SUPPLEMENTARY ONLINE MATERIALS

Study 1

Target Character Selection

As reported in the main text, we confirmed our target character selection by collecting data on how feminine and masculine each target was. Figure S1 depicts these data. Confirming character selection, stereotypical girls were evaluated as significantly more feminine than counterstereotypical girls, whereas stereotypical boys were evaluated as significantly less feminine than counterstereotypical boys.

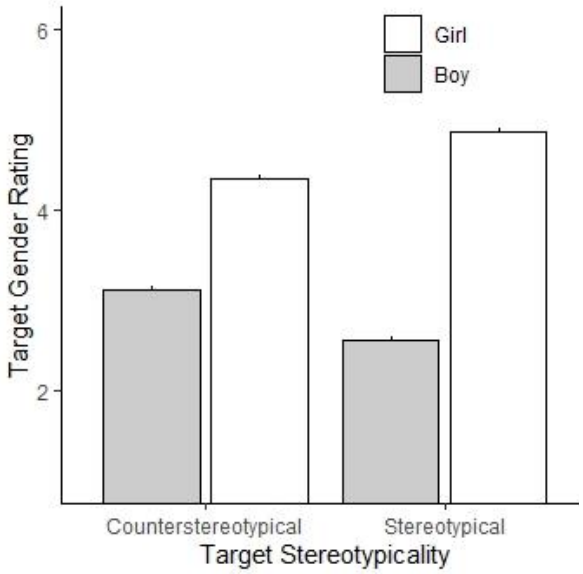

*Figure S1. R*atings of how feminine or masculine targets were based on targets' gender stereotypicality and gender. Higher scores on the y-axis indicate more femininity. Stereotypical (vs. counterstereotypical) girls (white bars) were rated as more feminine. Stereotypical (vs. counterstereotypical) boys (grey bars) were rated as more masculine.

Equivalence Test

As explained in the manuscript, we conducted an equivalence test to examine whether the effect of target stereotypicality on target positivity was meaningfully different from 0 (Lakens et al., 2018). To run the equivalence test, we needed to identify the smallest effect size of interest (i.e., an SESOI). Ideally, following recommended procedures (i.e., the *small telescopes* procedure), we would set this value to be the effect size that this study had 33% power to detect (Simonsohn, 2015). However, effect size calculations for mixed effects models, especially cross-classified models with continuous predictors, are still quite new and evolving (Correll et al., 2021) and therefore it is unclear what the typical heterogeneity of effect size in these models

looks like. Thus, we conducted an equivalence test using a *t*-test calculation. Although this is less ideal because we did not get to examine the effect accounting for variation among judges and clips, we felt we could make a more reliable estimate of SESOI. As reported above, expressers were significantly more positive towards stereotypical than counterstereotypical targets and, when data were aggregated across judges for each clip and analyzed using an independent samples *t*-test, it revealed an effect size of d =.23. Conversely, when target positivity data were aggregated across judges and analyzed using an independent samples *t*-test, gender-stereotypical targets did not exhibit significantly more positive behavior than gender-counterstereotypical targets; the effect size was d =.07. Based on calculations in G*Power (Faul et al., 2007), our study had 33% power to detect an effect of d =.15. Based on this calculation, the above effect of target stereotypicality on target positivity was smaller than the SESOI. Therefore, the analyses provide some support for the view that target emotion was not responsible for the observed effects: Expresser characters displayed more positive nonverbal behavior toward gender-stereotypical targets than counterstereotypical targets, but in our data, this relationship could *not* be explained by targets' own positive or negative behavior.

Indirect Effects

We had hypothesized that clip-induced differences in children's intersubjective norms would yield downstream effects on children's behaviors. However, bias condition did not directly influence gender-role behavior in Study 2 or 3, whether measured by what toys girls played with, how they customized their puppets, what toy they selected to take home with them, or how they behaved in a video recording for their peers (see statistics in main text). Nonetheless, it has been argued that the absence of such total effects should not preclude analysis of indirect pathways (Hayes, 2009; Kenny & Judd, 2014; MacKinnon et al., 2002; Rucker et al.,

2011; Shrout & Bolger, 2002; Zhao et al., 2010). In our case, we were interested in the pathway from nonverbal bias condition to behavior by way of intersubjective norms or personal beliefs. One reason for testing indirect pathways even in the absence of total effects is that tests of indirect effects have greater statistical power than tests of direct and total effects (Kenny & Judd, 2014; Rucker et al., 2011). This is especially true when the size of the "a" path (predictor to mediator) and "b" path (mediator to outcome) are similar (Kenny & Judd, 2014; Loeys et al., 2014) and larger than the "c" path (total effect). This could occur because the a and b pathways have lower standard error than the c path or because the mediator is more reliable or has less variance than the outcome variable. Therefore, we provide indirect tests of nonverbal bias on gender stereotypes and behaviors below. In each of these analyses, the conditions regarding variance were met—the a and b paths had roughly equivalent standard error that was larger than the c path, and the outcome measure had greater variance and less reliability than the mediator. See pathway variance in Tables S1 and S2 below.

Another reason that indirect effects may be obtained in the absence of a total effect is that there may be an unmeasured variable that is correlated with the mediator and has effects on an outcome that oppose those of the mediator (i.e., confounder bias; Bullock et al., 2010; MacKinnon & Pirlott, 2015). We examined the probability that unmeasured variables could account for the absence of a total effect (using the Left Out Variable Error Method; MacKinnon & Pirlott, 2015), finding that this probability was low (see Figure S1). Thus, we proceeded to focus only on the significance of the measured indirect effect in examining whether clip-induced changes in intersubjective norms yielded similar changes in gender-relevant attitudes, stereotypes, and behaviors. For each outcome, the nonverbal bias manipulation was effects-

coded (1 = Traditional), DANVA scores were centered and we examined the interactive influence of these variables on outcomes *through* intersubjective norms or personal beliefs.

Study 2

Results

Intersubjective norms.

Toy play. First, we examined the moderated indirect effect of condition on stereotypical toy play through intersubjective norms, which was significant: b = 112.47, se = 54.94, 95% CI [29.88, 254.34]. We broke this pattern down by examining simple (indirect) effects: the more that emotionally perceptive girls felt pressure to be feminine as a consequence of watching the Traditional (versus Reverse) clips, the more time they played with feminine than masculine toys, b = 18.10, se = 10.99, 95% CI [1.64, 46.45]. The same was not true for girls who were not yet skilled at reading emotion, b = -14.03, se = 10.79, 95% CI [-40.50, 4.18].

Puppet customization. We next examined the moderated indirect effect of nonverbal bias condition on gendered puppet customization, which was also significant: b = 2.15, se = 1.55, 95% CI [.22, 5.88]. However, simple effects were not significant. The more that emotionally perceptive girls felt pressure to be feminine as a consequence of watching the Traditional (versus Reverse) clips, they were no more likely to assign the girl and boy puppets stereotypical (vs. counterstereotypical) careers, b = .35, se = .28, 95% CI [-.02, 1.08]. The effect for girls who were not skilled at reading emotion was also not significant, b = -.27, se = .28, 95% CI [-.92, .14].

Toy selection. We next examined which toys the children selected at the end of the study to take home with them. The moderated indirect effect of nonverbal bias condition on toy selection was not significant, b = .45, se = .36, 95% CI [-.03, 1.42.]. The more that emotionally

perceptive girls felt pressure to be feminine from watching the Traditional (versus Reverse) clips, they were no more likely to select a feminine than masculine toy to take home with them, *b* = .07, *se*=.06, 95% CI [-.004, .26]. The effect for girls who were not yet skilled at reading emotion was also not significant, *b* = -.06, *se*=.06, 95% CI [-.22, .01].

Personal beliefs. Next, we examined the moderated indirect effect of condition on stereotypical toy play through each of the personal beliefs measured in Study 2, beginning with playmate preferences, followed by career aspirations and expectations. None of the indirect effects were significant. Nonetheless, we report the results of each indirect effect below.

Toy play. The moderated indirect effects of nonverbal bias condition on toy play through playmate preferences (*b* = 3.70, *se* = 16.64, 95% CI [-16.76, 59.98]), career aspirations (*b* = 3.24, *se* = 28.97, 95% CI [-43.53, 79.95]), and career expectations (*b* = 5.38, *se* = 22.21, 95% CI [-21.91, 80.18]) were not significant. Nonverbal bias condition did not impact girls' likelihood of playing with stereotypical or counterstereotypical toys by way of changing their personal beliefs about gender.

Puppet customization. The moderated indirect effects of nonverbal bias condition on puppet customization through playmate preferences (*b* = .12, *se* = .60, 95% CI [-.66, 1.99]), career aspirations (*b* = -.01, *se* = .37, 95% CI [-.98, .64]), and career expectations (*b* = -.26, *se* = 1.38, 95% CI [-3.21, 1.54]) were not significant. Nonverbal bias condition did not impact girls' likelihood of selecting stereotypical or counterstereotypical jobs for their puppets by way of changing their personal beliefs about gender.

Toy selection. The moderated indirect effects of nonverbal bias condition on toy selection through playmate preferences (*b* = -.003, *se* = .10, 95% CI [-.27, .14]), career aspirations (*b* = -.004, *se* = .09, 95% CI [-.25, .15]), and career expectations (*b* = -.02, *se* = .13, 95% CI [-.48, .14])

were not significant. Nonverbal bias condition did not impact girls' likelihood of selecting a stereotypical or counterstereotypical toy by way of changing their personal beliefs about gender.

Study 3

As in Study 2, nonverbal bias did not interact with emotional perceptivity to have a total effect on girls' gender-role behavior. Nonetheless, we report our pre-registered indirect effects of nonverbal bias condition on behavior through intersubjective norms and personal beliefs here.

Results

Intersubjective norms.

Competence. We observed a significant indirect effect of condition on nonverbal competence via felt pressure—as moderated by emotion perceptivity, b = -.24, se = .16, 95% CI [-.70, -.02] (see Figure S2). Emotionally-perceptive girls felt more pressure to be feminine in the Traditional (versus Reverse) condition and consequently conveyed less competence in their nonverbal behavior, b = -.04, se = .03, 95% CI [-.13, -.003]. Among girls who were not yet as emotionally-perceptive, there was no significant effect, b = .03, se = .03, 95% CI [-.01, .10].

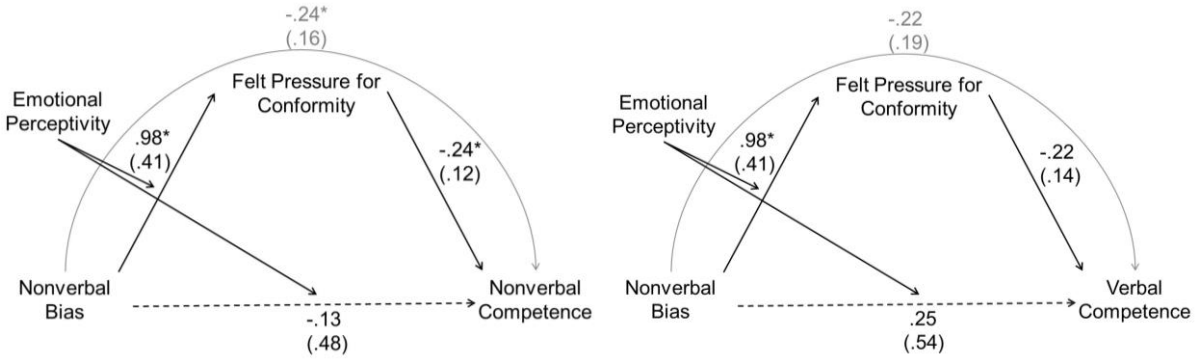

Figure S2. Moderated indirect effects of nonverbal bias on girls' competence in Study 3. Emotionally perceptive girls in the Traditional (versus Reverse) condition felt more pressure to be feminine and therefore conveyed significantly less nonverbal (left panel) but not significantly less verbal/paraverbal (right panel) competence.

The moderated indirect effect of condition on *verbal* competence was not significant, *b* = -.22, *se* = .19, 95% CI [-.78, .02] (see Figure S2). Thus, indirect effects of condition were stronger for nonverbal than verbal displays of competence.

Warmth. As in analyses on nonverbal competence, nonverbal bias did not interact with emotional perceptivity to have a direct impact on girls' nonverbal warmth (see statistics in main text). There were no moderated indirect effects of nonverbal bias condition on nonverbal warmth, *b* = .06, *se* = .21, 95% CI [-.36, .50], or verbal warmth, *b* = .01, *se* = .13, 95% CI [-.25, .27]. Thus, girls exhibited moderately-high verbal and nonverbal warmth regardless of condition or perceived pressure to be feminine.

Personal beliefs.

Competence. There was no moderated indirect effect of condition on nonverbal competence, *b* = .003, *se* = .05, 95% CI [-.07, .14], or verbal competence, *b* = -.02, *se* = .08, 95% CI [-.26, .08], via personal beliefs.

Warmth. There was no moderated indirect effect of condition on nonverbal warmth, *b* = .03, *se* = .12, 95% CI [-.14, .40], or verbal warmth, *b* = -.004, *se* = .04, 95% CI [-.13, .06], via personal beliefs.

Sensitivity Analyses

We conducted sensitivity analyses to estimate the possibility of confounder bias in the indirect effects reported in Study 2 and Study 3. These analyses indicate how strong the correlations would need to be between a potential confounding variable and the mediator (i.e., Felt Pressure) and between that same potential confounding variable and the outcome (e.g., nonverbal competence) for the indirect effect we observed to be zero. For example, if parent stereotypes predicted child felt pressure *and* nonverbal competence, sensitivity analyses tell us

how large those relationships would each need to be to render the indirect effect in this study null. To do this, we employed the Left Out Variables Error (LOVE) method (Mauro, 1990) using R script written by MacKinnon and Pirlott (2015). See Figure S3 for all sensitivity analysis plots from each moderated indirect effect. These analyses suggest that the indirect effects were fairly robust to confounder bias. For example, to render the indirect effect of condition on stereotypical puppet careers null, a confounder would have to be correlated with both the mediator (i.e., Felt Pressure) and the outcome (i.e., stereotypical puppet career choice) at a moderately strong degree (i.e., at about r=.55). Similarly, to render the indirect effect of condition on nonverbal competence null, a confounder would have to be correlated with both the mediator and the outcome at about r=.40.

Exploratory Outcome Variable in Study 3

After recording their video introductions in Study 3, we asked girls to reflect on how much the girls who were pictured would like them and how much girls in general would like them. This measure was exploratory and we added it after 10 participants had already completed the study. Therefore, we were cautious in overinterpreting these effects, but provide the results here for the interested reader. We anticipated that these reflected appraisals should correspond to how much girls modulated their behavior in the videos. We regressed the appraisal ratings on nonverbal competence, nonverbal bias condition, emotional perceptivity, and all interactions. Among girls skilled at reading subtle emotion, reflected appraisals should be highest for those in the Traditional condition with low nonverbal competence and those in the Reverse condition with high nonverbal competence. The three way interaction was not significant, b=5.03, se=4.49, $t(74)$=1.12, p=.267. However, the pattern was consistent with our hypotheses. The interaction of nonverbal competence and condition was significant for those high in emotional

perceptivity, *b*=1.82, *se*=.87, *t*(74)=2.10, *p*=.039, such that the more competent girls in the Reverse condition portrayed themselves, the more they thought other girls would like them, *b*=1.83, *se*=.68, *t*(74)=2.70, *p*=.009. There was no relationship between nonverbal competence and reflected appraisals for girls in the Traditional condition, *b*=.01, *se*=.54, *t*(74)=.01, *p*=.990.

We expected the same pattern of results with verbal and vocal competence. We regressed the appraisal ratings on verbal and vocal competence, nonverbal bias condition, emotional perceptivity, and all interactions. Again, the three way interaction was not significant, *b*=4.72, *se*=5.36, *t*(74)=.88, *p*=.382 and nor was the interaction of nonverbal competence and condition, *b*=1.62, *se*=.94, *t*(74)=1.72, *p*=.090. However, the more competently girls in the Reverse condition portrayed themselves, the more they thought other girls would like them, *b*=1.70, *se*=.82, *t*(74)=2.07, *p*=.042. There was no relationship between verbal/vocal competence and reflected appraisals for girls in the Traditional condition, *b*=.08, *se*=.46, *t*(74)=.17, *p*=.868. There were no meaningful relationships with nonverbal or verbal/vocal warmth. These results should be evaluated with caution given the non-significant omnibus tests, but may point towards important future directions for study.

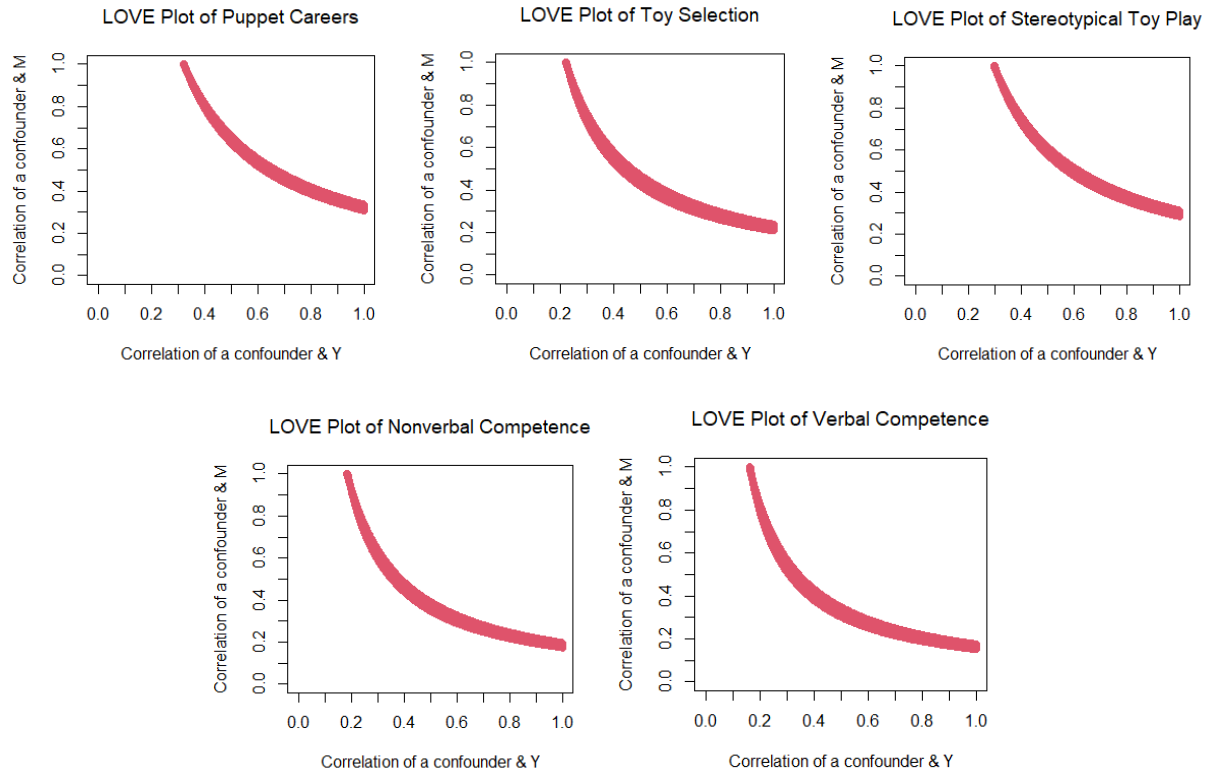

Figure S3. Sensitivity plots for indirect effects in Study 2 and Study 3 using the Left Out Variable Error (LOVE) method (Mauro, 1990). The red curve depicts the required correlations between a potential confounder and the outcome variable (on the X-axis) and the mediator (on the Y-axis) to render the indirect effect of nonverbal bias on the outcome variable by way of felt pressure for conformity 0.

Table S1
Statistics for a, b, and c' paths in PROCESS models reported in Study 2 and Study 3

	A						*b*						*c'*					
	b	***se***	***t***	***p***	**CI**		***b***	***se***	***t***	***p***	**CI**		***b***	***se***	***t***	***p***	**CI**	
STUDY 2																		
Stereotypical Playmate Preferences	1.13	.51	2.39	.029	.12	2.14	-.20	.63	-.31	.756	-1.45	1/06	.90	2.63	.34	.734	-4.36	6.16
Unstructured Toy Play	1.13	.51	2.39	.029	.12	2.14	99.27	33.82	2.94	.005	31.68	166.85	-259.15	142.16	-1.82	.730	-543.24	24.94
Puppet Customization	1.13	.51	2.39	.029	.12	2.14	1.89	.74	2.57[z]	.010	.45	3.34	-.56	2.32	-.24[z]	.810	-5.11	3.99
Career Aspirations	1.13	.51	2.39	.029	.12	2.14	.41	.14	2.93	.005	.13	.69	-.40	.59	-.68	.500	-1.58	.78
Career Prescriptions	1.13	.51	2.39	.029	.12	2.14	.07	.25	.27	.790	-.43	.57	-.47	1.06	-.42	.674	-2.56	1.66
Toy Selection	1.13	.51	2.39	.029	.12	2.14	.40	.22	1.81	.075	-.04	.84	.13	.92	.14	.890	-1.72	1.98
STUDY 3																		
Activity Preferences Scale	.98	.41	2.40	.018	.17	1.79	.26	.12	2.1	.039	.01	.50	-.12	.48	-.25	.802	-1.08	.84
Nonverbal Competence	.98	.41	2.40	.018	.17	1.79	-.24	.12	-1.99	.049	-.48	-.001	-.13	.48	-.26	.792	-1.07	.82
Verbal Competence	.98	.41	2.40	.018	.17	1.79	-.22	.14	-1.62	.108	-.49	.05	.25	.54	.47	.637	-.81	1.32
Nonverbal Warmth	.98	.41	2.40	.018	.17	1.79	.06	.19	.24	.733	-.31	.43	.15	.73	.21	.836	-1.30	1.60
Verbal Warmth	.98	.41	2.40	.018	.17	1.79	.01	.12	.06	.949	-.23	.25	-.13	.47	-.28	.780	-1.07	.81

Note: For dichotomous outcomes, the value listed in the column for *t* values is actually a *Z* value since we conducted a logistic regression. These values are indicated with a [z].

Table S2
Statistics for c paths in PROCESS models reported in Study 2 and Study 3

	C					
	b	***se***	***t***	***p***	**CI**	
STUDY 2						
Stereotypical Playmate Preferences	.68	2.52	.27	.789	-4.35	5.71
Unstructured Toy Play	-146.68	144.82	-1.01	.315	-435.98	142.63
Puppet Customization	1.22	2.12	.58[z]	.565	-2.93	5.38
Career Aspirations	.07	.60	.11	.914	-1.13	1.26
Career Prescriptions	-.37	1.01	-.37	.715	-2.39	1.65
Toy Selection	.58	.91	.642	.523	-1.23	2.39
STUDY 3						
Activity Preferences Scale	.13	.48	.28	.783	-.82	1.08
Nonverbal Competence	-.36	.47	-.77	.441	-1.29	.57
Verbal Competence	.04	.53	.07	.945	-1.01	1.08
Nonverbal Warmth	.21	.70	.30	.762	-1.18	1.61
Verbal Warmth	-.13	.46	-.27	.784	-1.03	.78

Note: For dichotomous outcomes, the value listed in the column for *t* values is actually a *Z* value since we conducted a logistic regression. These values are indicated with a [z].

Table S3

Full Statistics of Nonverbal Bias Condition x Emotional Perceptivity x Target Gender Interaction in Study 2

	ΔR^2	F (1, 60)	p
Felt Pressure for Conformity	.002	.16	.693
Stereotypical Playmate Preferences	<.001	.001	.980
Unstructured Toy Play	.002	.15	.701
Puppet Customization	N/A	.33[z]	.740
Career Aspirations	.001	.07	.797
Career Prescriptions	<.001	.03	.870
Toy Selection	.006	.39	.535

Note: Puppet customization is a binary outcome, so the value in the F column actually reflects a Z score and there is no ΔR^2 reported. This value is indicated with a [z].