Cognition 224 (2022) 105073

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Thinking takes time: Children use agents' response times to infer the source, quality, and complexity of their knowledge

Emory Richardson *, Frank C. Keil

Yale University, New Haven, CT, United States of America

A R T I C L E I N F O

Keywords:
Cognitive development
Social cognition
Cognitive effort
Response time

A B S T R A C T

Limits on mental speed entail speed-accuracy tradeoffs for problem-solving, but memory and perception are accurate on much faster timescales. While response times drive inference across the behavioral sciences, they may also help laypeople interpret each othejognrs' everyday behavior. We examined children's (ages 5 to 10) use of agents' response time to infer the source and quality of their knowledge. In each trial, children saw a path-finding puzzle presented to an agent, who claimed to have solved it after either 3s or 20s. In Experiment 1 (n = 135), children used agents' response speed to distinguish between memory, perception, and novel inference. In Experiment 2 (n = 135), children predicted that fast responses would be inaccurate, but were less skeptical of slow agents. In Experiment 3 (n = 128), children inferred task complexity from agents' speed. Our findings suggest that the simple intuition that *thinking takes time* may scaffold everyday social cognition.

A colleague once posed a trick question to the polymath John von Neumann. Von Neumann was famous for his quick mental calculations, but the solution to this particular problem could either be found by brute-force calculation, or by a rarely noticed shortcut. 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 answered 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 logic of this exchange is intuitively obvious, which is remarkable given that it relies on a fairly intricate series of counterfactuals and violations of expectation. Even a von Neumann could not sum an infinite series so quickly; but while both guessing and memory can be nearly instantaneous, correctly *guessing* an infinite sum is extremely improbable — so improbable that we jump to the conclusion that von Neumann must have simply recalled the trick. Discovering that von Neumann *had* summed the infinite series reveals a prodigious mental speed — and by extension, extraordinary competence with numbers. This interpretation is effortless because we intuitively understand the relative speed-accuracy tradeoffs of different cognitive processes and the significance of violating those tradeoffs.

Yet, despite the widespread use of response time as an inductive tool in the behavioral sciences, attention to laypeople's own inferences about each others' response times has been sporadic and unsystematic. Nevertheless, the few existing studies suggest that timing is a rich and flexible cue. For example, adults who respond more quickly to trivia questions are also rated as more charismatic by peers (von Hippel, Ronay, Baker, Kjelsaas, & Murphy, 2016). Conversely, adults interpret longer latencies as *reluctance* in response to requests (Roberts, Francis, & Morgan, 2006), as *memory failure* in response to trivia questions (Brennan & Williams, 1995), and as *indecision* between equally desirable options in decision-making (Frydman & Krajbich, 2016; Gates, Callaway, Ho, & Griffiths, 2021). Moreover, in negotiation contexts, buyers' hesitation (or lack thereof) can reveal their price-point, allowing experienced sellers to adjust their selling strategy in response (Konovalov & Krajbich, 2017). Inferences like these often feel effortless despite their sophistication; yet, we are also notoriously bad at estimating the time required to complete a given task. Where do timing-based inferences come from, and how systematic are they? We suggest that even the most sophisticated inferences build on a simple intuition already present in early childhood: *thinking takes time*. On this account, seeing an agent spend more or less time on a task than expected demands explanation. Reasoning about the time costs agents incur to achieve their goals may

* Corresponding author at: Department of Psychology, Yale University, Box 208205, New Haven, CT, United States of America.
E-mail address: emory.richardson@yale.edu (E. Richardson).

https://doi.org/10.1016/j.cognition.2022.105073
Received 4 September 2021; Received in revised form 7 February 2022; Accepted 21 February 2022
Available online 3 March 2022

enable the same kinds of sophisticated inferences about beliefs and desires that we make by reasoning about the costs agents incur while pursuing goals in spatial environments (Baker, Jara-Ettinger, Saxe, & Tenenbaum, 2017). Indeed, timing may be such a flexible cue because *faster* or *slower* responses simply index *more* or *less* thought; the question of what the agent was thinking *about* must be determined by context. Our proposal suggests a developmental approach; while young children may expect thinking about unfamiliar problems to take more time than remembering the answer or seeing it directly, explaining why an agent took more or less time than expected in a specific context may be more difficult. We return to this point in the general discussion.

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

Children can also identify plausible explanations for differences in response times *between two agents*, such as competence or task difficulty. However, conflicting heuristics based on effort, speed, and outcomes often confound younger children's competence judgments until late childhood unless task difficulty is transparent (Leonard, Bennett-Pierre, & Gweon, 2019a; Nicholls, 1978). For example, when experimenters *explicitly described* agents in a story as (A) finishing a puzzle quickly or slowly, (B) thinking it easy or difficult, and (C) trying hard or not, preschoolers integrated difficulty, effort, and speed (Heyman & Compton, 2006). Of course, speed, effort, and task difficulty may rarely be explicitly described in the real world. Still, more recent work suggests children can also integrate these cues spontaneously under certain conditions: when presented with videos that varied agents' speed in building block towers and the relative difficulty of their task, preschoolers recognized the tradeoffs — but only when physical cues to difficulty were unambiguous (Leonard, Bennett-Pierre, & Gweon, 2019a; Gweon et al., 2017). However, cognitively challenging tasks frequently require no physical effort at all. Nevertheless, the difficulty of cognitive processes *themselves* may be detectable for toddlers, even when the objective difficulty of the task is less certain.

Consider the case of speech disfluencies like *uh* and *um:* disfluencies occur more frequently under high cognitive load, such as might be imposed by recalling a rare word or weak memory, or by planning a complex utterance (Clark & Fox Tree, 2002; Kidd, White, & Aslin, 2011b). By 30 months, children appear to interpret speakers' speech disfluencies as resulting from processing difficulties: they predictively look at hard-to-describe or unfamiliar objects at the onset of the filled pause (Arnold, Kam, & Tanenhaus, 2007; Kidd, White, & Aslin, 2011a; Orena & White, 2015). Children's inference may reflect implicit causal reasoning: hard-to-describe objects cause processing difficulties, which in turn cause disfluencies; hence, a disfluency signals that the speaker is preparing to refer to an unfamiliar or hard-to-describe object. When given an alternative cause for a speaker's disfluency, participants' 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: chronically forgetful speakers may just as disfluent in producing rare words as common words. On this account, the listener's reasoning begins not from beliefs about task difficulty per se, but from recognizing the signs of effortful cognitive processing. These cues then trigger a post-hoc search for an explanation of those difficulties.

To the extent that response time signals effortful cognitive processing, it could license a variety of inferences about everyday behaviors. However, psychological processes themselves vary in the time and effort involved. For example, 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 ourselves; if it turns out to be fire-engine red, we might wonder whether the person is colorblind. Memory retrieval may be slower than perception, but someone who takes tens of seconds to respond when asked their spouse's birthday might still elicit doubt or consternation, even before they produce an answer. In contrast, longer pauses are to be expected for questions that require explicit thought about complex relations: *thinking takes time*, on a scale that memory and perception only require under unusual circumstances. In short, violating the expected timescale may be conspicuous: why did the person need to *think* about what color they were seeing? Why *didn't* the person need to think about the answer to a complex calculation? Our success as individuals and as a species depends on our ability to quickly and accurately assess the knowledge, intentions, and competence of other agents; a response that is "too quick" may suggest very different inferences than a response that is "too slow".

1. General method

Here, we examined the development of explicit timing-based inferences in childhood. We initially focus on children ages 5–10 because younger children may struggle to evaluate time and difficulty simultaneously for cognitive tasks (e.g., Leonard, Bennett-Pierre, & Gweon, 2019a; Nicholls, 1978). In each experiment, participants were introduced to a pathfinding puzzle (Fig. 1). After learning the rules, participants watched other agents play the game one by one. After either ~3 s or ~ 20 s, the agent signaled that he thought he knew the solution. Participants were then asked to make a judgment. In Experiment 1, participants saw a complex puzzle presented to the agent, and judged whether the agent "figuring out the answer for the first time", or "remembering the answer from yesterday". We predicted that participants would categorize fast responses as memory, and slower responses as reasoning. Experiment 2 was identical, but participants judged whether the agent had "actually figured out" the answer or "made a mistake". We predicted that even the youngest children would expect fast responders to make mistakes, and to be more likely to make mistakes than slow responders. In Experiment 3, participants saw the agent draw a card with one of two puzzles (simple or complex), but the participants themselves could not see which; they were asked to guess which puzzle the agent was looking at, and then guess whether the agent's solution was accurate. We predicted that children would integrate response time and puzzle difficulty to infer which map the agent was looking at and whether their solution was accurate. Importantly, the puzzle difficulty and the agents' response time were never explicitly mentioned in any of the experiments; inferences based on time or difficulty were made spontaneously. All children participated through an online platform for developmental research (Sheskin and Keil, 2018). Participants came from 39 US states, were 51.7% female, 65% white, and had a median household income of $77,083, as estimated by US Census data for their reported postal code (US household median: $68,703). The pre-registrations, power analyses, data, and materials for each experiment are available on the first author's OSF repository. Though our preregistered analysis plan uses standard linear regressions and ANOVAs, we also provide analogous analyses using ordinal regressions, following recent recommendations against using standard regressions for ordinal data (Liddell & Kruschke, 2018), which were brought to attention during the review process. Because both analyses produce nearly identical results in each experiment but ordinal regressions are still atypical in the field, we present our preregistered analyses in the main text and provide the ordinal regressions as a point of comparison in the supplementary materials.

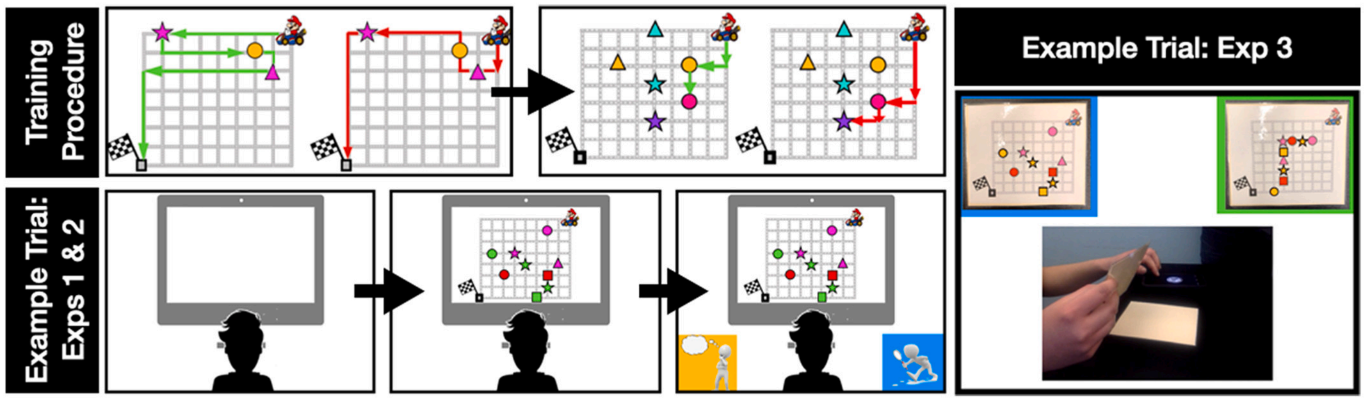

Fig. 1. Training Procedure: Children had three chances to answer 3 of 3 training questions correctly: (1) "Which road is shorter, red or green? Yes, red, great job!", (2) "Which road breaks the rule? Yes, green! Great job! And *why* does it break the rule?", (3) "Let's say Mario drives on the red road 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? Example Trials: After a Hard map appeared on the screen (Exps 1 & 2) or the agent drew a card with either an Easy or Hard map (Exp 3), the agent signaled that they had found the shortest road that followed the rules, after either 3s or 20s. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2. Experiment 1

2.1. Method

2.1.1. Participants

We recruited 45 adults through MTurk, as well as 90 children in two age groups (45 age 5–7, M = 6.5, SD = 0.94; 45 age 8–10, M = 9.45, SD = 0.90; 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.

2.1.2. Materials

We created six grid maps with simple geometric shapes of different colors scattered across them. 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).

2.1.3. Procedure

2.1.3.1. 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 Fig. 1). These were as follows. (1) Which of two example roads is shorter, (2) Which of two example roads breaks a rule, (3) Why does that road break a rule, (4) Identify an item to pick up next in an example 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.

2.1.3.2. First task: memory vs. inference. For each test item (Fig. 1), 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 thinks 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 s before the engine started, and three for ~20 s.

2.1.3.3. Second task: perception vs. inference. Like memory, perceptual processes are nearly instantaneous, making direct perceptual access another potential explanation for fast responses. In a second task, the experimenter introduced two new cartoon agents, one of which was wearing opaque goggles. The experimenter specified that *neither* had played the game before (and so could not be remembering the maps), but that the silhouette with goggles "likes to cheat"; a computer in 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 himself. The other 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". We expected participants to infer that only a cheater would have responded so quickly to a complex puzzle, but did not preregister the hypothesis for the second task.

2.2. Results

The three responses at each response speed were averaged to create a single score for each (Fig. 2). A repeated measures ANOVA revealed a significant effect of Speed ($F(1,132) = 202.09$, $p < .0001$, $\eta^2 = 0.605$) and an AgeGroup*Speed interaction ($F(2,132) = 17.51$, $p < .0001$, $\eta^2 = 0.210$), but no effect of AgeGroup ($F(2,132) = 0.425$, $p = .65$, $\eta^2 = 0.006$). 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, but is easily explained by having direct perceptual access ($M_{Young} = 71.1\%$, binomial $p = .003$, $M_{Old} = 84.4\%$, binomial $p < .0001$, binomial $M_{Adult} = 91.1\%$, $p < .0001$).

Next, we explored whether 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* ($\beta_7 = 0.66$, $p < .001$, $\beta_9 = 0.88$, $p < .001$, $\beta_{10} = 0.95$, $p < .001$), but the 5, 6, and 8 year olds did not differ from chance ($\beta_5 = 0.21$, $p = .226$, $\beta_6 = 0.23$, $p = .180$, $\beta_8 = 0.28$, $p = .111$). The 7, 8, 9, and 10 year olds were more likely to rate the *slow* responses as *inference* ($\beta_7 = 0.54$, $p < .001$, $\beta_8 = 0.37$, $p = .020$, $\beta_9 = 0.70$, $p < .001$, $\beta_{10} = 1.06$, $p < .001$), but the 5 and 6 year olds did not differ from chance ($\beta_5 = -0.08$, $p = .618$, $\beta_6 = 0.28$, $p = .075$).

These results evince an early-developing commonsense intuition that "thinking takes time", while perception and memory — even memory for a solution to a complex problem — are expected to be much faster. In Experiments 2 and 3, we ask whether children can use this intuition to predict the accuracy of an agents' response, and whether they modulate their judgments according to the difficulty of the problem.

3. Experiment 2

3.1. Method

3.1.1. Participants

We recruited 45 adults through MTurk, as well as 90 children in two age groups (45 age 5–7, M = 6.79, SD = 0.82; 45 age 8–10, M = 9.84, SD = 0.82; 51% girls). An additional twelve children and one adult were screened out and replaced before data collection for failing the training (6), losing internet connection (2), fussing out (2), parent interference (2), and colorblindness (1).

3.1.2. Procedure

3.1.2.1. First task: speed-accuracy tradeoffs. We made one change to our materials from Experiment 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*", again using a 4-point confidence scale.

3.1.2.2. Second Task: Speed & Competence. To compare our results to past work on children's timing-based inference, one trial at the end of the experiment asked participants to judge the relative competence of two agents who each accurately solved the same puzzle after either 3s or 20s. Because this task was included simply to compare our results with past work, the full method and results are described in the Supplemental Materials. In brief, younger children and adults were equally likely to judge the fast and slow agent as "better at this game", but older children believed the fast agent was better.

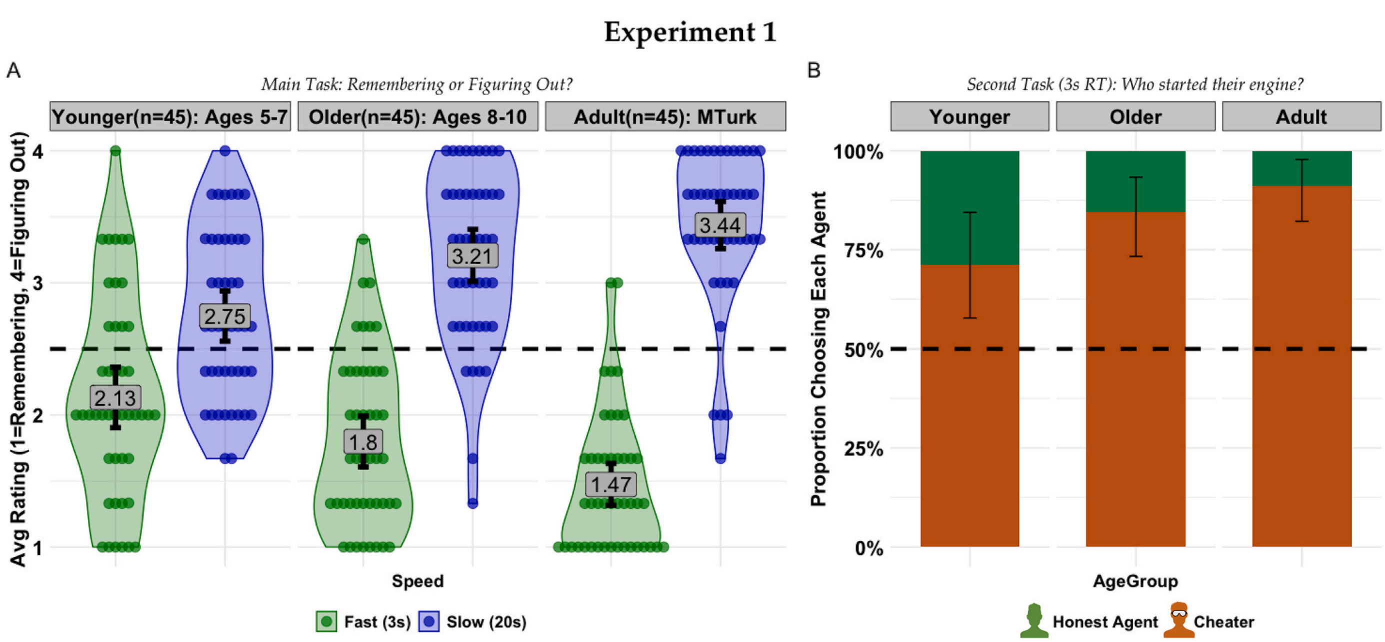

Fig. 2. **(A)** Violin plots and means with 95% CIs for Fast and Slow trials in Experiment 1. Each dot is the average of each participant's 3 fast trials (Green) or 3 slow trials (Blue). Participants rated each agent on a 4-point scale (1 = Definitely remembering, 2 = Probably remembering, 3 = Probably figuring out for the first time, 4 = Definitely figuring out for the first time). **(B)** Proportions inferring that a Fast (3s) agent had "*cheated*" by seeing the solution in their computerized goggles (brown) or "*played fair*" by solving the puzzle themselves (green). Error bars are 95% CIs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2. Results

Results are shown in Fig. 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 ($F(1,111) = 72.16$, $p < .0001$, $\eta^2 = 0.394$) and an AgeGroup*Speed interaction ($F(2,111) = 10.69$, $p < .0001$, $\eta^2 = 0.162$), but no effect of AgeGroup ($F(2,111) = 1.50$, $p = .23$, $\eta^2 = 0.026$). 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: $F(3,111) = 5.27$, $p = .002$, $\eta^2 = 0.125$; MapOrder_CB*Speed: $F(1,111) = 3.97$, $p = .049$, $\eta^2 = 0.034$). 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: $M_{Fast} = 1.77$, $M_{Slow} = 2.76$, $t(111) = -8.49$, $p < .0001$, bonferroni corrected). However, all age groups predicted that fast responses were likely to be wrong ($M_{Young} = 2.24$, $t(44) = -3.04$, $p = .004$, $M_{Old} = 2.19$, $t(44) = -3.55$, $p < .001$, $M_{Adult} = 1.77$, $t(44) = -9.47$, $p < .0001$). While adults and older rated the slow responses as likely to be right, younger children did not differ from chance ($M_{Young} = 2.49$, $t(44) = -0.12$, $p = .907$, $M_{Old} = 2.66$, $t(44) = 2.16$, $p = .036$, $M_{Adult} = 2.76$, $t(44) = 3.00$, $p = .004$). Children may have been right to be skeptical of accuracy on the slow trials: given the computational complexity of these problems, even 20s is too fast to solve them except by luck. Indeed, given past work suggesting that even older children can be unreasonably credulous towards confident speakers (Kominsky, Langthorne, & Keil, 2016), even a skepticism on the fast trials that emerges around age 6 or 7 may be precocious. However, since the difference between the fast and slow trials was not significant in the younger age group, interpretations of the youngest children's responses as reflecting skepticism of the agents' accuracy should be taken with a grain of salt. In Experiment 3, we examine the possibility of an early-but-nuanced skepticism more closely, asking whether children's judgments integrate both response time and task difficulty. (See Figs. 4 and 5.)

4. Experiment 3

If observers infer that "*more time* = *more effort*", then they may infer that agents who spend more time are solving more complex problems than agents who spend less time. We predicted that children would infer that (A) fast agents were more likely to be looking at *easy* maps than *hard* maps, (B) *fast* agents were more likely than the *slow* agents to be looking the easy maps, and (C) fast agents' *solutions* were *correct* for the easy maps but *incorrect* for the hard maps. Because Experiments 1 and 2 suggested that timing-based inferences appear to emerge around age 6 or 7, we focused on ages 6–8 in Experiment 3.

4.1. Method

4.1.1. Participants

We recruited 32 adults through MTurk, as well as 96 children (32 age 6, M = 6.46, SD = 0.31; 32 age 7 M = 7.55, SD = 0.31; 32 age 8, M = 8.59, SD = 0.30; 48 girls). An additional 13 children were screened out and replaced before data collection for failing the training (7), technical difficulties preventing videos from playing (5), and fussing out (1).

4.1.2. Materials

We generated a set of *Easy* puzzles by rearranging the treasures on the complex puzzles from Experiments 1 and 2 into a row of alternating colors and shapes, so that the shortest route passed directly through

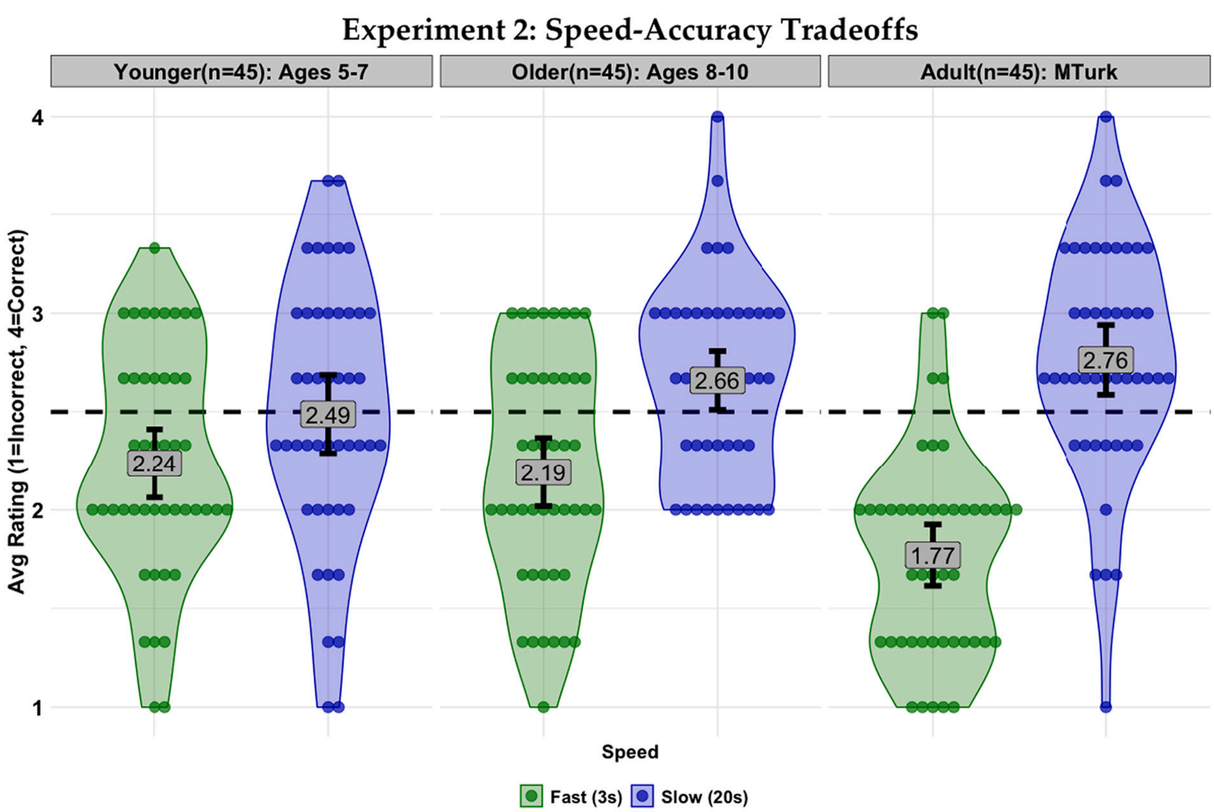

Fig. 3. Violin plots and means with 95% CIs for Fast and Slow trials in Experiment 2. Each dot is the average of each participant's1 3 fast trials (Green) or 3 slow trials (Blue). Participants rated each agent on a 4-point scale (1 = Definitely incorrect, 2 = Probably incorrect, 3 = Probably correct, 4 = Definitely correct). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

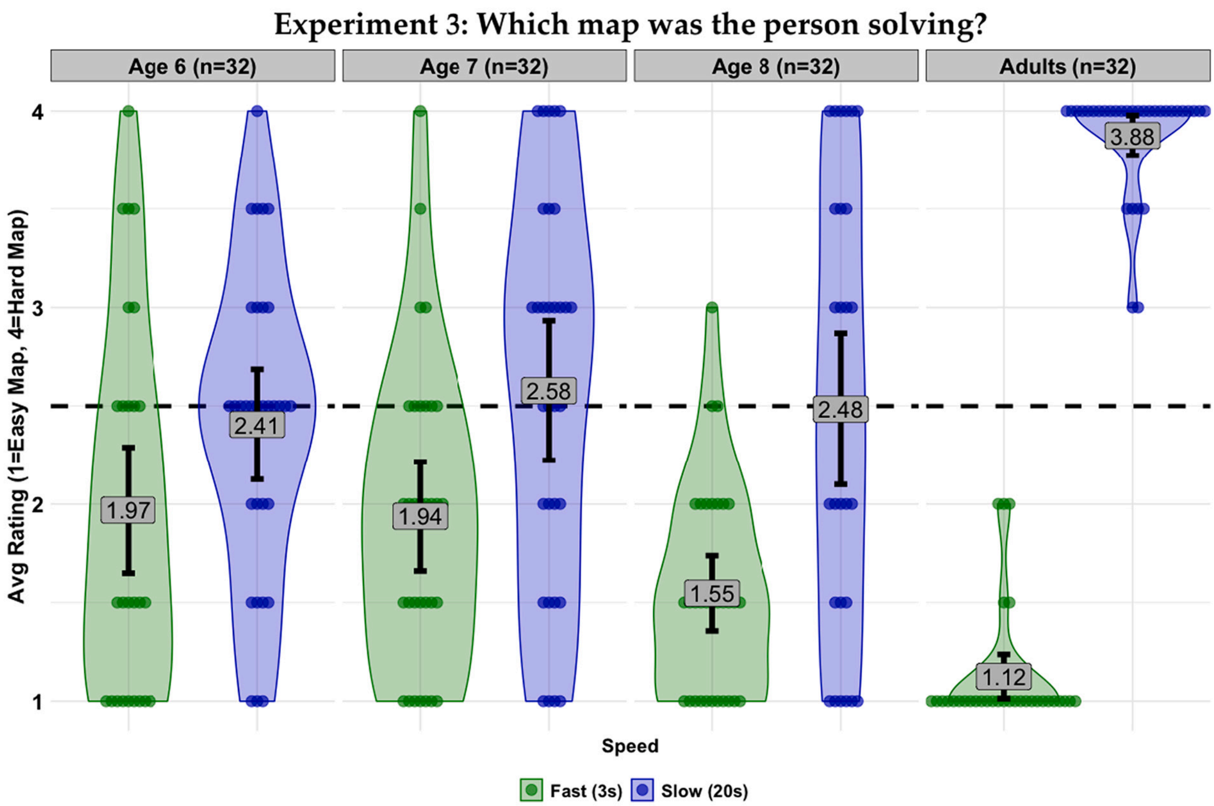

Fig. 4. Violin plots and means with 95% CIs for Fast and Slow trials in Experiment 3. Each dot is the average of each participant's 2 fast trials (Green) or 2 slow trials (Blue). Participants guess which map each agent was looking at, on a 4-point scale (1 = Definitely Easy, 2 = Probably Easy, 3 = Probably Hard 4 = Definitely Hard). Note that "Easy" and "Difficult" were implicit — map difficulty was never explicitly mentioned. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

them. This produced pairs of maps which were identical in the number and kind of treasures, but were *Easy* or *Hard* to solve. In each trial, participants saw an agent draw a card with one of the two maps; when the agent rang a bell to signal that they were ready, children guessed which map was on the card. As in previous experiments, neither time nor the difficulty of the maps was ever explicitly mentioned.

4.1.3. *Procedure*

The training phase was the same as in Experiments 1 and 2.

4.1.3.1. First Task: Which Map?. Agents were described as playing the game for the first time. The experimenter explained that each person would take a card with a map on it, and ring a bell when they thought they knew the shortest road. On each trial, the experimenter showed the participant a slide with an Easy map or a Hard map, and a video embedded between them, and reminded the participant of the task: "this person might get the card with this map [points at simple map] or they might get the card with this map [points at hard map]. Your job is to guess which map was on their card". After the agent rang the bell, the experimenter said "*They rang the bell, so that means that they think they've figured out the shortest road through the map. But which map was on the card they got?*". Children first chose one of the two alternatives and then were asked whether the agent was "*probably*" or "*definitely*" looking at the map; adults used a 4-point scale directly. In two Fast trials, the agent rang the bell after 3s, and in two Slow trials the agent rang the bell after 20s. The order of the trials and the color of the answers was counterbalanced.

4.1.3.2. Second task: difficulty & accuracy. After the main task, participants completed two additional Fast trials. In one video, both cards had Hard maps. In the other, both had Easy maps. Participants were first asked which map the agent was looking at, but then also guessed whether the agent's solution was correct or not, on a 4-point scale. The order of the Easy and Hard trials was counterbalanced.

4.1.4. *Results*

The two ratings at each response speed were averaged to create a single score for each (Fig. 4). The primary question of interest was whether children would infer that the *Fast* agent was looking at the *Easy* puzzle. To test this, we centered children's average ratings on the Fast trials on chance performance (2.5 on a 4-point scale), and centered children's age on the average age of the sample. This makes the intercept of the regression equivalent to a t.test for the whole sample, but allows us to simultaneously check for age effects, using age as a continuous variable. As predicted, children were more likely to infer that the Fast agent must have been looking at the *Easy* puzzle than the *Hard* puzzle ($\beta_{Int} = -0.68$, SE = 0.076, $p < .0001$); the age effect was also significant, though smaller ($\beta_{Int} = -0.21$, SE = 0.093, $p < .026$). To examine the developmental pattern more closely, we also conducted one-sample t. tests comparing each age to chance separately; all ages were significantly more likely to infer that the *Fast* agent was looking at the *Easy* puzzle than the *Hard* puzzle ($M_{Age6} = -0.53$, t (31)= − 3.38, $p = .002$; $M_{Age7} = -0.56$, t (31)= − 4.13, $p = .00026$; $M_{Age8} = -0.95$ t (31)= − 10.2, $p < .0001$). The effect was similar for adults ($M_{Age8} = -1.38$ t (31)= − 25.0, $p < .0001$). Next, we compared children's inferences for *Fast* and *Slow* agents, using *AgeYears* and *Speed* as predictors. A repeated measures ANOVA revealed a significant effect of *Speed* (F(1,93) = 28.11, $p < .0001$, $\eta^2 = 0.232$) but no effect of *AgeYears* or *AgeYears**Speed* interaction (*AgeYears*: F(2,93) = 1.45, $p = .24$, $\eta^2 = 0.03$; *AgeYears**Speed*: F(2,93) = 1.31, $p = .27$, $\eta^2 = 0.027$). Paired-sample t.tests revealed that the effect was similar for all age groups individually, including adults (*Age6*: $M_{Fast} = 1.97$, $M_{Slow} = 2.41$, t (31) = 14.1, $p <$

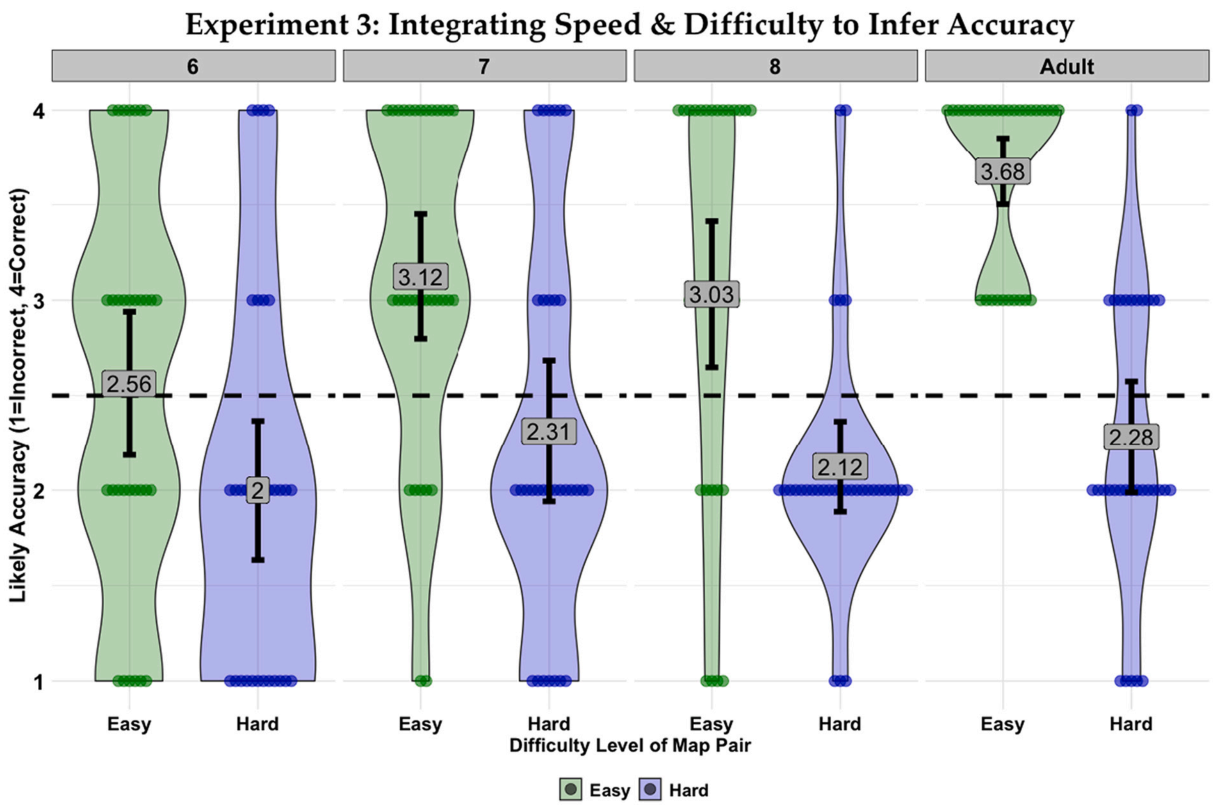

Fig. 5. Violin plots and means with 95% CIs for two Fast trials in Experiment 3 in which participants first inferred which of two Hard or two Easy puzzles the agent was solving, and then predicted the agent's accuracy on a 4 point scale (1 = Definitely incorrect, 2 = Probably Incorrect, 3 = Probably Correct, 4 = Definitely Correct). Though participants were equally likely to guess either puzzle when the difficulty was equalized, they predicted that the fast agent's solution was correct for Easy puzzles and incorrect for Hard puzzles. Note that "Easy" and "Difficult" were implicit — map difficulty was never explicitly mentioned.

.0001; *Age7*: $M_{Fast} = 1.94$, $M_{Slow} = 2.58$, $t(31) = 14.2$, $p < .0001$; *Age8*: $M_{Fast} = 1.55$, $M_{Slow} = 2.48$, $t(31) = 16.3$, $p < .0001$; *Adults*: $M_{Fast} = 1.12$, $M_{Slow} = 3.88$, $t(31) = 70.5$, $p < .0001$).

Children's judgments thus appear to integrate both the complexity of the puzzle and the agent's response speed. This conclusion is further corroborated by the results of the second task (Fig. 5), in which a *Fast* agent drew one of two *Easy* puzzles or one of two *Hard* puzzles: deprived of task difficulty as a cue, participants were no more likely to infer that the agent had drawn one than the other, in any age group (all *p's* = n.s.; see Supplemental Materials). However, all ages were more likely to say that the agent had solved the *Easy* puzzle than the *Hard* puzzle (Difference scores: $M_{Age6} = 0.56$, 95 CI: 0.03–1.09; $M_{Age7} = 0.81$, 95 CI: 0.25–1.31; $M_{Age8} = 0.91$, 95 CI: 0.50–1.31; $M_{Adult} = 1.35$, 95 CI: 1.06–1.61), suggesting that they recognized the relative difficulty of the two puzzles. Estimations of absolute difficulty were less clear. Adults and children ages 7 and 8, but not age 6, believed that the agent's solution was correct for the *Easy* puzzle ($M_{Age6} = 2.56$, 95 CI: 2.22–2.91; $M_{Age7} = 3.12$, 95 CI: 2.81–3.41; $M_{Age8} = 3.03$, 95 CI: 2.66–3.38; $M_{Adult} = 3.68$, 95 CI: 3.52–3.84), while children ages 6 and 8, but not adults or 7-year-olds, believed that the agent's solution was incorrect for the *Hard* puzzle ($M_{Age6} = 2.00$, 95 CI: 1.66–2.38; $M_{Age7} = 2.31$, 95 CI: 1.97–2.69; $M_{Age8} = 2.12$, 95 CI: 1.91–2.38; $M_{Adult} = 2.28$, 95 CI: 2.00–2.56).

5. General discussion

Response time has been a powerful inductive tool in the behavioral sciences. It has been used to infer preference strength (Konovalov & Krajbich, 2019), intelligence (Salthouse, 1996), the strength of memory traces (Singer & Tiede, 2008), and of course, diligence in online surveys. Response times have even been argued to impose bottom-up constraints on models of perception, by comparing the maximum transmission speed of a single neuron with the time typically sufficient for basic perceptual tasks (Feldman & Ballard, 1982). Less attention has been paid to how laypeople themselves interpret response times.

Our experiments provide evidence that from an early age, the commonsense intuition that "*thinking takes time*" may help us interpret everyday behaviors. Indeed, Experiment 3 suggests that children spontaneously integrate task difficulty to estimate *how much* time a task should take: all ages expected slower responses to harder problems than easier problems. These estimates may help children decide *how fast is too fast* for an agent solving a novel problem. Children appeared to recognize speed-accuracy tradeoffs and modulate their accuracy judgments according to task difficulty (Experiments 2 and 3). Moreover, when confronted with quick responses to hard problems, children believed that the agent must have recalled the answer from memory or seen it directly (Experiment 1). Given participants' propensity to explain away fast responses as inaccurate, memory-based, or simple, it may seem inconsistent that only 8–10-year-olds inferred that agents who quickly solved complex novel problems were more competent than slower agents (Experiment 2). However, given the complexity of the hard puzzles, adults' judgments may simply reflect the more sophisticated judgment that the fast agent had only "solved" the puzzle by a lucky guess.

The possibility of 'lucky guesses' illustrates an interesting contrast between reasoning about agents' allocation of time and reasoning about their navigation of space: costs measured in time may be more malleable than costs in distance, making reasoning about the utility of agents' actions on the basis of time more challenging — but also potentially more informative about the agent themselves. An agent that prefers a reward that is spatially distal over one that is spatially proximal *must* pay a higher cost to obtain the more valuable reward *every* time they do so; in this sense, space imposes a fixed cost to any physical action (Baker et al., 2017; Jara-Ettinger, Gweon, Schulz, & Tenenbaum, 2016; Jara-Ettinger, Gweon, Tenenbaum, & Schulz, 2015). For instance, if the toddler in the dining room prefers the cherries in the kitchen to the dinner on the table, the *distance* to the kitchen is fixed both for them and their parent: no walking, no cherries. The immutability of spatial costs may make them particularly useful for analyzing rational action even in early infancy (Gergely & Csibra, 2003; Liu, Ullman, Tenenbaum, & Spelke, 2017). In contrast, while it may take longer to *count* a bowl of 35 cherries than a bowl of 30 cherries, a lucky guess could get them the 35 cherries without having to count; moreover, developmental changes in counting skill and precision in approximate number estimation may lead to different expected values of each strategy for a toddler and their parent (Baer & Odic, 2019; Halberda & Feigenson, 2008). Analogously, if an expert gives a quick estimate instead of a time-consuming calculation, we might infer that an 'educated guess' is sufficiently precise; but a novice who gives a quick estimate may not even understand the parameters of the question. In other words, while guessing and reasoning have their characteristic time signatures, successfully reasoning about time-costs may require us to consider context-specific factors like the complexity of the problem, the methods available to solve it, and potentially the competence of the agent. Thus, even with hard constraints on mental speed, cost-based reasoning about time may require more sophisticated inferences than cost-based reasoning about space.

Timing's sensitivity to context may also help explain why competence judgments show a protracted developmental trajectory in the existing literature (Heyman & Compton, 2006; Nicholls, 1978; Stipek & Iver, 1989): integrating multiple cues can be challenging for children. Neither speed, effort, nor accuracy alone signals competence: a competent agent must *outperform* the expected speed-accuracy tradeoffs *because of* their abilities. When simple heuristics like "more effort = better outcomes", "better outcomes = more competent", and "faster = better" imply conflicting competence judgments, children may find it difficult to weigh the relative importance of each dimension. Accounting for differences in motivation and attention adds to the challenge: conscientiousness and distraction can both increase response time, just as genius and haste can decrease it. However, recent computational work has suggested that by adulthood, people can distinguish distraction from focused thought by integrating the response time and complexity of the most likely topic of focus (Berke & Jara-Ettinger, 2021).

Children's inferences in our experiments were less sophisticated than competence judgments, but also more general: thinking takes more time than remembering or seeing, more complex problems require more thinking, and some problems are impossible to solve immediately. However, while our aim was not to establish the earliest age at which children can reason about response time, task demands still limit our conclusions about younger children's abilities. Though the rules to the puzzle game we used were simple enough that most children had no trouble learning them, the game was novel to children, and the procedure provided little reinforcement of the rules after the training. While our participants displayed precocious skepticism and sensitivity to task difficulty, novelty and low incentives may have hindered younger children's performance. If children learn to simulate others' mental processes through experience of their own mental processes in similar contexts, our study may have underestimated their capacities simply by giving them little experience solving the maps themselves (Kano, Krupenye, Hirata, Tomonaga, & Call, 2019; Meltzoff & Brooks, 2008; Sommerville, Woodward, & Needham, 2005). While the task was also novel for older children and adults, they may have also found it easier to simulate solving the maps for themselves before answering. Future work could test children's performance in a more familiar context, or compare their performance with and without additional practice solving similar puzzles.

Future work could explore the impact of inferences about time on children's learning strategies. Some of these inferences may come from monitoring their own response time. For example, children 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). By adulthood, people's problem-solving strategies not only weigh time costs in the hundreds of milliseconds, but may integrate both cognitive and physical costs (Feghhi & Rosenbaum, 2019; Gray, Sims, Fu, & Schoelles, 2006). Choices between different strategies can be thought of in terms of opportunity costs: in addition to costs and benefits of each strategy individually, an individual who uses one strategy foregoes the opportunity to benefit from the other strategy (Boureau, Sokol-Hessner, & Daw, 2015). Thus, an understanding of how children learn to allocate time effectively may need to consider the both effectiveness of the problem-solving strategies available to them and the cognitive and physical tradeoffs between those strategies. For instance, in contexts that provide immediate accuracy feedback and little penalty for mistakes, it may be more rational to learn by trial-and-error than attempting to solve a problem through thinking alone. Future work could explore how children allocate time when the costs of error are high or low. Future work could also explore the use of response time in combination with other common social learning strategies. For instance, novices may be generally slow; but delays from experts may indicate a complex problem, a valuable solution, or a gap in the field's knowledge. Thus, experts' time allocation in particular may help learners estimate the value of persistence, either generally or for a specific task or problem-solving method. Indeed, children persist longer at physical tasks after observing adults spend more time and effort, but only if the adult's persistence paid off (Leonard, Bennett-Pierre, & Gweon, 2019a; Leonard, Lee, & Schulz, 2017).

Even in early childhood, the assumption that agents pursue goals efficiently by minimizing expected costs while maximizing expected rewards helps us reason about others' preferences, knowledge, and beliefs (Gergely & Csibra, 2003; Jara-Ettinger et al., 2015; Jara-Ettinger et al., 2016). Much of this work has focused on the costs imposed by navigating complex spatial environments. As a fundamental constraint on every cognitive process and social interaction, time imposes costs that are even more ubiquitous, but may be more challenging to evaluate because of their sensitivity to context. Nevertheless, our results suggest that by age 6, the commonsense intuition that 'thinking takes time' — more time than perception and memory — may allow us to infer *how* another agent knows something as well as the *quality* and *complexity* of their knowledge. As children learn to integrate contextual information such as agents' expertise and the difficulty of a problem, this simple intuition could scaffold more sophisticated reasoning about agents' knowledge, intentions, and reliability.

Author contributions

ER developed the study concept; ER designed the experiment with input from FCK; ER collected and analyzed the data. ER drafted the manuscript, and FCK provided critical revisions. All authors approved the final version of the manuscript for submission.

Acknowledgments

This research was supported by NSF grant DRL 1561143 awarded to Frank C. Keil.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2022.105073.

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. https://doi.org/10.1037/0278-7393.33.5.914

Baer, C., & Odic, D. (2019). Certainty in numerical judgments develops independently of the approximate number system. *Cognitive Development, 52*, Article 100817. https://doi.org/10.1016/j.cogdev.2019.100817

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour, 1*(4). https://doi.org/10.1038/s41562-017-0064

Berke, M., & Jara-Ettinger, J. (2021, May 20). Thinking about thinking through inverse reasoning. In *Proceedings of the cognitive science society*. https://doi.org/10.31234/osf.io/r25qn

Boureau, Y.-L., Sokol-Hessner, P., & Daw, N. D. (2015). Deciding how to decide: self-control and meta-decision making. *Trends in Cognitive Sciences, 19*(11), 700–710. https://doi.org/10.1016/j.tics.2015.08.013

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., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition, 84*(1), 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3

Droit-Volet, S., & Wearden, J. H. (2001). Temporal bisection in children. *Journal of Experimental Child Psychology, 80*(2), 142–159. https://doi.org/10.1006/jecp.2001.2631

Feghhi, I., & Rosenbaum, D. A. (2019). Judging the subjective difficulty of different kinds of tasks. *Journal of Experimental Psychology. Human Perception and Performance, 45*(8), 983–994. https://doi.org/10.1037/xhp0000653

Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science, 6*(3), 205–254. https://doi.org/10.1207/s15516709cog0603_1

Frydman, C., & Krajbich, I. (2016). Using response times to infer others' beliefs. *An Application to Information Cascades., 68*. https://doi.org/10.2139/ssrn.2817026

Gates, V., Callaway, F., Ho, M. K., & Griffiths, T. (2021). A rational model of people's inferences about others' preferences based on response times. *psyArXiv*. https://doi.org/10.31234/osf.io/25zfx

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences, 7*(7), 287–292. https://doi.org/10.1016/S1364-6613(03)00128-1

Gray, W. D., Sims, C. R., Fu, W.-T., & Schoelles, M. J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review, 113*(3), 461–482. https://doi.org/10.1037/0033-295X.113.3.461

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

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology, 44*(5), 1457–1465. https://doi.org/10.1037/a0012682

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. https://doi.org/10.1111/j.1467-7687.2006.00540.x

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. https://doi.org/10.1016/j.cognition.2015.03.006

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

Kano, F., Krupenye, C., Hirata, S., Tomonaga, M., & Call, J. (2019). Great apes use self-experience to anticipate an agent's action in a false-belief test. *Proceedings of the National Academy of Sciences, 116*(42), 20904–20909. https://doi.org/10.1073/pnas.1910095116

Kidd, C., White, K. S., & Aslin, R. N. (2011a). 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).

Kidd, C., White, K. S., & Aslin, R. N. (2011b). Toddlers use speech disfluencies to predict speakers' referential intentions: Toddlers use disfluencies to predict referential intentions. *Developmental Science, 14*(4), 925–934. https://doi.org/10.1111/j.1467-7687.2011.01049.x

Kominsky, J. F., Langthorne, P., & Keil, F. C. (2016). The better part of not knowing: Virtuous ignorance. *Developmental Psychology, 52*(1), 31–45. https://doi.org/10.1037/dev0000065

Konovalov, A., & Krajbich, I. (2017). On the strategic use of response times. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3023640

Konovalov, A., & Krajbich, I. (2019). Revealed strength of preference: Inference from response times. *Judgment and Decision making, 14*.

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. https://doi.org/10.1126/science.aan2317

Leonard, J. A., Bennett-Pierre, G., & Gweon, H. (2019a). Who is better? Preschoolers infer relative competence based on efficiency of process and quality of outcome. In *Proceedings of the cognitive science society* (p. 7).

Leonard, J. A., Garcia, A., & Schulz, L. E. (2019b). How Adults' actions, outcomes, and testimony affect Preschoolers' persistence. *Child Development*. https://doi.org/10.1111/cdev.13305

Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology, 79*, 328–348. https://doi.org/10.1016/j.jesp.2018.08.009

Liu, S., Ullman, T. D., Tenenbaum, J. B., & Spelke, E. S. (2017). Ten-month-old infants infer the value of goals from the costs of actions. *Science, 358*(6366), 1038–1041. https://doi.org/10.1126/science.aag2132

Meltzoff, A. N., & Brooks, R. (2008). Self-experience as a mechanism for learning about others: A training study in social cognition. *Developmental Psychology, 44*(5), 1257–1265. https://doi.org/10.1037/a0012888

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. https://doi.org/10.1111/cdev.12421

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. https://doi.org/10.1016/j.specom.2006.02.001

Salthouse, T. A. (1996). *The processing-speed theory of adult age differences in cognition* (p. 26).

Sheskin, M., & Keil, F. (2018). TheChildLab.com a video chat platform for developmental research. *PsyArxiv.*. https://doi.org/10.31234/osf.io/rn7w5

Singer, M., & Tiede, H. L. (2008). Feeling of knowing and duration of unsuccessful memory search. *Memory and Cognition, 36*(3), 588–597. https://doi.org/10.3758/MC.36.3.588

Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition, 96*(1), B1–B11. https://doi.org/10.1016/j.cognition.2004.07.004

Stipek, D., & Iver, D. M. (1989). Developmental change in children's assessment of intellectual competence. *Child Development, 60*(3), 19.

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. https://doi.org/10.1177/0956797615616255

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