32$, $p = .040$). No other simple effects of pair perceived race/ethnicity composition in the *baseline* condition were significant (all $p$s > .10).

Next, to test for differences in condition effects based on the perceived race/ethnicity composition of children in a pair, we tested for the effects of perceived race/ethnicity composition and its interaction with condition on each dependent variable. We regressed the dependent variable on C1, C2, perceived race/ethnicity composition, and the two-way interactions between perceived race/ethnicity composition and each of the condition contrasts.

There was a significant interaction between perceived race/ethnicity composition and C1 on coder-rated connection, such that the difference between perceived different-race/ethnicity pairs and perceived same-race/ethnicity pairs was larger in *baseline* condition than in the two scaffolded conditions ($b = -0.26$, $F(1, 127) = 4.32$, $p = .040$). In the *baseline* condition, perceived same-race/ethnicity pairs appeared more connected during the *free talk* phase than perceived different-race/ethnicity pairs (statistics provided above); however, this perceived race/ethnicity-composition effect was not seen in the *active* ($b = .20$, $F(1, 132) = .39$, $p = .53$) or *passive* ($b = -0.23$, $F(1, 132) = .52$, $p = .472$) conditions.

**Pair-Level Personality Effects.** We also explored the role of pair personality composition in children's connection with one another between conditions. We measured

participants' personalities via parent responses to the Openness to Experience, Shyness, and Sociability subscales from the Inventory of Child Individual Differences-Short Version (ICID-S; Deal et al., 2007). Parents responded to 12 items on a 1 (*Much less than the average child or not at all*) to 7 (*Much more than the average child*) scale. Items included, for example, "My child makes friends easily" (full set of items are provided below). Responses to all items were summed (with the Shy subscale reverse coded) to calculate a total personality score ranging from 12 to 84, with higher scores indicating higher extraversion. In order to create a "personality composition" score for each pair, we median-split participants (*Median* = 62) such that children who scored lower than the median were coded as "introverted" and children who scored at the median or higher were coded as "extraverted." Pairs containing one introvert and one extravert were scored as *mixed personality* pairs ($N$ = 70), pairs with two introverts were scored as *introverted* pairs ($N$ = 34), and pairs with two extraverts were scored as *extraverted* pairs ($N$ = 37; see Table S3). Analyses on the effect of pair personality composition were based on the 133 pairs (for the free talk measures) or 141 pairs (for the rest of the dependent variables) based on availability of video recordings and parent-reported personality for each child in a pair.

First, we tested the effects of pair personality composition at *baseline*. We first created two dummy codes (dc) for the three personality compositions: dc1 (*mixed personality* = 0, *extraverted* = 1, *introverted* = 0) and dc2 (*mixed personality* = 1, *extraverted* = 0, *introverted* = 0). Dc1 evaluated whether *introverted* pairs and *extraverted* pairs differed from one another, and dc2 evaluated whether *introverted* and *mixed personality* pairs differed from one another. We then regressed the dependent variable on an independent variable for condition centering around *baseline*, dc1, dc2, and the two-way interactions between the condition variable and each dummy code of personality composition.

In the *baseline* condition, there was a significant effect of dc1 on coder-rated connection and word count, such that *extraverted* pairs appeared more connected and talked more during the *free talk* phase than *introverted* pairs (coder-rated connection: $b = 1.14$, $F(1, 127) = 6.40$, $p = .013$; word count: $b = 62.13$, $F(1, 127) = 4.01$, $p = .047$). Additionally, there was a significant effect of dc1 ($b = 1.68$, $F(1, 135) = 6.40$, $p = .013$) and dc2 ($b = 1.51$, $F(1, 135) = 6.36$, $p = .013$) on Behavioral Intentions, respectively. Lastly, there was a significant effect of dc 1 ($b = 0.97$, $F(1, 135) = 7.56$, $p = .007$) and dc2 ($b = 0.91$, $F(1, 135) = 8.27$, $p = .005$) on the Sharing Task, respectively. These results suggest that in the *baseline* condition, *extraverted* and *mixed personality* pairs were more connected than *introverted* pairs, evident from both *free talk* performance and children's self-reports. No other simple effects of pair personality composition in the *baseline* condition were significant (all $p$s > .08).

Next, to evaluate whether condition effects varied based on pair personality composition, we tested for its interaction with condition (C1 and C2) on each dependent variable. We regressed the dependent variable on C1, C2, dc1, dc2, and the two-way interactions between each condition contrast and dummy code of personality composition.

There were significant interactions between C1 and dc1on Behavioral Intentions and the Sharing Task ($b = -0.63$, $F(1, 132) = 5.61$, $p = .019$ and $b = -0.32$, $F(1, 132) = 5.39$, $p = .022$, respectively), such that the difference between *introverted* and *extraverted* pairs was larger in the *baseline* condition than in the two scaffolded conditions. There were also significant interactions between C1 and dc2 on Behavioral Intentions and the Sharing Task ($b = -0.53$, $F(1, 132) = 4.98$, $p = .027$ and $b = -0.36$, $F(1, 132) = 8.46$, $p = .004$, respectively), such that the difference between *mixed personality* and *introverted* pairs was larger in the *baseline* condition than in the two scaffolded conditions. In the *baseline* condition, children in *extraverted* pairs and *mixed*

*personality* pairs expressed greater intentions to approach their partners outside of the study than those in *introverted* pairs (statistics provided above); however, in the two *experimental* conditions, no such difference was observed (all *p*s > .72). A similar pattern emerged for the Sharing Task: In the *baseline* condition, children in *extraverted* pairs and *mixed personality* pairs shared more funny pictures with their partner compared to those in *introverted* pairs (statistics provided above); however, in the *experimental* conditions, no such difference was observed (all *p*s > .07). Thus, these results suggest that scaffolding closed gaps between *introverted* pairs and *mixed* or *extraverted* pairs.

**Child-Level Personality Effects.** To check whether *individual* children's personalities (as opposed to *pairs* of children's personality combinations) impacted their performance in the *free talk* phase and responses to the child-reported measures, we tested the effect of total personality score (described in the section above) on each dependent variable. Specifically, we estimated a linear mixed effects model where we regressed the dependent variable on C1, C2, total personality score, the two-way interactions between child-in-pair and each of the two condition contrasts, and a by-pair random intercept. Note that, because three out of the five items that make up the coder-rated connection were only rated at the pair level, we only analyzed the comfort and fun items which were rated at the individual child level.

We found a significant main effect of personality on the free talk ratings and number of words spoken during free talk, such that the more extraverted the child, the more they appeared to be comfortable ($b$ = .02, $F(1, 184) = 13.09$, $p < .001$) and having fun ($b$ = .01, $F(1, 161) = 6.02$, $p = .015$), and the more they talked ($b$ = .77, $F(1, 254) = 9.89$, $p = .002$) during *free talk*. Additionally, there was a significant interaction between C2 and personality on perceived comfort during *free talk* ($b$ = -0.02, $F(1, 187) = 8.78$, $p = .003$). Personality had a significant

effect on *free talk* comfort in the *baseline* ($b$ = .02, $F(1, 181) = 6.45$, $p$ = .012) and *passive* conditions ($b$ = .03, $F(1, 185) = 14.47$, $p < .001$) but not in the *active* condition ($b$ = -0.002, $F(1, 190) = .06$, $p$ = .810). Children's individual personalities did not impact how they responded to child-reported measures (all $p$s > .16). Overall, these results suggest that children's individual personalities influence how they interact with peers that they meet for the first time, but adult scaffolding that encourages children's active self-disclosure during initial encounters with peers may help attenuate these individual differences.

**Pair Conversation Content.** In order to explore the content of children's conversations during the *free talk* phase, videos of children's conversations were transcribed by coders unaware of condition assignment. Transcripts were analyzed using Linguistic Inquiry and Word Count (LIWC-22) software (Boyd et al., 2022). Similar to the analyses above, we regressed the LIWC-22 scores on the contrast between the *experimental* conditions (collapsed) and the *baseline* condition (C1), the contrast between the two *experimental* conditions (C2), pair gender composition, and the two-way interactions between pair gender composition and each of the two condition contrasts.

Constructs analyzed were: word count (i.e., number of words spoken); analytical thinking (i.e., logical, formal thinking); clout (i.e., leadership and status); authenticity (i.e., honesty); emotional tone (i.e., degree of positive vs. negative tone); personal pronouns (e.g., "I", "you"); affiliation (e.g., "we", "our"); positive tone (e.g., "good", "love"); social behavior (e.g., "said", "care"); politeness (e.g., "thank", "please"); communication (e.g., "tell", "said"); family (e.g., "parent", "brother"); friends (e.g., "friend", "dude"); leisure (e.g., "game", "play"); home (e.g., "house", "room"); food (e.g., "eat", "dinner"); need (e.g., "need", "have to"); want (e.g., "want", "wish"); curiosity (e.g., "look", "wonder"); space (e.g., "in", "up"); feeling (e.g., "hard", "cool");

time (e.g., “when”, “now”); future focus (e.g., “will”, “may”); and nonfluencies (e.g., “uh”, “um”). Other categories were not analyzed due to lack of relevance to hypotheses or age-inappropriateness (e.g., politics; death; swear words).

**Table S5**

*Correlations between adult-rated free talk measures*

| Measure | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1. Global | - | | | | |
| 2. Connection | .90*** | - | | | |
| 3. Comfort | .92*** | .83*** | - | | |
| 4. Fun | .91*** | .85*** | .86*** | - | |
| 5. Talkativeness | .90*** | .84*** | .84*** | .92*** | - |

*Note.* Values are Pearson’s *r*s. p>.05, *.05>p>.01, **p<.01, ***p<.001. p-values have been adjusted for multiple comparisons using the Holm–Bonferroni method.

**Table S6**

*Correlations between child-reported measures*

| Measure | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. Liking | - | | | | | |
| 2. Fun | .65*** | - | | | | |
| 3. Closeness | .63*** | .66*** | - | | | |
| 4. Future Interest | .16 | .10 | .13 | - | | |
| 5. Approach Intentions | .56*** | .44*** | .53*** | .13 | - | |
| 6. Sharing | .20 | .17 | .12 | .10 | .30** | - |

*Note.* Values are Pearson’s *r*s. *.05>p>.01, **p<.01, ***p<.001. p-values have been adjusted for multiple comparisons using the Holm–Bonferroni method.

Overall, pairs’ conversations were similar across conditions. There were no differences in conversation content between pairs in the two *experimental* conditions (*p*s > .19). However, there were differences in conversation content between *experimental* condition pairs and *baseline*

condition pairs. Pairs used more positive tone, polite words, and more feeling words in the *baseline* condition than the *experimental* conditions (positive tone: $b$ = -1.17, $F(1, 126) = 4.45$, $p$ = .04; politeness: $b$ = -1.09, $F(1, 126) = 4.16$, $p$ = .04, feeling words: $b$ = -0.10, $F(1, 126) = 5.04$, $p$ = .03). No other condition effects were significant ($p$s > .05). These results suggest that pairs in the *baseline* condition may have engaged in more formal conversation (e.g., "how are you?" or "nice to meet you") than pairs in *experimental* conditions.

**Measure Correlations.** Measure correlations were computed via Pearson method for adult-rated *free talk* phase measures (see Table S4) and child-reported measures (see Table S5). Based on the high correlations among adult-rated *free talk* phase measures, measures were combined into a single composite score ("coder-rated connection").

**Experimenter Familiarity Effects.** Due to the design of our study (see Method section in the manuscript), during the scaffolded encounter phase, one child ("Child A") always interacted with the same experimenter ("Experimenter A"), whereas the other child ("Child B") interacted with Experimenter B during the setup and learning phases and with Experimenter A during the scaffolded encounter phase. To determine whether these differences in familiarity with Experimenter A impacted the results, we examined the effect of child-in-pair (i.e., Child A vs Child B) on each dependent variable. Specifically, we estimated a linear mixed effects model where we regressed the dependent variable on C1, C2, child-in-pair (Child A = -0.5, Child B = 0.5), the two-way interactions between child-in-pair and each of the two condition contrasts, and a by-pair random intercept. Because three out of the five items comprising coder-rated connection were rated at the pair level, we only analyzed the comfort and fun items, which were rated at the individual child level.

We observed a main effect of child-in-pair only in the sharing task, such that Child A shared more "funny pictures" with their partner than Child B ($b$ = -0.26, $F$(1, 141) = 5.08, $p$ = .026). There was a significant interaction between child-in-pair and Contrast1 on the coder rating on how much fun the child seemed to be having during *free talk* ($b$ = .08, $F$(1, 133) = 4.47, $p$ = .036) and the sharing task ($b$ = .18, $F$(1, 140) = 4.77, $p$ = .031). In the *baseline* condition, Child A appeared to be having more fun during the *free talk* and shared more than Child B (free talk fun: $b$ = -0.20, $F$(1, 134) = 4.33, $p$ = .039; sharing task: $b$ = -0.63, $F$(1, 141) = 9.58, $p$ = .002); however, this difference was not seen in the *active* (free talk fun: $b$ = 0.09, $F$(1, 134) = .82, $p$ = .366; sharing task: $b$ = -0.15, $F$(1, 141) = .51, $p$ = .478) or the *passive* (free talk fun: $b$ = .01, $F$(1, 134) = .01, $p$ = .910; sharing task: $b$ = -0.02, $F$(1, 143) = .01, $p$ = .919) conditions. No other main effects of child-in-pair or interactions between child-in-pair and condition were significant (all $p$s > .05). Thus, Child A vs Child B differences were not robust across conditions and dependent variables.

**Boy-Boy vs. Girl-Girl Pairs.** We examined same-gender pairs ($n$ = 72, half girl-girl pairs and half boy-boy pairs) and tested whether two boys paired together (i.e., boy-boy pairs) and two girls paired together (i.e., girl-girl pairs) differed in their connectedness. Specifically, we regressed each dependent variable on C1, C2, pair type, and the two-way interactions between pair type and each of the two condition contrasts. We did not find any main effects of pair type or interactions between pair type and condition (all $p$s > .15), suggesting that pairs of boys and pairs of girls did not differ in their initial connections, nor in how they were impacted by the *experimental* condition manipulations.

**Partner Response Comparisons.** Our hypotheses and accompanying analyses focus on outcomes at the pair level to match the level at which key manipulations (i.e., condition and

gender composition) were implemented. Nevertheless, we also tested alignment between partners' responses (within a pair) in two ways. First, the ICCs between children within dyads for each dependent variable are as follows: free talk fun (ICC = .80), free talk comfort (ICC = .65), liking (ICC = .14), fun (ICC = .18), closeness (ICC = .16), future interest (ICC = .05), approach intentions (ICC = .09), sharing (ICC = .10). Second, the absolute difference scores between the ratings of children within a pair (means and standard deviations) are as follows: free talk fun ($M_{dif}$ = .42, $SD_{dif}$ = .50), free talk comfort ($M_{dif}$ = .64, $SD_{dif}$ = .61), liking ($M_{dif}$ = .79, $SD_{dif}$ = .81), fun ($M_{dif}$ = .94, $SD_{dif}$ = .91), closeness ($M_{dif}$ = 1.04, $SD_{dif}$ = 1.04), future interest ($M_{dif}$ = .81, $SD_{dif}$ = .77), approach intentions ($M_{dif}$ = 1.99, $SD_{dif}$ = 1.86), sharing ($M_{dif}$ = .97, $SD_{dif}$ = 1.06).

The ICC values suggest that 1) children were aligned in their signals of connection during the *free talk* phase, as perceived by independent raters unaware of condition assignment; however, 2) children were less aligned in their responses to the child-reported measures. Conversely, the small difference score values suggest that, on average, children within a pair provided qualitatively similar responses. For example, if Child A rated their liking of Child B as 5, whereas Child B rated their liking of Child A as 4, both children felt positively about one another (despite providing different ratings).

**Supplemental Material References**

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). The development and psychometric properties of LIWC-22. *Austin, TX: University of Texas at Austin*, 1–47.

Deal, J. E., Halverson Jr, C. F., Martin, R. P., Victor, J., & Baker, S. (2007). The Inventory of Children's Individual Differences: Development and validation of a short version. *Journal of Personality Assessment, 89*(2), 162–166. https://doi.org/10.1080/00223890701468550

Rodriguez, F. (2020). Passing vs Non-Passing: Latina/o/x Experiences and Understandings of Being Presumed White. *Electronic Theses, Projects, and Dissertations. 1121*. https://scholarworks.lib.csusb.edu/etd/1121