Research Article

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science
1–12

DOI: 10.1177/0956797618823670
www.psychologicalscience.org/PS
SAGE

Subtle Linguistic Cues Increase Girls' Engagement in Science

Marjorie Rhodes[1], Sarah-Jane Leslie[2], Kathryn M. Yee[1], and Katya Saunders[1]
[1]Department of Psychology, New York University, and [2]Department of Philosophy, Princeton University

Abstract
The roots of gender disparities in science achievement take hold in early childhood. The present studies aimed to identify a modifiable feature of young children's environments that could be targeted to reduce gender differences in science behavior among young children. Four experimental studies with children ($N = 501$) revealed that describing science in terms of actions ("Let's do science! Doing science means exploring the world!") instead of identities ("Let's be scientists! Scientists explore the world!") increased girls' subsequent persistence in new science games designed to illustrate the scientific method. These studies thus identified subtle but powerful linguistic cues that could be targeted to help reduce gender disparities in science engagement in early childhood.

Keywords
cognitive development, language, science education, sex, social cognition, open data, open materials

Received 5/14/18; Revision accepted 10/30/18

The persistent underrepresentation of women in science limits women's intellectual and economic opportunities and impedes scientific progress by constraining the available talent pool (Beede et al., 2011). The roots of this disparity take hold in early childhood (Bian, Leslie, & Cimpian, 2017; Chapin, 2006; Weinburgh, 1995): Gender stereotypes about who can or should do science, and gendered patterns of interest in science, emerge by the time children begin formal schooling and widen across development (Newton & Newton, 1992; Nosek et al., 2009; Zhai, Jocz, & Tan, 2014). The present studies tested whether subtle linguistic cues—describing science as an action (e.g., asking children to "do science") instead of an identity (e.g., encouraging children to "be scientists")—increase girls' persistence in new science activities. Our aim was to identify modifiable elements of young children's environments that could help to bolster early science engagement.

In early childhood, children begin to develop beliefs that some variations in human behavior mark fundamentally distinct kinds of people whereas other variation is more temporary and incidental (Gelman, 2003; Rhodes & Mandalaywala, 2017). Thus, when it comes to science, children may view variations in interest, knowledge, or abilities as indications that some people are fundamentally "science people" or "true scientists deep down" and some people are not (Knobe, Prasada, & Newman, 2013; Rattan, Good, & Dweck, 2012); alternatively, they may perceive such variations as reflecting flexible differences in mood or previous experiences.

During the preschool years, when children are just beginning to develop beliefs about their own capacities for academic success (Eccles & Wigfield, 2002), subtle linguistic cues can powerfully shape whether children view variations as marking fundamentally distinct kinds of people. Specifically, category labels and generic claims lead children to view particular features as marking fundamentally distinct kinds of people when they would not otherwise do so (Baron, Dunham, Banaji, & Carey, 2014; Rhodes, Leslie, Bianchi, & Chalik, 2017; Waxman, 2010). For example, when 4- to 5-year-olds were given descriptions of new people using category labels (e.g., "Rose is a carrot-eater"), they expected

Corresponding Author:
Marjorie Rhodes, New York University, Department of Psychology, 6 Washington Place, Room 301, New York, NY 10003
E-mail: marjorie.rhodes@nyu.edu

behavior (e.g., eating carrots) to be stable across time and contexts. They did not exhibit that expectation if they heard the same information presented as a behavioral description (e.g., "Rose eats carrots whenever she can"; Gelman & Heyman, 1999).

As another example, when children, again between the ages of 4 and 5, were introduced to a new, made-up way of categorizing people ("Zarpies") through a series of generic claims (e.g., "Zarpies climb fences"), they expected the category to mark people who are fundamentally similar to each other and different from others (Rhodes, Leslie, & Tworek, 2012). In contrast, if children were introduced to the same information using nongeneric language (e.g., "This Zarpie climbs fences"), they instead viewed whether or not one was a Zarpie as more incidental and flexible.

As these examples illustrate, the implications of category labels and generic claims go far beyond the specific content that they communicate. There is nothing in the sentence "Zarpies climb fences" that explicitly communicates that climbing fences is something only Zarpies do, that Zarpies have other properties in common beyond those mentioned, or that climbing fences is part of the innate character of a Zarpie. Yet children began to develop these beliefs after relatively brief exposure to generic language (Gelman, Ware, & Kleinberg, 2010). We and other researchers have proposed that this occurs because young children are actively trying to sort out which variation reflects fundamental and stable differences among people and which is more incidental; consequently, they are highly attuned to signals from their cultural community that a particular feature marks an important distinction (Gelman, Taylor, Nguyen, Leaper, & Bigler, 2004; Segall, Birnbaum, Deeb, & Diesendruck, 2015). Thus, language can powerfully shape children's beliefs—sometimes in ways not easily predicted from an adult communicator's perspective—because language interacts with the basic processes underlying how children construct their understanding of the world.

Here, we considered the implications for early science behaviors of language that children might interpret as indicating that a capacity for science marks fundamentally distinct kinds of people. Such language is common in curricula, books, museums, and other media (Rhodes & Bushara, 2015). For example, we recently analyzed the language used to describe science in PBS children's television shows (33 different TV series, 993 episodes; Rhodes & Leslie, 2018). Category labels and generic claims (e.g., "You scientists did a great job!" "Scientists think that Einiosaurus may have lived in herds," "A great scientist can solve any problem") were the most frequent way of talking about science across these programs. In contrast, discussions of science as an activity (e.g., "I'm not going to use magic to make a rainbow, I am going to use science," "Science is all action!") were less common.

There is nothing explicitly negative or discouraging in the sentences using category labels and making generic claims about scientists; in fact, they often sound fun and engaging. Nevertheless, on the basis of previous research in this area, we suspected that young children might interpret these sentences as meaning that the category *scientists* marks a fundamentally distinct kind of person. We hypothesized that this language could lead to problematic consequences if children have reason to question whether they themselves are the kind of people who fit in the scientist category (e.g., after experiencing setbacks in science or developing stereotypes about scientists), because children might disengage if they no longer view science as consistent with their own identities (Master, Cheryan, & Meltzoff, 2017; McPherson, Park, & Ito, 2018). In contrast, describing science as an activity that people do—instead of as something that marks membership in a special group—would not have these problematic consequences.

Here, we tested whether describing science as an activity (instead of with category labels and generic claims) increases engagement in science, particularly among children who do not share stereotypic qualities with scientists (in this case, among girls). We focused primarily on early childhood because at this age, children are just beginning to develop concepts of science and scientists (Barman, 1999) as well as beliefs about their own capacities in these areas (Cain & Dweck, 1995; Smiley & Dweck, 1994), and because they are also highly susceptible to the influence of subtle linguistic cues. Thus, identifying modifiable features of young children's environments could be particularly important and consequential.

During the preschool years, one key component of early science education is helping children to understand science as a tool for asking and answering questions about the world. For example, a core goal of science education of the "Pre-K for All" program in New York City (where this research was conducted) is for children to become skilled at making observations, making predictions, and learning how to test predictions through further observation and experimentation (NYC Department of Education, 2018). With this in mind, we focused our test activities on children's use of the basic scientific method. To do so, we adopted science activities commonly used in early childhood classrooms to create our experimental paradigms.

Study 1

Method

Participants. Participants were 84 children (mean age = 4.84 years; 39 female, 45 male; 53.6% European American, 4.8% African American, 13.1% Hispanic, 4.8% Asian American, 20.2% multiethnic, 3.5% unknown) whose families had signed up to participate in developmental research at a campus laboratory. The only inclusion criterion (for all studies reported in this article) was that children had to be within the target age range and fluent in English. Our target sample size for Study 1 was 20 boys and 20 girls in each of our language conditions, and we stopped data collection on the weekend when we reached this sample size. This sample size gave us the power to detect consequences of the language manipulation that influence the likelihood of a participant disengaging from the presented science game by approximately 50% (relative to a participant of the same gender assigned to the other language condition; Schoenfeld, 1983). Because we were interested in identifying language manipulations that could have possible real-world relevance, we focused only on those that could have this level of strong influence on children's behavior. The institutional review board of New York University approved the procedure, parents provided written consent for their children to participate, and children provided oral assent prior to testing. All data and analytic code for the studies reported here are available at our repository on the Open Science Framework (https://osf.io/p8f7w/).

In addition to the participants included in analyses, an additional 22 children began testing but were excluded. The majority of these (n = 16) were excluded because of experimenter mistakes (e.g., not adhering to the condition-specific script and thus presenting a mixture of "do science" and "be scientist" language) or equipment failure involving the science game. (The task, as described in the Procedure section that follows, involved smelling particular items in cups and making guesses; problems occasionally arose with this procedure, including instances in which the cups were set up incorrectly or were spilled and revealed their contents during critical phases of the study.) The additional 6 children were excluded either because of parental interference (a parent insisting that their child complete more trials of the science game; n = 1) or because the child failed to respond to prompts during the introductory phase (n = 5). All research sessions were videotaped. The experimenter recorded the children's responses live, and then all videotapes were scored from video by a separate coder, who also rated any deviations from the experimental protocol (e.g., equipment failure or experimenter errors with the scripts) and indicated whether a problem occurred that warranted exclusion. These decisions were made prior to analyses.

Procedure. Prior to the start of testing, each child was randomly assigned prior to the start of testing to the "be-scientist" or "do-science" language conditions. To begin the research session, children completed a 3-min introduction to science in which an experimenter described the scientific method (i.e., making observations, making predictions, checking, and recording) either using category labels and generic claims (e.g., "Today we're going to be scientists! Scientists use their five senses to learn about the world") or describing science in terms of actions (e.g., "Today we're going to do science! When people are doing science they use their five senses to learn about the world"). As described earlier, we focused on understanding and using the scientific method because understanding science as a tool for asking and answering questions about the world is fundamental to early science education. Also, we wanted to focus on an aspect of science that is general and broad enough not to be strongly associated with particular stereotypes. For example, we avoided using things like toy chemistry sets, about which children may already hold specific gender-related beliefs. The content that children heard was identical across conditions, but whether science was described as an identity or an activity systematically varied by condition.

After the introductory phase, a new experimenter, who was blind to which language condition children had received, entered the room and asked children to practice what they had just learned about the scientific method by smelling the hidden contents of a covered cup, making a guess about the contents, and then checking to see whether the guess was accurate. From this point on, all of the language that children heard was identical across conditions. All children completed four initial trials that were controlled by the experimenter. These four trials were "rigged": Children first completed two easy trials, in which the smell was obvious and they had to select between only two answer choices, followed by two difficult trials, in which the smell was purposefully misleading and children had to select among five answer choices (e.g., the cup contained a sponge soaked in lemon juice, and both a lemon and a sponge were answer choices). We did this so that children would experience setbacks as they began to practice science. The task itself (smelling covered cups) was adapted from one that is often used to teach young children about observation and prediction in preschools and informal science learning environments. The full experimental protocol for this and all studies is available at https://osf.io/p8f7w/.

After the four experimenter-controlled trials, children were told that they had the opportunity to continue the science game that they had just played, but that it was up to them how much they wanted to do. Before each trial they were asked, "Do you want to keep making guesses, or do you want to do something else?" If they wanted to keep making guesses, they were given another covered cup to smell, and they made a guess about the contents, which the experimenter recorded. In this phase, however, children were told that they would check all of the guesses when they were all done playing (this was done to avoid a situation in which individual children experienced different levels of success and failure).

Children were given the opportunity to complete up to 10 additional trials. We measured how many trials children chose to complete as an indicator of their willingness to continue engaging in science activities after having experienced setbacks and so we could test how this form of persistence varied by participant gender and by the language they had heard. This measure of persistence was beneficial because it allowed us to test for the effects of our manipulation on children's actual behavior (how many trials of the task children chose to do) instead of relying on self-reports of interest or attitudes, which can be difficult to assess reliably among children in this age range. Further, task engagement is an important indicator of the success of early childhood curricula: Young children cannot learn new skills if they are unwilling to practice them, so the amount of time that young children in our studies chose to spend trying out the task and practicing the steps of the scientific method that they had just learned was an externally valid indicator of the consequences of language for early science education.

Analysis plan. We modeled these data using survival curve analysis, using the *survival* and *survminer* packages in R programming environment (R Core Team, 2018). Although more commonly used in studies of health-related outcomes (e.g., to model rates of morbidity after patients receive one of two cancer treatments), these packages have also been used to model persistence in psychological studies (e.g., to predict the risk of quitting on delay-of-gratification tasks; McGuire & Kable, 2012). Here, we used these models to predict the probability of children choosing to stop playing the science game across the 10 possible trials. These models are well suited for predicting events that occur over time (in this case, the event is a participant's choice to stop the game) and can also account for the fact that the event never occurs for some subset of the population (in other words, some children persist throughout the whole task). Because we were interested in interaction effects, we implemented Cox proportional-hazards models using the *coxph* function in R. We tested for the main and interactive effects of participant gender and language condition on the likelihood of choosing to stop the task. We report regression coefficients from these models, along with standard errors, Wald statistics—a test of the significance of each coefficient, $z = \beta/SE(\beta)$—and associated p values. For the regression coefficients, a positive sign means that the likelihood of stopping is higher than in the reference group, whereas a negative sign means that the likelihood of stopping is lower. The exponentiated coefficients for these analyses are referred to as the hazard ratios and are indicators of the effect sizes. For ease of interpretation, we report these as percentages (the percentage change associated with being in one group or another), along with 95% confidence intervals (CIs).

Results

Girls in the do-science condition were less likely to stop playing the game than girls in the be-scientist condition, $\beta = -0.92$, $SE = 0.41$, $z = -2.28$, $p = .023$—for girls, the do-science condition reduced the likelihood of stopping by 60% (95% CI = [12%, 82%])—whereas this pattern was reversed for boys, $\beta = 0.79$, $SE = 0.35$, $z = 2.22$, $p = .026$—for boys, the be-scientist condition reduced the likelihood of stopping by 65%). The Gender × Condition interaction was reliable, $\beta = 1.79$, $SE = 0.54$, $z = 3.33$, $p \leq .001$. As shown in Figure 1, many girls in the be-scientist condition chose to complete no additional trials of the persistence task after experiencing the two difficult trials.

The substantial drop from Trial 0 to Trial 1 seen in Figure 1 suggests that many children chose to play no additional trials of the game after the four experimenter-controlled trials, and the drop in the number of children playing the science game was particularly pronounced among girls in the be-scientist condition. Tests of the proportional-hazards assumption confirmed that this assumption was met for the model as a whole, and for the key Gender × Condition interaction, but that this assumption was violated for the subsumed main effect of gender. We addressed this issue by examining effects of condition separately by gender, but also by conducting follow-up analyses: A Gender × Time interaction in the model (recommended for addressing this issue; Fox & Weisberg, 2011) revealed an identical pattern of findings. See https://osf.io/p8f7w/ for the full output of that model.

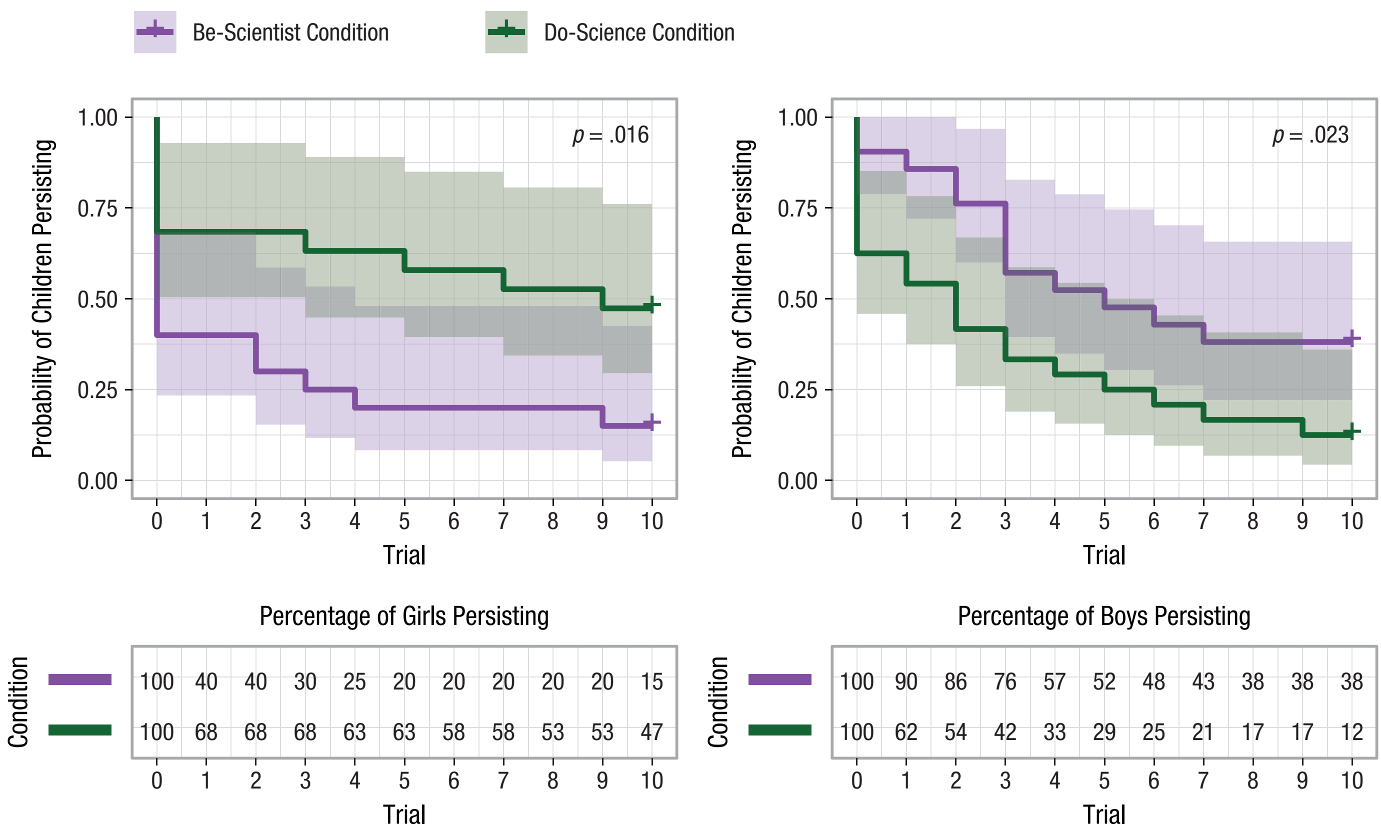

Fig. 1. Results from Study 1: survival curves for girls (left) and boys (right; top row) and the percentage of girls and boys persisting across trials (bottom row), separately for the be-scientist and do-science conditions. For the survival curves, p values are from log-rank tests. Trial 0 indicates that everyone had the opportunity to play the first trial of the game.

Study 2

Method

In Study 2, we attempted to replicate the results of Study 1 using a different science game—predicting whether objects would sink or float in water—and more naturalistic patterns of success and failure. Because this task required more background knowledge, we recruited slightly older children to participate.

Participants. Participants were 89 children (45 male, 44 female; age: M = 6.14 years, range = 4.99–9.37; 29.2% Caucasian, 6.7% African American, 14.6% Hispanic, 6.7% Asian American, 16.9% multiethnic, 25.9% unknown) who were recruited and tested in a children's museum (n = 41) and a public elementary school (n = 48). Sample sizes were determined as in Study 1; age did not differ by condition, $t(87) = -0.22$, p = .82. One additional child began testing but was excluded because of equipment failure (involving the iPad used to present the science game). Sessions in the museum were video recorded, and sessions in the school were audio recorded. The institutional review boards of both New York University and the New York City Department of Education approved the study procedures.

Procedures. Procedures were very similar to those in Study 1; children first completed an introductory phase in which, by random assignment, they heard about "being scientists" or "doing science" and then were given an opportunity to play a science game. In Study 2, however, the science game asked children to use their sense of sight to look at the objects and make predictions about whether they would sink or float. Because Study 2 was conducted in community-based settings in which we had limited control over the testing environment, a single experimenter conducted the entire session (as is common for developmental research in such settings). To reduce concerns that experimenter expectations could influence the findings, we presented the dependent measure on an iPad, using a commercially available game, instead of having the experimenter present it. We found it was possible to reduce, but not eliminate, experimenter–child interactions with this method. The tablet-based game presented each object and prompted children to guess whether it would sink or float. Children responded verbally, and the experimenter assisted with pushing the button on the game to register the guess. The experimenter drew the child's attention to whether each guess was right or wrong (but did not control the outcome—so experiences of success varied across children) by saying "Your prediction was right! It did float!" or "Your prediction was wrong! It sank!" After each trial, the experimenter asked, "Do you want to keep playing the science game, or do you want to do something else?" Experimenters were trained to use the same tone across trials (regardless of the outcome) and across conditions. Children were allowed to play for as long as they wished.

Analysis plan. We modeled the data as in Study 1, except that we included children's level of accuracy (calculated as the percentage of correct guesses that they made) as a continuous covariate in the models. Overall, children's accuracy level was 58%, indicating that this was a challenging task. Accuracy did not vary significantly by gender (girls: M = 57% correct; boys: M = 61% correct), $t(87) = -0.79$, $p = .433$, and the results were virtually identical when the accuracy variable was excluded from the model. Although we had a larger age range in Study 2 than in Study 1, we did not have an adequate sample size to examine how these effects change across age, and participant age was not included in the statistical models. Thus we tested for only for main and interactive effects of participant gender and language condition, controlling for children's individual levels of accuracy.

Results

As in Study 1, the effect of language varied by gender, $\beta = 1.23$, $SE = 0.45$, $z = 2.72$, $p = .0065$; girls had a lower likelihood of stopping the science game in the do-science condition than in the be-scientist condition, $\beta = -0.82$, $SE = 0.33$, $z = -2.46$, $p = .014$ (for girls, the do-science condition reduced the likelihood of stopping by 56%); in this study, there was no effect for boys, $\beta = 0.4$, $SE = 0.32$, $z = 1.25$, $p = .21$ (see Fig. 2).

Study 3

Study 3 examined a larger sample of young children to test how the gendered patterns found in Studies 1 and 2 emerge across the preschool years. Because young children are just being introduced to science, age-related trajectories could inform future research regarding underlying mechanisms. For example, if the do-science language becomes increasingly beneficial for girls with age, this might indicate that these subtle linguistic cues become particularly important as children begin to acquire gender stereotypes.

Method

Participants. Participants were 4- and 5-year-old children (N = 160; 80 girls, 80 boys; age: M = 4.73 years, range = 4.01–5.97; 48.1% Caucasian, 7.5% African American, 7.5% Hispanic, 11.3% Asian American, 20.6% multiethnic, 5% unknown) who were recruited and tested at seven public preschool sites (n = 77), an elementary school (n = 7), and a children's museum (n = 76). An additional 17 children began testing but were excluded from analyses because of experimenter errors (n = 8), because the video or audio failed to record (n = 5), or because the child refused to answer study questions (n = 4). The sample size for Study 3 was doubled, relative to Studies 1 and 2, to 40 boys and 40 girls in each language condition because we planned to include an additional continuous predictor (child age).

Procedures. Procedures were similar to those in Study 1, except that we developed a new science game that we thought would be particularly engaging to these young children—guessing what foods particular animals liked to eat (presented with toy figurines). Also, in order to increase our sample size (and the extent to which our sample represented our community), we ran this study (and Study 4) in a variety of community settings, including public prekindergarten centers, instead of in our campus-based laboratory. As in Study 2, in these settings, it was not possible to keep experimenters blind to language condition. All experimenters were blind to child age at the time of testing, however, as a key focus of this study was the emergence of experimental effects across age. Further, 60% of the data across Studies 3 and 4 was collected by undergraduate research assistants who were blind both to hypotheses and to the previous findings in this article (the remaining 40% was run by the third author). There were no main or interactive effects of type of experimenter in any analysis.

As in the previous studies, the game that served as the dependent measure was presented as an opportunity to practice the scientific method by observing, predicting, and then checking the accuracy of one's guesses. As in previous studies, children first completed an introductory phase similar to that in Study 1, in which children, by random assignment, heard about either "being scientists" or "doing science." During the introductory phase, children completed two trials with experimenter-controlled feedback in which they experienced one success (correct guess) and one failure (incorrect guess). Subsequently, as in Study 1, they were

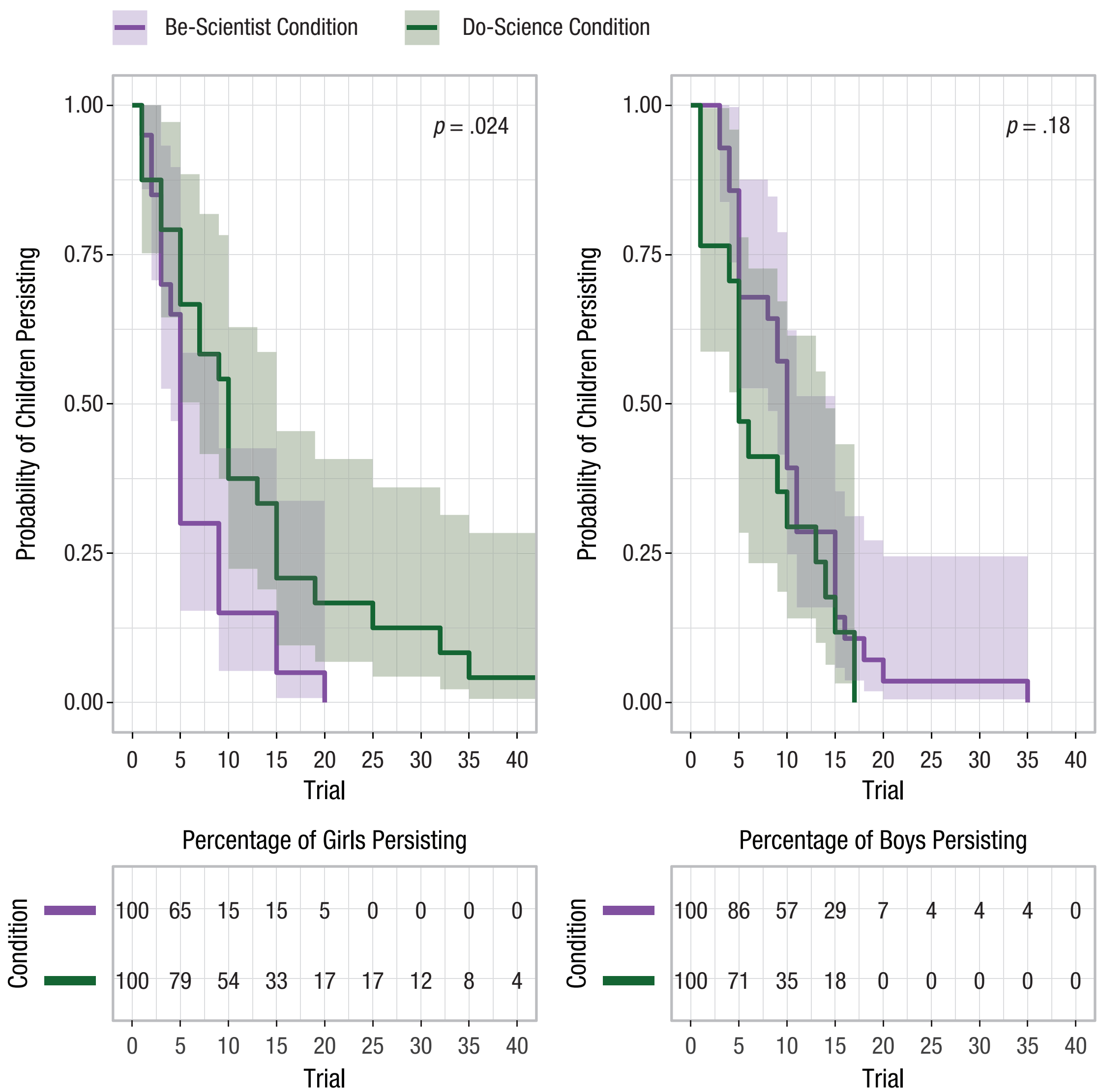

Fig. 2. Results from Study 2: survival curves for girls (left) and boys (right; top row) and the percentage of girls and boys persisting across trials (bottom row), separately for the be-scientist and do-science conditions. For the survival curves, p values are from log-rank tests. Trial 0 indicates that everyone had the opportunity to play the first trial of the game. Unlike in Study 1, children were allowed to play the game for as long as they wished.

given the option to continue playing the science game or do something else. Children could complete up to 10 trials, and as in Study 1, no additional feedback about the accuracy of children's guesses was provided during the test trials.

Analysis plan. We first modeled the data using the *coxph* procedure to examine the survival curves, as in Studies 1 and 2, but we also included children's exact ages as a continuous predictor. After finding a three-way interaction involving participant age (see below), we then also modeled these data using quasi-Poisson regression models, predicting the number of trials that children chose to complete. The quasi-Poisson models were not a perfect fit for our data because they did not directly account for the fact that some children completed all 10 trials (and therefore never experienced "the event" of stopping). However, these models provided an easier (more intuitive) way of visualizing the three-way interaction. Further, the two approaches to the analyses— looking at the survival curves and implementing the Cox proportional-hazards models and the quasi-Poisson regression models—yielded virtually identical patterns of significant effects (codes for all models are available at https://osf.io/p8f7w/).

Results

The survival analyses revealed that the effect of gender and condition interacted with participant age, β = 1.82, *SE* = 0.81, *z* = 2.26, *p* = .024. For ease of interpretation given the interaction with age (a continuous variable), Figure 3 plots the number of trials completed, instead of the survival curves, as described above. Similar to the survival-curve analyses, these quasi-Poisson regression models revealed a three-way interaction among gender, language condition, and child age, β = −1.39, *SE* = 0.69, *t*(152) = −2.03, *p* = .044. Overall, girls completed more trials in the do-science condition than in the be-scientist condition, β = 0.47, *SE* = 0.22, *t*(76) = 2.11, *p* = .038, whereas the effect for boys interacted with age, β = −0.91, *SE* = 0.46, *t*(76) = −1.98, *p* = .051. The be-scientist condition became more beneficial to boys' persistence, β = 0.78, *SE* = 0.33, *t*(38) = 2.37, *p* = .023, and more detrimental to girls' persistence, β = −0.46, *SE* = 0.45, *t*(37) = −1.03, *p* = .31, across this age range; within the be-scientist condition, the Gender × Age interaction was reliable, β = 1.24, *SE* = 0.56, *t*(75) = 2.24, *p* = .028.

Study 4

Finally, we investigated whether the gendered patterns observed in Studies 1 through 3 were particular to a domain that is the target of relevant cultural stereotypes or arise more broadly for identity- versus action-focused language.

Method

Participants. Participants were 4- and 5-year-old children (*N* = 168; 85 girls, 83 boys; mean age: = 4.83 years, range = 4.01–5.97; 36.9% Caucasian, 7.7% African American, 13.7% Hispanic, 14.9% Asian American, 17.3% multiethnic, 9.5% unknown) who were recruited and tested at seven public preschool sites (*n* = 58), a campus laboratory (*n* = 1), and a children's museum (*n* = 109). Sample size was chosen to match that of Study 3. An additional 29 children began testing but were excluded from analyses because of experimenter errors (*n* = 13), because the video or audio failed to record (*n* = 4), because of parental influence (*n* = 5), or because the child refused to answer study questions (*n* = 7).

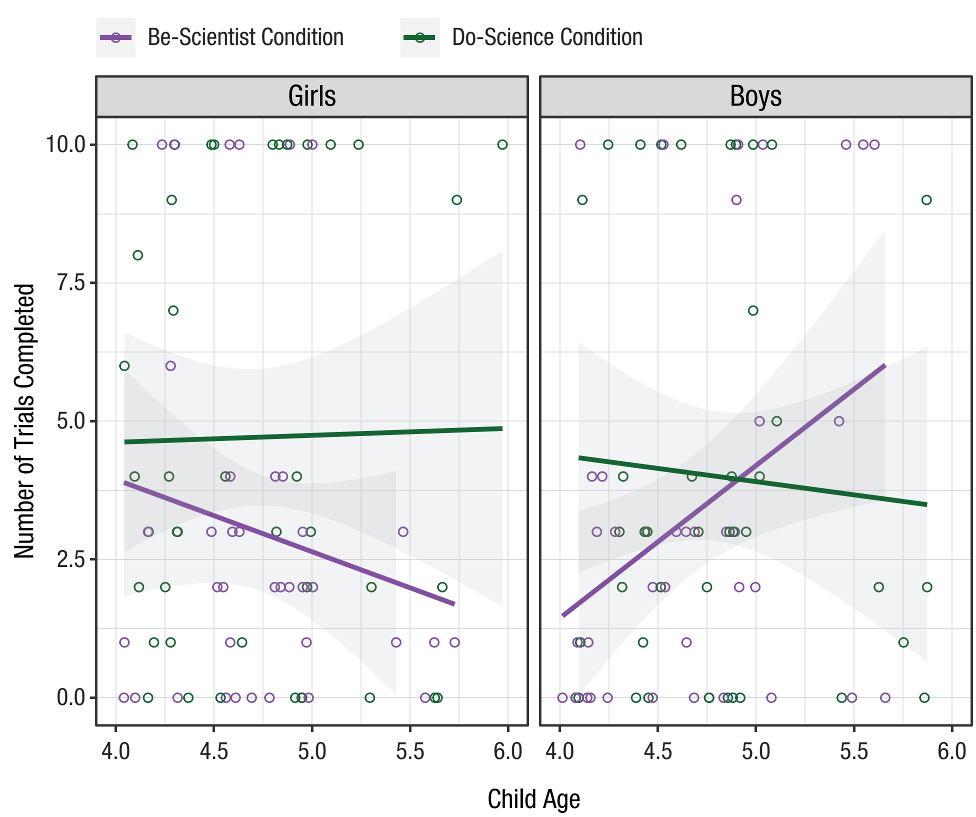

Fig. 3. Results from Study 3: scatterplots showing the association among number of trials completed, child age, and language condition, separately for girls and boys. The lines show the predicted values from a quasi-Poisson model predicting the total number of trials completed; the circles represent the data of individual participants.

Procedures. Procedures were nearly identical to those used in Study 3 and involved all of the same materials and the same game (guessing which foods to feed particular animals), except that the experimenter presented the task as a caring game rather than a science game. Thus, the children were asked to "be a carer" or "do caring" and decide which foods to give the animals. As in the condition training in Studies 1 through 3 involving science, children here first completed a training activity using condition-specific language. (In this case, they were shown a doll and asked to use their senses to make guesses about what the baby needed—e.g., a bottle or a diaper change; for the full script, see https://osf.io/p8f7w/.) This was all described as part of "caring" or "being a carer." After the condition-specific training, they completed the dependent measure in which they were asked to care for animals by figuring out which foods they needed to eat.

As in Study 3, they were first asked to make one guess that they were told was accurate and one that they were told was inaccurate, and then they could complete up to 10 additional trials of the game without further feedback. During this phase, the task domain was not mentioned at all in either Study 3 or 4; children were simply asked, "Do you want to keep making guesses, or do you want to do something else?" in both studies, with no mention of science or caring. Thus the dependent measure was implemented in exactly the same way in Study 4 as in Study 3, allowing us to test whether receiving identity- versus action-focused language influenced children's performance on the same task they completed in Study 3, when beliefs about science were not task relevant.

Analysis plan. We analyzed the data in the same manner as in Study 3, implementing both the Cox proportional-hazards models and the quasi-Poisson regression models. We followed up these analyses with a series of Bayesian analyses, as described below.

Results

Neither the Cox proportional-hazards models nor the quasi-Poisson regression models revealed main or interactive effects of language condition (all p values associated with main or interactive effects of condition > .4; for full models, see https://osf.io/p8f7w/). The absence of these effects is notable given that the sample size and dependent measure used here were identical to those in Study 3, suggesting that this study was adequately powered to detect effects had they been present.

To further facilitate interpretation of the pattern of findings across studies, including these null results, we next pooled the data across Studies 1 and 3 and implemented a Bayesian regression model to generate a robust estimate of the interaction between gender and condition in the studies focusing on science. (We excluded Study 2 from this analysis because that study included older children and a dependent measure on a very different scale.) We used the *brms* package in R to model the number of trials completed before quitting with a cumulative distribution and gender, condition, age (as a centered continuous variable), and all of their interactions as predictors. We set weakly informative priors for all predictors (a normal distribution with a mean of 0, indicating no relationship between the predictors and the dependent measure, and SD = 1). The code and full outputs of these models are available at https://osf.io/p8f7w/. In this analysis, the parameter estimate for the Gender × Condition interaction was −0.99 with a 95% credibility interval of [−1.78, −0.20] (meaning that the credible values for this parameter differed from 0). Comparing this model (using the *bayes_factor* function in *brms*) to one that does not account for the variable effect of language by gender (by removing both the Gender × Condition interaction and the three-way interaction involving age) revealed that the estimated Bayes factor in favor of the model that incorporated the Gender × Condition interactions is approximately 6, indicating support for the hypothesis that the effect of language on children's persistence varies by gender (Dienes, 2014).

In comparison, the parameter estimate for the Gender × Condition interaction when children were asked to think about taking caring of animals (and science was not mentioned) was −0.22 (95% credibility interval = [−.1.16, 0.71]), and the parameter estimate for the interaction involving age was −0.04 (95% credibility interval = [−1.39, 1.32]). Bayes factor analysis comparing this model to one without the interactions among gender and condition indicated that the observed data were 2.63 times more likely to arise under the null hypothesis than under the alternative in the domain of caring. There is ongoing debate about how to best interpret Bayes factors (e.g., Kruschke & Liddell, 2018); in this case, however, consideration of both the credibility intervals and the Bayes factors compared across models supports the interpretation that action- versus identity-focused language influences persistence in a manner that varies by gender for science, but not across all task domains.

Discussion

These studies reveal how subtle linguistic cues influence girls' science engagement. Language describing science as action, rather than in terms of identities, led to more persistence for girls across three different, ecologically valid science games designed to teach children about the scientific method. The present studies set the foundation for future research to test for long-term consequences of language exposure in children's daily lives and to consider how this modifiable feature of children's environments could be targeted to boost girls' persistence before gendered trajectories in science begin to diverge.

We consistently found that identity-focused language has a gendered effect on persistence in science, but Study 4 revealed that it does not do so in all domains. We suspect that identity-focused language is problematic for children when they have reason to doubt that they are members of the relevant category, because this then reduces motivation to engage in category-relevant behavior. Reasons to doubt could stem from the separate or interactive effects of social stereotypes (e.g., beliefs that scientists are usually men; e.g., Master, Cheryan, & Meltzoff, 2016), from their own experiences of difficulty (e.g., Dweck, 2006), from a belief that success requires innate talent (e.g., Bian et al., 2017), and from general beliefs that membership in a group is important and rare. Consistent with this interpretation, results showed that in the prosocial domain, asking children to "be helpers" (instead of "to help") interferes with subsequent helping behavior after children experience highly salient setbacks, regardless of child gender (Foster-Hanson, Cimpian, Leshin, & Rhodes, 2018). Thus, the effect of language does not depend on gender alone; instead, gender can operate as one factor that interacts with children's other task-related beliefs and experiences.

In the present studies, none of the language that we provided was evaluative. Unlike the literature on praise, we did not examine the consequences of praising children for being "good scientists" or "smart scientists" (e.g., Cimpian, Arce, Markman, & Dweck, 2007; Kamins & Dweck, 1999). We examined the implications of language that can be used even more broadly (e.g., through curricula, to whole groups of children at a time) and, indeed, was the most common way of introducing science to young children in the samples of children's media that we analyzed (Rhodes & Leslie, 2018).

Action-focused language consistently led to increased persistence among girls but had more variable consequences for boys. Positive in-group stereotypes can increase task engagement (Master et al., 2017; Walton & Cohen, 2003), provided that such stereotypes are elicited subtly, so as to avoid creating anxiety about living up to expectations (Cimpian, 2013; Cimpian, Mu, & Erickson, 2012; Park, Schaeffer, Nolla, Levine, & Beilock, 2017). Thus, positive in-group stereotypes activated by the "be-scientist" language could have benefited boys in Study 1, and done so more with age in Study 3 (presumably as children acquire these stereotypes). From this perspective, it is important to consider whether identity-focused language could benefit girls over boys in other domains. Study 4 (which considered caregiving) might appear inconsistent with this possibility, but because this study involved taking care of animals (something children might associate with farmers, for example) and did not use a familiar category label, the extent to which identity-focused language has positive consequences for girls in domains where they hold positive in-group stereotypes remains an open question.

These differing patterns by gender complicate efforts to build interventions based on these findings. There are several caveats to consider. First, there was a general, nongendered advantage of the "do-science" language among the youngest children in this research (in Study 3). Second, because there are also racial, ethnic, and economic stereotypes about scientists, it is possible that boys from more diverse backgrounds than those sampled here could benefit from action-focused language. Interestingly, we found no effect of language for boys in Study 2, which included the most ethnically, racially, and economically diverse sample in this research. We did not have sufficient power to test for predictors of individual variation in boys' responses, but doing so will be important in future work. Finally, it will be important to consider how language might influence children's behavior over time, particularly once they encounter more serious setbacks in science (which could lead even children who are part of positively stereotyped groups to question their own capacities). Future research will need to address these issues to identify communities of children that are most likely to benefit from intervention promoting action-focused language about science.

Action Editor

Bill von Hippel served as action editor for this article.

Author Contributions

M. Rhodes and S.-J. Leslie designed the studies and wrote the manuscript. M. Rhodes supervised data collection and analyzed the data. K. M. Yee and K. Saunders contributed to the research protocols, collected the data, and supervised data entry and processing.

Acknowledgments

We thank the Children's Museum of Manhattan and the New York City Department of Education for their support of this work and Robert Rehder, Barbara Sarnecka, Jeff Rouder, and Zachary Horne for their input on analyses.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health under Award Number R01HD087672. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Open Practices

All data, materials, and analytic code (including that used to produce the graphs in this article) have been made publicly available via the Open Science Framework and can be accessed at https://osf.io/p8f7w/. These studies were not formally preregistered, as data collection began before this practice was typical in developmental science. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797618823670. This article has received the badges for Open Data and Open Materials. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

References

Barman, C. R. (1999). Students' views about scientists and school science: Engaging K-8 teachers in a national study. *Journal of Science Teacher Education*, *10*, 43–54. doi:10.1023/A:1009424713416

Baron, A. S., Dunham, Y., Banaji, M., & Carey, S. (2014). Constraints on the acquisition of social category concepts. *Journal of Cognition and Development*, *15*, 238–268. doi:10.1080/15248372.2012.742902

Beede, D., Julian, T., Langdon, D., McKittrick, G., Khan, B., & Doms, M. (2011). *Women in STEM: A gender gap to innovation* (ESA Issue Brief #04-11). Washington, DC: U.S. Department of Commerce.

Bian, L., Leslie, S. J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, *355*, 389–391. doi:10.1126/science.aah6524

Cain, K. M., & Dweck, C. S. (1995). The relation between motivational patterns and achievement cognitions through the elementary school years. *Merrill-Palmer Quarterly*, *41*, 25–52. doi:10.1111/j.1467-8624.1998.tb06197.x

Chapin, J. (2006). The achievement gap in social studies and science starts early: Evidence from the Early Childhood Longitudinal Study. *The Social Studies*, *97*, 231–238. doi:10.3200/TSSS.97.6.231-238

Cimpian, A. (2013). Generic statements, causal attributions, and children's naive theories. In M. R. Banaji & S. A. Gelman (Eds.), *Navigating the social world: What infants, children, and other species can teach us* (pp. 269–274). New York, NY: Oxford University Press.

Cimpian, A., Arce, H. C., Markman, E. M., & Dweck, C. S. (2007). Subtle linguistic cues affect children's motivation. *Psychological Science*, *18*, 314–316. doi:10.1111/j.1467-9280.2007.01896.x

Cimpian, A., Mu, Y., & Erickson, L. C. (2012). Who is good at this game? Linking an activity to a social category undermines children's achievement. *Psychological Science*, *23*, 533–541.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*, Article 781. doi:10.3389/fpsyg.2014.00781

Dweck, C. S. (2006). *Mindset: The new psychology of success*. New York, NY: Random House.

Eccles, J., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*, 109–132. doi:10.1146/annurev.psych.53.100901.135153

Foster-Hanson, E., Cimpian, A., Leshin, R., & Rhodes, M. (2018). Asking children to "be helpers" backfires after setbacks. *Child Development*. Advance online publication. doi:10.1111/cdev.13147

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (2nd ed.). Thousand Oaks, CA: SAGE.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford, England: Oxford University Press.

Gelman, S. A., & Heyman, G. D. (1999). Carrot-eaters and creature-believers: The effects of lexicalization on children's inferences about social categories. *Psychological Science*, *10*, 489–493. doi:10.1111/1467-9280.00194

Gelman, S. A., Taylor, M., Nguyen, S. P., Leaper, C., & Bigler, R. S. (2004). Mother-child conversations about gender: Understanding the acquisition of essentialist beliefs [Target article and commentaries]. *Monographs of the Society for Research in Child Development*, *69*(1), 1–142. doi:10.1111/j.1540-5834.2004.06901002.x

Gelman, S. A., Ware, E. A., & Kleinberg, F. (2010). Effects of generic language on category content and structure. *Cognitive Psychology*, *61*, 273–301. doi:10.1016/j.cogpsych.2010.06.001

Kamins, M. L., & Dweck, C. S. (1999). Person versus process praise and criticism: Implications for contingent self-worth and coping. *Developmental Psychology*, *35*, 835–847. doi:10.1037/0012-1649.35.3.835

Knobe, J., Prasada, S., & Newman, G. E. (2013). Dual character concepts and the normative dimension of conceptual representation. *Cognition*, *127*, 242–257. doi:10.1016/j.cognition.2013.01.005

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*, 178–206.

Master, A., Cheryan, S., & Meltzoff, A. N. (2016). Computing whether she belongs: Stereotypes undermine girls' interest and sense of belonging in computer science. *Journal of Educational Psychology*, *108*, 424–437. doi:10.1037/edu0000061

Master, A., Cheryan, S., & Meltzoff, A. N. (2017). Social group membership increases STEM engagement among preschoolers. *Developmental Psychology*, *53*, 201–209. doi:10.1037/dev0000195

McGuire, J. T., & Kable, J. W. (2012). Decision makers calibrate behavioral persistence on the basis of time-interval experience. *Cognition*, *124*, 216–226. doi:10.1016/j.cognition.2012.03.008

McPherson, E., Park, B., & Ito, T. A. (2018). The role of prototype matching in science pursuits: Perceptions of scientists that are inaccurate and diverge from self-perceptions predict reduced interest in a science career. *Personality and Social Psychology Bulletin*, *44*, 881–898. doi:10.1177/0146167217754069

Newton, D. P., & Newton, L. D. (1992). Young children's perceptions of science and the scientist. *International Journal of Science Education*, *14*, 331–348. doi:10.1080/0950069980200909

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., . . . Greenwald, A. G. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences, USA*, *106*, 10593–10597. doi:10.1073/pnas.0809921106

NYC Department of Education. (2018). *PK–8 science scope & sequence 2018*. Retrieved from https://www.weteachnyc.org/media2016/filer_public/c8/f8/c8f8c331-596f-4705-b445-b35dc3e2adb3/science_scope_and_sequence.pdf

Park, D., Schaeffer, M. W., Nolla, K. M., Levine, S. C., & Beilock, S. L. (2017). How do generic statements impact performance? Evidence for entity beliefs. *Developmental Science*, *20*(2), Article e12396. doi:10.1111/desc.12396

Rattan, A., Good, C., & Dweck, C. S. (2012). It's ok, not everyone can be good at math: Instructors with an entity theory comfort (and demotivate) students. *Journal of Experimental Social Psychology*, *48*, 731–737. doi:10.1016/j.jesp.2011.12.012

R Core Team. (2018). *R: A language and environment for statistical computing*. Retrieved from http://www.R-project.org

Rhodes, M., & Bushara, L. (2015). Learning about science and self: A partnership between the Children's Museum of Manhattan and the Psychology Department at New York University. In D. Sobel & J. Jipson (Eds.), *Fostering cognitive development in children's museums* (pp. 104–119). New York, NY: Psychology Press.

Rhodes, M., & Leslie, S.-J. (2018). Children's media communicates essentialist views of science and scientists. Retrieved from PsyArXiv website: https://psyarxiv.com/s6chb/

Rhodes, M., Leslie, S. J., Bianchi, L., & Chalik, L. (2017). The role of generic language in the early development of social categorization. *Child Development*, *89*, 148–155. doi:10.1111/cdev.12714

Rhodes, M., Leslie, S.-J., & Tworek, C. M. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences, USA*, *109*, 13526–13531. doi:10.1073/pnas.1208951109

Rhodes, M., & Mandalaywala, T. M. (2017). The development and developmental consequences of social essentialism. *WIREs Cognitive Science*, *8*(4), Article e1437. doi:10.1002/wcs.1437

Schoenfeld, D. A. (1983). Sample-size formula for the proportional-hazards regression model. *Biometrics*, *39*, 499–503.

Segall, G., Birnbaum, D., Deeb, I., & Diesendruck, G. (2015). The intergenerational transmission of ethnic essentialism: *How* parents talk counts the most. *Developmental Science*, *18*, 543–555. doi:10.1111/desc.12235

Smiley, P. A., & Dweck, C. S. (1994). Individual differences in achievement goals among young children. *Child Development*, *65*, 1723–1743. doi:10.2307/1131290

Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology*, *39*, 456–467. doi:10.1016/S0022-1031(03)00019-2

Waxman, S. (2010). Names will never hurt me? Naming and the development of racial and gender categories in preschool-aged children. *European Journal of Social Psychology*, *40*, 593–610. doi:10.1002/ejsp.732

Weinburgh, M. (1995). Gender differences in student attitudes toward science: A meta-analysis of the literature from 1970 to 1991. *Journal of Research in Science Teaching*, *32*, 387–398. doi:10.1002/tea.3660320407

Zhai, J., Jocz, J. A., & Tan, A.-L (2014). 'Am I like a scientist?': Primary children's images of doing science in school. *International Journal of Science Education*, *38*, 553–576. doi:10.1080/09500693.2013.791958