Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Reverse-engineering the process:Adults' and preschoolers' ability to infer the difficulty of novel tasks

Permalink

https://escholarship.org/uc/item/3k3724jw

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 39(0)

Authors

Gweon, Hyowon
Asaba, Mika
Bennett-Pierre, Grace

Publication Date

2017

Peer reviewed

Reverse-engineering the process:
Adults' and preschoolers' ability to infer the difficulty of novel tasks

Hyowon Gweon (hyo@stanford.edu)
Mika Asaba (masaba@stanford.edu), Grace Bennett-Pierre (gbp@stanford.edu)
Department of Psychology, Stanford University, Stanford, CA 94305

Abstract

The ability to reason about the difficulty of novel tasks is critical for many real-world decisions. To decide whether to tackle a task or how to divide labor across people, we must estimate the difficulty of the goal in the absence of prior experience. Here we examine adults' and preschoolers' inferences about the difficulty of simple block-building tasks. Exp.1 first established that building time is a useful proxy for difficulty. Exp.2 asked participants to view the initial and final states of various block-building tasks and judge their relative difficulty. While adults were near-ceiling on all trials, children showed varying levels of performance depending on the nature of the dimensions that varied across structures. Exp. 3 replicated the pattern. These results suggest that children can reverse-engineer the process of goal-directed actions to infer the relative difficulty of novel tasks, although their ability to incorporate more nuanced factors may continue to develop.

Keywords: Difficulty; Physical reasoning; Social cognition

Introduction

We often think about how easy or difficult it is to achieve a goal. From a child trying a new jungle gym to a scientist building a research team, the ability to reason about task difficulty is critical for many real-world decisions; it informs decisions about the self (e.g., deciding whether to tackle a task or seek help), about others (e.g., understanding who needs help), and even about groups (e.g., assigning tasks in a collaborative project). Although these decisions might seem "easy", they involve more than simply remembering and retrieving our past experiences; they often require estimates and predictions about novel tasks. Such sophisticated inferences might be especially challenging for young children, who frequently face tasks they have never attempted or completed.

Indeed, having this ability does not mean that our estimates are always accurate. Even as adults, we often under- or overestimate the difficulty of certain tasks, failing to meet deadlines or suboptimally allocating time and effort. Nevertheless, our estimates are usually *accurate enough* to get by, suggesting that even these inaccurate estimates might be generated in systematic ways. Indeed, our accuracy and precision in estimating the difficulty of a task might improve with experience and knowledge about the task. However, the ability to predict the difficulty of a novel task (i.e., prior to the actual experience with the task) is crucial for making effective decisions about planning, learning, and even interacting with others. What are the cognitive mechanisms that underlie our ability to predict and estimate difficulty, and how does this ability develop in early childhood?

An intuitive understanding of difficulty

Judgments about perceived task difficulty, or task-related effort, have been mostly studied in terms of its effect on achievement, motivation, and performance attribution in personality and social psychology (e.g., Atkinson, 1957; Weiner, 1966). Early work has operationalized the notion of difficulty as the subjective probability of success or the introspective assessment of required effort (e,g., Atkinson, 1957; Heider, 1958), allowing the notion of difficulty to be measured in quantifiable terms. However, these definitions could easily be intertwined with other agent-dependent concepts such as competence, ability, or intelligence. Prior developmental work has also focused on children's perception of task difficulty and its relationship to motivation and performance in formal educational contexts (Crandall, Katkovsky, & Preston, 1962; Nicholls, 1978; Nicholls & Miller, 1983). These studies suggest that although children around age six consider task difficulty in selecting their own goals (Heckhausen, 1967) they still have trouble differentiating objective task difficulty from agent ability (Nicholls & Miller, 1983).

Some recent work provides indirect support for the idea that children ages 5 to 6 can differentiate objective difficulty from subjective competence. Given information about agents' decisions to pursue goals that vary in costs (i.e., climbing a high hill vs. a low hill) and subjective rewards, children infer agents' competence (subjective costs) (Jara-Ettinger, Gweon, Tenenbaum, & Schulz, 2015). Children also reason about the expected costs for discovering a causal mechanism, and prefer to teach someone a toy that would be *harder* (i.e., require more trial-and-error) for the person to figure out on her own (Bridgers, Jara-Ettinger, & Gweon, 2016) even though both toys are equally easy for them. These results suggest that children may be able to use the properties of the physical environment to estimate the costs of achieving a goal even without any prior experience.

Indeed, an intuitive understanding of task difficulty does not guarantee adult-like inferences. Numerous studies report children's failure in planning and problem-solving tasks that require sequential representation of task space (e.g., Tower of Hanoi; Klahr & Robinson, 1981). While children may successfully detect explicit, perceptual cues (e.g., height of hills, number of buttons on toys), average performance of others (Nicholls, 1978), or actual subjective experiences (e.g., solving standardized test problems such as Raven's matrices; Mueller & Dweck, 1998), they may still fail to infer the difficulty of novel tasks especially when it requires representing or simulating possible states of the world that are not readily observable. Despite the importance of effort estimation however, children's intuitive concept of difficulty has been rarely studied in its own right. Thus the mechanisms that underlie our ability to reason about difficulty and how they develop in early childhood still remain as important open questions.

Current approach Here we explore adults' and children's ability to estimate the difficulty of novel tasks. Given the early-emerging understanding of physical events (Baillargeon, 2004; Spelke, Breinlinger, Macomber, & Jacobson, 1992), and the costs of simple goal-directed actions (Liu & Spelke, 2016; Csibra, 2003), our approach is to ground the basic source of difficulty in agents' interventions on the physical world. We designed a novel task that asked participants to estimate the difficulty of simple engineering goals: building block structures. We explore the idea that humans, even early in life, can estimate the difficulty of novel tasks by reasoning about (1) what physical transitions are involved in the building process, and (2) how an agent might act on the physical states to cause these transitions.

One challenge with eliciting difficulty estimates is that there is no standard metric for measuring the *actual* difficulty. To establish an objective "ground truth" for our tasks, we used a variable that is often used to capture the lay notion of difficulty: time needed to complete a task. In Exp.1 we first establish that people's intuitive sense of difficulty is tightly correlated with their estimates of expected time and the actual time. In Experiments 2 and 3, we systematically vary the physical features of the block structures as well as other factors that influence properties of agents' actions in order to examine adults' and preschoolers' ability to judge relative difficulty of various building tasks.

Experiment 1

In Exp.1 we had two basic goals for investigating people's ability to estimate task difficulty. First, we wanted to verify that people's difficulty estimates systematically reflect a real-world property of the task that can be measured in standard metric (i.e., time). We thus recruited separate groups of participants to get (1) difficulty estimates and (2) building time estimates of various block structures, as well as their (3) *actual* building times, and explored the relationships among these variables. Next, we used these estimates to verify that the pairs of block structures (to be used as stimuli in subsequent experiments) varied in their relative difficulty.

Methods

Participants Separate groups of adults were recruited for the Difficulty Estimation task (N=57, Age: 20-56), Time Estimation task (N=60, Age: 21-68), and Build task (N=14, Age: 18-31). The Difficulty Estimation and Time Estimation tasks were conducted on Amazon's Mechanical Turk (AMT); we excluded participants who gave identical responses on all trials (Difficulty, N=3). The Build task was conducted in lab; one participant was dropped due to technical error.

Materials 28 photos (14 each for initial and final states) of various block structures were used for the task. Each structure had a photo of its initial state (e.g., scattered blocks) and final state (completed structure). Blocks were 1" plain, yellow, green, red, or blue wooden cubes. We designed seven pairs of structures that varied in specific dimensions: (1) *Number1* (3 blocks forming a triangle vs. 10 blocks forming a circle), (2) *Number2* (5 blocks forming a small cross vs. 13 blocks forming a larger cross), (3) *Stability1* (10 blocks in a horizontal line vs. 10 blocks stacked vertically), (4) *Stability2* (two piles of blocks divided by color (yellow and green) vs. a castle-like structure with levels of yellow and green blocks), (5) *Number&Stability* (2 long green blocks stacked vertically vs. 10 plain blocks stacked vertically, height matched), (6) *Probability* (5 red blocks taken out of a transparent box that contained approximately 85% red and 15% blue, or 15% red and 85% blue), and (7) *Process* (2 towers of 5 blocks from an initial state that was either near-complete or very incomplete).

Procedure Participants in all three tasks viewed the same initial and final state photos, but responded to different questions depending on the task. In the Difficulty Estimation task, participants were provided examples of very "easy" and "hard" structures in the beginning to anchor them appropriately on the scale (0 - 100). In each trial, they viewed the initial and the final state photos of a given block structure on the screen (with an arrow pointing from the initial to the final state photo to indicate the physical transition) and answered the question "How difficult would it be to do *this*?" with a sliding bar. In the Time Estimation task, participants saw the same example structures (presented as structures that take a short or a long time to make) to anchor them on the scale (0 - 100 seconds); the question in each trial was "How long would it take to make this?". Two structures within a pair were presented sequentially, but the order of presentation was counterbalanced both within each pair and across all pairs.

In the Build task, the experimenter laid out blocks in front of the subject as in the initial state photo in each trial, and asked to use the blocks to create the structure shown in the final state photo. We recorded how long the subject spent building the block structure from start to finish.

Results

First, we asked whether expected building time can be a good proxy for estimated task difficulty. Even though separate groups participated in the Difficulty Estimation and Time Estimation Tasks, these estimates were highly correlated (Fig.2 Right: $r = .923, t = 8.296, df = 12, p < .001$). This suggests that people's intuitive sense of difficulty can be directly mapped onto estimates of time, and that *actual* building times may be an approximate "ground truth" for difficulty. Given this result, we then asked how well people's estimated building times reflect actual building times. Although people generally overestimated the building times (Fig.2 Left: $intercept = 13.647, t = 4.227, p = .001$), the correlation was fairly high ($r = .780, t = 4.3167, df = 12, p = .001$). Sec-

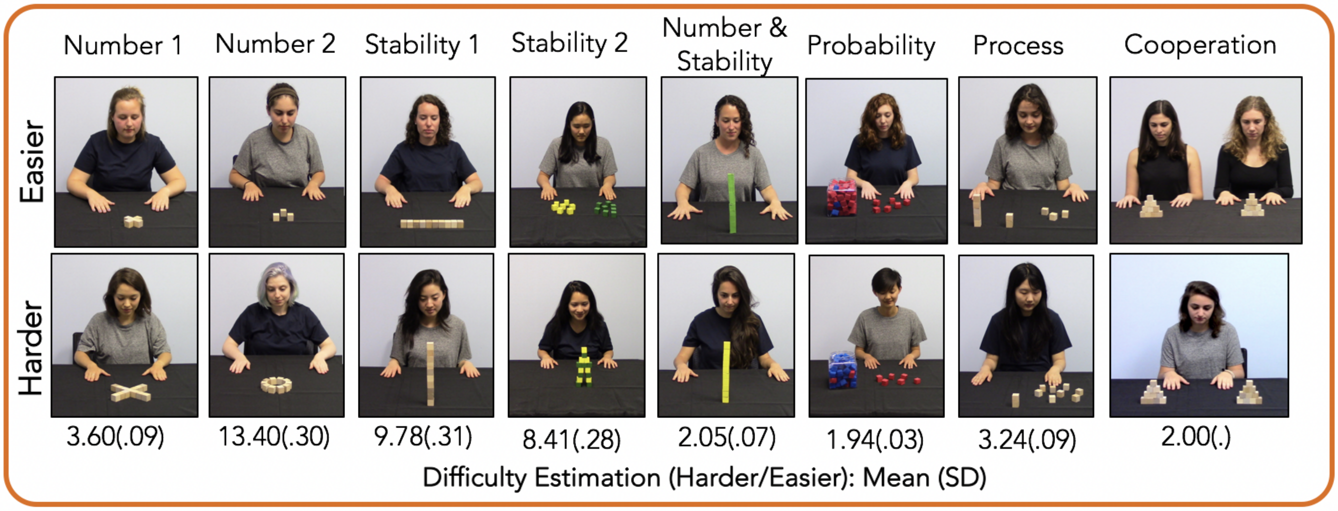

Figure 1: Stimuli used in Expts. 2-3 (final states, except for Process); Exp.1 stimuli did not include agents. Each pair (shown as columns) had an easier structure and a harder structure. Ratio of difficulty estimates are shown beneath each trial.

ond, we verified that the difficulty of block structures within a pair was significantly different in all pairs (paired t-tests, p's $< .002$); differences in estimated time and actual building times were also significant (paired t-tests, p's $< .001$). These graded measures of difficulty also allowed us to calculate the degree to which one structure was "harder" than the other. We calculated the ratio of estimated difficulty between the two structures (higher value indicates a larger difference) and report these in Figure 1.

Collectively these results suggest that adults can make reliable difficulty estimates of individual block structures in ways that systematically reflect some objective, quantifiable aspect of these tasks (i.e., how long it takes to build the structures). Furthermore, we were able to verify that within a pair of block structures, one was clearly more difficult than the other. Despite all pairs having a clear "answer", the magnitude of the difference between the structures varied across pairs.

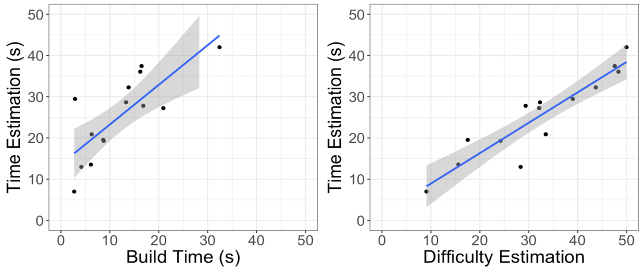

Figure 2: Results from Exp.1. Correlation between Actual Build Time and Estimated Build Time (left); correlation between Estimated Difficulty and Estimated Build Time (right).

Experiment 2

In Exp. 2, we used the stimuli from Exp. 1 to ask whether adults and children can infer the relative difficulty of block-building tasks. Given the results from Exp. 1, we expected adults to show high accuracy in these binary judgments, choosing the structure that was verified as "easier(harder)" in both estimated difficulty and the actual building times.

Our main goal was to examine how children's performance might differ from that of adults. Although adults' estimates indicated that all 7 pairs had a clearly "harder" structure, they varied in *why* the structures varied in difficulty. In Number and Stability trials, the final structures differed in their observable perceptual properties (size, height). The Number&Stability trial was matched on these perceptual cues, making the number of actions needed to complete the task the only determining factor for difficulty. To succeed in the Probability trial, children had to understand that relative difficulty is influenced by the availability of the required blocks (thus the ease of acquiring them) even when the final structures are identical. Success in the Process trial required an understanding that the overall difficulty of a task is easier when one starts from a partially complete state.

In light of prior work reviewed above (e.g., Nicholls, 1978, and Liu & Spelke, 2016), we could consider two extreme possibilities: preschool-aged children might fail to distinguish relative difficulty across the board, or they might successfully detect relative difficulty on all trials. However, a more plausible possibility is that children may succeed in some cases, but selectively fail on other cases. For instance, although it may be easier to detect the differences when a property of the block structures are clearly different (e.g., number, stability), children might struggle in cases where identical structures were built via different *processes*. In particular, in Probability and Process trials, one cannot rely on the number of blocks used in the structures or their final shapes; one must reason about the agents' actions involved in building the structures. Thus children might struggle selectively in these trials.

Indeed, it is also possible that children have a simple heuristic that difficulty depends entirely on the structure alone. Thus in Exp. 2 we added another trial: two identical sets of towers were built, but one was built by two agents (one tower each) while the other was built by a single agent. Success on this task might speak against the possibility that children fail simply because identical structures were built.

Methods

Participants Adults (N=45, Age: 21-59) were recruited on AMT. An additional 13 adults were excluded because they failed the warm-up task (N=8) or the attention check questions (N=5). Twenty-five preschoolers (17 female, $Mage(SD)$: 4.8(.4), Range: 4.1-5.4) were recruited from a laboratory preschool. Seven additional children were excluded due to failure to respond correctly in the warm-up task (N=6) or experimenter error (N=1).

Materials The materials were almost identical to those in Exp. 1, except that the photos now showed an agent looking neutrally at the blocks (initial state) or completed structure (final state). Children viewed these photos on a 15" Macbook Pro (using MATLAB and Psychtoolbox) and indicated their responses by placing their hands on a response pad. Adults viewed the stimuli on Qualtrics. See Fig.1 for stimuli.

Procedure All children were tested in a quiet room, seated next to the experimenter. Half of the children were always asked to indicate the "easier" one, and the other half were always asked to indicate the "harder" one. A warm-up task ensured children understood the meaning of the word "easier(harder)"; children were first presented with two identical boxes, which the experimenter had them first push and then lift, and were asked "Which one is easier(harder) to push?" and "Which one is easier(harder) to lift?" In the main task, children saw photos of green and yellow blocks presented side by side on the laptop screen. All children were able to identify the green (yellow) blocks by placing their left (right) hand on the response pad. In subsequent test trials, children were told: "Anne and Sally were playing with blocks today." as two initial state photos were presented on the screen. The two final state photos were then revealed below the initial state photos; the experimenter pointed to each photo and said, "This is what Anne made, and this is what Sally made. One of them was easier(harder) to make. Which one was easier(harder) to make?" Agents differed across trials and unique names were used for each agent. Trial order and the side of correct response (L/R) were counterbalanced across trials.

Adults participated in an almost identical task on AMT. Similarly to children, the initial states were presented first and then the final states were revealed below these photos. The only difference was that adults read the questions on the screen and answered by clicking on the correct answer.

Results

Adults: As expected, adults performed near-ceiling on all trials ($p < .001$). See Fig.3.

4-5 year-olds: Performance did not differ by question type (easier/harder) so we collapsed the responses throughout (χ^2=.784, df = 1, p = .376). Children showed above-chance performance in 6 of the 8 trials (Number1 ($77.3\%, p = .017$), Number2 ($73.9\%, p = .035$), Stability1 ($81.0\%, p = .007$), Stability2 ($90.5\%, p < .001$), NumStab ($80.0\%, p = .004$), Cooperation ($81.0\%, p = .007$)), while they did not show above chance performance on the remaining two, Process ($41.7\%, p = .541$) and Probability ($57.1\%, p = .664$).

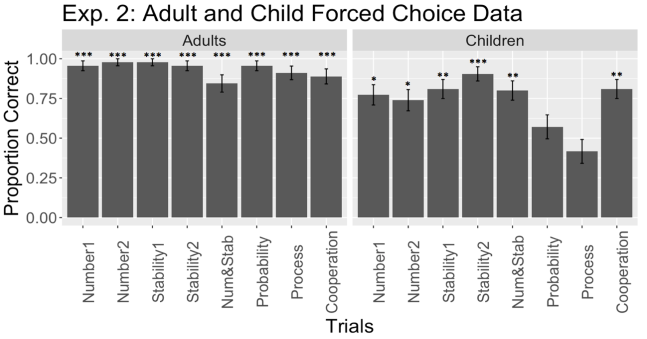

Figure 3: Exp. 2 results. Average % correct for each trial (error bars indicate 95% confidence intervals; ***p <.001, **p <.01, *p <.05).

We asked whether children's chance-level performance was lower than other trials with similar properties. Performance on Process trial was lower than the Number trials (χ^2 = 14.894, df = 1, p < .001), suggesting that even though the two structures differed in the overall number of actions, children failed if two identical structures were built from diffrent starting points. Similarly, children performed significantly worse in the Probability trial than the Stability trials(χ^2 = 20.313, df = 1, p < .001); even though the two structures were made of the same number of blocks, children failed when difficulty judgment relied on the process of sampling.

To examine whether children's performance improved with age, we conducted a logistic mixed-effects model with age and trial as fixed effects and subject as a random effect. Age and the Process trial predicted children's accuracy (age: $\beta = 1.523, z = 2.463, p = .014$; Process: $\beta = -1.765, z = -2.500, p = .013$), suggesting that children's accuracy improved with age but they struggled in the Process trial regardless of age. Finally, among trials where children were reliably above chance, children performed worse than adults in Number2 ($\chi^2(1) = 6.98$,df$= 1, p = .008$) Process ($\chi^2(1) = 12.57$,df$= 1, p < .001$) and Probability trials ($\chi^2(1) = 17.36$,df$= 1, p < .001$), and marginally for Number1 ($\chi^2(1) = 3.51$,df$= 1, p = .061$) and Stability1 ($\chi^2(1) = 3.64$,df$= 1, p = .057$), but not in other trials).

Overall, adults and children were able to judge the relative difficulty of simple physical tasks from just the initial and the final states, without any information about the intermediate processes. It is unlikely that participants had built identical structures in the past and simply recalled their prior experiences to answer these questions. Furthermore, our results suggest that participants did not rely on simple heuristics (e.g, number of blocks, sizes of the structures); their performance was above-chance even when the number and the shape of the structures were identical (Stability1) or their shape and height were matched (Number&Stability). These results suggest that adults and children were able to reason about the process of the physical transitions between the initial and the final states. Children were less accurate than adults on some but not on all trials; importantly, they showed a marked diffcuIty to detect the differences when identical structures were built and the only determining factor was the quality of the actions involved in the building process.

Experiment 3

Exp.3 replicated Exp. 2 with separate groups of children and smaller number of trials per child. In addition to successes, we were interested in replicating the failures in Process and Probability trials that presumably tested a more nuanced understanding of the building process.

Participants Thirty-five preschoolers (18 female, $Mage(SD)$: 4.7(.4), Range: 4.0 - 5.4) participated in Number, Stability, and Number&Stability trials (Group1). Another 35 children (15 female, $Mage(SD)$: 4.2(.7), Range: 4.0 - 5.8) participated in the Sampling, Process and Cooperation trials (Group2). Across groups, 17 additional children were dropped due to experimenter error (N=7), sibling interference (N=1), not speaking English (N=2), failing the warm-up task (N=6) or not finishing the game (N=1).

Materials & Procedure The task was almost identical to Exp. 2, except that in the warm-up task children were presented with simple line drawings and indicated which was easier(harder) to make, and the photos for Sampling, Process, Cooperation trials were presented on paper (8.5 x 11").

Results Children's performance was highly similar to the pattern in Exp.2: Again, accuracy was above-chance on the same 6 of 8 trials (Number1 (77.1%, $p = .002$), Number2 (68.6%, $p = .041$), Stability1 (85.7%, $p = .001$), Stability2 (74.3%, $p = .006$), Num&Stab (85.7%, $p = .001$), Cooperation (77.1%, $p = .002$); children were at chance on Probability (62.9%, $p = .176$) and Process (57.1%, $p = .500$) trials.

Although age did not predict performance in each group (Group 1:$\beta = .583, z = 1.271, p = .204$, Group 2:$\beta = .404, z = .952, p = .341$), collapsing across groups (similar in size to Exp.2), we again saw a trending relationship between age and accuracy ($\beta = .522, z = 1.676, p = .094$; collapsing across all data, age was a significant predictor of accuracy ($\beta = .693, z = 2.694, p = .007$).

General Discussion

In order to investigate the development of the intuitive sense of difficulty, we designed a concrete, manual activity that even young children enjoy and easily understand: building block structures. Across three experiments, we showed that (1) adults' intuitive sense of difficulty accurately reflects actual measures of difficulty (i.e., building time) in both graded estimates and binary judgments, (2) preschoolers show above-chance performance when the pair of structures varied in the expected number of required actions (due to number of blocks, stability, or the number of agents), but (3) they fail on trials in which identical trials were built, which presumably require them to reason specifically about the process of building and the property of actions involved. Collectively, adults and children made systematic judgments about the difficulty of physical tasks from visually observing their initial and final states, without prior experience with the exact building activity or explicit information about the intermediate processes. However, children are still developing these skills throughout the preschool years and possibly beyond.

We found that although adults had a tendency to overestimate the building time, it was strongly correlated with the actual build time, suggesting that these estimates systematically reflected some "ground truth" difficulty of these tasks. Furthermore, these time estimates were tightly linked to adults' difficulty estimates. Indeed, adults' binary judgments reflected the relative difficulty of pairs of structures, resulting in near-ceiling accuracy. This was in stark contrast to children's performance, which was similar to adults in some trials but at chance on some others.

What develops, and what makes us better? One possibility is that the accumulated experience of interacting with physical objects might support a more robust understanding of the underlying physics, increasing the precision of the simulation that might be necessary for generating these intermediate processes (Battaglia, Hamrick, & Tenenbaum, 2013). Another possibility is that experience improves children's understanding of the dynamics between the physical states and the actions required to cause appropriate transitions between these states. These are not mutually exclusive, and both might lead to more accurate representations of the intermediate processes and the effort (e.g., physical, mental) associated with these transitions. Having self-experience with objects helps infants understand others' goal-directed actions (Sommerville, Woodward, & Needham, 2005); it is possible that self-experience continues to help adults and children in making these everyday estimates. Future work may explore whether direct experience with these building tasks increases the precision of time and difficulty estimates.

One important question here is how and when children begin to utilize different dimensions of tasks (e.g., process, probability) when making judgments about difficulty. Despite recent work showing an early-emerging sensitivity to statistical distributions of objects (Xu & Garcia, 2008) and the process by which these objects are sampled by an agent (Gweon & Schulz, 2011), our results suggest that preschoolers may still fail to incorporate this understanding in reasoning about the relative difficulty of agents' sampling behaviors. Children's failure on Process trials parallels school-aged children's difficulty understanding the relationship between time, speed, and distance concepts (Siegler & Richards, 1979); when one train started to travel ahead of another train (but they travelled at equal speeds and stopped at the same place), children fail to answer that this train travelled for a shorter time. These observations are consistent with the possibility that children may struggle to discern the differences in difficulty when the tasks are highly similar in their physical properties. While these results suggest the role of a representational capacity that allows children to simulate multiple intermediate future states sequentially over time, further research is needed to understand the exact nature of their difficulty.

On the other hand, children's robust performance on most trials points to the possibility that the basic inferential ability to estimate difficulty may emerge early. Prior work has found remarkable sophistication in infants' understanding of physical events (e.g., Spelke et al., 1992; Stahl & Feigenson, 2015), as well as their understanding of agents' actions and interventions on the physical world, both for others (e.g., Liu & Spelke, 2016; Newman, Lockhart, & Keil, 2010) and their own (Upshaw & Sommerville, 2015). Thus it is possible that even younger children have the necessary inferential and representational prerequisites for an intuitive sense of difficulty that may manifest not only in their immediate motor plans but also in their predictions of future events. Due to the verbal demands (e.g., meaning of the words "easy" and "hard"), the current paradigm is unlikely to be useful for children under age 3. Future work might exploit building time (a proxy for difficulty in our tasks) in a predictive looking paradigm to address this possibility. Indeed, a time-consuming task is not always judged as harder than a less time-consuming task. Although here we looked at simple cases in which estimated difficulty directly maps onto time, it would be interesting to further investigate how objective and subjective aspects of physical effort (e.g., height of tower and an agent's building competence) as well as mental effort (e.g., careful placement of blocks) may dissociate time and difficulty estimates.

Difficulty is a difficult concept to investigate scientifically. The current work is a small step to understanding this intuitive yet incredibly complex concept. By first examining how people reason about simple, concrete tasks we may obtain clearer insights on how these intuitions arise, and how they develop into more abstract notions of difficulty that are embedded in people's lay use of this word.

References

Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review*, *64*, 359–372.

Baillargeon, R. (2004). Infants' physical world. *Current Directions in Psychological Science*, *13*(3), 89–94.

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, *110*(45), 18327–18332.

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.*

Crandall, V. J., Katkovsky, W., & Preston, A. (1962). Motivational and ability determinants of young children's intellectual achievement behaviors. *Child Development*, *33*, 643–661.

Csibra, G. (2003, March). Teleological and referential understanding of action in infancy. *Philos Trans R Soc Lond B Biol Sci*, *358*(1431), 447–458.

Gweon, H., & Schulz, L. (2011, June). 16-Month-Olds Rationally Infer Causes of Failed Actions. *Science*, *332*(6037), 1524–1524.

Heckhausen, H. (1967). *The anatomy of achievement motivation*. Academic Press: New York.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

Jara-Ettinger, J., Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2015). Children's understanding of the costs and rewards underlying rational action. *Cognition*, *140*, 14–23.

Klahr, D., & Robinson, M. (1981). Formal assessment of problem-solving and planning processes in preschool children. *Cognitive Psychology*, *13*, 113–148.

Liu, S., & Spelke, E. S. (2016, December). Six-month-old infants expect agents to minimize the cost of their actions. *Cognition*, *160*, 35–42.

Mueller, C. M., & Dweck, C. S. (1998). Praise for intelligence can undermine children's motivation and performance. *Journal of Personality and Social Psychology*, *75*(1), 33–52.

Newman, G. E., Lockhart, K. L., & Keil, F. C. (2010, February). "End-of-life" biases in moral evaluations of others. *Cognition*, 1–7.

Nicholls, J. G. (1978). The development of the concepts of effort and ability, perception of academic attainment, and the understanding that difficult tasks require more ability. *Child Development*, *49*(3), 800–814.

Nicholls, J. G., & Miller, A. T. (1983). The differentiation of the concepts of difficulty and ability. *Child Development*, *54*(4), 951-959.

Siegler, R. S., & Richards, D. (1979, February). Development of Time, Speed, and Distance Concepts. *Developmental Psychology*, *15*(3), 288–298.

Sommerville, J., Woodward, A., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, *96*(1), B1–B11.

Spelke, E., Breinlinger, K., Macomber, J., & Jacobson, K. (1992, October). Origins of knowledge. *Psychological Review*, *99*(4), 605–632.

Stahl, A. E., & Feigenson, L. (2015, April). Observing the unexpected enhances infants' learning and exploration. *Science*, *348*(6230), 91–94.

Upshaw, M. B., & Sommerville, J. A. (2015). Twelve-month-old infants anticipatorily plan their actions according to expected object weight in a novel motor context. *Frontiers in public health*, *3*(1), 32.

Weiner, B. (1966). Role of success and failure in the learning of easy and complex tasks. *Journal of Personality and Social Psychology*, *3*(3), 339–344.

Xu, F., & Garcia, V. (2008, April). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences*, *105*(13), 5012–5015.