Supplementary Materials for

Walking in her shoes: Pretending to be a woman role model increases young girls' persistence in science

Coding of experimenter behavior during the persistence task

Video recordings were available for 216 participants (103 girls, 113 boys). Two research assistants independently coded the videos for experimenter enthusiasm, encouragement, and engagement during the persistence task (see Table 1), and their codes were averaged.

Table S1. Coding categories of experimenter behavior during the persistence task.

Category	Explanation	Ratings
Enthusiasm	How happy and lively was the experimenter throughout the session?	1- Not at all enthusiastic 2- Slightly enthusiastic 3- Moderately enthusiastic 4- Very enthusiastic
Encouragement	How much did the experimenter encourage the child to persist in the game (either verbally or nonverbally)?	1- Not at all encouraging 2- Slightly encouraging 3- Moderately encouraging 4- Very encouraging
Engagement	How attentive was the experimenter during the session: did the experimenter look at the child and pay full attention to the session (vs. did the experimenter look away from the child and was distracted)?	1- Not at all engaged 2- Slightly engaged 3- Moderately engaged 4- Very engaged

Among girls, there were no significant condition differences in experimenter enthusiasm ($F(2,100) = 2.22$, $p = .114$), encouragement ($F(2,100) = .64$, $p = .527$), or engagement ($F(2,100) = 1.20$, $p = .306$) (Table 2).

Table S2. Descriptive statistics of experimenter behavior by condition for girls.

	Enthusiasm		Encouragement		Engagement	
	M	*SD*	*M*	*SD*	*M*	*SD*
Baseline	3.44	.56	2.59	.40	3.45	.38
Exposure	3.64	.31	2.65	.43	3.58	.37
Roleplay	3.46	.43	2.71	.41	3.49	.34

Among boys, there was no significant condition difference in experimenter enthusiasm ($F(2,110) = .17$, $p = .847$) or engagement ($F(2,110) = 1.38$, $p = .255$). However, there was a significant condition difference in experimenter encouragement ($F(2,110) = 3.98$, $p = .021$). Follow-up pairwise comparisons revealed that experimenters were rated as significantly more encouraging in the Roleplay condition than in the Baseline condition ($\beta = -.24$, 95% CI = [-.45, -.03], $p = .020$). Since boys in the Roleplay condition did not persist longer than boys in the Baseline condition, differences in experimenter encouragement did not seem to affect their behavior. There were no significant differences in experimenter encouragement between the Exposure and Baseline conditions ($\beta = -.07$, 95% CI = [-.28, .15], $p = .744$), or between the Exposure and Roleplay conditions ($\beta = -.17$, 95% CI = [-.39, .04], $p = .129$) (Table 3).

Table S3. Descriptive statistics of experimenter behavior by condition for boys.

	Enthusiasm		Encouragement		Engagement	
	M	*SD*	*M*	*SD*	*M*	*SD*
Baseline	3.42	.58	2.62	.44	3.33	.43
Exposure	3.49	.47	2.68	.42	3.43	.38
Roleplay	3.43	.44	2.86	.28	3.49	.41

Self-reported motivation: Game-specific vs. broader interests

As exploratory analyses, we probed children's self-reported motivation separately for game-specific questions and broader interest in being a scientist. The game-specific motivation score was calculated as the average of the standardized responses to the following questions: (1) "Do you like the science game or do you not like it?" (Follow up question: "Do you like/not like the science game a little, some, or a lot?", range: 0-5); (2) "How did playing the science game make you feel: happy or sad?" (Follow up question: "Did it make you a little happy/sad, happy/sad, or really happy/sad?", range: 0-5); and (4) "If you had a chance to do something tomorrow, would you play the science game, or would you do something else?" (score = 1 if they chose to play the game, 0 otherwise). The broader interest score was children's response to the question: "Do you want to be a scientist when you grow up?" (Follow-up question: "Do you want/not want to be a scientist a little/some/a lot?", range: 0-3).

We conducted two separate ANCOVA models. In the first model, we predicted game-specific motivation by gender, condition, age group, and their interaction, controlling for accuracy level. None of these effects was statistically significant (Fs < 3.30, ps > .070). The second model was identical to the first, except that broader interest was set as the dependent variable. The only significant effect was the interaction between age and gender ($F(1, 231) = 5.16$, $p = .024$). Follow-up pairwise comparisons showed that older (vs. younger) girls were less interested in becoming scientists when they grow up ($\beta = .68$, 95% CI = [.19, 1.17]), $p = .007$). However, there was no age difference among boys in wanting to be scientists ($\beta = -.11$, 95% CI = [-.59, .38]), $p = .663$). The interaction between age and gender on broader interest did not vary by condition ($F(1, 231) = .06$, $p = .813$).

Accuracy in the science game

We conducted an ANOVA analysis predicting accuracy in the science game by children's age, gender, condition, and the interaction between gender and condition. There was a main effect of age such that older (vs. younger) children were more accurate in the science game $F(1,233) = 11.50$, $p < .001$. However, accuracy did not vary by gender $F(1,233) = .08$, $p = .778$, condition $F(2,233) = .00$, $p = .999$, or their interaction $F(2,233) = 1.70$, $p = .185$ (Table 3). We also performed a survival curve analysis and found that accuracy was not related to persistence in the science game, $\beta = -.07$, $z = -.16$, $p = .871$.

Table S3. Descriptive statistics of accuracy scores by gender and condition.

		Baseline			Exposure			Roleplay		
		Mean	*SD*	*Median*	*Mean*	*SD*	*Median*	*Mean*	*SD*	*Median*
Accuracy Score	Girls	.72	.26	.75	.66	.23	.67	.68	.23	.74
	Boys	.65	.22	.67	.71	.23	.69	.68	.19	.67