Children use agents' response time to distinguish between memory and novel inference

Emory Richardson (emory.richardson@yale.edu), Frank Keil (frank.keil@yale.edu)
Yale Department of Psychology, 2 Hillhouse Ave. New Haven, CT 06511 USA

Abstract

Psychologists frequently use response time to study cognitive processes, but response time may also be a part of the commonsense psychology that allows us to make inferences about other agents' mental processes. We present evidence that by age six, children expect that solutions to a complex problem can be produced quickly if already memorized, but not if they need to be solved for the first time. We suggest that children could use response times to evaluate agents' competence and expertise, as well as to assess the value and relevance of information.

Keywords: mental speed, development, theory of mind, cognitive effort

Introduction

According to anecdote, a colleague once posed a trick question to the polymath John von Neumann at a cocktail party. von Neumann was famous for his quick mental calculations, but the solution to this particular problem could be found either by brute-force calculation, or by an insightful trick that was apparently rarely noticed. The question was "*Two cyclists twenty miles apart are moving towards each other at 10 mph; a fly caught between them is moving at 15 mph, from wheel to wheel and back again, until it is crushed between the wheels of the bikes. How far will the fly travel?*". The brute force approach is to sum the geometric series, adding up all of the fly's increasingly short trips between the wheels; alternatively, one could notice that the cyclists will meet in exactly one hour, when the fly will have travelled exactly fifteen miles. von Neumann produced the answer immediately, and the colleague disappointedly replied "oh, you've heard about the trick". von Neumann retorted "what trick? I simply summed the geometric series!". The anecdote illustrates an interesting feature of commonsense psychology: though "mental speed" can signal high competence, response latencies may also help identify which cognitive process generated the response — in this case, memory or inference.

Timing-based inferences have at least three parts: (1) the observer's expectations about how long a task will take to complete, (2) their representation of an agent's actual response time, and (3) access to a plausible explanation for any difference between the two times. By age three, children can accurately represent and compare two durations. While children's time perception is less precise than adults, even 3 year olds are more likely to class a 3s interval between two stimuli as more similar to a 4s interval than a 1s interval (Droit-Volet & Wearden, 2001; see Wearden, 2016, for review).

People also have ample opportunity to notice the regularities in the time needed to perform cognitive processes that would allow them to form expectations about how long a task should take — both in themselves and in other agents. Perceptual processes tend to be fast and automatic. Seeing someone take several seconds to respond to a question like "is this ball red or pink?" may tell us that the color they're looking at is an ambiguous case, even if we can't see it; if it turns out to be fire-engine red, we might infer that the person is colorblind. Memory retrieval may be similar; someone who takes tens of seconds to respond when asked what their high school homeroom teacher's name was may cause observers to doubt the strength of their memory even before they produce an answer (Brennan & Williams, 1995). In contrast, long pauses may be expected for questions that require explicit inferences about complex stimuli. If a person glances at a novel 9x9 Sudoku and claims to know the solution right away, you would likely doubt the accuracy of their answer more than that of a person who takes several minutes to respond. However, if they turn out to be correct, or even close to correct, you might look for evidence of extraordinary abilities. Because response time is part of every social interaction and cognitive process, it is potentially a rich source of information about other minds and the world we share with them.

Preschoolers expect difficult tasks to take more time to complete than simple tasks, but estimating how difficult the task itself is may be a challenge for young children — particularly for cognitive (as compared to physical) effort. In one study, 4-5 year olds were able to identify which of two block towers or arrangements took more effort, but did not distinguish between completing a mostly-finished tower and starting from scratch, or between selecting only red blocks from a box of mostly blue blocks as opposed to selecting them from a box of mostly red blocks (Gweon, Asaba, & Bennett-Pierre, 2017). The authors suggest that these latter tasks may be challenging because they provide fewer physical cues to difficulty.

If children are explicitly told how difficult other agents considered a task or how much time and effort other agents put into the task, their judgments of an agent's competence integrate both completion time and complexity (Heyman & Compton, 2006; cf. Nicholls, 1978), suggesting that children believe higher competence is a plausible explanation for faster response times. Yet, even without explicit cues, children will spontaneously integrate task difficulty and completion time if the task is sufficiently transparent. In one set of experiments, preschoolers watched two agents building block towers; one completed the task faster than the other. When the resulting towers were identical, children said the faster agent was better; when the faster agent completed a *much* easier block task in *slightly* less time, children said the slower agent was better; however, when physical cues were more ambiguous, children's judgments were at chance, though adults unanimously agreed that the agent who completed the more difficult task was more competent (Leonard, Bennett-Pierre, & Gweon, 2019). Thus, one possibility is that children map their understanding of physical effort and time onto cognitive tasks as they acquire more abstract representations of difficulty.

However, studies of children and adults' interpretation of filled pauses in speech suggest that even toddlers can reason about the difficulty of cognitive processes. Children and adults produce filler words like *uh* and *um* more frequently under cognitive load, such as might be imposed by recalling a rare word or weak memory, or planning a complex utterance (Clark & Fox Tree, 2002; Kidd, White, & Aslin, 2011b). Moreover, children as young as 30 months appear to interpret *others*' speech disfluencies as resulting from processing difficulties, as they predictively look at hard-to-describe or unfamiliar objects at the onset of the filled pause (Arnold, Hudson Kam, & Tanenhaus, 2007; Kidd, White, & Aslin, 2011a; Orena & White, 2015). On this account, listeners infer the referent by implicit causal reasoning: hard-to-describe objects cause processing difficulties, which in turn cause disfluencies; hence, a disfluency signals that the speaker is planning to refer to an unfamiliar or hard-to-describe object. If participants are given an alternative cause for a speaker's disfluency, the inference is blocked. For example, if a speaker frequently forgets the names of common objects, a speech disfluency may not imply that they are trying to recall a *rare* word in particular — for a chronically forgetful speaker, the disfluency is equally likely for rare and common words.

In short, even at 30 months, children's evaluations of others' behavior may incorporate (1) the connection between the difficulty of a cognitive task and difficulty of cognitive processing, and (2) processing difficulty and its external cues — for at least some tasks, processes, and cues.

In the present work, we suggest that the time it takes an agent to complete a task may be sufficient for children to infer the sort of cognitive process that generated it. If the agent's task requires complex sequential reasoning to complete successfully, limits on processing speed suggest that an accurate solution could not be achieved under some minimum threshold; however, once a solution is learned, one may expect an agent to recall the solution nearly instantaneously.

Study 1

In Study 1, we asked whether participants would use an agent's response time to determine whether the agent had produced the response from memory, or by inference. Participants were introduced to a pathfinding puzzle (Figure 1a-c) with a MarioKart character, in which they had to (A) figure out the *shortest* road to the finish line that (B) collected *all* of the prizes on the map, and (C) did *not* collect two in a row of the same color or two in a row of the same shape. After learning the rules, participants watched other agents play the game one by one. After either ~3 seconds or ~20 seconds, the agent would start his engine to signal that he thought he knew the "shortest road that follows the rules", and the participant would be asked to decide: was the agent "figuring out the answer for the first time", or "remembering the answer from yesterday"? Despite the simple constraints, problems of this sort can quickly become extremely challenging to solve. However, once a solution is found and memorized, the solution can be recalled quickly despite the complexity. Thus, our first prediction was that participants would categorize fast responses as memory, and slower responses as inference. Our second prediction was that three-second responses would be categorized as memory earlier in development than twenty-second responses would be categorized as inference. Three seconds in particular is an impossibly fast latency to solve a puzzle of this sort; twenty seconds may at least be sufficient time to produce *some* solution. The pre-registration and materials are available at the OSF repository.

Method

Participants. We recruited 45 adults through MTurk, as well as 90 children in two age groups (45 age 5-7, M=6.5, SD=.94; 45 age 8-10, M=9.45, SD=.90; 53% girls). An additional 8 children (3 age 5, 4 age 6, and 1 age 7) and 5 adults were excluded before data collection for answering training questions incorrectly; these were replaced with new participants. Children participated through an online platform for developmental research that allows researchers to video chat with families using pictures and videos on slides (Sheskin & Keil, 2018).

Materials. We created six grid maps with simple geometric shapes of different colors scattered across it at the intersections. Each map had 8 or 9 shapes of 3 or 4 different colors. A flag in the bottom left corner of the grid marked the finish line, and a MarioKart character at the top right marked the starting line.

During the test phase, participants saw the maps appear in front of a cartoon silhouette facing a computer screen. Children answered by using color-coded cartoon figures on the left and right of the screen (Fig. 1); presentation of these answer choices was counterbalanced (Color_CB). Adults answered using a scale slider. The order of the six maps was reversed for half the participants (MapOrder_CB). Finally, four counterbalances were created to vary the order of the agents' response times (TimeOrder_CB).

Procedure. Training Phase: Participants were told that they would play a racing game with Mario, and learned the object of the game and the rules. The experimenter described the task to the children over a video-chat; the same materials were presented to adults as a prerecorded voiceover slideshow. The experimenter told the participant that Mario wanted to collect all of the treasures on a map, and take the shortest road through the map that followed a rule. The rule was that Mario could "*not pick up two treasures in a row that are the same color,* or *two in a row that are the same shape. But he has to pick up* all *of the treasures*". Participants were then required to answer four comprehension questions correctly (see Figure 1a-b). These were as follows. (1) Which of two roads is shorter, (2) Which of two roads breaks a rule, (3) Why does that road break a rule, (4) Identify an item to pick up next in a sequence. Participants who answered each question correctly the first time proceeded to the test phase. Children's incorrect responses were gently corrected after each question, and they proceeded to the test phase only if they were able to answer all the comprehension questions correctly in two additional training rounds. Adults incorrect responses were not corrected, and adults proceeded to the test phase only if they correctly answered all the comprehension questions in the next round. After learning how to play the game, participants were told that they would watch other people playing the game, and that their job was to decide whether each person was (A) remembering the shortest road through the map from playing it the day before, or (B) figuring out the shortest road for the first time. At this point, MTurk participants also answered an attention check question in order to screen out participants who were skimming instructions.

Test Phase (Fig 1c): For each test item, the experimenter presented a new silhouette sitting in front of a blank screen, saying "*Here's the next person. We'll show him the map, and when he knows the shortest road that follows the rules, he'll start his engine*", at which point the map appeared on the screen. After ~3s or ~20s, an engine sound played, and the experimenter said "*now he's started his engine, so he thinks he knows the shortest route through the map*", and participants decided whether the agent had been "remembering the answer from yesterday" or "figuring it out for the first time". Children first chose one of two alternatives (*remembering* or *figuring out*) and then were asked whether the agent was "*probably*" or "*definitely*" [remembering / figuring it out]; adults used a 4-point scale directly. Three maps were presented for ~3 seconds before the engine started, and three for ~20 seconds.

Comprehension Question: Finally, participants were asked a comprehension question to more explicitly test their understanding of response speed. The experimenter introduced two new cartoon agents, one of which was wearing opaque goggles. The experimenter explained that the silhouette with goggles "likes to cheat"; his goggles would show him the shortest road when he looked at the map, and so he would not have to figure out the answer. The other

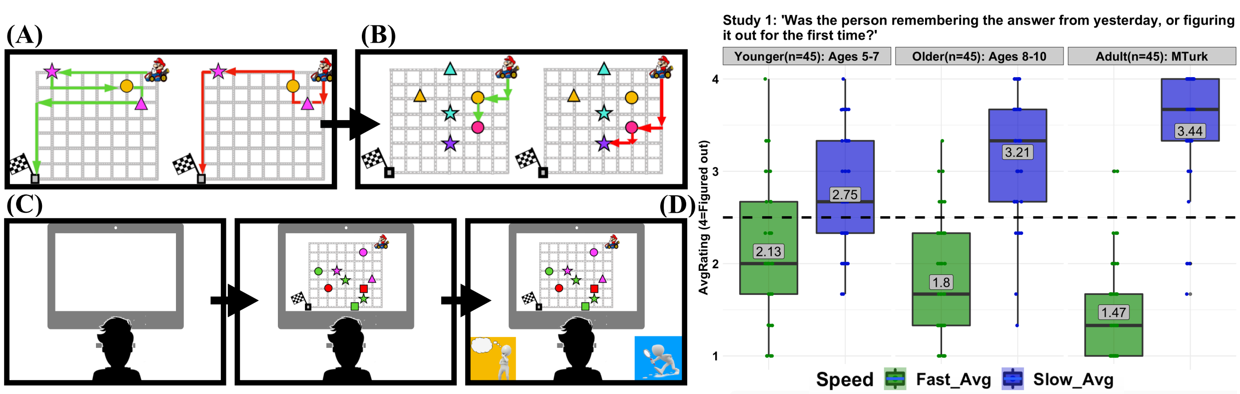

Figure 1. Training questions in Study 1. **(A)** "Which road is shorter, red or green? Yes, red, great job!" **(B) Part 1:** "Which road breaks the rule? Yes, green! Great job! And *why* does it break the rule?" **(B) Part 2:** "Let's say Mario drives on the red road instead. So first, he picks up the pink circle, and then he picks up the purple star. Which treasure should Mario pick up *next*, so that he's not breaking any rules?" **(C) Example Trial:** Participants saw a map appear on a screen for an agent to see for 3s or 20s; then, an engine sounded to signal that the agent thought he knew the answer. Participants were then asked whether the agent was remembering the answer or figuring it out. **(D) Results:** Box plots of averaged responses for Study 1, faceted by age group. Grey labels indicate means.

silhouette was described as playing fair. The experimenter presented a map to the two characters, and an engine sounded after ~3s. Participants were then asked *who* had started his engine: the one who “cheated with his special glasses and saw the answer”, or the one who “played fair”.

Results and Discussion

Results are shown in Figure 1d. The three responses at each response speed were averaged to create a single score for each. A repeated measures ANOVA revealed a significant effect of Speed ($F(1,132)=202.09$, $p<.0001$, $\eta^2 = .605$) and an AgeGroup*Speed interaction ($F(2,132)=17.51$, $p<.0001$, $\eta^2 = .210$), but no effect of AgeGroup ($F(2,132)=0.425$, $p=.65$, $\eta^2 = .006$). As predicted, all age groups categorized the fast response as memory ($M_{Young}=2.13$, $t(44) = -3.23$, $p=.002$, $M_{Old}=1.80$, $t(44) = -7.33$, $p<.0001$, $M_{Adult}=1.47$, $t(44) = -13.00$, $p<.0001$), and categorized the slow response as inference ($M_{Young}=2.75$, $t(44) = 2.63$, $p=.012$, $M_{Old}=3.21$, $t(44) = 7.23$, $p<.0001$, $M_{Adult}=3.45$, $t(44) = 10.58$, $p<.0001$). All age groups also identified the fast responder as having cheated (Fig. 2), including 80% of 6 year olds and 86.7% of 7 year olds, suggesting that even the youngest children recognized that three seconds is an impossibly fast latency to solve the puzzle for the first time ($M_{Young}=71.1\%$, binomial $p=.003$, $M_{Old}=84.4\%$, binomial $p<.0001$, binomial $M_{Adult}=91.1\%$, $p<.0001$). Next, we explored whether the children would rate the fast response as memory at younger ages than they rated slow responses as inference, but our prediction here was not supported. For each subject’s average rating for fast and slow responses, we calculated the deviation from chance responding. We then regressed these values on age in years, using contrast coding to compare each level to chance. The 7, 9, and 10 year olds were more likely to rate the *fast* responses as *memory* ($\beta_7 = .66$, $p<.001$, $\beta_9 = .88$, $p<.001$, $\beta_{10} = .95$, $p<.001$), but the 5, 6, and 8 year olds did not differ from chance ($\beta_5 = .21$, $p=.226$, $\beta_6 = .23$, $p=.180$, $\beta_8 = .28$, $p=.111$). The 7, 8, 9, and 10 year olds were more likely to rate the *slow* responses as *inference* ($\beta_7 = .54$, $p<.001$, $\beta_8 = .37$, $p=.020$, $\beta_9 = .70$, $p<.001$, $\beta_{10} = 1.06$, $p<.001$), but the 5 and 6 year olds did not differ from chance ($\beta_5 = -0.08$, $p=.618$, $\beta_6 = .28$, $p=.075$).

Our results suggest that by age 7, children recognize that memory and novel inference require different amounts of time. Moreover, even young children expressed skepticism that an agent could quickly solve a complex inference without cheating.

Study 2

In Study 2, we explored children's skepticism of implausibly fast responses to complex inference problems more directly. Routing problems similar to those we used in Study 1 can be computationally challenging even for computers, and certainly

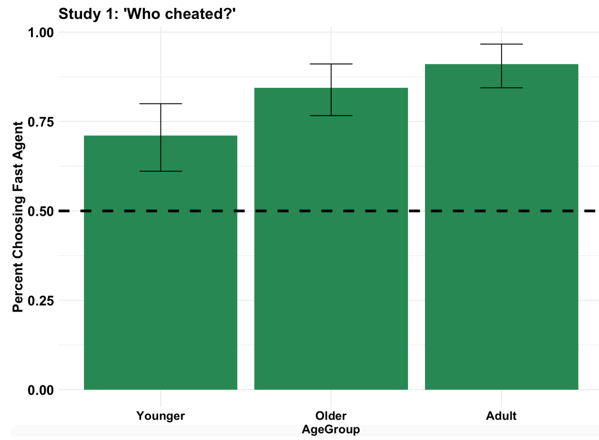

Figure 2. Percent of participants in Study 1 identifying the fast respondent as the cheater.

impossible for a human to solve in just a few seconds. Children in Study 1 seemed to understand this complexity by age 6, believing that the agent who immediately claimed to know must have cheated. Thus, we predicted that even the youngest children would see the speed-accuracy tradeoffs in these problems: they would expect fast responders to make mistakes, and be more likely to make mistakes than slow responders. Alternatively, because past work has suggested that children think that “fast = smart” (Heyman & Compton, 2006), younger children in particular may reason that since faster agents are smarter, and smarter agents are better at solving puzzles, a faster response is more likely to be accurate. However, because children in Heyman & Compton’s (2006) experiments were *told* that the puzzles were easy or difficult and that the agents were fast or slow, rather than making those judgments for themselves, we believed that children might recognize the limits on processing speed and the puzzle’s complexity when allowed to evaluate time and difficulty themselves. The pre-registration and materials are available at the OSF.

Method

Participants. We recruited 45 adults through MTurk, as well as 90 children in two age groups (45 age 5-7, M=6.79, SD=.82; 45 age 8-10, M=9.84, SD=.82; 51% girls). Children participated through an online platform for developmental research that allows researchers to video chat with families using pictures and videos on slides (Sheskin & Keil, 2018).

An additional 12 children and 1 adult were screened out and replaced before data collection. 5 children (three age 5, two age 6) and one adult were screened out for failing the training questions. Two children were excluded due to a lost internet connection; two because excited parents interfered; two quit before completing the activity; and one because his mother reported that he was partially colorblind when he misnamed a color during the training stage (though we did not separately test children for colorblindness, participants were required to distinguish green from red and yellow from blue to answer the training questions, including naming colors spontaneously; participants with either of the two common forms of colorblindness would have been unlikely to pass the training by chance).

Procedure. We made one change to our materials from Study 1. Agents were now described as playing the game for the first time, and participants were asked to guess whether each agent had "*actually* figured out the shortest road, or if they *made a mistake*". Though one could construe "made a mistake" as referring to a mistake made in the process while leaving open the possibility that the end result of the process was accurate, it is unlikely that children or adults interpreted the "mistake" option in that way given the rest of the sentence; moreover, that interpretation would work against our hypotheses, and so cannot explain the results. The wording was chosen for the sake of brevity without sacrificing clarity. Children were then asked whether the agent had "*probably* [made a mistake / figured it out]" or "*definitely* [made a mistake / figured it out]". Adults answered on a 4-point scale directly. Additionally, at the end of the experiment, we asked children to make a competence judgment. Children were shown two new figures and told that these agents would "*each* start their engines when they think they've *figured out* the shortest road that follows the rules". One agent started his engine at 3s, and the other at 20s. However, unlike in the accuracy test items, we then told the children that both agents had "actually figured out the shortest road". Children were then asked which of the two agents was "better at this game". If children have the "fast = better" bias observed in prior studies (Heyman & Compton, 2006), then of two agents who both accurately solve a problem, children should believe that the faster agent is more competent.

Results and Discussion

Results are shown in Figure 3. The three responses at each response speed were averaged to create a single score for each. A repeated measures ANOVA revealed a significant effect of Speed ($F(1,111)=72.16$, $p<.0001$, $\eta^2 = .394$) and an AgeGroup*Speed interaction ($F(2,111)=10.69$, $p<.0001$, $\eta^2 = .162$), but no effect of AgeGroup ($F(2,111)=1.50$, $p=.23$, $\eta^2 = .026$). There were two unexpected order of presentation effects, both suggesting that the predicted effect was larger for one counterbalance than the others; however these effects were smaller than the main effect of Speed, and subsequent analyses suggested that they could not explain the focal findings (TimeOrder_CB: $F(2,111)=5.27$, $p=.002$, $\eta^2 = .125$; MapOrder_CB*Speed: $F(2,111)=3.97$, $p=.049$, $\eta^2 = .034$).

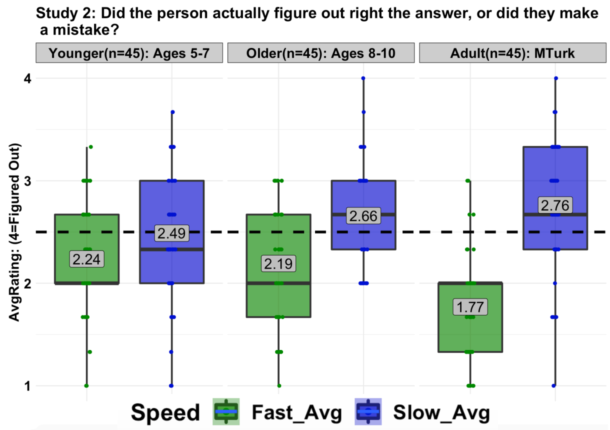

Figure 3. Box plots of averaged responses for Study 2, faceted by age group. Grey labels indicate means.

Post hoc comparisons of the Age*Speed interaction revealed that while the adults and older children distinguished between 3s and 20s responses as predicted, the difference for younger children was not significant (Young: $M_{Fast}=2.24$, $M_{Slow}=2.49$ $t(111) = -2.18$, $p=.473$; Older: $M_{Fast}=2.19$, $M_{Slow}=2.66$, $t(111)=-4.02$, $p=.0016$; Adult: $M_{Fast}=1.77$, $M_{Slow}=2.76$, $t(111) = -8.49$, $p<.0001$, with bonferroni corrections). However, all age groups rated fast responses as likely to be wrong ($M_{Young}=2.24$, $t(44)=-3.04$, $p=.024$, $M_{Old}=2.19$, $t(44) = -3.55$, $p=.006$, $M_{Adult}=1.77$, $t(44) = -9.47$, $p<.0001$, with bonferroni corrections). While adults rated the slow responses as likely to be right, children did not differ from chance ($M_{Young}=2.49$, $t(44)=-0.12$, p=ns, $M_{Old}=2.66$, $t(44)=2.16$, $p=.20$, $M_{Adult}=2.76$, $t(44) = 3.00$, $p=.026$, with bonferroni corrections). For the Ability question, younger children and adults did not differ from chance when asked whether the fast or slow agent was "better" at the game, but older children were more likely to choose the faster agent. ($M_{Young}=46.7\%$, binomial $p=ns$, $M_{Old}=28.9\%$, binomial $p=.007$, binomial $M_{Adult}=40.0\%$, p = ns). Past research using vignettes about agents solving puzzles has suggested that while children have a "faster=better" bias, children under 7 also confound effort, ability, and outcome, while older children begin to attribute outcomes to some combination of effort and ability (Heyman & Compton, 2006; Nicholls, 1978; Stipek & Iver, 1989). Children's judgments in our competence task are consistent with that developmental trajectory: told that the outcome was the same for the fast and slow agent, the 5-7 year olds considered them equally competent. Nevertheless, when outcome was left unspecified and children were simply asked whether or not the person had accurately answered the question, even the younger children recognized a limit on how quickly the puzzle could be solved.

In short, Study 2 suggests that children are skeptical of fast responses to complex inference questions;

however, their intuitions may be fragile until late in development.

General Discussion

By at least age 6 or 7, children can use response times to infer whether an agent is solving a problem for the first time, or retrieving the solution from memory. In our studies, people did this without any knowledge of the agent and without hearing the agent's reply. Moreover, 71.1% of younger children, including 80% of 6 year olds and 86.7% of 7 year olds, believed that an agent's ~3s response latency to a question that required a complex inference was better explained by cheating than by high competence. These may be potentially useful intuitions about cognitive effort and the latencies of two important cognitive processes. Study 2 provides converging evidence that young children are preemptively skeptical about the accuracy of a response if an agent claims to have solved a complex novel problem too quickly. Finally, children's and adult's judgments of agents' competence in Study 2 are intriguing. While children's judgments do suggest a "fast=better" bias, adults may have been justifiably skeptical. In reality, correctly solving a problem of that sort so quickly is improbable if not impossible. Thus, though we told participants that both the 3s and 20s responses were correct, many adults may have considered the 3s response a lucky guess.

The development of pragmatic inferences on the basis of response time is an under-researched aspect of humans' commonsense psychology; for adults, timing supports much more sophisticated inferences in everyday life than the comparatively simple intuition that "thinking takes time" addressed in our experiments. For example, adults interpret longer latencies in response to requests as reluctance (Roberts, Francis, & Morgan, 2006); adults interpret longer latencies in response to trivia questions as uncertainty in the answer (Brennan & Williams, 1995); and adults who respond more quickly to trivia questions are also rated as more charismatic by peers (von Hippel et al., 2016). The simple intuition that "thinking takes time" may enable many of the more sophisticated inferences that adults make on the basis of response time. Yet, because time itself conveys no additional information, the question of what the agent is thinking *about* in the "extra" time — not to mention the judgement about when "extra" time begins — must be determined by context. This sensitivity to context may partially account for the protracted developmental trajectory of children's judgments of response time and competence in the developmental literature (Nicholls, 1978; Stipek & Mac Iver, 1989; Heyman & Compton, 2006; but see Leonard et al., 2019). These studies tested a variety of simple heuristics such as "more effort = better outcomes", "better outcomes = more competent", and "faster = better", finding that children frequently confounded multiple factors. Yet, prior to weighing factors like effort and outcome, an adult-like competence judgement may require the child to first compare the agent's actual and expected response times; and to estimate the expected response time to begin with, the child must weigh the nature and difficulty of the task itself. While participants in our experiments drew increasingly sharp distinctions between memory and inference on the basis of response time from early childhood to adulthood, and also explained faster-than-expected responses to a complex novel inference as "cheating" by age 6, neither the adults nor the youngest children were significantly more likely to identify the faster agent as more competent, suggesting that competence judgments depend on more than just speed and outcome.

Response time may enable a variety of higher-order inferences that can help orient learners towards efficient information acquisition. Future work can investigate the role of speaker-specific characteristics in interpreting response times. For example, a longer response latency may provide more information about the value of a question if a domain-expert is responding than if a non-specialist is answering. A pause from a non-specialist can be written off as a simple lack of background knowledge or the ability to use that knowledge; if a domain-expert pauses, the longer latency might also suggest a gap in the field's knowledge, or that an answer depends on a particularly complex set of factors. Similarly, the amount of time an agent does decide spend on a task may signal how much time the specific task is worth, and help learners rationally balance explore-exploit tradeoffs. For example, children persist longer at difficult tasks after observing adults spend more time and effort, but only if the adult's persistence paid off (Leonard, Lee, & Schulz, 2017; Leonard, Garcia, & Schulz, 2019).

Timing-based inferences may also be relevant to research on metacognition and developing theories of agency. Participants' metacognitive judgments about the difficulty of the task and how long it would take them personally may have informed their judgments of how long the task would take others. Metacognitive judgements about the difficulty of a task also help adults decide how to allocate study time in everyday life (Son & Metcalfe, 2000). From ages 6 to 12, children also increasingly modulate the time spent on easy versus difficult items in the Raven's Progressive Matrices battery with age, and the degree of modulation is strongly correlated with performance (Perret & Dauvier, 2018). Of course, a learner may also fall victim to sunk costs by continuing to struggle with an item that is too difficult; as Metcalfe (2002) has argued, a rational learner may do well to focus on items that are neither too easy, nor too hard for them. Our experiments did not examine children's ability to assess or regulate their own performance. However, future research could ask whether reasoning about cognitive processes in others contributes to children's own metacognitive control.

We are constantly solving problems of varying degrees of novelty and complexity. Some of these problems are fairly concrete in nature: how to get over an obstacle, how to open a jar, or how to make an artifact work as intended. Others are more abstract, involving ratios and differences, conflicting constraints, or generalizations on the basis of sparse patterns. Social learning allows us to minimize the costs of solving these problems on our own, and may help us decide which problems are worth solving; yet, learning from others also requires us to quickly evaluate what others know, how they know it, and the likely quality of their knowledge. Our studies suggest that an agent's response time may help our commonsense psychology interpret their behavior.

Acknowledgments

We thank the members of the Yale Cognition & Development lab for helpful feedback. This research was supported by NSF grant DRL 1561143 awarded to Frank C. Keil.

References

Arnold, J. E., Kam, C. L. H., & Tanenhaus, M. K. (2007). If you say thee uh you are describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(5), 914–930.

Brennan, S., E., & Williams, M. (1995). The feeling of Another's Knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, *34*, 383–398.

Clark, H. H., & Tree, J. E. F. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73-111.

Droit-Volet, S., & Wearden, J. H. (2001). Temporal Bisection in Children. Journal of Experimental Child Psychology, 80(2), 142–159.

Gweon, H., Asaba, M., & Bennett-Pierre, G. (2017). Reverse-engineering the process: Adults' and preschoolers' ability to infer the difficulty of novel tasks. *Proceedings of the Cognitive Science Society*, 6.

Heyman, G. D., & Compton, B. J. (2006). Context sensitivity in children's reasoning about ability across the elementary school years. *Developmental Science*, *9*(6), 616–627.

Heyman, G. D., Gee, C. L., & Giles, J. W. (2003). Preschool Children's Reasoning About Ability. *Child Development*, *74*(2), 516–534.

Kidd, C., White, K. S., & Aslin, R. N. (2011a). Toddlers use speech disfluencies to predict speakers' referential intentions: Toddlers use disfluencies to predict referential intentions. *Developmental Science*, *14*(4), 925–934.

Kidd, C., White, K. S., & Aslin, R. N. (2011b). Learning the Meaning of "Um": Toddlers' developing use of speech disfluencies as cues to. In I. Arnon & E. V. Clark (Eds.), *Experience, Variation, and Generalization: Learning a First Language* (p. 28).

Leonard, J. A., Bennett-Pierre, G., & Gweon, H. (2019). Who is better? Preschoolers infer relative competence based on efficiency of process and quality of outcome. *Proceedings of the Cognitive Science Society*, 7.

Leonard, J. A., Lee, Y., & Schulz, L. E. (2017). Infants make more attempts to achieve a goal when they see adults persist. *Science*, *357*(6357), 1290–1294.

Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? Journal of Experimental Psychology: General, 131(3), 349–363.

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), 16.

Orena, A. J., & White, K. S. (2015). I Forget What That's Called! Children's Online Processing of Disfluencies Depends on Speaker Knowledge. *Child Development*, *86*(6), 1701–1709.

Perret, P., & Dauvier, B. (2018). Children's Allocation of Study Time during the Solution of Raven's Progressive Matrices. *Journal of Intelligence*, 6(1), 9.

Roberts, F., Francis, A. L., & Morgan, M. (2006). The interaction of inter-turn silence with prosodic cues in listener perceptions of "trouble" in conversation. *Speech Communication*, 48(9), 1079–1093.

Sheskin, M., & Keil, F. (2018). TheChildLab.com A Video Chat Platform for Developmental Research. *PsyArXiv*.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and Control Strategies in Study-Time Allocation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 26(1), 18.

Wearden, J. (2016). The psychology of time perception. Palgrave Macmillan.

von Hippel, W., Ronay, R., Baker, E., Kjelsaas, K., & Murphy, S. C. (2016). Quick Thinkers Are Smooth Talkers: Mental Speed Facilitates Charisma. Psychological Science, 27(1), 119–122.