Problems in Pedagogy: Teaching Deficits in an In-Lab Sample

Victor C. Hunt

Advised by Julian Jara-Ettinger, Assistant Professor of Psychology

A thesis submitted in partial fulfillment of the requirements for the degree of

Bachelor of Science in Cognitive Science at Yale University

New Haven, Connecticut

April 20, 2018

Table of Contents

Abstract

A ubiquitous facet of human life, pedagogy is crucial to the success and growth of our species (Tomasello, 2009). Despite its significance, humans struggle with pedagogy well into adulthood. Prior work has begun investigating situations where teaching breaks down, in order to analyze the sources of teachers' failures. In a recent set of experiments, Aboody, Velez-Ginorio, Santos & Jara-Ettinger (under review) found that teachers struggle because they fail to understand the kinds of hypotheses naïve learners consider likely. Therefore, teachers did not provide enough evidence for naïve participants to successfully learn from. This study has one important limitation, however: in contrast to other recent pedagogy research, data were collected entirely online. It is possible that the less naturalistic setting (or lack of teacher motivation) caused teachers to perform poorly. In the current work, we replicate Aboody et al.'s initial experiment with a lab-based sample of teachers. We successfully replicate the results of Aboody et al. (under review), therefore demonstrating that teachers failures in the task cannot be explained by the sample or online methods utilized.

Introduction

Humans spend a considerable amount of time learning from each other, and this has been critical to our survival and success. While we share many traits with our closest evolutionary relatives, we are by far the most prolific and effective social learners (Sherwood et. al., 2008; Tomasello, 2009). This ability may be one reason our species has been so successful (Bandura, 1977): Despite minimal changes to our biological evolution, our ability to relate to other minds in complex ways has enabled us to make significant progress over the past several millennia (Sherwood et. al., 2008; Tomasello, 2009). We are able to “ratchet” our knowledge over generations (Tomasello, 2009), devoting less time to rediscovering ideas, and more time building on top of previous knowledge.

Teaching is a fairly unique human trait, although several other species have also been observed engaging in “teaching-like” activities (Skerry, Lambert, Powell & McAuliffe, 2013). It is possible that we are such prolific teachers and social learners because of a sophisticated ability to reason about other people’s knowledge and beliefs, called Theory of Mind (Ghrear, Birch, and Bernstein, 2016). Theory of Mind allows us to "reason about our own and others' mental states" (Ghrear, Birch, and Bernstein, 2016), and therefore it may facilitate our social learning and teaching abilities.

Although young children are effective social learners (Gweon, H., Shafto, P., & Schulz, L., 2014; Bridgers, S., Jara-Ettinger, J., & Gweon, H., 2016) and teachers (Shafto, Goodman & Griffiths, 2014; Rhodes, gelman & brickman, 2010), even adults struggle to teach effectively in some situations (Chi, Siler, Jeong, Yamauchi & Hausmann, 2001; Chi, Siler & Jeong, 2004), especially in unconstrained, naturalistic interactions (Chi, Siler, Jeong, Yamauchi & Hausmann 2001; Chi, Siler & Jeong, 2004).

Why do adults teach well in some tasks, but struggle in others? In order to reconcile these findings, Aboody et. al. (under review) investigated why teaching sometimes breaks down. They proposed three possible hypotheses that could account for participants' struggles in some tasks, and success in others. The first possibility is that the naturalistic tasks typically used to measure how well we teach are too complex. These tasks may present too many teaching options, making it difficult for teachers to decide what to teach. A second possibility is that teachers have difficulty representing specific hypotheses that learners would consider plausible. As tasks become more complex, it becomes more challenging to conceptualize what are the most plausible ideas from a naive perspective. Finally, the third possibility was that both of these difficulties affect how we teach in complicated tasks.

Ultimately, Aboody et. al. (under review) found that while participants in teaching tasks made rational choices in attempts to communicate information, they failed to infer the breadth of hypotheses learners considered, and how much data learners would require to reject these incorrect hypotheses. Put differently, teachers overestimated what learners knew, and therefore didn't provide enough information. This research, however, was conducted online, on Amazon Mechanical Turk. While it is possible that teachers struggled due to legitimate reasons, it is also possible that these online workers were simply distracted or unmotivated, and struggled due to these factors. Or, perhaps online teaching tasks are simply too artificial, and therefore do not accurately reveal adults' capacities. If any of these possibilities are true, Aboody et al.'s original results may not generalize to a more naturalistic version of the task, conducted with a motivated, undistracted in-lab sample of participants. In our first experiment, we test these possibilities by replicating Aboody et al.'s first experiment with an in-lab sample of learners.

Experiment

This experiment consisted of a two-part, between-subjects design. First, participants took part in a teaching task. These participants learned how to activate a new machine, and then selected examples to teach naïve learners. Next, these examples were shown to naïve learners, and their understanding of the machine was assessed.

Participants in the learning task were first taught how the machine worked in general terms (e.g., that when it activated, it played music), without being told specifically what would activate the machine. Participants were then shown a set of examples (generated by one of the teachers), and asked to deduce what rule activated the machine. As in the original study, teachers' performance was derived from the success rate of the participant learners that were assigned to them. If a teacher's learners failed to learn how the machine worked, and thus performed poorly, that teacher was categorized as having performed poorly (and vice versa). Aboody et al (under review) found that teachers did not perform optimally (e.g., many learners struggled to figure out how the machine worked from teachers' demonstrations).

It is possible that learners in the original task struggled only because the teachers recruited from Amazon Mechanical Turk were distracted when producing their examples, or unmotivated to teach. Or, perhaps the task was so artificial that participants couldn't teach effectively. In the current experiment, we test this possibility by utilizing an in-lab sample of participants for our teachers, and making our task more naturalistic. If the original findings replicate, this will provide evidence that the original effect observed is indeed robust.

Methods

Participants

Adult participants (n = 20) were recruited from the Spring 2018 Yale University Introduction to Psychology subject pool to take part in the teaching task. A second group of participants (n = 200; mean age = 35.46 years; range = 19–68 years) were recruited from Amazon's Mechanical Turk platform to take part in the learner task.

Stimuli

Stimuli consisted of five 1.5 x1.5x1.5 inch (3.38 $inch^3$) blocks painted red, orange, blue, purple, and green. Each block was labeled with a black letter on one face: A, B, C, D, and E, respectively. In addition, a music machine was constructed from a 4" by 6" by 10" box. The machine activated (played music) only when two specific blocks (orange "B" and green "E") were on top of the machine together. Neither the order of the blocks, nor the presence or absence of other blocks affected whether the machine activated. Although participants were quite convinced that the blocks caused the machine to activate, the experimenter actually surreptitiously activated the machine when demonstrating its function to participants.[1]

[1] Note: the machine used in the previous study conducted online used light to indicate the machine being activated. The present study uses music as a more salient indication of activation in an in-person setting.

Procedure

Teacher condition.

First, adult participants learned about a new machine called a “music machine”. They were told that the music machine activates when certain blocks are placed on top, and were introduced to five blocks (lettered A-E). Participants were told that the machine activated only when both the orange block (B) and the green block (E) were placed on top. These two blocks caused the machine to activate regardless of their order, or the presence of other blocks. Participants were shown two examples of these blocks activating the machine (one example containing just blocks B and E, and one example containing all the blocks, in a mixed-up order). Participants’ understanding was then assessed in several different ways. Participants were first asked to indicate which blocks activated the machine (multiple choice). Then they were asked to indicate whether the machine was on or off in 3 separate depictions of the machine with blocks on top of it. Any participants who failed any of these four questions were automatically excluded from the study.

Once it was clear that participants understood how the machine worked, teachers were then asked to communicate this information to naïve learners. Teachers could select examples of blocks to demonstrate on top of the machine. Participants understood that learners would see whether or not each demonstration activated the machine. Teachers were asked to provide between a minimum of three to a maximum of twenty unique examples. We assessed teachers’ understanding of their task, and of the learner’s knowledge state; teachers who answered these check questions wrong were corrected.

Next, the experimenter told participants he was turning the machine off (so they did not receive feedback from the machine for each example they gave). Then, participants were able to begin providing examples, by placing block(s) on top of the machine. Participants were asked not to speak or explain themselves during the process, but were told that they would have a chance to explain each example later on. The experimenter indicated when they took a photograph of the demonstration by saying, "Okay," after which the teacher would either provide another example or alert the experimenter that they were satisfied with the examples they had provided.

After participants indicated they were done providing examples, they responded to several post-test questions. They were asked to rate on a Likert scale how confident they were in their demonstrations being sufficient to teaching a naïve learner. Then they were asked whether there were examples they wanted to provide but were unable to and if they felt they had provided any extra examples. Finally, teachers were asked to explain why they had stopped providing examples.

Learner condition.

Participants in the learner condition were introduced to the music machine and the blocks. They were taught only what the machine looked like when it activated (when the machine activated, a music symbol appeared above the top left corner. See Figure 1). Next, learners viewed a teacher's demonstrations. 10 learners were assigned to view each teacher's examples. Learners viewed their teacher's examples in the order that the teacher provided them (see Supplemental Materials). Next, their understanding of the machine was assessed. First, learners were asked to explain qualitatively how the machine worked.

Then, we showed learners every possible way to combine the five blocks (5 blocks = 31 possible combinations). For each combination, learners had to identify whether the machine was on, or off. This provided a thorough picture of what learners thought, and how well they understood the rule.

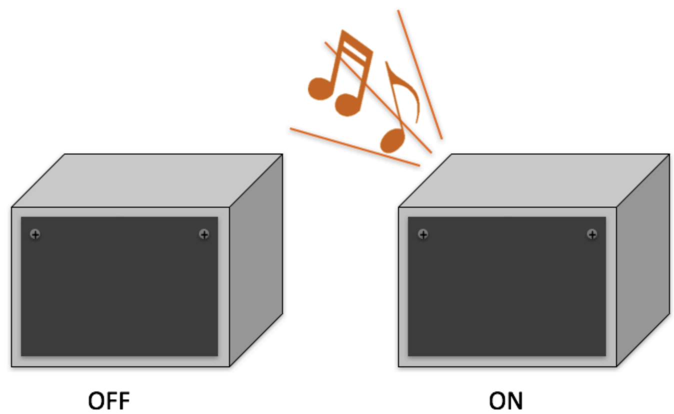


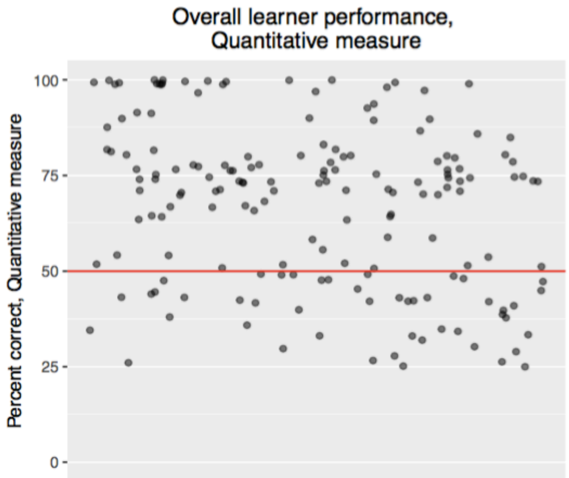
Figure 1. [Learners were taught what the machine looked like when it was turned on or turned off with the images above. When the machine was on it had an image of music notes on the upper left side and when it was off it did not. To see how this looked on images of teacher demonstrations, see 3a-3b in Supplemental Materials]

Results

Participants in the teacher condition provided an average of 7.2 examples (range = 4–15; SD = 3.4). After choosing these examples, teachers were quite confident they would effectively teach a naïve learner about the machine, reporting an average confidence rating of 5.2 on a 7-point scale (range = 2–7; SD = 1.3). These results are comparable to the results of Aboody et. al. (under review) which found that teachers provide an average of 8.2 examples (range = 3-20; SD=4.4) and reported a confidence rating of 6.05 on a 7-point scale (range = 5-7; SD = 0.76).

Although teachers were quite confident in the quality of their demonstrations, this did not translate into effective learning outcomes. A mere 26.5% of learners ($n= 53$) performed at or near ceiling in the quantitative task – in contrast, in the original sample, half of learners performed at or near ceiling (50%, $n=100$). Although many participants struggled to understand exactly how the machine worked, overall participants still performed above chance on our quantitative measure ($M = 72.2\%$, $t(199) = 13.7$, $p < .001$). Even when participants who performed perfectly are excluded from these analyses, participants still performed above chance ($M = 63\%$, $t(150) = 8.5$, $p < .001$).

Perhaps learners struggled, not because of teachers' abilities, but because they just didn't attend to the task. In this case, we would expect all learners to perform equally poorly, regardless of the teacher whose examples they learned from. However, this is not the case: the teacher learners were assigned to could predict learners' quantitative performance (Monte Carlo permutation test, $p = .004$, 10,000 samples). In fact, this relationship is even stronger than that in the original paper (Monte Carlo permutation test, $p = .03$, 10,000 samples). This provides evidence that our results cannot be explained simply by assuming that a randomly distributed subset of participants did not attend to the task.



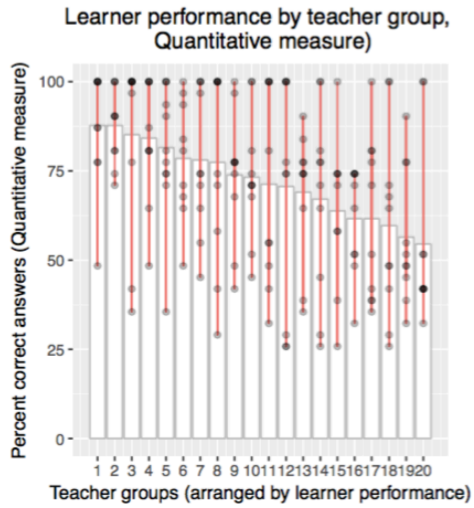


Figure 2a. In this figure, we plot learners' overall performance on the quantitative test questions, as a proportion of the number of correct answers each learner gave. Performance was above chance (indicated by the red line). The x-axis was extended to avoid overplotting and has no relevant value. Figure 2b. In this figure, the bars represent the mean performance of each group of learners (e.g., the 10 learners assigned to each teacher), arranged from highest to lowest mean performance. The red lines show the range of learner performance, and the dots plot learners' actual performance (the darker the points, the more learners they represent).

Discussion

While common, teaching is a difficult task to accomplish effectively. This study found that only a quarter of participants performed close to perfectly, but most did not. Strong learner performance was also not distributed evenly among teachers as some teachers were able to predict how well learners would do, which we show through our Monte Carlo permutation test. This demonstrated that learners' difficulties went beyond challenges attending the task but with the teacher they were assigned. Thus, teacher failed to generate sufficient examples which is consistent with Aboody et. al. (under review).

Aboody et al. (under review) proposed that teachers either struggled deciding what information would be most helpful to share when given many choices or they would choose information that did not accurately represent learners' conceived possibilities. In a two-part study, they ultimately found evidence for the latter, as leaners predicted how to activate the machine more accurately when their hypothesis spaces were constrained than when they were not. Teachers did not fail to give informative data, but failed to consider a wide enough breadth of hypotheses that learners might consider.

A more naturalistic setting allowed teachers to consider more possible considerations for how the machine worked. In fact, some teachers in lab provided examples that online participants could not produce. For example, some teachers turned blocks around to hide the letter to indicate that the blocks' letters did not matter (Figures 4a-4c). Other participants used block position to emphasize the blocks that did activate the machine (see Supplemental). However, although some participants did consider a broad range of potential hypotheses, these participants do not appear to have provided enough evidence to clarify for learners what these examples were intended to communicate. In fact, these examples sometimes misled learners.

"I could not figure out what made the machine turn on. I was paying attention to the examples but they seemed to follow no rhyme or reason," commented one learner. "The way for it to turn on is to have block E and block B next to each other. It does not matter which is on what side, but they must be next to each other," explained another learner, when prompted to explain how the machine worked. This learner *almost* understood how the machine worked – but misunderstood the role that block orientation played. Although teachers clearly anticipated certain non-obvious hypotheses learners might entertain (although failing to anticipate others), teachers also did not seem to provide enough data to truly make their point. For example, turning the blocks around (to demonstrate that the letters don't matter) may rule out certain letter-based hypotheses – but may introduce or support hypotheses about block orientation mattering.

The in-person replication in this study was designed to rule out several important potential confounds with the original study. The first difference is the origin of the subject pool. In Aboody et. al. (under review), teacher participants came from a group of Amazon Mechanical Turk workers around the United States. The teachers in this in-person replication all came from Yale University's 2018 Introduction to Psychology course. In the in-lab replication, experimenters had much greater control over participants' environment. Experimenters could ensure that the students were undistracted, and focused on the task. Furthermore, students in the subject pool are generally quite motivated to help their fellow students, and so may have been more motivated than online participants. Additionally, these two groups may have had different levels of understanding about teaching strategies in general. All these things considered, the results from this study replicated those of Aboody et. al. (under review), suggesting that the findings are consistent across most backgrounds.

Another difference between the two designs was how teachers learned the rule. This difference is the independent variable that was changed between Aboody et. al. (under review) and the present study. Our goal was to make the task more naturalistic and so receiving the instructions in person versus digitally was a part of the design. As mentioned above, the more naturalistic setting allowed the in-person subject pool greater flexibility. Teachers in this study had the opportunity to ask clarifying questions during the task, which may have contributed to their understanding and eventually their example choices. However, the fact teachers continued to struggle even in this more naturalistic setting is noteworthy.

Interestingly, teachers in the online study conducted by Aboody et. al. (under review) provided *more* examples than the teachers in the in-person replication. Rather than being too "artificial," the online task imposed greater constraints that may have led to more effective teachers. With greater restraints, teachers focused on a narrower window of plausible ideas about how the machine activated and worked harder to clarify what learners needed to know about those ideas. Learners also performed well in the online study, adding to teachers' effectiveness.

It is important to note that our results are consistent with previous research on the curse of knowledge. In this work, researchers have found that we overestimate others' knowledge as a function of our own knowledge in a domain. Thus, our own previous knowledge in a subject inhibits our ability to reason about what other minds think about the same topic (Birch & Bloom, 2004; Nickerson, R.S., 1999; Nickerson, R.S., 2001). In the current work, although teachers had a high confidence in their ability, learners were not as successful as teachers predicted. Prior work has found that our own privileged knowledge inhibits our ability to reason about what other people think about the same topic. Additionally, we often do not sufficiently adjust our assumptions about another person's mental state, and thus we fail to accurately infer what someone else does or does not know (Epley & Gilovich, 2001; Furnham & Boo, 2011).

This curse of knowledge can occur across a broad range of knowledge-acquisition modes and stages of human development. Seemingly, once we attain knowledge, regardless of how it is obtained, we become hindered by it. For example, children show clear signs of a curse of knowledge whether information is acquired through testimony, observation, or direct interactions (Bhandari & Barth, 2010). And although we become better at overcoming this curse of knowledge as we age, possessing information still hinders our ability to communicate and teach others into adulthood. In adults, Birch et. al. (2017) demonstrated how either inhibition errors or fluency misattributions are sufficient to create a curse of knowledge. Respectively, these mechanisms describe our difficulty suppressing the relevant knowledge we possess when reasoning about a less informed mind and mistaking the subjective ease with which ideas to come to our own mind as a shared ease (Birch et. al., 2017).

Previous studies have investigated how the curse of knowledge may be impacted by the mode in which the knowledge is gained. Similar to Bhandari & Barth (2010), which looked at how acquiring information either through testimony or observation affected curse of knowledge in children, this study examined whether teaching information either digitally or in-person contributes to teachers' ability to reason about other minds and communicate knowledge with the goal of teaching.

Crucially however, while previous work has produced many descriptions of the curse of knowledge, no quantitative definitions exist. Therefore, it is impossible to predict with precision how performance generalizes across task variations. In our future and ongoing research, we begin to try to formalize our understanding of the curse of knowledge, investigating how we come up with possible hypotheses ourselves, and how we decide which possibilities others may consider.

In our current ongoing work, we introduce participants to a new game, called the "dice game". Combining 2 dies in the right way earns players a point. Participants are not told how the game works; they are simply shown one example of a dice pair that would earn a point. Participants are then asked to come up with possible hypotheses regarding the game's rules. We investigate whether participants come up with different hypotheses when asked to reason for themselves than when they are asked to reason in the third person, and list hypotheses that *another person* might come up with. With this manipulation, we can reveal whether even at baseline, when participants themselves have no privileged knowledge, whether we represent others' minds as different from our own in a systematic way. If this is the case, this work may begin to shed light on how precisely the curse of knowledge operates.

Conclusion

The results of the in-person study reveal that teachers' ability to represent learners' hypothesis spaces does not change in more naturalistic settings. While differences in participant motivation may increase, their performance is no better – and in fact, performance of in-lab teachers was worse in many ways! We find that teachers do take advantage of the naturalistic settings by asking more questions and considering other possible hypothesis spaces only available in more naturalistic settings, but still not enough to accurately teach naïve learners. These results motivate us to look further into the curse of knowledge effects in a follow up to this study, which will uncover whether the same reasoning limitations occur when the curse of knowledge is not present.

Author Contributions

Aboody and Professor Jara-Ettinger designed the study. Hunt collected the data with Aboody. Aboody analyzed the data and produced Figure 2a and 2b. Hunt produced an outline and final draft. Aboody provided feedback on the outline and final draft. Aboody, Hunt, and Jara-Ettinger then worked on drafts of the paper, with small bits of text introduced by each.

Acknowledgements

I would like to express my gratitude to God for getting me to this point and beyond. I would also like to thank Professor Julian Jara-Ettinger, Rosie Aboody, Madison Flowers, and the Computation Development Lab at Yale University. Additionally, I would like to thank Professor Joshua Knobe, Professor Mark Sheskin, and Aaron Resnick for their guidance throughout my time in the Cognitive Science Department. I would also like to thank P.S. 138, Prep for Prep, and Phillips Exeter Academy. Lastly, I would like to thank these individuals without whom, I would not have been able to get to this point and complete this degree: Angela Hunt, Arline Parks, Karen Summers, Eona Summers, Temperance Gaines, Sondra Peeden, Mishael Eques, and Khymaya Perkins – I dedicate this achievement to you all.

Supplemental Materials

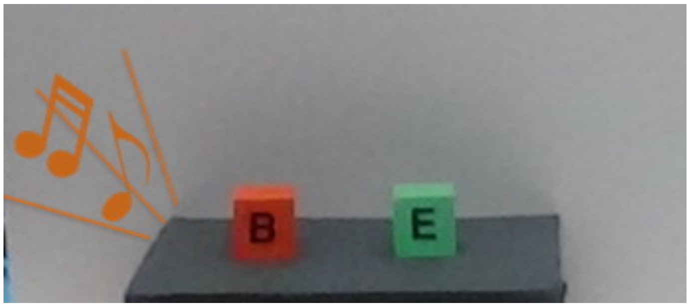


Figure 3a.

Figure 3b.

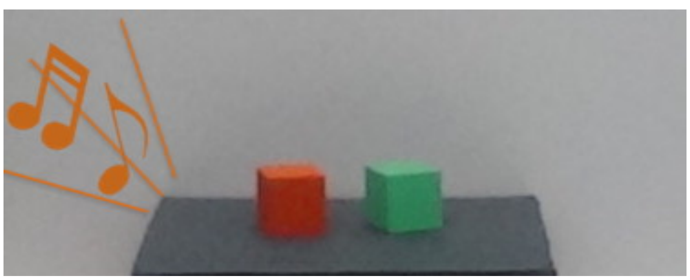


Figure 4a.

Figure 4b.

Figure 4c.

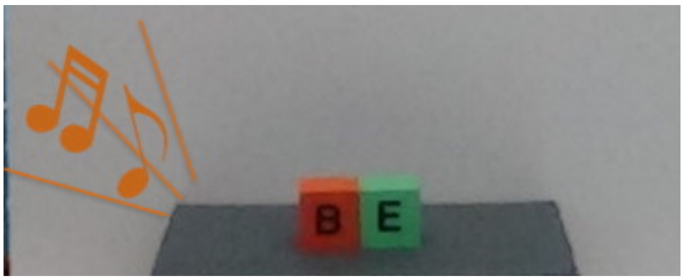


Figure 5a.

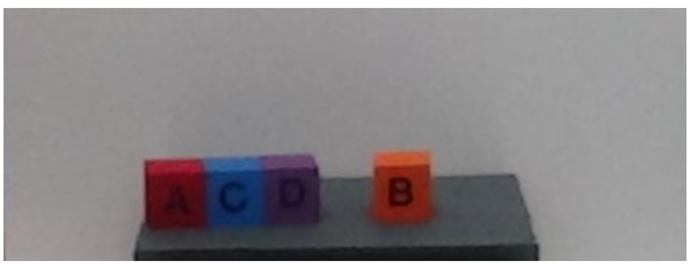


Figure 5b.

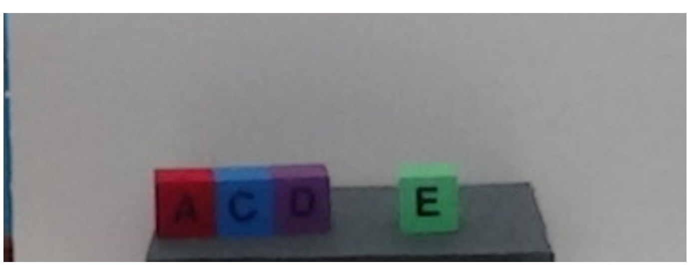


Figure 5c.

References

Aboody, R., Velez-Ginorio, J., Santos, L. R., Jara-Ettinger, J. (under review) When teaching breaks down: Teachers rationally select what information to share, but misrepresent learners' hypothesis spaces. *In Preparation*

Bandura, A. (1977). Social learning theory. Englewood Cliffs, NJ: Prentice-Hall.

Birch, S. A., & Bloom, P. (2004). Understanding childrens and adults limitations in mental state reasoning. *Trends in Cognitive Sciences, 8*(6), 255-260.

Birch, S. A., & Bloom, P. (2007). The Curse of Knowledge in Reasoning About False Beliefs. *Psychological Science, 18*(5), 382-386.

Birch, S. A., Brosseau-Liard, P. E., Haddock, T., & Ghrear, S. E. (2017). A 'curse of knowledge' in the absence of knowledge? People misattribute fluency when judging how common knowledge is among their peers. *Cognition,166*, 447-458.

Bhandari, K., & Barth, H. (2010). Show or tell: Testimony is sufficient to induce the curse of knowledge in three- and four-year-olds. *Quarterly Journal of Experimental Psychology, 63*(2), 209-215.

Bridgers, S., Jara-Ettinger, J., & Gweon, H. (2016). Children consider others' expected costs and rewards when deciding what to teach. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 559-564). Cognitive Science Society.

Chi, M. T., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. Cognitive Science, 25(4), 471-533.

Chi, M. T., Siler, S. A., & Jeong, H. (2004). Can tutors monitor students' understanding accurately? Cognition and instruction, 22(3), 363-387.

Epley, N., & Gilovich, T. (2001). Are Adjustments Insufficient? *PsycEXTRA Dataset*.

Epley, N. (2015). *Mindwise: how we understand what others think, believe, feel and want*. London: Penguin Books.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics,40*(1), 35-42.

Ghrear, S. E., Birch, S. A., & Bernstein, D. M. (2016). Outcome Knowledge and False Belief. *Frontiers in Psychology, 7*.

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.

Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin, 125*(6), 737-759.

Nickerson, R. S. (2001). The Projective Way of Knowing: A Useful Heuristic That Sometimes Misleads. *Current Directions in Psychological Science, 10*(5), 168-172.

Pezzo, M.V. (2003). Surprise, defence, or making sense: What removes hindsight bias? Memory, 11, 421–441.

Pohl, R. (1998). The effects of feedback source and plausibility of hindsight bias. European Journal of Cognitive Psychology, 10, 191–212.

Shafto, P., Goodman, N. D., & Griffiths, T. L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. Cognitive psychology, 71, 55-89.

Sherwood, C. C., Subiaul, F., & Zawidzki, T. W. (2008). A natural history of the human mind: tracing evolutionary changes in brain and cognition. *Journal of Anatomy, 212*(4), 426-454.

Skerry, A. E., Lambert, E., Powell, L. J., & Mcauliffe, K. (2013). The Origins of Pedagogy: Developmental and Evolutionary Perspectives. Evolutionary Psychology, 11(3)

Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology, 73*(3), 437-446.

Tomasello, M. (2009). Why we cooperate: Based on the 2008 Tanner lectures on human values at Stanford University. Cambridge, Mass.: MIT Press.

Rhodes, M., Gelman, S. A., & Brickman, D. (2009). Children's attention to sample composition in learning, teaching and discovery. Developmental Science, 13(3), 421-429.

Ronfard, S., & Corriveau, K. H. (2016). Teaching and preschoolers' ability to infer knowledge from mistakes. *Journal of Experimental Child Psychology, 150*, 87-98.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-Analysis of Theory-of-Mind Development: The Truth about False Belief. *Child Development, 72*(3), 655-684.