Is recursive "mindreading" really an exception to limitations on recursive thinking?

Wilson, Ross; Hruby, Ales; Perez-Zapata, Daniel; van der Kleij, Sanne W; Apperly, Ian A

DOI:
10.1037/xge0001322

Journal of Experimental Psychology: General

Is Recursive “Mindreading” Really an Exception to Limitations on Recursive Thinking?

Ross Wilson, Ales Hruby, Daniel Perez-Zapata, Sanne W. van der Kleij, and Ian A. Apperly

Online First Publication, March 9, 2023. https://dx.doi.org/10.1037/xge0001322

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

https://doi.org/10.1037/xge0001322

Is Recursive "Mindreading" Really an Exception to Limitations on Recursive Thinking?

Ross Wilson, Ales Hruby, Daniel Perez-Zapata, Sanne W. van der Kleij, and Ian A. Apperly
School of Psychology, University of Birmingham

The ability to mindread recursively—for example, by thinking what person 1 thinks person 2 thinks person 3 thinks—is a prime example of recursive thinking in which one process, representation, or idea becomes embedded within a similar one. It has also been suggested that mindreading is an exceptional example, with five recursive steps commonly observed for mindreading, in comparison with just one or two in other domains. However, conceptual analysis of existing recursive mindreading tasks suggests that conclusions about exceptional mindreading are insecure. Revised tasks were devised to provide a more rigorous test of recursive mindreading capacity. Study 1 ($N = 76$) found significantly worse performance at level-5 recursive mindreading on the revised tasks (17% correct) compared with the original tasks (80% correct), and no effect of moderate financial bonuses for good performance. Study 2 ($N = 74$) replicated poor performance at level-5 recursive mindreading on the revised tasks (15% correct) in the absence of bonuses, but found better performance (45% correct) when participants were offered large bonuses for accuracy, encouraged to take as much time as needed, and assisted with a strategy for recursive reasoning. These findings suggest that, like recursive thinking in other domains, recursive mindreading is effortful and limited. We discuss how the proposed role for high levels of recursive mindreading in communication, culture, and literature might be reconciled with these limitations.

Keywords: mindreading, theory of mind, mentalizing, recursion, higher-order

Supplemental materials: https://doi.org/10.1037/xge0001322.supp

Recursive thinking is often held to be a critical and distinctive feature of human cognition, underwriting human-unique abilities for language, social interactions, and social institutions and culture (e.g., Camerer, 2003; Corballis, 2014; Dunbar, 2003; Hauser et al., 2002; Scott-Phillips, 2014; Sperber & Wilson, 1986; Tomasello, 2014). For example, in language, it is claimed that recursive embedding of structures makes it possible to say an unlimited number of things with limited linguistic elements (e.g., Hauser et al., 2002; though see e.g., de Boer et al., 2012). In strategic reasoning and negotiation, the right thing for person 1 to do depends upon what person 2 will do, which itself depends on person 2's judgment of what person 1 will do, and so on (e.g., Camerer, 2003; De Freitas et al., 2019). Moreover, in literature, plots frequently appear to depend upon characters varying in what they know or think and in what they know and think about each other, and so forth (e.g., van Duijn et al., 2015; Zunshine, 2006).

When interpreting claims in any of these literatures, it is useful to distinguish between the recursive potential of the representational systems, the level most commonly evident in performance, and the capacity of humans' use of recursion beyond what is most frequently observed. The recursive *potential* of mindreading, embedded grammatical constructions, or levels of strategic thinking has no principled limit. However, *performance* often appears severely limited. Evidence from large linguistic corpora suggests that embedding of grammatical clauses beyond a single recursive step is extremely rare in human speech and writing (Karlsson, 2007). Evidence from a variety of economic games converges to suggest that people most often reason at either one or two levels of recursion (e.g., Bosch-Domènech et al., 2002; Camerer et al., 2004). Finally, in terms of the range of humans' *capacity* for recursion, corpus data suggest that clausal embedding does not exceed two levels of recursion in spoken language or three levels in written language (Karlsson, 2007). For strategic reasoning, group mean performance up to level 4 or 5 has been observed in exceptional circumstances when participants are selected for high analytic skills or incentivized with high rewards for success (Camerer et al., 2004). In summary, while there is no principled limit on the number of recursive embeddings, evidence from linguistics and behavioral economics converges to suggest that people rarely exceed 1–2 embeddings—a finding attributed to rapidly increasing processing complexity at higher levels (e.g., Camerer, 2003; Karlsson, 2007; Klindt et al., 2017; Levinson, 2013).

Ian A. Apperly https://orcid.org/0000-0001-9485-563X

Data for the two studies are openly available at the following address: https://osf.io/ac738/?view_only=6ea1d767f9d442bc9d4d2356539fd226. No aspect of the studies was preregistered.

This work was supported by a grant from the Economic and Social Research Council, UK (ES/R005028/1).

Open Access funding provided by University of Birmingham: This work is licensed under a Creative Commons Attribution 4.0 International License (CC-BY). This license permits copying and redistributing the work in any medium or format, as well as adapting the material for any purpose, even commercially.

Correspondence concerning this article should be addressed to Ian A. Apperly, School of Psychology, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom. Email: i.a.apperly@bham.ac.uk

A notable exception to this pattern is recursive "mindreading" about other people's mental states, where it is claimed that most adults achieve four or five embeddings, with a significant minority achieving 6–8 levels (e.g., Oesch & Dunbar, 2017; O'Grady et al., 2015). Such highly recursive mindreading has been suggested to explain social phenomena as diverse as human pragmatic abilities (Scott-Phillips, 2014; Sperber & Wilson, 1986), the complexity of human social networks (Dunbar, 2003), and the existence of religion (Dunbar, 2017) and literary fiction (Oatley, 2011; Zunshine, 2006). In the present studies, we re-examine the evidence for this exceptional capacity and find it lacking. We develop new methods that are capable of detecting varying levels of recursive reasoning and find that most participants progress beyond two levels of recursion only when offered high levels of incentives and support.

The Empirical Case for Exceptional Recursive Mindreading

In a seminal paper, Kinderman et al. (1998) devised narratives involving multiple characters and questions about their thoughts and feelings at increasing levels of recursion. The stories and questions have been refined over subsequent studies (e.g., Paal & Bereczkei, 2007; Stiller & Dunbar, 2007), most recently by Oesch and Dunbar (2017). For example:

> Frank, Betty and Brian all share a three-bedroom house situated close to the high street. Frank is very lazy and never helps with the housework. One morning Brian and Betty decide that they should make him go and do the weekly shop. Brian walks into Frank's bedroom and wakes him up and tells Frank what they need. Betty would like some eggs and bacon for a cooked breakfast and Brian would like lettuce and tomatoes to make a salad for lunch. Frank says he can remember the list so Brian gives Frank £10 and leaves the house and goes to work. Betty reminds Frank that she only wants him to buy free range or organic produce. Frank becomes confused so Betty writes down the shopping list for him and tells him that Brian will not mind what type of lettuce Frank buys. Frank gets dressed and walks to the shops. When Frank gets to the supermarket he finds he has left the shopping list in his dressing gown and becomes worried that Brian might shout at him.

Test questions ranged from one to nine levels of recursion and required participants to judge whether a statement was true or false. For example:

> Betty knew that Brian was worried that Frank might think he was stupid and shout at him for incompetence. [Intended to require 3 levels of recursion; correct answer = "false"]

> Frank realised that Betty wanted Brian to think that Frank believed that Betty knew Frank was worried that Brian knew Frank was confused with respect to Betty's desire that Frank buy free range or organic produce for a cooked breakfast with bacon and eggs for Betty and a lunch salad with lettuce and tomatoes for Brian. [Intended to require 9 levels of recursion; correct answer = "false"]

Performance on such questions frequently exceeds performance on questions designed to test recursive syntax and memory for the story (Oesch & Dunbar, 2017; Stiller & Dunbar, 2007; though see O'Grady et al., 2015) supporting the idea that recursive mindreading might be exceptional. Early versions of the task have been criticized for inadequately testing recursive abilities (O'Grady et al., 2015), but amended tasks designed to address these concerns have demonstrated similarly high levels of recursive ability (Oesch & Dunbar, 2017; O'Grady et al., 2015). Moreover, people scoring higher on these tasks tend to have larger social networks (Lewis et al., 2011; Powell et al., 2010; Stiller & Dunbar, 2007), larger gray matter volumes in brain areas associated with social ability (Lewis et al., 2011), are more cooperative (Paal & Bereczkei, 2007), and have lower traits for primary and secondary psychopathy (Vonk et al., 2015), suggesting that performance on the tasks captures variance that is relevant for social behavior. However, it may still be that the variance in performance on these tasks reflects something other than the variance in recursive mindreading.

Why Recursive Mindreading Might Be More Difficult Than Is Sometimes Supposed

Despite current empirical evidence, consideration of the combinatorics of recursion provides reason to be surprised by participants' apparent success. Consider one of the simplest possible cases of two agents, A1 and A2, playing hide-and-seek between two locations, L1 and L2. A spectator of this game has four possible "level-1" mental states to consider: A1 thinks L1 is the hiding location; A1 thinks L2 is the hiding location; and likewise A2 can think either L1 or L2 is the hiding location.[1] If the spectator now imagines that agents might also consider what each other thinks, each level-1 possibility yields two further "level-2" possibilities: For example, A1 (thinking L1 is the hiding location) thinks that A2 thinks it is L1; and A1 (thinking L1 is the hiding location) thinks that A2 thinks it is L2. The number of possibilities admitted by this scenario doubles with each level of recursion that the spectator entertains, so while a spectator reasoning at level 1 has just four possibilities to entertain, a level-5 reasoner has 64. Moreover, the rate of increasing complexity grows rapidly as the number of agents and locations (or other belief contents)[2] increases. The general case is described by the following formula: $\text{Possibilities} = AL(AL - L)^{(R-1)}$, where A is the number of agents, L is the number of locations, and R is the level of recursion.

Consider the example above from Oesch and Dunbar (2017), with three people (Frank, Betty, and Brian), and let us make the (greatly) simplifying assumption that there are only two belief contents that anybody could entertain: either "Frank [should] buy free range or organic produce for a cooked breakfast with bacon and eggs for Betty and a lunch salad with lettuce and tomatoes for Brian," or he should not. Even with this simplifying assumption, a level-9 reasoner is confronted with 393,216 unique combinations of mental states, 262,164 of which are level-9 combinations. Previous research may not have taken full account of these combinatorics, and may not have intended to claim that participants were entertaining such a possibility space when they succeeded in giving a correct answer to level-9 questions. However, it is also unclear what alternative claim was being made. Spelling out the combinatorics of recursive mindreading illustrates the complexity inherent in these tasks, and serves two important purposes. First, it highlights the care that is necessary in understanding how participants arrive at their answers when asked recursive mindreading questions—a problem we

[1] A more natural way to describe this might be that A will decide to hide at either 1 or 2, and B will decide to seek at either 1 or 2. This "behavioural" redescription of the problem nonetheless encounters the same recursive complexity as a "mindreading" description.

[2] The notion of "locations" serves to elaborate the case of a hide-and-seek game. The same principles extend to any other belief contents.

address next. Second, it highlights the need to reconsider whether recursive mindreading is as widespread as some theories suggest and consider how participants handle the vicious combinatorics when they do engage in recursive mindreading—a challenge we address in the General Discussion.

Criteria for Testing Recursive Mindreading and Problems With Existing Tasks

The idea that participants must entertain the full possibility space for recursively embedded mental states sets a high bar for what it might mean to reason recursively, which might be both difficult to meet, and difficult to test. We propose a pragmatic approach, which is to require that a test of recursive mindreading should be sensitive to detect participants who are not reasoning at the intended level of recursion. This resembles the criterion used in standard work on level-1 and level-2 recursive mindreading in children (e.g., Perner & Wimmer, 1985; Wimmer & Perner, 1983). Aside from guessing, a plausible unintended way of answering recursive mindreading questions would be for participants to reason only part way through the recursive chain. For example, in the level-9 example above participants might progress no further than successfully identifying what Frank realizes Betty wants Brian to think (i.e., starting from the first mental state and working forward), or alternatively whether Brian knew Frank was confused with respect to Betty's desire (i.e., starting from the last mental state and working backward). To test recursive mindreading, it is essential that such "partial-chain" strategies would lead participants to an error, which could therefore be distinguished from a correct answer. In Table 1, we illustrate this point with the complete set of partial chains for this level-9 question. We have offered glosses with minor variations in wording to aid readability while preserving the underlying meaning.

In this example, the great majority of responses based upon incorrect "partial-chain" reasoning would yield the answer "False" and so would not be distinguishable from a correct response to the original level-9 test question. Similar problems affect more than half of all questions employed by Oesch and Dunbar (2017) and O'Grady et al. (2015). Interpretation of the remaining questions is affected by other problems such as ambiguous pronouns in the test question or insufficient clarity in the story about characters' knowledge or intentions in relation to one another (see supplemental materials). The false positives that could result from these problems would serve to inflate the apparent level of recursion achieved by participants.

In summary, despite its widespread influence, existing work fails to provide sound evidence that recursive mindreading is an exception to the general pattern of more limited recursion observed in other domains of cognitive science.

New Stimuli to Address the Criteria

We created new stimuli that included more perspective differences between characters and so greatly increased our ability to detect errors arising from partial-chain responses during recursive

Table 1
Illustration of Each "Forward" and "Backward" Partial Chain Derived From a Question Designed to Require Nine Levels of Recursion

Intended level of recursion	Original question	Correct answer
9	Frank realized that Betty wanted Brian to think that Frank believed that Betty knew Frank was worried that Brian knew Frank was confused with respect to Betty's desire that Frank buy free range or organic produce for a cooked breakfast with bacon and eggs for Betty and a lunch salad with lettuce and tomatoes for Brian.	False
	Forward partial chains	
1	Frank realized that he should buy free range…	True
2	Frank realized that Betty wanted him to buy free range…	False
3	Frank realized that Betty wanted Brian to think that Frank should buy free range…	False
4	Frank realized that Betty wanted Brian to think that Frank believed that Frank should buy free range…	False
5	Frank realized that Betty wanted Brian to think that Frank believed that Betty knew that Frank should buy free range…	False
6	Frank realized that Betty wanted Brian to think that Frank believed that Betty knew that Frank was worried that Frank should buy free range…	False
7	Frank realized that Betty wanted Brian to think that Frank believed that Betty knew that Frank was worried that Brian knew that Frank should buy free range…	False
8	Frank realized that Betty wanted Brian to think that Frank believed that Betty knew that Frank was worried that Brian knew that Frank was confused that Frank should buy free range…	False
	Backward partial chains	
1	Betty desired Frank to buy free range…	True
2	Frank was confused with respect to Betty's desire that he buy free range…	False
3	Brian knew that Frank was confused with respect to Betty's desire that he [Frank] buy free range…	False
4	Frank was worried that Brian knew that Frank was confused with respect to Betty's desire that he [Frank] buy free range…	False
5	Betty knew that Frank was worried that Brian knew that Frank was confused with respect to Betty's desire that he [Frank] buy free range…	False
6	Frank believed that Betty knew that Frank was worried that Brian knew that Frank was confused with respect to Betty's desire that he [Frank] buy free range…	False
7	Brian thought that Frank believed that Betty knew that Frank was worried that Brian knew that Frank was confused with respect to Betty's desire that he [Frank] buy free range…	False
8	Betty wanted Brian to think that Frank believed that Betty knew that Frank was worried that Brian knew that Frank was confused with respect to Betty's desire that he [Frank] buy free range…	False

Note. Note how 14 of the 16 possible simplifications yield the same answer as the intended level-9 question.

mindreading. Conceivably, increasing the number of perspectives could make the task easier by decreasing the possibility for confusion among characters with the same perspective. Conversely, if previous methods generated false positives by conflating the answers for correct and incorrect levels of recursive reasoning, then increasing the number of perspectives should result in decreased performance. Our first study compared performance on the Original and Revised tasks using level-5 recursive mindreading questions and examined the effect of moderate variation in the level of financial incentives for successful task performance. We predicted that participants would perform less well on the Revised tasks, but that incentives would improve performance disproportionately on these Revised tasks by incentivizing the effort necessary to perform well in the condition that we expected to be harder. In Study 2, we sought to replicate the pattern observed on Revised tasks and to examine whether performance on the tasks could be improved with a combination of encouragement, larger incentives, and scaffolding a strategy for success at recursive reasoning.

Study 1

Data from Studies 1 and 2 are openly available via a link in the References section.

Method

Full study materials are presented in appendices. The present studies were not preregistered.

Participants

Our primary objective was to detect whether participants could respond correctly to level-5 questions at levels above chance in either the Original or Revised tasks. We expected a baseline rate of 14.3% correct answers if participants selected from the seven alternative choices at random. Power analysis implemented in G*Power (Faul et al., 2007) indicated that to detect the performance of 50% or more correct answers with 95% power would require at least 18 participants in each cell of the design.

Seventy-six people were recruited using MTurk via CloudResearch: 31 identified as female, 45 as male; age range 18–66 years, $M = 31$ years. Participation was limited to persons within the United Kingdom; among these, 51 identified as White from the United Kingdom, 2 as White Polish, 1 as White Estonian, 1 as White Italian, 1 as White European, 1 as White and Black African, 2 as White and Asian, 2 White British and Asian, 1 White British and Native American, 3 as African, 1 as Arab, 2 as Chinese, 1 as Korean, 1 as Hongcongese, 2 as Indian, 1 as Bangladeshi, 2 as Pakistani. Participants were given a total of $8 (£6) for completing the whole study.

Stimuli

Original Stories. The "Date story" and "Interview story" were adapted from Oesch and Dunbar (2017; see Appendix A). The main text for each story was used unaltered, but a single level-5 recursion question was presented because Oesch and Dunbar identified this as the most frequent maximum level for participants. Additionally, whereas Oesch and Dunbar's questions employed completed statements that participants had to judge true or false, we employed unfinished statements that participants had to complete by choosing from seven possible alternatives. This format matched our Revised Stories and substantially reduced the rate of false positives due to guessing from 50% to 14.3%. The Date story plus level-5 question and seven alternative answers comprised 289 words. The Interview story and questions comprised 290 words. True/false check questions were used to check participants' engagement with the task. These comprised six and 11 words for the Date and Interview stories, respectively.

Revised Stories. The "Garden story" and "Chef story" were devised to create a larger number of perspective differences between characters than in previous work on recursive mindreading, and so to reduce the chances that partial-chain reasoning strategies would produce the same answer as correct, full-chain reasoning. We give one example story and level-5 question below, and in Appendix A provide both stories together with explanations of how the forced choice alternative answers corresponded to different partial-chain reasoning strategies.

Garden Story.

> This is a story about two siblings, Peter and Mary, and their parents and Gran. One day, Peter and Mary are playing in the back garden of their house enjoying the attention of the adults who are sat in deck chairs watching them. Peter is enthusiastically drawing. "I'm going to the loo" says Peter. He leaves his drawing in a little box and goes into the house. When Mary sees that Peter has left, she decides to play a trick. She takes Peter's drawing from the box and puts it behind a bush, and then continues playing with her back to the bush. Mum, Dad, and Gran are amused by the trick, when suddenly, a strong wind blows the drawing and moves the picture under the family car without Mary seeing. The doorbell rings and Mum goes inside to answer. Dad tells Mary it would be better to pick up the picture and leave it in the kitchen, and Gran says she'll go with her. On their way inside they agree that the picture is too crumpled so instead Gran suggests that Mary takes it to Mum's study while she makes a cup of tea. Instead Mary finds Peter on his way out of the toilet, and gets him to help her find a heavy book to press the picture in, and they leave it on the shelf in the living room. After Mary has gone Peter sees Mum, and tells her how Mary helped him press his picture, but now he has hidden it safely in his bedroom.
>
> Level-5 question: Where does Peter think Mary thinks Mum thinks Dad thinks Gran thinks the picture is?

Our Revised stories enabled us to pose a level-5 question for which partial-chain responses yielded unique answers for levels 1–4 when working backward from the last mental state, and unique answers for levels 1–2 when working forward from the first mental state in the chain. Reasoning forward to the remaining levels 3 and 4 yielded the same answer as the correct level-5 response, with the result that these new tasks still tend to overestimate true performance levels, though much less severely than previous methods. We failed to overcome this limitation without adversely increasing the length and complexity of the stories. As will become apparent, this had only a minor impact on interpreting the effects we observed. The Garden story plus level-5 question and seven alternative answers comprised 298 words. The Chef story and questions comprised 459 words. The Garden and Chef check questions comprised 11 and seven words, respectively.

Design and Procedure

Participants took part in an online study comprising two sets of tasks, with separate instructions before each set. The four recursion stories comprised the first set of tasks, and only these tasks will be described here. The experiment was approved by the University of Birmingham ethics committee under program ERN_09_719.

Every participant completed both Original stories and both Revised stories. Stories were presented in a fixed order, beginning with the two Original stories (Date then Interview) and followed by the Revised stories (Chef then Garden). Presenting the Original stories first enabled us to gauge performance on these stories without any possibility of influence from our Revised stories. Performance on the Revised stories could have been influenced by prior completion of the Original stories, and this limitation was addressed in Study 2 in which participants only completed Revised stories. In the No-bonus condition participants were instructed to take their time and answer questions to the best of their ability. In the Bonus condition, participants were additionally told that they would receive a $4 bonus if they answered all questions correctly, and were guided to invest an appropriate level of effort with information about the average time taken by participants who had answered correctly in pilot work (Date = 2 min, Interview = 2.5 min, Chef = 4.5 min, Garden = 2 min). Advertisements were posted separately and simultaneously for the bonus condition, with the possibility of earning extra money being made explicit during recruitment to the bonus condition.

After each story participants were shown one level-5 recursive question and presented with seven possible answers, only one of which was correct. Incorrect answers were always plausible scenarios involving the characters and plot of the story. For the revised stories, incorrect answers also corresponded to the outcomes of partial-chain reasoning, as described above. For the reasons already discussed, this arrangement was impossible for the Original stories. The only exception to this arrangement occurred in the Chef story for Study 1, where an editing error resulted in one of the alternative answers being a location ("storage room") that did not exist in the story. Selecting this answer would be incorrect, and so if there were any effect of this editing error it would be to make it easier for participants to exclude this answer when trying to select the correct answer.

For each story, we randomized the order in which the possible answers were listed. To maximize participants' opportunities for success, each story remained on the screen when the recursive question was presented, obviating any need to memorize the story. Note that this arrangement differs from the majority of previous studies, in which the story or video scenario was inaccessible while participants answered the critical questions. Following the recursion question, and with the story no longer accessible, participants were posed a simple check question to be judged correct or incorrect. This afforded a basic check on participants' engagement with the task and text.

The time taken to answer all questions was recorded along with answers. Participants were removed from further analysis if they: answered two or more check questions incorrectly (3 from the Bonus condition); answered one or more story questions in under 20 s (2 from the No-bonus condition); or took longer than the mean time +3 standard deviations to answer one or more of the recursive questions (1 from No-bonus; 4 from Bonus). This resulted in 33 participants per condition for analysis. *M* (*SD*) total completion time for the Bonus condition was 857 (405) s, and for the No-bonus condition was 806 (373) s.

Results and Discussion

We first examined whether participants' rate of correct answers varied between the Original and the Revised recursive mindreading tasks, and whether this pattern was influenced by the presence of a bonus for good performance. We fitted a mixed effects model using the *glmer* function within the lme4 package (Bates et al., 2015) in R (R Core Team, 2020), with Accuracy on each trial as dependent variable, Condition (Original, Revised) and Bonus (no incentive, incentive) as fixed effects and a by Participant random slope for Condition. A by-Item random effect resulted in estimation problems and was therefore not included in the model. Contrasts for both predictors were defined using deviation coding, Bonus (−0.5, 0.5) and Condition (−0.5, 0.5). The interaction term between Bonus and Condition was coded by multiplying contrasts for the two factors.

The final model included a significant main effect for Condition, $\chi^2(1) = 113.59$, $p < .001$, indicating that participants performed better in the Revised condition than in the Original condition. The effect of Bonus was not significant, $\chi^2(1) = 0.02$, $p = .88$. However, there was a significant interaction between Condition × Bonus, $\chi^2(1) = 5.97$, $p = .02$.

We followed up on the interaction using the emmeans package in R (Lenth, 2021), using Tukey-adjusted *p*-values. Performance on the Revised condition was better than the Original for both the no-bonus, $b = 2.46$, $SE = 0.48$, $z = 5.13$, $p < .001$, and bonus condition, $b = 4.10$, $SE = 0.62$, $z = 6.61$, $p < .001$. The effect of Bonus was not significant in the Revised, $b = 0.94$, $SE = 0.53$, $z = 1.78$, $p = .28$ and the Original condition, $b = -0.70$, $SE = 0.48$, $z = -1.45$, $p = .47$. For continuity with the analysis strategy used in Study 2, we conducted a second follow-up analysis in which the number of correct answers in each cell of the design was summed, yielding a range from 0 to 2. Consistent with the results above, separate analysis of performance on the Original and Revised tasks showed no effect of Bonus: Original task, No-bonus (74% correct; $M = 1.48$) versus Bonus (84% correct; $M = 1.70$), $U(N_{\text{no-bonus}} = 33, N_{\text{bonus}} = 33) = 637$, $Z = 1.40$, $p = .162$; Revised task No-bonus (23% correct; $M = 0.45$) versus Bonus (11% correct; $M = 0.21$), $U(N_{\text{no-bonus}} = 33, N_{\text{bonus}} = 33) = 438$, $Z = -1.70$, $p = .09$. Thus, although the omnibus analysis revealed a significant interaction between Condition and Bonus, this was due to numerically better performance when a bonus was available in the Original condition and numerically worse performance when a bonus was available in the Revised condition. Since neither effect was individually significant, and since this pattern was not predicted, we will not consider this further.

Finally, since performance in the Revised condition was notably poor in comparison with the Original condition, we examined whether participants were performing better than expected if they had guessed the correct answer in each story. Using the binomial distribution for 33 participants and a guessing rate of 1/7 (for the seven alternative response options), above-chance performance would be indicated by nine or more participants answering correctly. This was achieved for just one of the two Revised stories (Chef story) in the No-bonus condition, where nine participants gave the correct answer.

In summary, in line with previous research, we found high levels of success on level-5 questions in the Original condition. However, performance was significantly poorer on the Revised stories, on which we had taken steps to reduce the chances of false positive answers. Performance was not influenced by the availability of a bonus for accuracy in either condition. These results are consistent with our task analysis that shows the Original tasks to be prone to a high level of false positive responses, meaning that they could systematically overestimate recursive abilities. However, the fact that level-5 performance was above chance for only one of our Revised tasks is also consistent with the possibility that either our tasks or our design did not yield a fair test. It is possible that we had not succeeded in constructing tasks that were solvable in principle, or that performance was disrupted by prior completion of the Original tasks. Therefore, the first objective of Study 2 was to check whether the poor performance on the Revised stories would replicate in a second sample in which participants only viewed the Revised stories. The second objective checked that the tasks were solvable in principle by testing whether participants could give correct answers in circumstances designed to be highly favorable, where there was a combination of encouragement, higher incentives, and scaffolding.

Study 2

Method

Participants

Our primary objective was to detect whether participants could respond correctly to level-5 questions at levels above chance in either version of the Revised task. We expected a baseline rate of 14.3% correct answers if participants selected from the seven alternative choices at random. Power analysis implemented in G*Power (Faul et al., 2007) indicated that to detect performance of 50% or more correct answers with 95% power would require at least 18 participants in each cell of the design.

A total of 74 people were recruited through Prolific.co. The availability (or not) of a bonus was not apparent to participants at the point that they chose to enter the study. Fifty-seven participants identified as female, 16 male, 1 nonbinary; age range 18–63 years, M = 33 years. Participation was limited to English-speaking people within the United Kingdom; among these, 64 identified as White from the United Kingdom, 2 as having a multiple ethnicity background, 1 White and Black, 1 White Irish, 1 White Jewish, 2 Indian, 1 Pakistani, 1 Chinese, 1 Caribbean. Participants were paid £3.

Stimuli

The Revised stories were very similar to those used in Study 1, with minor amendments to correct the editorial error described for Study 1 and to ensure that there were clear correct answers for recursive questions at levels 1–4, as well as at level 5 (see Appendix B for full text).

Design and Procedure

Participants took part online, and completed one of two conditions via Qualtrics: revised stories with a bonus, encouragement, and scaffolding to promote correct answers—for brevity, we label this the "Bonus" condition (34); revised stories without a bonus (40)—the "No-bonus" condition. The experiments were approved by the University of Birmingham ethics committee under program ERN_09_719.

The No-bonus condition was carried out in the same manner as study 1, with the exception that participants only viewed the two Revised stories (Chef then Garden).

In the Bonus condition, we sought to encourage good performance by acknowledging the complexity of the questions and encouraging participants to take their time and use a pen and paper if they wished (see Appendix B for full instructions). To incentivize good performance, participants were told that they would receive a bonus of £10 for correctly answering all questions. We sought to scaffold good performance in two ways. First, participants read an example story that was simpler than the Revised stories and were asked relatively simple level-1 and level-2 recursive questions and a check question. Second, for each Revised story, participants built up gradually to the level-5 question. This began with a level-1 question concerning the last mental state in the recursive chain of the level-5 question (e.g., where Michael thought the prawn cocktail was located) followed by a level-2 question in which the next mental state was added (e.g., where Nik thinks Michael thought the prawn cocktail was located), and so on. Although each question had a different correct answer, and each could be answered entirely independently, getting participants to start from level 1 illustrated one effective strategy that participants could build upon for solving higher-level questions. For each question, participants chose one answer from among the same seven alternative forced choice responses. Each question had a unique response, but since participants did not know this and received no feedback, it was difficult to use responses to simpler questions to eliminate possible answers to more complex questions.

We used the same check questions as in Study 1, and since the instructions were more elaborate in Study 2, an additional attention check was placed before the stories began in which participants were asked not to choose either of the options given ("correct/incorrect") to show they were paying attention to the instructions. Participants were removed from further analysis if, for either story, they answered in less than 20 s (2 participants from the No-bonus condition), or if they took longer than mean +3 standard deviations to answer (1 removed from No-bonus, 1 from Bonus). Participants were also removed if they selected any answer to the attentional check question (4 from No-bonus) or incorrectly answered any specific story check question including the practice question (3 from No-bonus, 3 from Bonus), leaving 30 participants in each condition. M (SD) total completion time for the Bonus condition was 1,027 (457) s, and for the No-bonus condition was 395 (328) s.

Results

Our first question was whether the poor performance we observed on the Revised Task in the No-bonus condition of Study 1 would be replicated in a second sample who had not previously completed the Original task. Using the binomial distribution for 30 participants and a guessing rate of 1/7 (for the seven alternative response options), above-chance performance would be indicated by nine or more participants answering correctly. As in Study 1, performance was

relatively poor (Garden Story 7/30 correct; Chef Story 2/30 correct) and, in this case, was not above-chance levels for either story.

Our second question was whether the performance could be improved through a combination of incentives and scaffolding. The number of correct answers on level-5 questions for the two stories was summed for participants who received no bonus, and for those receiving bonus and scaffolding, yielding a range from 0 to 2. A Mann–Whitney *U* test showed a significant difference between performance in the No-bonus (15% correct; $M = 0.30$) and the Bonus (45% correct; $M = 0.90$) conditions: $U(N_{\text{no-bonus}} = 30, N_{\text{bonus}} = 30) = 648$, $Z = 3.25$, $p = .001$.

Our third question was how participants performed across level-1 to level-5 questions in the presence of bonus and scaffolding. The number of participants giving correct answers to each question for the two stories is plotted in Figure 1. Participants performed above-chance levels (9/30 correct answers) for every question apart from the level-5 question for the Chef Story (8/30 correct answers). However, there was also a clear trend for lower performance with increasing levels of recursion, with a majority of participants giving the correct answer on both stories for levels 1 and 2, but not for higher levels.

General Discussion

It has been proposed that recursive mindreading is an exceptional case, in comparison with evidence of much more limited recursive thinking in other cognitive domains (e.g., Corballis, 2014; Dunbar, 2003, 2017; Scott-Phillips, 2014). In the current studies, we critically evaluated the methods used to support these claims and provided new empirical evidence that success on tasks that require mindreading at high levels of recursion is much less common than previously suggested.

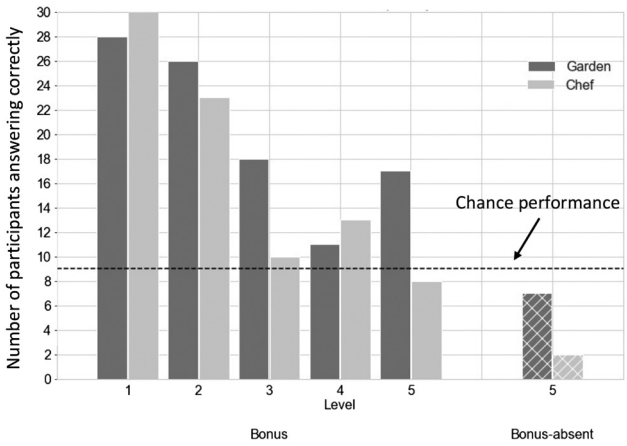
Previous studies have sought to pose mindreading questions at up to nine levels of recursion (e.g., Oesch & Dunbar, 2017; O'Grady et al., 2015) and found that the great majority of adults succeed at level-4 and -5 questions, while a notable minority perform well even on level-8 questions. In Study 1, we replicated similarly high levels of success (80% overall) on level-5 questions modeled on the Original tasks used in previous research. However, in the introduction, we also described how participants could often succeed on these questions even if they reasoned with shorter recursive chains (also see supplemental materials). This potential for false positives means that previous studies are prone to systematic bias that may have led recursive mindreading abilities to be overestimated.

We addressed this problem by developing new stimuli for a test of level-5 recursive mindreading. These stimuli greatly reduced the potential for false positives by ensuring that participants would be less likely to arrive at the correct "level-5" answer by reasoning with attenuated versions of the intended level-5 recursive chains. In Study 1, our Revised task was significantly harder than the Original task, with an overall success rate of just 17%, above-chance performance on just one of the two stories, and no effect of financial incentives for success. Study 2 replicated findings from Study 1, with participants providing correct answers 15% of the time on the Revised task in the "no-bonus" condition. In the presence of bonus, scaffolding, and encouragement in the "bonus" condition, participants succeeded 45% of the time on level-5 questions, though the performance was above chance on level-5 questions for just one of the two stories. Overall, Studies 1 and 2 suggested that the Original tasks were substantially overestimating recursive mindreading abilities, and the Revised tasks provided only limited evidence of success on level-5 recursive mindreading even in highly favorable circumstances.

These findings for mindreading converge with other evidence on recursion in language and strategic reasoning: each has the *potential* for unlimited recursion, but the number of levels observed in typical *performance* is limited to a very few recursive steps. In the current study, a minority of participants demonstrated a *capacity* for at least level-5 recursion when motivated and supported to do so, which is also consistent with findings from strategic reasoning (e.g., Camerer et al., 2004). Yet higher levels are surely possible in principle given sufficient motivation, time, and scaffolding, for example by using pen and paper, spreadsheets, or other external symbolic media to help construct relevant partial chains, track relationships between characters, and map relevant parts of the possibility space. However, the conditions for demonstrating such capacity are unlikely to be representative of everyday activities. Previous work in linguistics and economics has interpreted similar patterns of limited performance despite unlimited recursive potential in terms of the severe limitations imposed by human working memory capacity to cope with the nonlinear increase in possibilities with each recursive step (e.g., Camerer et al., 2004; Karlsson, 2007). The present findings suggest that the same is likely to be true for mindreading.

Figure 1

Number of Participants (Out of 30 in Total for Each Group) Answering Correctly for the Recursion Questions at Each Level in the Two Conditions of Study 2



Do the Revised Tasks Underestimate Recursive Mindreading Ability?

Even if the Original tasks are an unreliable indicator of recursive mindreading ability, could it also be that our Revised task underestimates recursive mindreading? It is notable that the Revised task involved more characters (five per Revised story vs. four per Original story), and while the two Original stories and one Revised story had word counts ranging from 289 to 298, the second Revised stories was substantially longer, at 459 words. However, there are also reasons for thinking that our methods in the present studies encouraged success. First, we gave participants every chance

to succeed by allowing them unlimited time to read and re-read the scenarios while answering test questions. This contrasts with most previous studies that did not allow participants to revisit the story or video while answering. We also encouraged participants to take their time and use pen and paper to support their reasoning. Second, we employed stringent criteria by only including data from participants who performed perfectly on attention check questions. Third, we offered financial incentives for success. Fourth, in the "bonus" condition of Study 2 participants answered a series of related mindreading questions that scaffolded a recursive chain starting with level-1 and building toward the final level-5 question. Fifth, as noted in the introduction our tasks retain some bias toward false positives, because two of the eight possible partial versions of the intended level-5 reasoning chain would yield the same answer as correct level-5 reasoning. Finally, there is clear evidence that the tasks are solvable in principle, since participants *did* succeed at above-chance levels at level 5 on the Chef story in the no-bonus condition of Study 1, and on levels 1–5 for the Garden story and levels 1–4 for the Chef story when correct performance was incentivized and scaffolded in Study 2.

It is notable that performance on the level-2 questions was less good than might be expected from the evidence that many 5- to 6-year-old children succeed at level-2 recursive mindreading (e.g. Sullivan et al., 1994). Note, however, that standard level-2 tasks are designed with the minimum complexity necessary to ask a level-2 question. In contrast, our Revised tasks were designed with the minimum complexity necessary to ask a valid level-5 question. The required number of characters, locations, and narrative steps resulted in a much larger possibility space even for level-2 questions in these tasks. We think this is likely to explain why adults were not simply at the ceiling on level-2 questions. The fact that performance was nonetheless far above chance on these questions, the level-1 questions, and on the questions from the Original task, is reassuring evidence that our online samples of participants were capable of giving reliable responses.

How Do People Handle Situations That Entail Recursive Mindreading?

One of the motivations for thinking that humans might be exceptionally good at recursive mindreading comes from conceptual analysis of the pragmatics of communication. Classic work by Grice (1969) established the idea that even the simplest acts of ostensive communication rest upon hearers representing not only what speakers might know, think, or feel, but also their intention to communicate this information, and their intention that this communicative intention be recognized by the hearer. How should the present findings be reconciled with these and other considerations about the role of recursive mindreading?

As Grice recognized, this analysis of communication risks making onerous demands on the recursive mindreading abilities of communicators. This has been met with three kinds of response. First, some researchers have questioned whether pragmatic inferences really do require recursive mindreading (e.g., Clark, 1996; Moore, 2017; van Duijn, 2016). Second, some researchers accept that communication requires recursive mindreading but propose that these needs are met by a specialized cognitive "module" (e.g., Sperber, 2000; Sperber & Wilson, 1986). To the extent that modular processing is inaccessible for other purposes (see e.g., Carruthers, 2006; Fodor, 1983), the present findings might be considered irrelevant because they do not concern the kind of recursive mindreading that is involved in communication. Third, some researchers have argued that general-purpose recursive mindreading is indeed sufficient to meet the demands of Gricean communication (e.g., Scott-Phillips, 2014), as well as the potential demands of narrative production and comprehension, social interaction, and culture (e.g., Dunbar, 2003, 2017; Oatley, 2011; Zunshine, 2006). The present findings remove empirical support for extraordinary recursive mindreading, and so call this third interpretation into question.

Of course, this lack of empirical support does not invalidate conceptual arguments for high levels of recursive mindreading in language pragmatics, narrative production and comprehension, social interaction, and culture. However, unless there are specialized modules for recursive mindreading supporting each of these functions, it does suggest that attention should be paid to how this role is served with limited rather than exceptional recursive mindreading abilities. Here, we briefly discuss the potential importance of dodging, motivation, iterative bootstrapping, and scaffolding.

Dodging

We have highlighted that existing stimuli cannot reliably identify what level of recursion participants have used to give a "correct" answer. But could it be that in this respect such stimuli are more realistic than our revised stimuli? If this were so then might it be possible for people to "dodge" the challenge of recursive mindreading, and rely on partial-chain reasoning being an adequate strategy much of the time? Consistent with this line of reasoning, social abilities are supported by many processes other than mindreading, and there is a tendency for researchers to overestimate the necessity for any kind of mindreading—recursive or otherwise—in our everyday activities (e.g., I. A. Apperly, 2010). Therefore, it is plausible that mindreading may sometimes be avoided, and even when a situation or stimulus *can* be analyzed in terms of mindreading - for example saying "thank you" to the bus driver who has just sold you a ticket—it remains an empirical question whether mindreading is what participants actually do when they encounter that situation or stimulus. On the other hand, mindreading is clearly necessary in some circumstances, and when it is necessary the present analyses suggest that "dodging" via partial-chain reasoning is a poor strategy. This is because partial-chain reasoning does not consistently deliver the correct answer, and the only reliable solution is to check the correct answer via full-chain reasoning. We, therefore, consider other factors that might influence or support recursive mindreading when it cannot be dodged.

Motivation

Camerer et al. (2004) observed more recursion in participants' strategic reasoning when financial rewards for success were high. Superior performance in the "bonus" condition of Study 2 above is consistent with recursive mindreading also being affected by extrinsic motivators (though note that this condition additionally included a scaffolding manipulation). Moreover, recursive mindreading may benefit from intrinsic motivation, whether that is motivation for social processing in general (Chevallier et al., 2012), or motivation for mindreading in particular (Carpenter et al., 2016). Even if previous tests are an unreliable indicator of recursive capacity, social motivation may nonetheless influence participants' willingness to engage with these complex social tasks, which may

explain how they capture variance that is relevant for social behavior (Lewis et al., 2011; Paal & Bereczkei, 2007; Powell, 2010; Stiller & Dunbar, 2007; Vonk et al., 2015).

Iterative Bootstrapping

It is possible that repeated experience with a particular kind of recursive reasoning problem allows for iterative bootstrapping. Camerer et al. (2004) illustrate this possibility for "beauty contest games," in which each player must pick a number from 0 to 100, with the winner being the player whose pick is closest to 2/3 of the average of all players. On a first assumption that picks of other players will be randomly distributed from 0 to 100, a player should themselves pick 100*2/3. However, a player should also assume that other players will reason similarly, adjust their pick in accordance, but then realize that other players may reason in the same way as this too, and so on, until concluding that zero is the only rational pick. Arriving at a figure below 1 requires a minimum of 12 such recursive steps. Empirically, in one-shot games, most players pick between 20 and 35, consistent with just two recursive steps. This is true unless players have the opportunity to learn over multiple trials, in which case picks do tend toward zero over time. This illustrates the possibility of iterative bootstrapping whereby solutions based on limited recursive reasoning in previous rounds become the inputs for limited recursive reasoning on later rounds, but ultimately resulting in a solution that would have required much higher levels of recursive reasoning in a one-shot scenario. Such a process may account for unusually high levels of recursion observed in some neuroscientific studies that use computational modeling to infer recursion that is implicit in gameplay over many repeated trials (e.g, Charpentier & O'Doherty, 2018; Lee & Seo, 2016; Yoshida et al., 2008, 2010). Note that computational neuroscientists typically take these findings as evidence of recursive mindreading, whereas behavioral economists using similar or identical tasks typically assume that participants model each other's recursive behavioral strategies and not recursive representations of mental states. It is very difficult to distinguish between these possibilities on the basis of participants' nonverbal behavior. However, there seems no reason in principle why iterative bootstrapping might not support recursive mindreading when the situation affords the necessary learning.

Scaffolding

Finally, it is possible that mindreading earlier in a recursive chain scaffolds further recursive mindreading. This is illustrated by the "bonus" condition in Study 2, in which participants built up to a Level-5 question by answering questions about the entire series of embedded mental states, starting from Level 1. Such a scenario of chunking earlier recursive inferences to serve as scaffolding for later inferences seems particularly plausible for mindreading scenarios that unfold over time, as is often the case in literature and drama. In accord with van Duijn et al. (2015), we conjecture that part of the art of skilled plot writing is to achieve a similar effect more elegantly than in Study 2 and that such devices relieve the reader or viewer of the need to entertain the full recursive chain at once.

Limitations

The present work is limited in its ability to cast empirical light on motivation, bootstrapping, and scaffolding. Study 1 showed no effect of a bonus designed to raise participants' motivation, while Study 2 showed more accurate performance in a condition that combined higher levels of bonus with encouragement to use strategies such as writing, and an attempt to scaffold participants' reasoning. The present work cannot tell us which of these factors was responsible for more accurate performance. Moreover, recruitment for Studies 1 and 2 advertised the presence or absence of a bonus for good performance, and so it is conceivable that participants who were more responsive to intrinsic motivation self-selected for the no-bonus conditions, while those more responsive to extrinsic motivation self-selected for the bonus conditions. Importantly, such self-selection could not have caused the critical differences between Original and Revised stories. It is also uncertain what impact it could have had upon performance in the bonus and no-bonus conditions, but this is clearly a suboptimal design for investigating effects of motivation. Future work could address these limitations by deconfounding rewards from scaffolding and encouragement to provide clear evidence on the role of motivation in recursive mindreading.

Conclusion

Theorists have made a compelling case that high levels of recursive mindreading may be necessary to explain human pragmatic abilities (Scott-Phillips, 2014; Sperber & Wilson, 1986), the complexity of human social networks (Dunbar, 2003), and the existence of religion (Dunbar, 2017) and literary fiction (Zunshine, 2006), and have drawn upon previous evidence of exceptional recursive mindreading in support of these claims. Our analysis of the problems inherent in previous tests of recursive mindreading calls previous evidence into question, and our new empirical evidence from revised tasks is consistent with the view that recursive mindreading is unexceptional, and limited in a way that is similar to recursive grammar and strategic reasoning. We do not contest the theorists' case about the importance of recursive mindreading, but we suggest that it may be productive for psychologists to consider how these needs are met with limited recursive abilities, through processes such as iterative bootstrapping and scaffolding.

Context

Mindreading (or "theory of mind") began as a topic of interest to comparative and developmental psychologists (e.g., Premack & Woodruff, 1978; Wimmer & Perner, 1983), but mindreading in adults has become a major focus for researchers interested in the cognitive and neural basis of social interaction and communication (e.g., I. A.Apperly, 2010; Ferguson & Bradford, 2021; Gilead & Ochsner, 2021). One challenge is to understand ways in which adults—who have already acquired critical mindreading concepts during childhood—might continue to vary in their capacity for mindreading (e.g., I. Apperly & Wang, 2021). The ability to mindread recursively is an important contender for such a capacity, because previous evidence suggests that it varies between adults and captures variance that is relevant for social behavior (Lewis et al., 2011; Paal & Bereczkei, 2007; Powell et al., 2009; Stiller & Dunbar, 2007; Vonk et al., 2015), because recursive mindreading features in prominent theories of communication, language/literature, and social ability (Dunbar, 2003, 2017; Scott-Phillips, 2014; Sperber & Wilson, 1986; Zunshine, 2006), and because previous

evidence suggested that adults' capacity for recursive mindreading was exceptional compared with other recursive thinking. Thus, in the process of developing new and better methods for examining individual differences in adults' mindreading, we had cause to look closely at existing measures of recursive mindreading. This led us to question whether existing measures were capable of producing reliable results, which gave rise to the present series of experiments.

References

Apperly, I., & Wang, J. (2021). Mindreading in adults: Cognitive basis, motivation, and individual differences. In H. J. Ferguson & E. E. Bradford (Eds.), *The cognitive basis of social interaction across the lifespan* (pp. 96–116). Oxford University Press.

Apperly, I. A. (2010). *Mindreaders: The cognitive basis of "theory of mind"*. Psychology Press/Taylor & Francis.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H., Singmann, H., & Dai, B. (2015). *lme4: Linear mixed-effects models using Eigen and S4. R package Version 1.1-7*. 2014.

Bosch-Domènech, A., Montalvo, J. G., Nagel, R., & Satorra, A. (2002). One, two, (three), infinity, …: Newspaper and lab beauty-contest experiments. *American Economic Review*, *92*(5), 1687–1701. https://doi.org/10.1257/000282802762024737

Camerer, C., Ho, T. H., & Chong, J. K. (2004). A cognitive hierarchy model of games. *The Quarterly Journal of Economics*, *119*(3), 861–898. https://doi.org/10.1162/0033553041502225

Camerer, C. F. (2003). *Behavioral game theory: Experiments on strategic interaction*. Princeton University Press.

Carpenter, J. M., Green, M. C., & Vacharkulksemsuk, T. (2016). Beyond perspective-taking: Mind-reading motivation. *Motivation and Emotion*, *40*(3), 358–374. https://doi.org/10.1007/s11031-016-9544-z

Carruthers, P. (2006). *The architecture of the mind*. Oxford University Press.

Charpentier, C. J., & O'Doherty, J. P. (2018). The application of computational models to social neuroscience: Promises and pitfalls. *Social Neuroscience*, *13*(6), 637–647. https://doi.org/10.1080/17470919.2018.1518834

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences*, *16*(4), 231–239. https://doi.org/10.1016/j.tics.2012.02.007

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Corballis, M. C. (2014). *The recursive mind*. Princeton University Press.

de Boer, B., Sandler, W., & Kirby, S. (2012). New perspectives on duality of patterning: Introduction to the special issue. *Language and Cognition*, *4*(4), 251–259. https://doi.org/10.1515/langcog-2012-0014

De Freitas, J., Thomas, K., DeScioli, P., & Pinker, S. (2019). Common knowledge, coordination, and strategic mentalizing in human social life. *Proceedings of the National Academy of Sciences*, *116*(28), 13751–13758. https://doi.org/10.1073/pnas.1905518116

Dunbar, R. I. (2003). The social brain: Mind, language, and society in evolutionary perspective. *Annual Review of Anthropology*, *32*(1), 163–181. https://doi.org/10.1146/annurev.anthro.32.061002.093158

Dunbar, R. I. (2017). What's missing from the scientific study of religion? *Religion, Brain & Behavior*, *7*(4), 349–353. https://doi.org/10.1080/2153599X.2016.1249927

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Ferguson, H. J., & Bradford, E. E. (Eds.). (2021). *The cognitive basis of social interaction across the lifespan*. Oxford University Press.

Fodor, J. A. (1983). *The modularity of mind*. MIT Press.

Gilead, M., & Ochsner, K. N. (Eds.). (2021). *The neural basis of mentalizing*. Springer Nature.

Grice, H. P. (1969). Utterer's meaning and intentions. *The Philosophical Review*, *78*, 147–177.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Karlsson, F. (2007). Constraints on multiple center-embedding of clauses. *Journal of Linguistics*, *43*(2), 365–392. https://doi.org/10.1017/S0022226707004616

Kinderman, P., Dunbar, R., & Bentall, R. P. (1998). Theory-of-mind deficits and causal attributions. *British Journal of Psychology*, *89*(2), 191–204. https://doi.org/10.1111/j.2044-8295.1998.tb02680.x

Klindt, D., Devaine, M., & Daunizeau, J. (2017). Does the way we read others' mind change over the lifespan? Insights from a massive web poll of cognitive skills from childhood to late adulthood. *Cortex*, *86*, 205–215. https://doi.org/10.1016/j.cortex.2016.09.009

Lee, D., & Seo, H. (2016). Neural basis of strategic decision making. *Trends in Neurosciences*, *39*(1), 40–48. https://doi.org/10.1016/j.tins.2015.11.002

Lenth, R. (2021). *emmeans: Estimated marginal means, aka least-squares means (R Package version 1.5.4)*. https://CRAN.R-project.org/package=emmeans

Levinson, S. C. (2013). Recursion in pragmatics. *Language*, *8*(1), 149–162. https://doi.org/10.1353/lan.2013.0005

Lewis, P. A., Rezaie, R., Brown, R., Roberts, N., & Dunbar, R. I. (2011). Ventromedial prefrontal volume predicts understanding of others and social network size. *Neuroimage*, *57*(4), 1624–1629. https://doi.org/10.1016/j.neuroimage.2011.05.030

Moore, R. (2017). Gricean communication and cognitive development. *The Philosophical Quarterly*, *67*(267), 303–326. https://doi.org/.org/10.1093/pq/pqw049

Oatley, K. (2011). *Such stuff as dreams: The psychology of fiction*. Wiley.

Oesch, N., & Dunbar, R. I. M. (2017). The emergence of recursion in human language: Mentalising predicts recursive syntax task performance. *Journal of Neurolinguistics*, *43*(Part B), 95–106. https://doi.org/10.1016/j.jneuroling.2016.09.008

O'Grady, C., Kliesch, C., Smith, K., & Scott-Phillips, T. (2015). The ease and extent of recursive mindreading, across implicit and explicit tasks. *Evolution and Human Behavior*, *36*(4), 313–322. https://doi.org/10.1016/j.evolhumbehav.2015.01.004

Paal, T., & Bereczkei, T. (2007). Adult theory of mind, cooperation, Machiavellianism: The effect of mindreading on social relations. *Personality and Individual Differences*, *43*(3), 541–551. https://doi.org/10.1016/j.paid.2006.12.021

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that…" attribution of second-order beliefs by 5-to 10-year-old children. *Journal of Experimental Child Psychology*, *39*(3), 437–471. https://doi.org/10.1016/0022-0965(85)90051-7

Powell, J. L., Lewis, P. A., Dunbar, R. I., García-Fiñana, M., & Roberts, N. (2010). Orbital prefrontal cortex volume correlates with social cognitive competence. *Neuropsychologia*, *48*(12), 3554–3562. https://doi.org/10.1016/j.neuropsychologia.2010.08.004

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind. *Behavioral and Brain Sciences*, *1*(4), 515–526.

R Core Team. (2020). *R: A language and environment for statistical ## computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Scott-Phillips, T. (2014). *Speaking our minds: Why human communication is different, and how language evolved to make it special*. Macmillan International Higher Education.

Sperber, D. (2000). 'Metarepresentations in an evolutionary perspective'. In D. Sperber (Ed.), *Metarepresentations: An interdisciplinary perspective* (pp. 117–137). Oxford University Press.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (Vol. 142). Harvard University Press.

Stiller, J., & Dunbar, R. I. M. (2007). Perspective-taking and memory capacity predict social network size. *Social Networks*, *29*(1), 93–104. https://doi.org/10.1016/j.socnet.2006.04.001

Sullivan, K., Zaitchik, D., & Tager-Flusberg, H. (1994). Preschoolers can attribute second-order beliefs. *Developmental Psychology*, *30*(3), 395–402. https://doi.org/10.1037/0012-1649.30.3.395

Tomasello, M. (2014). *A natural history of human thinking*. Harvard University Press.

van Duijn, M. J. (2016). *The lazy mindreader: A humanities perspective on mindreading and multiple-order intentionality* [PhD thesis]. Leiden University.

Van Duijn, M. J., Sluiter, I., & Verhagen, A. (2015). When narrative takes over: The representation of embedded mindstates in Shakespeare's Othello. *Language and Literature*, *24*(2), 148–166. https://doi.org/10.1177/0963947015572274

Vonk, J., Zeigler-Hill, V., Ewing, D., Mercer, S., & Noser, A. E. (2015). Mindreading in the dark: Dark personality features and theory of mind. *Personality and Individual Differences*, *87*, 50–54. https://doi.org/10.1016/j.paid.2015.07.025

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

Yoshida, W., Dolan, R. J., & Friston, K. J., & Behrens, T. (2008). Game theory of mind. *PLoS Computational Biology*, *4*(12), Article e1000254. https://doi.org/10.1371/journal.pcbi.1000254

Yoshida, W., Seymour, B., Friston, K. J., & Dolan, R. J. (2010). Neural mechanisms of belief inference during cooperative games. *Journal of Neuroscience*, *30*(32), 10744–10751. https://doi.org/10.1523/JNEUROSCI.5895-09.2010

Zunshine, L. (2006). *Why we read fiction: Theory of mind and the novel*. Ohio State University Press.

Appendix A

Instructions, Stimuli, and Questions for Study 1

Main Task Instructions

No-Bonus Condition

You will be asked to read a set of four short stories depicting various social situations. Afterward, you will be asked to answer one key question related to the story, plus one check question. IMPORTANT: For the key question you will be able to refer back to each story as much as necessary, but please note that once you submit your answer, you will not be able to revisit the previous section. Most questions offer you two or more possible answers, while some require short typed answers to check that you have followed the materials while completing the survey. Take your time and answer the questions to the best of your ability, preferably in a quiet room away from any distractions.

Bonus Condition

You will be asked to read a set of four short stories depicting various social situations. Afterward, you will be asked to answer one key question related to the story, plus one check question. IMPORTANT: For the key question you will be able to refer back to each story as much as necessary, but please note that once you submit your answer, you will not be able to revisit the previous section. Most questions offer you two or more possible answers, while some require short typed answers to check that you have followed the materials while completing the survey. Take your time and answer the questions to the best of your ability, preferably in a quiet room away from any distractions. If you answer all questions correctly you will get a $4 BONUS.

Original Stories

The correct answer to the Original stories is shown in bold. All other possible answers were incorrect.

Interview Story

Gavin, Peter, Fiona, and Sophie are all waiting for a job interview. They are applying to become an assistant editor of a high-profile cookery magazine based in London. Gavin is a trained chef, Fiona has a degree in food technology, and Sophie has a degree in English literature. Peter has no experience in editing or cooking but does have a diploma in computing. Gavin engages them all in conversation to assess the competition. He tells them he is a chef in a high-profile restaurant and worked with food for the past 10 years. Fiona is next to respond and describes her degree course in food technology. Sophie mentions that she enjoys cooking and has a degree in English. Gavin laughs at Sophie and wishes her luck. Sophie, Peter, and Fiona all think Gavin is rather rude. Gavin asks Peter what qualifications he has and Peter refuses to answer as he is embarrassed by his lack of relevant skills. Gavin suggests that Peter is refusing to answer, as Gavin believes he is obviously the most qualified. Peter again refuses to reply so as not to give any information away and ignores Gavin. The two women tell Gavin to be quiet.

Level-5 question: Gavin knows that Peter understands that Fiona supposes that Gavin knows that Sophie thinks…

Alternative choices:

That Gavin has been speaking too quietly

That Gavin is a nice guy

Peter is enjoying the robust competition

Fiona has no chance of being offered the position

That Sophie is not good enough at writing

That it would be better if Gavin got the job so they could all avoid embarrassment if they bumped into him later

It would be helpful if Gavin adopted a less condescending and superior attitude toward the other applicants **[Correct answer]**

[Note, this would also be the correct answer if participants use the simplifying strategy of starting at either end of the recursive chain and following any number of recursive steps less than 5]

***Check Question**—shown on separate screen*

Fiona and Sophie are the only female characters in the story.

True/False

Date Story. Charlotte and Simon decide to arrange a blind date for their two friends. Simon's friend Martin is a florist and Charlotte's friend Jane is a doctor. Charlotte and Simon decide to go to the pub on the edge of town, as it is within walking distance of their house and a convenient place to meet Jane and Martin, after their blind date over dinner at a nearby restaurant. When Charlotte and Simon arrive at the pub, they greet Martin and Jane, and Martin buys a round of drinks. Charlotte and Jane both have a glass of red wine while Martin has a pint of beer and Simon has a whiskey. Martin and Simon leave the women at the bar to play a game of pool. Jane tells Charlotte that she did not like Martin as she thinks he is a chauvinist. Meanwhile, Martin tells Simon that the date went well, Jane fancied him, and that he was the perfect gentleman. Simon wins the game and they return to the girls. Jane thanks Martin for a lovely date and decides to go home early.

Level-5 question: Simon believes that Martin thinks that Charlotte supposes that Jane knows that Simon thinks…

Alternative choices:

They had an awful date at the restaurant

That Jane did not like Martin and thought he was a chauvinist

That he ruined Martin's date because he did not let him win the game of pool

That he should have arranged the blind date with Charlotte instead

That Charlotte likes wine while Jane prefers beer

That after walking a long way to the restaurant Charlotte was upset to have to walk even further to the pub

That they had a lovely date at the restaurant and a relaxing evening at the bar **[Correct answer]**

[Note, this would also be the correct answer if participants use the simplifying strategy of judging what Simon believes/thinks, or what Michael believes Simon thinks. For other simplifying chains the correct answer is not always perfectly clear because the story does not spell out what Charlotte or Jane know about what Michael or Simon think, but if anything it is reasonable to suppose that they accurately believe that Michael and Simon think the evening went well, meaning that this remains the best answer to select in all cases]

***Check Question** [shown on separate screen]* None of the characters drink alcohol. True/False

Revised Stories [Study 1]

For the Revised stories, there was a single level-5 recursive question, with seven forced choice answers. Each of the answers relates to a specific recursive level and direction (forward or backward). The correct answer to the question is shown in bold, and the direction and recursive level for each of the other answers are shown in square brackets.

Chef Story

This is a story about a group of trainee chefs called Jerome, Yasmin, Rachel, Nik, and their supervisor Michael. The trainees are competing to win a permanent job in the restaurant, and everyone is well aware that the other chefs may be watching. Each chef has been allocated particular tables to serve, and feedback from the guests will be used to assess the chefs' performance. One evening Jerome is asked to talk to a customer at one of his tables to discuss whether the food is appropriate for her dietary requirements. Before he goes to the table, he dishes up his prawn cocktail starter, and leaves it on the counter ready to finish when he comes back. Right after he leaves, Yasmin, who is worried about Jerome's success, decides to take the prawn cocktail and put it in the fridge to serve as her own. Yasmin returns to preparing her dish at her workstation, with her back to the fridge. Rachel, Nik, and Michael saw the whole thing and could not believe their eyes. Suddenly, without Yasmin seeing, one of the waiters rushes in and takes the prawn cocktail out of the fridge to make room for the dish he was carrying, and without thinking, he puts the prawn cocktail down on a heated worktop. Rachel is then called to one of her tables to deal with a customer complaint. Being the least competitive of the group, Nik decides to tell Yasmin where the prawn cocktail is and tells her to take the dish and put it in the bin as it will be warm, and cannot be served anymore. Michael tells Yasmin he will go with her. On their way out of the kitchen, Michael tells Yasmin that she really should recreate a new prawn cocktail and put it back on the counter before Jerome comes back. He tells her he's going to check the restaurant, but expects everything to be fixed when he gets back. However, moments later Yasmin meets Jerome, who is returning back to the kitchen. In an instant she changes her mind and decides to give the prawn cocktail to him to serve to his table, table 12, and then goes back to the kitchen. After Yasmin leaves, Jerome realizes that there is something wrong because the dish feels warm. When he sees Rachel coming back to the kitchen, he tells her how Yasmin tried to trick him into serving a ruined dish, but he served it to one of Yasmin's tables instead.

Level-5 question: Where does Jerome think Yasmin thinks Rachel thinks Nik thinks Michael thinks the prawn cocktail is?

Alternative choices:

In the storage room [backward 1, i.e. corresponding to: "Where does Michael think…"] *Note that due to an editing error, the location "storage room" does not exist in the Chef story for Study 1.

In the bin [backward 2, i.e. corresponding to: "Where does Nik think Michael thinks…"]

On the heated worktop [backward 3, i.e. corresponding to: "Where does Rachel think Nik thinks Michael thinks…"]

In the fridge [backward 4, i.e. corresponding to: "Where does Yasmin think Rachel thinks Nik thinks Michael thinks…"]

On the counter **[level 5 = Correct answer]**

At one of Yasmin's tables [forward 1, i.e. corresponding to: "Where does Jerome think…"]

At table 12 [forward 2, i.e. corresponding to: "Where does Jerome think Yasmin thinks…."]

***Check Question** [shown on separate screen]:* Prawn cocktail is a kind of drink. True/False

Garden Story

This is a story about to siblings, Peter and Mary, and their parents and Gran. One day, Peter and Mary are playing in the back garden of their house enjoying the attention of the adults who are sitting in deck chairs watching them. Peter is enthusiastically drawing. "I'm going to the loo" says Peter. He leaves his drawing in a little box and goes into the house. When Mary sees that Peter has left, she decides to play a trick. She takes Peter's drawing from the box and

puts it behind a bush, and then continues playing with her back to the bush. Mum, Dad, and Gran are amused by the trick, when suddenly, a strong wind blows the drawing and moves the picture under the family car without Mary seeing. The doorbell rings and Mum goes inside to answer. Dad tells Mary that it would be better to pick up the picture and leave it in the kitchen, and Gran says she'll go with her. On their way inside they agree that the picture is too crumpled so instead Gran suggests that Mary takes it to Mum's study while she makes a cup of tea. Instead, Mary finds Peter on his way out of the toilet and gets him to help her find a heavy book to press the picture in, and they leave it on the shelf in the living room. After Mary has gone Peter sees Mum, and tells her how Mary helped him press his picture, but now he has hidden it safely in his bedroom.

Level-5 question: Where does Peter think Mary thinks Mum thinks Dad thinks Gran thinks the picture is?

Alternative choices:

In the study [backward 1. i.e. corresponding to: "Where does Gran think…"]

In the kitchen [backward 2, i.e. corresponding to: "Where does Dad think Gran thinks…"]

Under the family car [backward 3, i.e. corresponding to: "Where does Mum think Dad thinks Gran thinks…"]

Behind a bush [backward 4, i.e. corresponding to: "Where does Mary think Mum thinks Dad thinks Gran thinks…."]

In a little box **[level 5 = correct answer]**

In the bedroom [forward 1, i.e. corresponding to: "Where does Peter think…"]

In the living room [forward 2, i.e. corresponding to: "Where does Peter think Mary thinks…"]

Check Question *[shown on separate screen].* Peter and Dad are the only male characters in the story. True/False

Appendix B

Instructions, Stimuli, and Questions for Study 2

Main Task Instructions

No-Bonus Condition

There will be two stories for you to read, each with two questions. The first question for each story is about how characters understand one another's point of view. You can refer back to the story at any time. The second question is a simple memory test relating to an element of the story. For this, you will not be able to refer back to the story.

Bonus Condition

There will be two stories for you to read, each with six questions. The first five questions for each story are like complicated puzzles about how characters understand one another's point of view. The best way to answer these is to be methodical. You can use a pen and paper to work them out if you wish. Take your time and keep referring back to the story. The last question for each story is a simple memory test relating to an element of the story. For this, you will not be able to refer back to the story. Our experience is that people find the first five questions for each story really challenging, but they are able to do it if they take time to work through it.

To encourage you to do this, and to thank you for your efforts you will get a £10 BONUS if you are able to answer all of the questions correctly.

First there will be a practice question to give you an idea of what to expect."

Practice Story *[presented before the test stories in the Bonus condition]*

John and Mary are two children who are helping the ice-cream man sell ice cream in the park. Mary realizes she has to go home. She would have liked to stay and sell ice cream with John, because they get on so well. "Don't worry," says John, "you can come back in the afternoon. I'll be here again to sell ice cream." "That's all right then," says Mary, "maybe I'll be able to join you in the afternoon then. In any case I know where you'll be." So Mary goes home.

Now only the ice-cream man and John are left in the park. "Know what, John," says the ice-cream man suddenly to John, "I've changed my mind. I'll drive my van to the church. There is nobody here in the park who wants to buy ice cream. Maybe I'll be able to sell some more at the church. Do you want to come with me?" "Yes," says John, "I'll help you again in the afternoon. Now I've to go home for lunch. But wait," says John, "Mary won't know that I'll be at the church. Could you call at her home on your way and tell her that in the afternoon we will sell ice cream at the church? " "Sure," says the ice-cream man, "I'll tell Mary."

While John is at home, the ice-cream man drives over to the church. He positions his van next to the church and sells ice cream. The ice-cream man has forgotten to tell Mary that John will be at the church in the afternoon to help sell ice cream. John doesn't know that the ice-cream man had forgotten to tell Mary.

Now, John has to go home. After lunch, he is doing his homework. He can't do one of the tasks. So he goes over to Mary's house to ask for help. Mary's mother answers the door. "Is Mary in?" asks John. "Oh," says Mary's mother, "that's funny, because Mary just left to meet you."

Level-1 question. Where is Mary going?

Level-2 question. Where does John think Mary is going?

Both questions had a dropdown menu with the three possible answers:

Home

The park

The church

Check Question *[shown on separate screen]:* Mary lives with her mother. True/False

Revised Stories [Study 2]

For each story, there were seven questions, working from recursive levels 1 to 5. Each of the questions had seven forced choice answers. Minor edits were made to each story to ensure that each question had a unique correct answer and that all alternative answers corresponded to plausible locations mentioned in the story. Next to

each question, the correct answer is shown in square brackets. The seven forced choice answers are listed under the questions with the direction and recursive level shown in square brackets.

Chef Story

This is a story about a group of trainee chefs called Jerome, Yasmin, Rachel, Nik, and their supervisor Michael. The trainees are competing to win a permanent job in the restaurant, and everyone is well aware that the other chefs may be watching. Each chef has been allocated particular tables to serve, and feedback from the guests will be used to assess the chefs' performance.

One evening Jerome is asked to talk to a customer at one of his tables to discuss whether the food is appropriate for her dietary requirements. Before he goes to the table, he dishes up his prawn cocktail starter and leaves it on the counter ready to finish when he comes back.

Right after he leaves, Yasmin, who is worried about Jerome's success, decides to take the prawn cocktail and put it in the fridge to serve as her own. Yasmin returns to preparing her dish at her workstation, with her back to the fridge. Rachel, Nik, and Michael saw the whole thing and could not believe their eyes.

Suddenly, without Yasmin seeing, one of the waiters rushes in and takes the prawn cocktail out of the fridge to make room for the dish he was carrying, and without thinking, he puts the prawn cocktail down on a heated worktop. Rachel is then called to one of her tables to deal with a customer complaint.

Being the least competitive of the group, Nik decides to tell Yasmin how he, Rachel, and Michael saw her put the prawn cocktail in the fridge. He tells her that he just saw a waiter move the prawn cocktail onto the heated worktop and that she should take the dish and put it in the bin as it will be warm, and cannot be served anymore. Michael is listening and tells Yasmin he will go with her.

On their way out of the kitchen, Michael tells Yasmin that she should put the prawn cocktail in the storage room, for any of the staff to eat after the restaurant has closed. Then, she really should recreate a new prawn cocktail and put it on the counter before Jerome comes back. He tells her he's going to check the restaurant, but expects everything to be fixed when he gets back.

However, moments later Yasmin meets Jerome, who is returning back to the kitchen. In an instant, she changes her mind and decides to give the prawn cocktail to him to serve to his table, table 12, and then goes back to the kitchen.

After Yasmin leaves, Jerome realizes that there is something wrong because the dish feels warm. When he sees Rachel coming back to the kitchen, he tells her how Yasmin tried to trick him into serving a ruined dish, but he served it to one of Yasmin's tables instead. Rachel doesn't have time to explain everything that happened and goes back to work.

Questions from 1 to 5 levels of recursion:

Where does Michael think the prawn cocktail is? [level 1—In the storage room]

Where does Nik think Michael thinks the prawn cocktail is? [level 2—In the bin]

Where does Rachel think Nik thinks Michael thinks the prawn cocktail is? [level 3—On the heated worktop]

Where does Yasmin think Rachel thinks Nik thinks Michael thinks the prawn cocktail is? [level 4—In the fridge]

Where does Jerome think Yasmin thinks Rachel thinks Nik thinks Michael thinks the prawn cocktail is? [level 5—On the counter]

Each question had a dropdown menu with seven possible answers corresponding to the set of correct answers for all five questions, plus the answer that would result from forward partial-chain reasoning for one and two levels of recursion:

In the storage room
In the bin
On the heated worktop
In the fridge
On the counter [level 5]
At one of Yasmin's tables [forward 1]
At table 12 [forward 2]

Check Question *[shown on separate screen]:* Prawn cocktail is a kind of drink. True/False

Garden Story

This is a story about two siblings, Peter and Mary, and their parents and Gran. One day, Peter and Mary are playing in the back garden of their house enjoying the attention of the adults who are sitting in deck chairs watching them. Peter is enthusiastically drawing. "I'm going to the loo" says Peter. He leaves his drawing in a little box and goes into the house.

When Mary sees that Peter has left, she decides to play a trick. She takes Peter's drawing from the box and puts it behind a bush, and then continues playing with her back to the bush. Mum, Dad, and Gran are amused by the trick, when suddenly, a strong wind blows the drawing and moves the picture under the family car without Mary seeing.

The doorbell rings and Mum goes inside to answer. Dad tells Mary that he and Gran saw the picture blow under the car and that it would be better to pick up the picture and leave it in the kitchen. Gran says she'll go with her.

On their way inside, they agree that the picture is badly crumpled so instead Gran suggests that Mary takes it to the study while she makes a cup of tea. Instead, Mary finds Peter on his way out of the toilet. She tells him that the picture got crumpled but doesn't explain how. She gets him to help her find a heavy book to press the picture in. They leave it on the shelf in the living room.

After Mary has gone Peter sees Mum, and tells her how Mary helped him press his picture, but now he has hidden it safely in his bedroom.

Questions from 1 to 5 levels of recursion:

Where does Gran think the picture is? [level 1—in the study]

Where does Dad think Gran thinks the picture is? [level 2—In the kitchen]

Where does Mum thinks Dad thinks Gran thinks the picture is? [level 3—Under the family car]

Where does Mary thinks Mum thinks Dad thinks Gran thinks the picture is? [level 4—Behind a bush]

Where does Peter think Mary thinks Mum thinks Dad thinks Gran thinks the picture is? [level 5—In a little box]

Each question had a dropdown menu with seven possible answers corresponding to the set of correct answers for all five questions, plus the answer that would result from forward partial-chain reasoning for one and two levels of recursion:

In the study
In the kitchen

Under the family car
Behind a bush
In a little box [level 5]
In the bedroom [forward 1]
In the living room [forward 2]

Check Question *[shown on separate screen]:* Peter and Dad are the only male characters in the story. True/False

Received March 23, 2022
Revision received September 17, 2022
Accepted September 26, 2022 ■