There must be another way! Girls are disadvantaged when divesting from inaccurate teaching is required

Mia Radovanovic (m.radovanovic@mail.utoronto.ca)
Ece Yucer (ece.yucer@mail.utoronto.ca)
Jessica A. Sommerville (jessica.sommerville@utoronto.ca)
Department of Psychology, University of Toronto

Abstract

While research has documented that children can compensate for overt cues to teaching inefficacy through exploration of novel solutions, an important question is whether children *use* exploration to detect inefficacy. Further, to move beyond ineffective teaching, learners must prioritize their own ideas. Thus, girls could be disadvantaged due to a greater emphasis on people-pleasing in their socialization. We tested 7- to 10-year-olds using a novel, video-game paradigm. Children were shown ineffective instruction but could only discover its inefficacy by independently attempting the solution. Children generally attempted the taught solution successfully and rationally traded-off between instruction and exploration. However, gender differences emerged in exploration, solving, and learning even after controlling for video game experience and teacher gender. These results have important implications, as girls may have a greater need to move beyond ineffective teaching when exposed to sexist content or beliefs.

Keywords: child development; pedagogy; exploration; gender differences; informant reliability; resource rationality

Introduction

From early in life, teaching is an integral tool for learning. Teaching allows for efficient transfer of knowledge through its strong communicative cues, like eye gaze and gesture (Csibra & Gergely, 2006, 2009). This is particularly true when content is difficult or impossible to learn for children, for example, when solving logical problems or computer programming problems (Mayer, 2004). Because of its importance, humans also make assumptions about teaching. For example, children assume that teaching is complete and will explore a novel object less after receiving pedagogical instruction from a seemingly knowledgeable teacher, compared to when observing a naïve learner (Bonawitz et al., 2011). Likewise, over-imitation studies have demonstrated that when a clear instrumental purpose for an action is missing, humans often assume a cultural or social purpose to the action and nonetheless imitate it faithfully (Hoehl et al., 2019). Together, this work elucidates rational and efficient adaptations to optimize the acquisition of cultural and instrumental knowledge. However, how does this process become disrupted when teaching is ineffective or inaccurate?

Teaching is not the only way children acquire information: children are also adept explorers. Even young children can utilize exploration to gain causal information (Schulz & Bonawitz, 2007; Schulz & Gopnik, 2004; Sim & Zu, 2017), and as children get older, they are able to optimize their exploratory actions to maximize information gain (Cook, Goodman, & Schulz, 2011; Ruggeri et al., 2019). These skills are especially evident in environments where search is emphasized. For instance, Schulz et al. (2018) demonstrated that children's problem-solving behaviors focused on exploring to reduce uncertainty, while adults tend to maximize rewards instead. Thus, children are quite capable of learning things on their own in the absence of teaching.

Moreover, children are able to adapt in the face of ineffective teaching when provided with explicit cues to evaluate the efficacy of an informant, sometimes by prioritizing more reliable informants and sometimes by increasing independent exploration. For instance, 4- and 5-year-old children are consistently able to choose between reliable and unreliable informants in an object-labeling task, choosing to learn novel labels from previously reliable informants (Corriveau, Pickard, & Harris, 2011). Similarly, these adaptations translate to exploratory behavior. For example, when a teacher provides a counterintuitive claim (e.g., that the largest of a set of objects is the lightest), elementary-aged children will explore the objects for longer than if provided with an intuitive claim (Ronfard, Chen, & Harris, 2018). Likewise, 6- and 7-year-olds will explore an object more after receiving instruction if an informant had previously been under-informative relative to completely informative (Gweon et al., 2014). Together, this evidence suggests that children are able to use explicit cues to identify ineffective teaching and that they trade off learning from others and learning from exploration to optimize learning.

However, children are not always provided with overt cues of a teacher's efficacy to evaluate teaching. This is especially true in real-world classrooms, where political agendas and stereotypes often get filtered into the education system. For instance, generational inequalities are transmitted through curricular omissions and erasures (King, 2017; Solomon, 2002; Yosso, 2005) and explicitly incorrect information is even present in the "hard sciences" for controversial topics like evolution or sexual health (Griffith & Brem, 2004; Fuller et al., 2021). In these cases, some teachers are unlikely to provide overt cues to teaching inefficacy because they themselves were educated in these traditions and have not always been trained to appreciate alternative frameworks (Bartolomé, 2004). Thus, in the absence of overt cues, it is worth interrogating children's ability to identify inaccuracies *through* their independent exploration rather than simply to compensate for inaccuracies *with* exploration.

This ability to identify ineffective teaching may also be particularly important but complicated at the intersection of

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society*.

exploration and gender. This is to say that in some cases, an ability to identify ineffective teaching will be especially important for girls and gender minorities. For instance, when children are provided with inaccurate sexual and reproductive education, girls disproportionately suffer the harms (e.g., Madkan et al., 2006). Yet, sexual education is largely inaccurate in the United States (Guttmacher Institute, 2022). Thus, in cases in which information pertains to girls' experiences of marginalization, identifying inaccuracies may be particularly important.

Simultaneously, female socialization in North America may make girls particularly unlikely to ignore teaching due to people-pleasing pressure. Mickelson's (1989) sex-role socialization hypothesis proposes that girls may be disadvantaged through socialization to obey authority figures and "be good." For instance, from a young age, greater pressure is placed on girls than boys to feel responsible for others' emotions (Zahn-Waxler, Cole, & Barrett, 1991) and to consider others' needs in their decisions (Jordan et al., 1991; Letendre, 2007). Thus, in situations where girls are provided inaccurate teaching, they may hesitate to explore their own ideas out of fear of hurting the feelings of the teacher. Likewise, prioritizing one's own exploration in the face of ineffective teaching requires the learner to be willing to stand out and innovate. Yet, relative to men, women are likely to avoid self-promoting for fear of repercussions (Brescoll, 2011; Daubman, Heatherington, & Ahn, 1992; Moss-Racusin & Rudman, 2010), and associations between girls, conformity, and obedience are especially evident in classrooms. For instance, teachers have rated boys as more non-conforming than girls (Gralewski & Karwowski, 2013), and they are more likely to praise girls' compliance (Jones & Myhill, 2004), while highlighting boys' non-conformity and independence (Gralewski, 2019). Thus, even if girls are able to identify inaccuracies, classroom settings may discourage them from prioritizing their own actions.

Thus, the present study sought to understand whether children could evaluate ineffective teaching in the absence of overt cues and, importantly, whether there are differences in girls' and boys' ability to calibrate to this teaching. As compensatory exploratory behavior can be observed in the elementary years (Gweon et al., 2014; Ronfard et al., 2018), and gender stereotypes emerge at a similar time (Bian, Leslie, & Cimpian, 2017; Lei et al., 2019), 7- to 10-year-olds were introduced to a novel search task: a platforming video game. They received instructions on how to solve the game then played a test game in which the taught solution did not work. However, children were not provided with any information to suggest the solution was ineffective. Instead, they had to discover its inefficacy by attempting the taught solution. Our first goal was to evaluate how children as a group responded to this situation. Given their broad exploratory skills, we hypothesized that children would successfully attempt the taught solution, and after attempting, children who prioritized exploring would succeed at greater rates. Our second goal was to evaluate differences between boys and girls. We did not expect to find differences in boys' and girls' attempting behaviors (and if anything, girls may be more likely to attempt the solution given that girls generally outperform boys in school; Duckworth & Seligman, 2006; Matthews, Ponitz, & Morrison, 2009; Orr, 2011) but that girls would struggle when required to divest from teaching, exploring less than boys. Consequently, we hypothesized that girls would be less likely to succeed and learn about the game than boys.

Methods

Participants

Our sample consisted of 150 7- to 10-year-olds (n = 150, M_{AGE} = 9.02 years, 76 girls, 74 boys) recruited from a database of families that had previously volunteered to participate in research. Sample size was determined utilizing *a priori* power analyses of preliminary data (≥ 80% power for all hypothesized effects). Parents identified their children as white (n = 57), multiracial (n = 33), East Asian (n = 24), South Asian (n = 10), Arab (n = 4), Southeast Asian (n = 3), Latin, Central, or South American (n = 2), Native American (n = 1), a race not listed (n = 3), or did not report (n = 13). For each child, at least one parent held at least a high school diploma (n = 8), college degree (n = 10), bachelor's degree (n = 34), master's degree (n = 52), doctoral or professional degree (n = 33), or did not report (n = 13). Thus, our sample was somewhat racially diverse but not educationally diverse. Children were individually tested in a single session using Zoom videoconferencing, lasting roughly 45 minutes. Prior to participating, parents provided informed consent, and children provided assent. Parents received a $5 gift card. Data were excluded from 14 additional children due to parental or sibling interference (n = 6), failure during practice (n = 4), technological issues (n = 2), refusal to comply (n = 1), and experimenter error (n = 1).

Procedure

To understand children's ability to detect incorrect teaching and its influence on exploratory behavior, children played a platforming video game (Figure 1). In the game, participants navigated as a frog character, and the goal to win was to find a trophy. To encourage exploration, collectibles and enemies were spread across the map, and different textures of platforms were utilized. While these elements did not systematically indicate the presence of trophies, they did introduce multiple possibilities that participants could explore. The game had four phases: practice, instruction, test, and post-test. Children learned about the game in the practice and instruction phases, and their problem-solving and reaction to instruction were measured during the test phase.

Practice. To ensure that children had sufficient knowledge of the game controls, participants completed a short practice trial in which they interacted with all of the basic elements of the game: collectibles, enemies, jumping, platform types, and trophies. Participants were given feedback from the experimenter while playing to help them learn the game

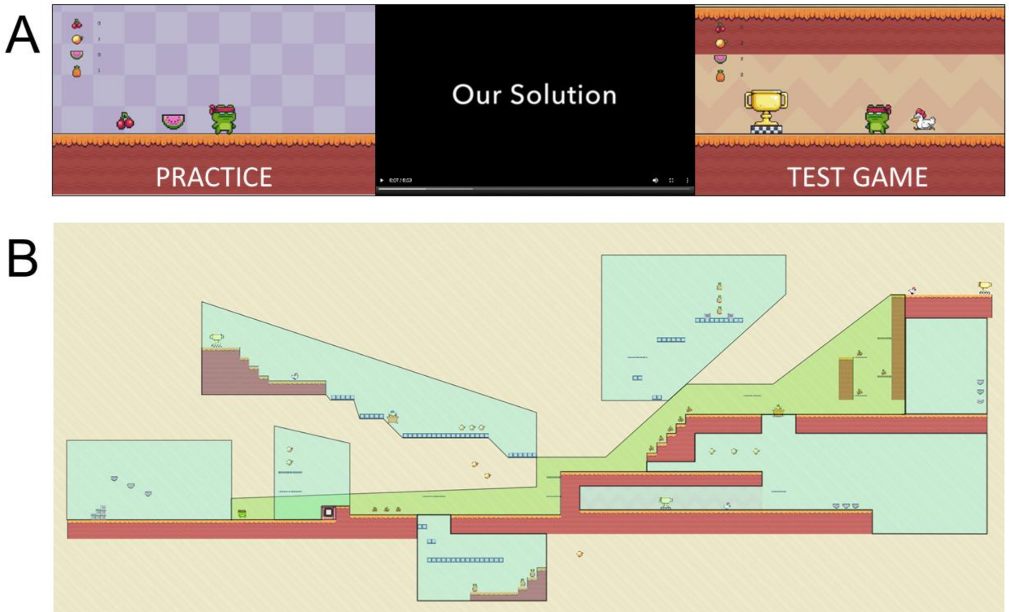

Figure 1: Frog game specifications. (A) Participants began with a practice, viewed an instructional video, and then were given seven minutes to solve a test game, (B) For the purposes of behavioral coding, the game map was divided into exploratory regions, depicted in blue, and imitative regions, depicted in green.

controls, and parents were able to provide instructions but not to complete the practice for their children. Only participants that were able to independently complete the practice within 10 minutes were included in the final sample. Importantly, it was emphasized to participants that the main goal was to find a trophy once they succeeded in the practice game.

Instruction. Following practice, children were shown an instructional video featuring a seemingly knowledgeable adult navigating the test game to find a trophy. Importantly, the game map in the practice game was different in several ways (e.g., background color, structure) from the test game, while the instructional video and test game were ostensibly the same. The adult providing instructions was not visible in the video but rather narrated the instructions while showing a video of themselves playing the frog game. To avoid the possibility that the gender of the teacher disproportionately influenced boys and girls, the voice of the teacher was matched to the gender of the child (i.e., a woman for girls, a man for boys). Prior literature indicates that children are very sensitive to pedagogical instruction (Bonawitz et al., 2011; Csibra & Gergely, 2006, 2009; Hoehl et al., 2019); thus, our instructional video balanced pedagogical and non-pedagogical elements. All experimenter directions emphasized that children would be shown a solution, but the instructions were not deterministic, stating that children "could win the game with any trophy [they] could find."

In the video, the frog character moved on a linear path from the left of the map to the right. Eventually, the frog navigated to six vertical platforms and jumped to a trophy. Critically, there was a slight discrepancy between the video and test games, such that two platforms were missing in the test game, rendering the teaching ineffective. Thus, children were not told that the taught solution was ineffective. They had to attempt the solution to generate evidence of its inefficacy.

Test Game. After instruction, participants completed the test game, which appeared identical to the instructional video except for the missing platforms. While the solution shown in the video did not work, the participants could solve the game by finding either of two alternate trophies in areas branching from the main path. Before the participant played the game, they were told by the experimenter, "For this part of the game, we want to see what you do on your own, so I'm going to do my best not to answer any questions, and we ask that parents do the same. I'll set a timer for seven minutes and let you know when seven minutes pass. Remember, the main goal of the game is to find a trophy, but you can win with any trophy that you find!" Participants were given up to seven minutes to solve the game, which began once the experimenter finished giving directions. During the test game, the experimenter's video was visible, but they did not provide feedback and maintained neutral affect.

Post-Test. After completing the game, participants answered several questions to assess their learning and experience with video games. To assess learning, participants were shown a series of four images portraying various game map areas (some which appeared in the game and some which did not) and asked whether they were part of the game. Thus, the maximum learning score was 4. To generate a control measure of gaming skill, participants were also asked to rate how often they played video games on a 5-point Likert scale (ranging from "never" to "daily").

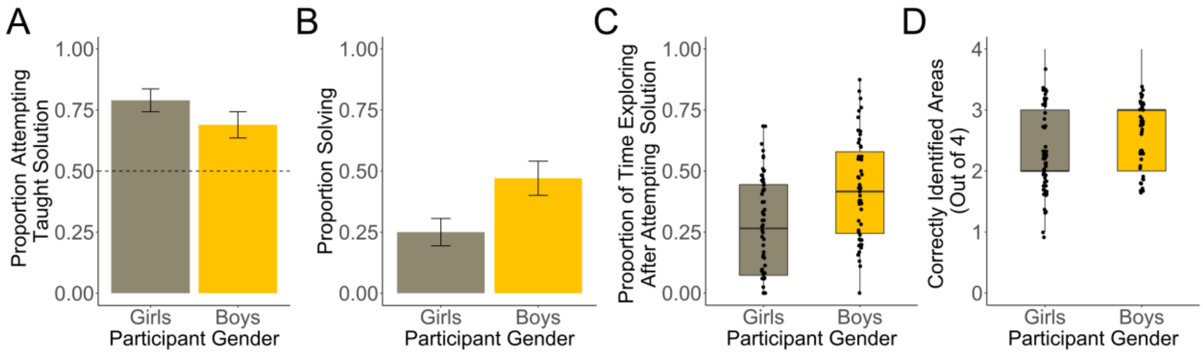

Figure 2: Visualizations of gender and key behaviors and outcome measures: (A) attempting the taught solution, (B) solving, (C) the proportion of time spent exploring after attempting, and (D) performance on post-test learning questions. Error bars represent standard errors of the mean. Dashed lines represent chance performance.

Behavioral Coding

Coding was performed by human raters watching raw footage of testing sessions to classify whether participants attempted the taught solution and solved the game, as well as time spent exploring and imitating.[1] Raters were blind to study hypotheses, but participants images were visible in videos. A primary coder was randomly assigned to each participant. Data generated by primary coders was utilized in analyses. In addition, a secondary coder was randomly assigned such that 50% of all data was double coded (to estimate interrater reliability). Agreement between primary and secondary coders was high across tasks (all *ICC's* > .94, all *p's* < .001).

Attempting the Taught Solution. A participant was considered to have attempted the taught solution if they followed the path in the instructional video and completed one full jump where the platforms were missing. At this point, participants should have evidence that the solution would not work as shown.

Solving. A participant was considered to have solved the game if they found a trophy within seven minutes.

Coding Imitation & Exploration. Classifications were made using the physical location of the frog on the game map (Figure 1). If the frog was in the portion of the game map that was used in the path of the instructional video, their behavior was coded as imitation starting from the frame they entered the area until the frame before they exited. If the frog was in the areas branching off from this path, exploration was instead coded from the frame they entered the area until the last frame before they exited. Thus, these definitions generated two mutually exclusive codes. As the absolute time children played the game varied, proportions were used in analyses. Importantly, the proportion of time engaging in exploration and imitation (respectively) was calculated both before participants attempted the taught solution and after so that behavior could be compared once participants gained evidence of the solution's inefficacy.

Results

Our first goal was to understand how children, as a group, interacted with instruction. Participants were not told that the instruction was ineffective and instead had to utilize their own exploratory actions to discover its inefficacy. Thus, an optimal strategy would be to first exploit instruction to try to solve the game quickly, then increase exploration after evidence of failure. Therefore, we hypothesized that children would initially favor imitation, increasing exploration after attempting the taught solution.

Thus, we first performed a one-sample t-test to compare the proportion of time children spent imitating prior to attempting the taught solution to chance (i.e., 0.5). Indeed, children initially imitated significantly more than they explored (M = 0.72, $t(148)$ = 11.88, p < .001, d =3.19). Likewise, we hypothesized that children would generally be able to attempt the taught solution given children's broad exploratory skills. We performed a binomial test comparing the proportion of children who attempted to chance (i.e., 0.5); 74% of children attempted the taught solution, a rate greater than chance (p < .001). Finally, a paired-sample t-test was performed to compare children's rates of exploration before and after attempting the taught solution. As a group, children explored more after attempting than before (M_{Diff} = .15, $t(110)$ = 6.03, p < .001, d = 0.57). Thus, children generated evidence that the solution was ineffective utilizing only their own exploration, and on average, adjusted their strategies based on this information.

We further sought to understand how exploratory behavior after attempting the solution related to success (i.e., solving and learning). We hypothesized that those who solved would spend a greater proportion of time exploring after attempting the taught solution. A two-sample t-test revealed that children who solved the game spent a significantly greater proportion of time exploring after attempting (M = .53) than children

[1]Outliers were defined *a priori* as points more than 3 SDs from the mean and removed using pairwise deletion.

who did not solve (*M* = .23; *t*(109) = 8.87, *p* < .001, *d* = 1.76). Likewise, we hypothesized that greater exploration after attempting would relate to greater learning. To this end, a Pearson correlation revealed that increases in the proportion of time spent exploring after attempting were modestly, positively associated with learning (*r* = .32, *p* < .001). Thus, increasing exploration was critical to achievement.

Having elaborated general trends, we turned to the role of gender in children's performance. Critically, we did not expect to find initial differences between boys and girls. However, we expected that girls would explore proportionally less after attempting the taught solution because they would feel more bound to the teacher's instructions due to differences in socialization. We further expected this difference to lead to decreased solving and learning in girls. To account for potential confounding factors, we entered age and our measure of gaming skill as control measures in all models predicting performance from gender. To understand the influence of gender on exploration and learning, we constructed linear regression models. As attempting the taught solution and solving were both binary outcomes (i.e., yes or no), logistic regressions with binomial link functions were used to model attempting and solving.

Indeed, we did not observe a significant difference in attempting the solution between boys and girls (*p* > .25). However, we did observe departures between boys and girls after this point. First, gender predicted exploration such that girls explored for a significantly smaller proportion of time after attempting than boys (β = -0.16, *SE* = .04, *t*(107) = -4.24, *p* < .001). Likewise, we observed a main effect of gender on solving such that girls were significantly less likely to solve the game than boys (β = -1.08, *SE* = 0.44, *z*(107) = -2.45, *p* = .01, Odds Ratio = 0.34:1). This implies that girls were about one-third as likely to solve as boys, supporting our hypothesis that girls would succeed less. Likewise, we observed a main effect of gender on learning such that girls answered significantly fewer questions about the game correctly than boys (β = -0.36, *SE* = 0.11, *t*(107) = -3.34, *p* = .001). Together, these results suggest that girls were not worse at the game, as we controlled for video game exposure, and we did not observe differences in attempting behavior. Instead, difficulties began for girls when divesting from the ineffective instruction, as we observed differences in their subsequent behaviors: exploration, solving, and learning.

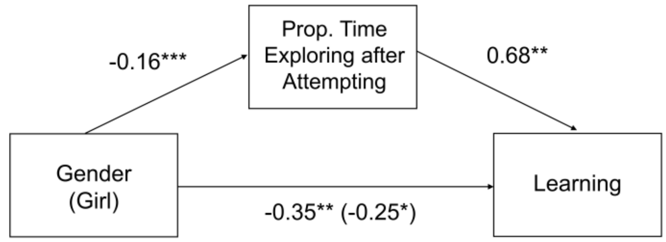

Figure 3: Standardized path estimates presented. The parenthetical value is the direct effect.

Finally, we sought to directly test whether the disparities between boys and girls were due to lower exploration in girls. In particular, we focused our analysis on learning, as decreased learning could lead to the greatest downstream consequences in realistic contexts (relative to solving). We hypothesized that girls exhibited lower learning because the proportion of time spent exploring after attempting the taught solution covaried with gender. This mediational hypothesis was tested with a bootstrap procedure to determine the significance of the indirect effect (Preacher & Hayes, 2004). 5000 bootstrap resamples and a random seed of 65336 were used to estimate the direct, indirect, and total effects using the PROCESS v4.0 (Hayes, 2022) macro. 95% confidence intervals were determined from the bootstrap resamples, and any interval that did not include 0 was considered to be significantly different from 0. This analysis revealed that gender affected learning as a function of its relationship with exploration (*ab* = 0.11, *SE* = 0.05, 95% CI [0.02, 0.22]). However, there was still a direct effect of gender on learning when the indirect path through exploration was considered (*c*' = 0.25, SE = 0.12, 95% CI [0.02, 0.48]). Thus, the relationship between gender and learning is partially accounted for by the co-occurrence of gender and decreased

Table 1: Models predicting key behaviors and outcome measures.

	Attempting Taught Solution		Solving Game	
Predictors	β (SE)	OR [95% CI]	β (SE)	OR [95% CI]
Intercept	0.76 (1.55)	2.13 [0.10 – 44.52]	-6.01 (1.91)**	0.002 [<.001, 0.10]
Age	0.13 (0.16)	1.14 [0.83 – 1.57]	0.47 (0.19)*	1.61 [1.10, 2.34]
Video Game Experience	-0.27 (0.18)	0.77 [0.54 – 1.08]	0.39 (0.20)†	1.48 [1.00, 2.21]
Gender (Girl)	0.44 (0.38)	1.55 [0.73 – 3.29]	-1.08 (0.44)*	0.34 [0.14, 0.81]
	Proportion of Time Exploring		Learning	
Predictors	β (SE)	[95% CI]	β (SE)	[95% CI]
Intercept	-0.20 (0.17)	[-0.50, 0.09]	1.33 (0.43)**	[0.49, 2.17]
Age	0.06 (0.02)***	[0.03, 0.09]	0.13 (0.05)**	[0.04, 0.22]
Video Game Experience	0.02 (0.01)	[-0.005, 0.05]	0.03 (0.04)	[-0.06, 0.11]
Gender (Girl)	-0.16 (0.04)***	[-0.23, -0.09]	-0.36 (0.11)**	[-0.57, -0.15]

exploration after attempting the solution, but gender still predicted decreased learning, suggesting additional pathways may account for girls' learning differences (Figure 3).

Discussion

This paper's primary conceptual objective was to understand how elementary-aged children engaged with ineffective teaching in the absence of explicit cues to ascertain efficacy. To this end, we designed a video game that required participants to evaluate the efficacy of a taught solution using only their own exploratory actions. Our findings revealed that overall, children calibrated to the evidence they generated and effectively balanced between learning from instruction and exploration. However, significant and robust gender differences emerged once children had to divest from the taught solution and explore their own solutions.

By creating a context in which children had to utilize their own actions to ascertain the quality of teaching, we observed several optimized behaviors. As instruction is generally correct and efficient (e.g., Kirschner, Sweller, & Clark, 2006; Stockard et al., 2018), it was reasonable for children to first utilize instruction to try to win the game. This was reflected in children's high initial rates of imitation and high rates of attempting the taught solution. Once children attempted the solution, they generated evidence that the solution was ineffective, which prompted behavioral change in the form of increased exploration. Likewise, children who adjusted to this evidence were more likely to solve and learn.

These findings are consistent with prior work pertaining to children's resource rationality. That is to say; humans seek to optimize effort expenditure and expect other agents to do so as well (Jara-Ettinger et al., 2016). Prior work has demonstrated that children will balance the effort they exert, prioritizing exploration over imitation when a solution does not work (Solby, Radovanovic, & Sommerville, 2021), balancing problem-solving strategies (Lucca, Horton, & Sommerville, 2020), and persisting less when they know teaching is available (Rett & Walker, 2020). The results elaborated here are consistent with these trends, demonstrating that children first sought to exploit teaching under the pretense that it was correct and efficient and generally divested from this teaching when it garnered diminishing returns and continued to fail.

On the other hand, while we did not observe significant differences in boys' and girls' attempts to use the taught solution, differences between boys and girls emerged at the point at which children needed to divest from instruction. Specifically, girls explored for a smaller proportion of time relative to boys after attempting the taught solution, were less likely to solve the game, and answered fewer learning questions correctly. High rates of attempting the taught solution suggest that girls did not underperform due to a mere lack of skill, as does the fact that we controlled for video game experience in our models. Thus, our results are instead consistent with work on girls' socialization which suggests that girls are more likely to people-please than boys (e.g., Mickelson, 1989; Zahn-Waxler et al., 1991). To this end, we observed that girls' decreased learning was mediated by their decreased exploration, suggesting that they learned less specifically because they did not divest from instruction even after they generated evidence of its inefficacy.

Simultaneously, even when the indirect pathway between gender and exploration was accounted for, gender predicted learning such that girls learned less than boys. Thus, an open question remains regarding the remaining pathways between gender and learning. While we controlled for video game experience, the game utilized was stereotypically "masculine." For this reason, girls may have been sensitive to expectations that they would underperform on the task relative to boys (e.g., Kaye & Pennington, 2016). Additional work should seek to understand whether these effects replicate across contexts that are stereotypically "feminine."

Moreover, the paradigm utilized introduces several future directions. For instance, the learning questions we utilized were closely connected to exploration of our specific game (i.e., knowledge of the game map). However, in real-world contexts, knowledge is often hierarchical and builds on itself. Thus, an ability to generalize learned information to new contexts is especially important (Fiorella & Mayer, 2016). Future work would benefit from assessing learning in applied contexts (e.g., mathematics) with questions focusing on rule-learning and generalization, in addition to the concrete and self-contained learning assessed here (e.g., DeCaro & Rittle-Johnson, 2012). Importantly, the type of ineffective teaching we aimed to target (i.e., wherein the teacher themselves is unaware that the teaching is ineffective) typically pertains to structural knowledge relating to historical oppression (e.g., Bartolomé, 2004). As this was the first known psychological exploration of such teaching, we purposely utilized a conceptual paradigm stripped of real-world elements. However, marginalization is apt to compound the difficulties children face when interacting with this type of teaching, and future work should study contexts, which are more realistic and consider that girls may respond differently across intersections of race and gender (Mickelson, 1989).

Of course, there are many situations in which focusing deeply on instruction and persisting through difficulty would be beneficial. Particularly when information is difficult or impossible for children to extract independently, expert instruction can be more effective for learning than children's own exploration (Mayer, 2004). In these contexts, girls may be advantaged for the same reasons they are disadvantaged in the context of our study. Indeed, girls generally outperform boys academically (Duckworth et al., 2006; Matthews et al., 2009; Orr, 2011). Likewise, exploring novel solutions can be time-consuming and costly. Thus, while populations benefit from having explorers, exploring is not always the most adaptive to the individual (Wisdom & Goldstone, 2011). Thus, our contention is not that children should always divest from teaching and explore their own solutions, but rather that we ought to understand the factors that limit children's ability to move beyond ineffective teaching in particular and the factors that create disparities between children in these contexts.

Acknowledgements

We thank the researchers of the Toronto Early Cognition Lab at the University of Toronto for their help in participant recruitment, data collection, and coding. In particular, we thank Denise Arefhaghi, Christie Lai, Yang (Leona) Liu, Aafiya Somani, Mahima Tirunelveli Santhakumar, Tiffany Tse, Justine Vorvis, and Grace Zheng. We also thank Arnav Verma for his assistance creating an online platform to host our video game for data collection. This work was supported by grants from the Canadian Foundation for Innovation and Ontario Research Fund titled "The Origins of Social Learning and Behavior" awarded to J.A.S. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

References

Bartolomé, L. I. (Winter 2004). Critical pedagogy and teacher education: Radicalizing prospective teachers. *Teacher Education Quarterly*, *31*(1), 97-122.

Bian, L., Leslie, S.-J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, *355*(6323), 389-391. https://doi.org/10.1126/science.aah6524

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, *120*, 322-330. https://doi.org/10.1016/j.cognition.2010.10.001

Brescoll, V. L. (2011). Who takes the floor and why. *Administrative Science Quarterly*, *56*(4), 622-641. https://doi.org/10.1177%2F0001839212439994

Cook, C., Goodman, N. D., & Schulz, L. E. (2011). Where science starts: Spontaneous experiments in preschoolers' exploratory play. *Cognition*, *120*(3), 341-349. https://doi.org/10.1016/j.cognition.2011.03.003

Corriveau, K. H., Pickard, K., & Harris, P. L. (2010). Preschoolers trust particular informants when learning new names and new morphological forms. *British Journal of Developmental Psychology*, *29*(1), 46-63. https://doi.org/10.1348/2044-835X.002009

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In Y. Munakata & M. H. Johnson (Eds.), *Processes of change in brain and cognitive development* (Vol. XXI, pp. 249-274). Oxford University Press.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148-153. https://doi.org/10.1016/j.tics.2009.01.005

Daubman, K. A., Heatherington, L., & Ahn, A. (1992). Gender and the self-presentation of academic achievement. *Sex Roles*, *27*(3-4), 187-204. https://doi.org/10.1007/BF00290017

Decaro, M. S., & Rittle-Johnson, B. (2012). Exploring mathematics problems prepares children to learn from instruction. *Journal of Experimental Child Psychology*, *113*(4), 552-568. https://doi.org/10.1016/j.jecp.2012.06.009

Duckworth, A. L., & Seligman, M. E. P. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology*, *98*(1), 198-208. https://doi.org/10.1037/0022-0663.98.1.198

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, *28*(4), 717-741. https://doi.org/10.1007/s10648-015-9348-9

Fuller, K., Clonan-Roy, K., Goncy, E., & Naser, S. (2021). The omission and minimisation of sexual decision-making skills in US sex education textbooks. *Sex Education*, 1-15. https://doi.org/10.1080/14681811.2021.1949974

Gralewski, J. (2019). Teachers' beliefs about creative students' characteristics: A qualitative study. *Thinking Skills and Creativity,* 31, 138-155. https://doi.org/10.1016/j.tsc.2018.11.008

Gralewski, J., & Karwowski, M. (2013). Polite girls and creative boys? Students' gender moderates accuracy of teachers' ratings of creativity. The Journal of Creative Behavior, 47(4), 290-304. https://doi.org/10.1002/jocb.36

Griffith, J. A., & Brem, S. K. (2004). Teaching evolutionary biology: Pressures, stress, and coping. *Journal of Research in Science Teaching*, *41*(8), 791-809. https://doi.org/10.1002/tea.20027

Guttmacher Institute. (2022, January 1). *Sex and HIV education*. https://www.guttmacher.org/state-policy/explore/sex-and-hiv-education

Gweon, H., Pelton, H., Konopka, J. A., & Schulz, L. E. (2014). Sins of omission: Children selectively explore when teachers are under-informative. *Cognition*, *132*(3), 335-341. https://doi.org/10.1016/j.cognition.2014.04.013

Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (3rd ed.). The Guilford Press.

Hoehl, S., Keupp, S., Schleihauf, H., Mcguigan, N., Buttelmann, D., & Whiten, A. (2019). 'Over-imitation': A review and appraisal of a decade of research. *Developmental Review*, *51*, 90-108. https://doi.org/10.1016/j.dr.2018.12.002

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences*, *20*(8), 589-604. https://doi.org/10.1016/j.tics.2016.05.011

Jones, S., & Myhill, D. (2004). 'Troublesome boys' and 'compliant girls': Gender identity and perceptions of achievement and underachievement. *British Journal of Sociology of Education*, 25(5), 547-561. https://doi.org/10.1080/0142569042000252044

Jordan, J. V., Kaplan, A. G., Miller, J. B., Stiver, I. P., & Surrey, J. L. (1991). *Women's growth in connection: Writings from the Stone Center*. Guilford Press.

Kaye, L. K., & Pennington, C. R. (2016). "Girls can't play": The effects of stereotype threat on females' gaming performance. *Computers in Human Behavior*, *59*, 202-209. https://doi.org/10.1016/j.chb.2016.02.020

King, L. J. (2017). The status of Black History in U.S. schools and society. *Social Education*, *8*(1), 14-18.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75-86. https://doi.org/10.1207/s15326985ep4102_1

Letendre, J. (2007). "Sugar and spice but not always nice": Gender socialization and its impact on development and maintenance of aggression in adolescent girls. *Child and Adolescent Social Work Journal*, *24*(4), 353-368. https://doi.org/10.1007/s10560-007-0088-7

Lucca, K., Horton, R., & Sommerville, J. A. (2020). Infants rationally decide when and how to deploy effort. *Nature Human Behaviour, 4*(4), 372-379. https://doi.org/10.1038/s41562-019-0814-0

Madkan, V. K., Giancola, A. A., Sra, K. K., & Tyring, S. K. (2006). Sex differences in the transmission, prevention, and disease manifestations of sexually transmitted diseases. *Archives of Dermatology*, *142*(3). https://doi.org/10.1001/archderm.142.3.365

Matthews, J. S., Ponitz, C. C., & Morrison, F. J. (2009). Early gender differences in self-regulation and academic achievement. *Journal of Educational Psychology*, *101*(3), 689-704. https://doi.org/10.1037/a0014240

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist*, *59*(1), 14-19. https://doi.org/10.1037/0003-066x.59.1.14

Mickelson, R. A. (1989). Why does jane read and write so well? The anomaly of women's achievement. *Sociology of Education*, *62*(1), 47. https://doi.org/10.2307/2112823

Moss-Racusin, C. A., & Rudman, L. A. (2010). Disruptions in women's self-promotion: The backlash avoidance model. *Psychology of Women Quarterly*, *34*(2), 186-202. https://doi.org/10.1111%2Fj.1471-6402.2010.01561.x

Orr, A. J. (2011). Gendered capital: Childhood socialization and the "Boy crisis" in education. *Sex Roles*, *65*(3-4), 271-284. https://doi.org/10.1007/s11199-011-0016-3

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 717-731. https://doi.org/10.3758/BF03206553

Rett, A., & Walker, C. M. (2020). Knowing when to quit: Children consider access to solutions when deciding whether to persist. *Proceedings of the 42st Annual Conference of the Cognitive Science Society*.

Ronfard, S., Chen, E. E., & Harris, P. L. (2018). The emergence of the empirical stance: Children's testing of counterintuitive claims. *Developmental Psychology*, *54*(3), 482-493. https://doi.org/10.1037/dev0000455

Ruggeri, A., Swaboda, N., Sim, Z. L., & Gopnik, A. (2019). Shake it baby, but only when needed: Preschoolers adapt their exploratory strategies to the information structure of the task. *Cognition*, *193*, 104013. https://doi.org/10.1016/j.cognition.2019.104013

Schulz, E., Wu, C. M., Ruggeri, A., & Meder, B. (2019). Searching for rewards like a child means less generalization and more directed exploration. *Psychological Science*, *30*(11), 1561-1572. https://doi.org/10.1177/0956797619863663

Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: Preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology*, *43*(4), 1045-1050. https://doi.org/10.1037/0012-1649.43.4.1045

Schulz, L. E., & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology*, *40*(2), 162-176. https://doi.org/10.1037/0012-1649.40.2.162

Sim, Z. L., & Xu, F. (2017). Learning higher-order generalizations through free play: Evidence from 2- and 3-year-old children. *Developmental Psychology, 53*(4), 642-651. https://doi.org/10.1037/dev0000278

Solby, H., Radovanovic, M., & Sommerville, J. A. (2021). A new look at infant problem-solving: Using DeepLabCut to investigate exploratory problem-solving approaches. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.705108

Solomon, R. P. (2002). School leaders and antiracism: Overcoming pedagogical and political obstacles. *Journal of School Leadership*, *12*(2), 174-197. https://doi.org/10.1177%2F105268460201200205

Stockard, J., Wood, T. W., Coughlin, C., & Rasplica Khoury, C. (2018). The effectiveness of direct instruction curricula: A meta-analysis of a half century of research. *Review of Educational Research*, *88*(4), 479-507. https://doi.org/10.3102/0034654317751919

Wisdom, T. M., & Goldstone, R. L. (2011). Innovation, imitation, and problem solving in a networked group. *Nonlinear Dynamics-Psychology and Life Sciences*, *15*(2), 229. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.299.5858&rep=rep1&type=pdf

Yosso, T. J. (2005). Whose culture has capital? A critical race theory discussion of community cultural wealth. *Race Ethnicity and Education*, *8*(1), 69-91. https://doi.org/10.1080/1361332052000341006

Zahn-Waxler, C., Cole, P. M., & Barrett, K. C. (1991). Guilt and empathy: Sex differences and implications for the development of depression. In J. Garber & K. A. Dodge (Authors), *The development of emotion regulation and dysregulation* (pp. 243-272). Cambridge University Press.