Preschoolers and Infants Calibrate Persistence from Adult Models

Julia A. Leonard (jlnrd@mit.edu), Max Kleiman-Weiner (maxkw@mit.edu), Yuna Lee (ynlee@mit.edu), Joshua Tenenbaum (jbt@mit.edu) & Laura E. Schulz (lschulz@mit.edu)
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

Abstract

Perseverance, above and beyond IQ, predicts academic outcomes in school age children, however, little is known about what factors affect persistence in early childhood. Here, we propose a formal Bayesian model of how children might learn how to calibrate effort from observing adult models and then explore this idea behaviorally across two experiments in children and infants. Results from Experiment 1 show that preschoolers persist more after watching an adult persist, but only if the adult is successful at reaching their goal. Experiment 2 and a pre-registered replication extend these findings, showing that even infants use adult models to modulate their persistence, and can generalize this inference to novel situations. These results suggest that both preschoolers and infants are sensitive to adult persistence and use it to calibrate their own effort in far-reaching ways.

Keywords: social learning; motivation; Bayesian inference

Introduction

Many parents tell their children: "It doesn't matter what grade you get; what matters is how hard you try!". This message is substantiated by scientific research: grit and self-control predict academic outcomes above and beyond IQ (Duckworth & Seligman, 2005; Eskreis-Winkler, Shulman, Beale, & Duckworth, 2014). While past work has looked at achievement motivation and persistence in school-aged children (e.g. Bandura, 1977; Eccles & Wigfield, 2002; Yeager & Dweck, 2012), less is known about the factors that affect persistence in infancy and early childhood.

One plausible source of evidence that children may use to guide their effort is their own past experience of success and failure. However, in novel situations, young children may rely on what others say or do. Indeed, work with first and second-graders suggest that children persist longer at both identical and transfer tasks when an adult model's persistence and communicates information consistent with a belief in self-efficacy (Zimmerman & Blotner, 1979; Zimmerman & Ringle, 1981).

What computations might support children's ability to use observations of another person's effort to calibrate their own? Inferences from observable data (How long did the adult try? Did they succeed?) about hidden states (How hard is the task? How skilled is the adult?) may be ambiguous. If, for example, a child sees an adult struggle to open a door and then succeed, the child might infer that either this door is difficult to open or that this adult is not very skilled. Insofar as young children understand that adults are more knowledgeable then they are (e.g. Lutz & Keil, 2002) children might decide that the door is hard to open. However, in order to inform their own effortful intervention on the same, or similar, doors, the child also needs to decide how different their own abilities are from the adult's. If they think they are just a little less skilled than the adult, then they should also try really hard to open the door. However if they are *much* less skilled than the adult, perhaps they shouldn't try at all. Thus, learning about effort from an adult model requires rich inferential reasoning. However, considerable work suggests that preschoolers, and even infants consider the efficiency (Scott & Baillargeon, 2013; Gergely & Csibra, 2003) and costs and rewards associated with goal-directed actions (Jara-Ettinger et al., 2015; Liu & Spelke, 2017). Thus it may well be that children can learn the utility of persistence from adult models even in very early childhood.

Here, we first formalize the inferences described above with a computational model based on recursive probabilistic reasoning. We then test the model predictions empirically by measuring preschoolers' persistence at baseline and in a 2 x 2 design where they first see adults' exert either high or low effort and succeed or fail at a similar task (Experiment 1). We also test the qualitative predictions of the model in an experiment and pre-registered replication looking at whether infants generalize the value of persistence to a novel task after seeing an adult succeed at two different tasks after expending great or little effort (Experiment 2).

Computational Framework

Since we are interested in how social learning ultimately impacts children's behavior, we model both the learner's social inference and its affect on rational behavior. The first challenge is to specify what a child learns when she observes both an adult's persistence on a task and whether or not the adult succeeds or fails at the task. Given empirical work showing that children expect adults to act as rational agents (Scott & Baillargeon, 2013), we propose that children expect adults to calibrate their effort based on their probability of a successful outcome (O). The probability of a successful outcome (P(O)) depends on the agent's skill (S), the task difficulty (D), and the amount of effort the agent puts into the task (E). Equation 1 shows how these quantities are related:

$$(1)\ P(O|E,S,D) = 1/(1 + e^{(-E*S/D)})$$

Effort and skill improve the odds of success. Task difficulty reduces the probability of success. With knowledge of their own skill and the task difficulty, an agent can evaluate their expected return on an amount of effort:

$$(2)\ \mathrm{EU}_{S,D}[E] = R(O)P(O|E,S,D) - C(E)$$

where R(O) is the reward anticipated from a given outcome and C(E) is the cost of that effort level. Using this model, a rational agent chooses their level of effort proportional to the agent's expected utility at that given effort level:

(3) $P(E|S,D) \propto \mathrm{EU}[E]$

This is a model of self-calibrated effort.

The challenge of social learning in the current task is that the child observer wants to learn information they can use to calibrate their own effort, but many of these variables are not directly observable. In the current model we propose that children learn about latent *task difficulty* from the observable features of data: the modeler's effort and outcome. We propose that children invert the self-calibrated effort model to jointly infer latent properties based on a single observation of the modeler's effort and success in a task using Bayesian inference (Equation 4, Figure 1):

(4) $P(D, S_a|E_a, O_a) \propto P(O_a|E_a, S_a, D)P(E_a|S_a, D)P(S_a)P(D)$

where $P(E_a | S_a, D)$ is the same effort calibration described in Eq. 1-3 and $P(S_a)$ and P(D) are the child's priors over the relative skill of the adult exemplar and the difficulty of the task. The prior over difficulty was a bimodal beta distribution to reflect the fact that tasks can be either easy or hard and the prior over skill was a unimodal beta biased toward high skill to reflect the relative skill of the adults compared to children[1]. Bayes rule allows the learner to rationally combine these quantities into an updated belief. The social learning component of our full model is shown as inference over the S_a and D nodes in Figure 1.

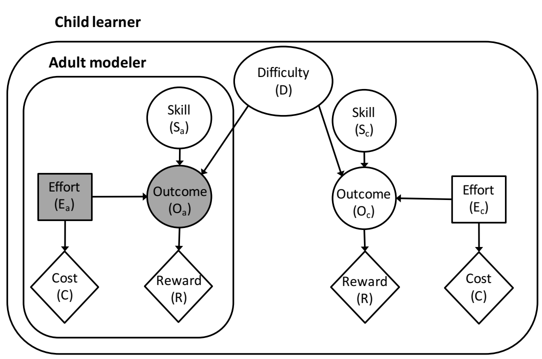

Figure 1: Causal structure of children's decision about how much effort to put into a task based on a adult's effort and success. Squares indicate decisions which are chosen by the agent and diamonds indicate costs and rewards. Shaded nodes indicate states observable to the child.

Using this model, we can capture commonsense inferences a child might make about the difficulty of a task (marginalizing out the skill of the adult) from the observable adult data of effort and outcome (Figure 2). When the child observes an adult expend high effort and succeed, the model infers that the task is difficult, but not outrageously so. When the child observes the adult expend low effort and succeed, the model infers that the task is easy. In both of the conditions where the adult fails to reach their goal, the model infers that the task was very difficult.

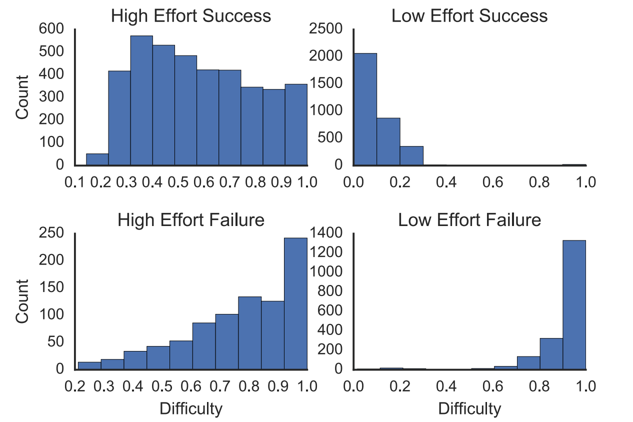

Figure 2: Posteriors over difficulty of the task after an observation for each condition A-D.

The second part of model describes how the child learner exploits this information to calibrate their own effort expenditure. Specifically, the inferred difficulty of the task guides the amount of effort children should employ as they themselves are assumed to be rational effort calibrators. By recursively embedding the rational inference shown in Equation 4 inside the rational model of effort calibration shown in Equations 1-2 we can model how a child will apply effort after observing an adult. Equation 5 shows the complete learning process:

(5) $$\mathrm{EU}[E_c] = R(O_c)\sum_D [P(O_c|E_c, S_c, D)P(D|E_a, O_a)] - C(E_c)$$

Next, we can look at how the model predicts children will integrate observations of adult's effort with their own skill to calibrate their effort allocation (Figure 3). In the High Effort Success condition, we can see that as the child's estimate of her own skill increases, the model predicts that children should try hard. In the Low Effort Success condition, the model predicts that children should try moderately hard even if they believe that they are relatively unskilled, because they believe that the task is easy. In both of the failure conditions, the model predicts children should not try very hard unless they believe they are extremely skilled since the task is probably very hard. In the baseline condition, with no adult model to learn from, children simply sample difficulty from their prior, predicting that they should increase their effort linearly with skill.

Although it is interesting to see if the model qualitatively captures the responses of children, we stress that our interest here is not in providing a precise test of the model (currently we have insufficient data to do so as we did not index skill perception in children and infants). Rather, the model can serve as a precise formulation of factors that might affect children's persistence. Put another way, although the data cannot directly test the validity of the model, the model can inform our predictions about how children might behave in response to the data. We now turn to the empirical design.

[1] See the code for the details of inference used in the simulations: https://github.com/jlnrd/Effort.

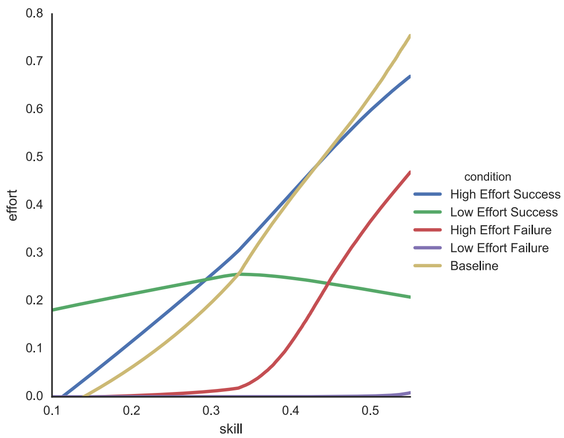

Figure 3: Model predictions of expected effort per skill for each condition.

Experiment 1

Methods

Participants and Materials Children were recruited from an urban children's museum. A power analysis indicated that 26 subjects were required to find large differences in planned post-hoc t-tests (d = 0.8, power = 0.8). Thus, we collected data on 130 children (26/condition; mean: 57.28 months; range: 48 - 71 months). Fourteen additional children were recruited but excluded from analysis due to parental interference (n = 3), no video recording (n = 2), not reaching criteria with the 'all done playing' bell (n = 1), not touching the box before ringing the bell (n = 5), successfully opening the box (which was supposed to be impossible; n = 1), or experimental error (n = 2). Children were randomly assigned to a Low Effort Success, High Effort Success, Low Effort Failure, High Effort Failure or Baseline. Ages were matched across conditions (β = 0.13, 95% CI [-0.91, 1.20][2]).

Two 18.49 x 8.51 x 8.51 cm wooden boxes were used. The boxes could only be opened through a secret sliding notch. A marble was hidden in the experimenter's box and a rubber frog was hidden in the child's box. These produced different sounds when the box was shaken and were used to indicate that the boxes were different. A bell was used for the child to indicate that she was 'all done playing' and a toy bear was used to demonstrate the use of the bell.

Procedure Children were tested individually in a private testing room. In all conditions, the experimenter pretended to play with the stuffed bear, and then said, "I'm all done playing" and rang the bell. The child was asked to play with the bear and indicate when she was done playing by ringing the bell. If the child failed to use the bell correctly the procedure was repeated. If the child failed to ring the bell after three repetitions, they were excluded from the study.

In all conditions except Baseline, the experimenter then brought out their wooden box and shook it, saying, "I think there's something inside of there!" In the Low Effort Success condition, the experimenter took approximately 5 seconds to identify the sliding notch and opened the box. In the High Effort Success condition, the experimenter made repeated attempts to open the box over 30 seconds before locating the sliding notch and opening it. In the Low Effort Failure condition, the experimenter manipulated the box for 5 seconds and then said, "I can't do it. Okay, I'm done." In the High Effort Failure condition, the experimenter performed the same actions as in the High Effort Success condition except that at the end, instead of opening the box, she said, "I can't do it. Okay, I'm done". In the Baseline condition there was no experimental modeling.

Next, the experimenter told the child that she needed to review some paperwork with the parent and that the child would get to play with a toy by herself. Children were told to ring the bell to indicate when they were done playing because the experimenter would be on the other side of the room talking with the parent. The child was then given a box that looked identical to the experimenter's box but had a different toy inside and was intended to be impossible to open. In the Baseline condition only, the experimenter introduced, the child's box to them, shaking it and saying, "It sounds like something is inside of there. I wonder if it can come out!" The experimenter then moved out of the child's line of sight to talk to their parents. The experiment was terminated when the child rang the bell or after four minutes, whichever came first. The experimenter always ended the study by saying, "Oops, I gave you the wrong box to open!," giving the child the original box, and working with the child to open the box successfully in the end.

Results

Behavioral results Because all children spent virtually all the time prior to ringing the bell trying to open the box, we used latency to ring the bell as the dependent measure indicating children's persistence. This was coded from videotape by a coder blind to condition; a second coder blind to condition rescored 80% of the videos (inter-rater reliability r=.99, p<.001.).

There was a main effect of condition on children's persistence[3] ($F(1, 125) = 11.05$, $p < .01$, η^2=.26; see Figure 4). Planned post hoc linear models revealed that children in the High Effort Success condition persisted significantly longer than children in the Low Effort Success condition (β = -0.58 log seconds, $t(50)$= -2.87, p =.006, r^2 = .12, 95% CI [-1.00, -0.19]). That is, as predicted, children tried harder when the adult tried hard and succeeded than when she succeeded effortlessly. This cannot be explained as mere

[2] Confidence intervals reported throughout from bootstrap with 10,000 samples.

[3] The dependent variable was transformed into log space so the distribution would better adhere to a normal distribution

imitation because when the adult failed to achieve her goal, there was no effect of the adults' persistence; the High and Low Effort Failure conditions did not differ (β = 0.01 log seconds, $t(50)$= 0.03, p =.97, r^2 = -.02, 95% CI [-0.44, 0.42]). There was also a trend for children in the High Effort Success condition to persist longer than children at Baseline (β = 0.42 log seconds, $t(50)$= 1.64, p =.11, r^2 = .03, 95% CI [-0.07, 0.91] but no difference between persistence in the Low Effort Success and Baseline conditions (β = -0.16 log seconds, $t(50)$= -0.68, p =.50, r^2 = -.01, 95% CI [-0.67, 0.32]). Children performed below Baseline in both Failure conditions (Low Effort Failure β = 0.83 log seconds, t(50)= 3.12, p =.003, 95% CI [-1.34, -0.35]; High Effort Failure: β = 0.41 log seconds, t(50)= 3.29, p =.002, 95% CI [-1.30, -0.30]).

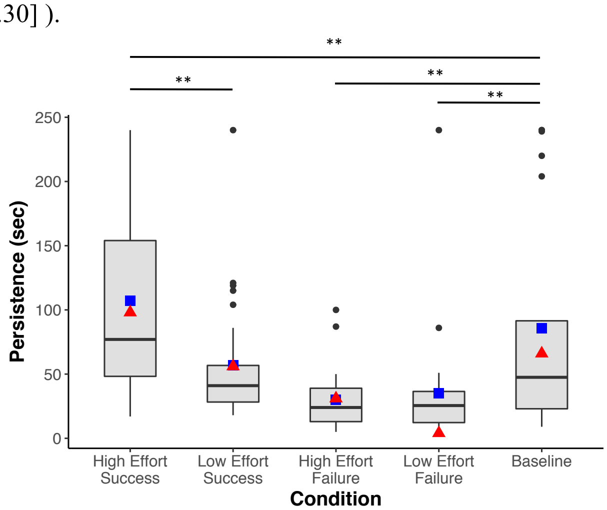

Figure 4: Time spent playing with the toy by condition in Experiment 1 (error bars represent 95% CI intervals from bootstrap with 10,000 samples). ** p<.01. Blue squares are the experimental means and red triangles are the model predicted means/ condition.

Comparing the model predictions to children's behavior Although the model cannot be directly compared to the data, it can inform our predictions about how children might behave in each condition. Assuming for instance that children start out believing that they are about half as skillful as the adults at opening boxes, we can compare the model mean predictions to children's persistence (Figure 4). Under this assumption, the model accurately predicts the human data in the success trials, with a somewhat less accurate fit in the failure trials (although the trends are in the correct direction). More interestingly, the model suggests that the variance in effort by trial is due to differences in children's perception of their skill (thus for instance, the few children who tried really hard in the failure conditions may have thought they were very good at opening boxes). As is clear from the different slopes of the lines connecting skill estimates with persistence in Figure 3, the model accurately predicts that differences in skill perception will lead to the most variance in persistence in the High Effort Success condition and Baseline (cf: the green High Effort Success condition and the yellow Baseline condition with the other conditions in Figure 3; and the variance across conditions in Figure 4).

The results of Experiment 1 suggest that although adult models may not greatly change children's persistence relative to baseline, seeing an adult succeed effortlessly leads to significantly less persistence than seeing an adult work hard to succeed. In Experiment 2, we extend this investigation in two ways. First, we test much younger children: 13-18 month-old infants, to see if adult models affect persistence even earlier in development. Second we show the infant an adult modeling persistence (or succeeding readily) across two different tasks with two distinct goals. We then give the infants a third task, with yet another goal. Given that infants can draw rich generalizations from sparse data (Gweon, Tenenbaum, & Schulz, 2010), especially in pedagogical contexts (Csibra & Gergely, 2009) we predicted that infants who saw a couple of examples of an adult working hard to succeed at two different goals would generalize the value of effort in this context, and persist longer on a novel task than those who saw an adult succeed effortlessly. Finally, we eliminated the Failure conditions in Experiment 2; pilot work suggested that when the experimenter persistently struggled in vain, infants became too fussy and inattentive to explore themselves.

Experiment 2

Methods

Participants and Materials Infants were recruited at an urban children's museum and tested individually in a quiet testing room off the museum floor. Assuming a relatively large effect size, a power analysis indicated that 34 participants per condition would allow a relatively high probability of finding any differences, between each of the three conditions in planned t-test comparisons (d = 0.8, power = 0.9). A total of 24 infants were excluded (11 in High Effort condition, 7 in Low Effort condition and 6 in Baseline) for the following reasons: never pressing the button on the toy (n=8), experimental error due to stimuli breaking or not getting child's date of birth (n=3), and parental interference (n=13). The remaining 102 infants (mean: 15.4 months; range: 13-18 months) were randomly assigned to the High Effort, Low Effort, and Baseline conditions (n = 34 per condition; ages were matched between conditions, β = -0.15, 95% CI [-0.50, 0.21]).

To ensure the robustness of the results, we also ran a pre-registered replication of the High Effort/Low Effort contrast (https://osf.io/g3aws/). In the pre-registered replication, a total of 30 infants were excluded (14 in High Effort condition and 16 in Low Effort condition). The remaining 80 infants (mean: 15.2 months; range: 13-18 months) were randomly assigned to the Effort, and Low Effort conditions (n = 40/ condition; ages were matched between conditions, β = -0.43, 95% CI [-1.14, 0.30]).

The experimenter stimuli consisted of 1) a transparent plastic tomato container with a rubber frog inside that looked as though it could be opened by removing a plastic lid on the bottom of the container, but actually opened by peeling off a sticker at the top of the container, and 2) a carabineer with a toy cow keychain attached. The child toy consisted of a red felt-covered box (6.35^3 cm) with a large round (3.81 cm diameter) easily pressed but inert button on top, and a hidden, difficult to activate button concealed under the felt. When the hidden button was activated, the toy played music. A 30 second timer was also used.

Procedure Infants sat next to their parent in a high chair or booster seat. The child seat had a tray on it. In the Low Effort condition the experimenter picked up the container with the toy inside, announced her intention ("Look, there's something inside of there! I want to get it out!") and then successfully accomplished her goal within 10 seconds; she repeated the task twice more, for a total of three demonstrations over 30 seconds. She removed the container and demonstrated the carabineer with the toy keychain attached. She announced her intention ("How do I get this off?") and again succeeded within 10 seconds, repeating the demonstration for a total of three times in 30 seconds. The High Effort condition was identical except that the experimenter worked at each task for the entire 30 seconds, narrating her attempts as she proceeded ("Hmm … I wonder how I can get my toy out of here?"), succeeding only at the end. In the Baseline condition, the adult did not model any behavior and the infant proceeded directly to the test trial.

In the test trial, the experimenter introduced the infant to the music box saying, "Now it's your turn to play with a toy. See this toy! This toy makes music!" She activated the toy out of the infant's sight using the hidden button. The experimenter then handed the toy to the infant and left the room. The parent was instructed to refrain from interacting with the infant except to return the toy to the child (up to three times) if the child dropped or handed off the toy. The experiment was terminated after two minutes or after the baby handed the toy to her parent and/or threw the toy down a total of three times. At the end of the experiment, the experimenter helped the infant successfully activate the music toy. All trials were videotaped.

Results

Persistence was operationalized as the number of times infants pressed the inert button. We looked both at how many times infants pressed the button overall and, as a potentially more nuanced measure, at the number of the times the infants pressed the button before first handing off or throwing the toy. These measures were coded from videotape by two coders blind to hypotheses and condition (inter-rater reliability r=.99, p<.001). Both the total number of times infants pressed the button, and the number of times they pressed the button before first handing off or discarding the toy, differed by condition[4] (Total button presses: $F(2, 99) = 5.27$, $p = .007$, $\eta^2 = .10$; Presses before first handoff: $F(2, 99) = 5.10$, $p = .008$, $\eta^2 = .10$; See Figure 5). Planned follow-up analyses revealed that, as predicted, children in the High Effort condition pressed the button more times than children in the Low Effort condition and children at Baseline; there was no significant difference in total button presses between the Low Effort condition and the Baseline condition. The same pattern of results held looking only at the total number of button presses before the first hand-off (See Table 1). All results were obtained through linear models but remained statistically significant when tested with non-parametric Wilcoxon rank-sum test.

Table 1: Linear models of total button presses and button presses before handoff for each condition difference in Experiment 1 and 2. The 95% confidence intervals are from a bootstrap with 10,000 samples for the two main outcome measures for each condition.

	Experiment 1	R^2	β	t	df	p	95% CI
Effort vs. Baseline	Total button presses	.10	1.30	2.85	66	.006	0.42, 2.20
	Button presses before handoff	.10	1.23	2.78	66	.007	0.37, 2.10
Effort vs. No Effort	Total button presses	.09	-1.24	-2.71	66	.008	-2.11, -0.35
	Button presses before handoff	.07	-1.14	-2.48	66	.02	-2.03, -0.24
No Effort vs. Baseline	Total button presses	-.01	0.06	0.13	66	.90	-0.77, 0.92
	Button presses before handoff	-.01	-0.09	0.24	66	.81	-0.65, 0.85
	Experiment 2						
Effort vs. No Effort	Total button presses	.05	-0.82	-2.08	78	.04	-1.61, -0.05
	Button presses before handoff	.04	-0.79	-1.91	78	.06	-1.60, -0.002

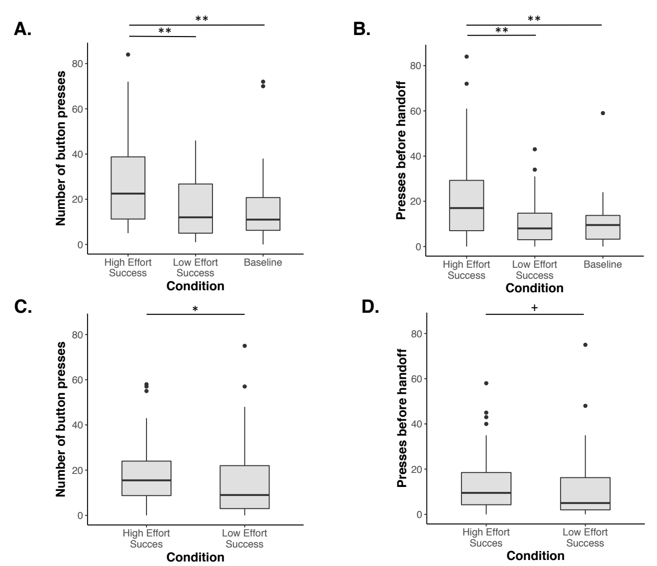

Figure 5: Infants in the Effort condition pressed the button more times in total (A) and before first handoff (B) than children in the No Effort condition or Baseline condition in Experiment 1. In a pre-registered replication, Experiment 2 replicated the results of Experiment 1 showing that children in the Effort condition pressed the button more times in total (C) and before first handoff (D) than children in the No Effort condition. * $p < .05$, ** $p < .01$, $^+ p < .1$.

To assess the robustness of this effect, we ran a pre-registered replication of the contrast between the High Effort and Low Effort conditions. Again, infants in the High

[4] The dependent variables were transformed to the 0.5 power so that the distribution would adhere better to a normal distribution.

Effort condition pushed the button both more times overall and more often before the first hand-off than infants in the Low Effort condition (see Table 1). These results were also obtained with a non-parametric Wilcoxon rank-sum test (Total button presses: $W = 1043.5$, $p = .02$, $r = -.26$; Button presses pre-handoff: $W = 1026.5$, $p = .03$, $r = -.26$), which more accurately reflect the effects due to the non-normal nature of the data[5]

Discussion

These results suggest that both preschool age children and infants learn about persistence from adult models. Behavioral and computational results from Experiment 1 suggest that adult models convey at least two kinds of information that affect children's persistence: whether the goal requires high or low effort and whether the goal is achievable or not. Preschoolers persisted most when an adult persisted successfully, significantly less when an adult succeeded effortlessly, and less still when the adult failed to achieve the goal. The computational model suggests a possible explanation for the variance of our child data, especially in the High Effort Success condition: the child's representation of her own skill level may affect how children act upon evidence. Future studies might specifically test this hypothesis by assessing children's initial estimates of their own skills. Future studies might also provide a precise, quantitative test of the model predictions.

Experiment 2 suggested that infants also use adult models to modulate their persistence, and that they do so in relatively far-reaching ways. Given just two examples of an adult working hard to achieve her goals, infants tried harder at a novel task relative both to baseline and to infants who saw an adult succeed effortlessly. This data both conforms to the behavioral and computational results of Experiment 1 and extends them by suggesting that children may learn to generalize from adult models with just a few examples. These results are in line with previous work suggesting that children draw rich, abstract generalizable inferences from sparse data, especially in pedagogical contexts (Csibra & Gergley, 2009).

Here we show an immediate effect of adult models on children's persistence. Future work might look at how adult effort affects children's persistence across contexts, over time, and with respect to children's explicit attitudes towards challenges (e.g., Yeager & Dweck, 2012). Note however, that persisting on a task is not always a good idea: it is reasonable to work hard only when there is a good chance that the hard work will pay off. Consistent with this, preschoolers were very sensitive to whether the adult succeeded or failed, persisting only in the former case. Similarly, a number of features of the task made it reasonable for the infants to be optimistic: the toy was a developmentally appropriate infant toy, they had heard the toy activate, they were given the toy pedagogically by a friendly adult, and they had previously seen the adult successfully achieve her own goals. Absent any of these factors, the infants might have been less influenced by the adults' persistence; however, precisely such factors might help children assess the relative costs and rewards of working hard at a goal. The current study suggests that at least when success is near at hand, it may be good for children to see adults struggle: if you try hard, they will too.

[5] Unlike in Experiment 1, the residuals of the linear models in Experiment 2 were not normal.

References

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavior change. *Psychological Review, 84,* 191-215.

Duckworth, A. L., & Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents, *Psychological Science, 16*, 939–944.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual review of psychology*, *53*(1), 109-132.

Eskreis-Winkler, L., Shulman, E. P., Beal, S. a, & Duckworth, A. L. (2014). The grit effect: Predicting retention in the military, the workplace, school and marriage. *Frontiers in Psychology*, *5.*

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naıve theory of rational action. *Trends in cognitive sciences*, *7*(7), 287-292.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in cognitive sciences*, *13*(4), 148-153.

Gweon, H., Tenenbaum, J.B., & Schulz, L.E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences,* 107(20), 9066-9071.

Jara-Ettinger, J. Tenenbaum, J. & Schulz, L. E. (2015). Not so innocent: Toddlers' inferences about costs and culpability. *Psychological Science, 26*(5), 633-640

Liu, S., & Spelke, E. S. (2017). Six-month-old infants expect agents to minimize the cost of their actions. *Cognition*, *160*, 35-42.

Lutz, D. J., & Keil, F. C. (2002). Early Understanding of the Division of Cognitive Labor. *Child Development, 73*, 1073–1084.

Scott, R. M., & Baillargeon, R. (2013). Do infants really expect agents to act efficiently? A critical test of the rationality principle. *Psychological science*, *24*(4), 466-474.

Yeager, D. S., & Dweck, C. S. (2012). Mindsets that promote resilience: When students believe that personal characteristics can be developed. *Educational Psychologist*, *47*, 302–314.

Zimmerman, B. J., & Blotner, R. (1979). Effects of model persistence and success on children's problem solving. *Journal of Educational Psychology, 71,* 508-513.

Zimmerman, B. J., & Ringle, J. (1981). Effects of model persistence and statements of confidence on children's self-efficacy and problem solving. *Journal of Educational Psychology, 73,* 485-493.