Evolution and Human Behavior 36 (2015) 313–322

Contents lists available at ScienceDirect

Evolution and Human Behavior

journal homepage: www.ehbonline.org

Original Article

The ease and extent of recursive mindreading, across implicit and explicit tasks

Cathleen O'Grady [a,1], Christian Kliesch [b,1], Kenny Smith [a], Thomas C. Scott-Phillips [c,*,2]

[a] School of Philosophy, Psychology, and Language Sciences, University of Edinburgh
[b] Department of Psychology, Lancaster University, Lancaster LA1 4YF
[c] Evolutionary Anthropology Research Group, Durham University

ARTICLE INFO

Article history:
Initial receipt 6 May 2014
21 December 2014
Final revision received 22 January 2015

Keywords:
Mindreading
Recursive mindreading
Mentalizing
Theory of mind
Metarepresentation
Intentionality
Social cognition

ABSTRACT

Recursive mindreading is the ability to embed mental representations inside other mental representations e.g. to hold beliefs about beliefs about beliefs. An advanced ability to entertain recursively embedded mental states is consistent with evolutionary perspectives that emphasise the importance of sociality and social cognition in human evolution: high levels of recursive mindreading are argued to be involved in several distinctive human behaviours and institutions, such as communication, religion, and story-telling. However, despite a wealth of research on first-level mindreading under the term Theory of Mind, the human ability for recursive mindreading is relatively understudied, and existing research on the topic has significant methodological flaws. Here we show experimentally that human recursive mindreading abilities are far more advanced than has previously been shown. Specifically, we show that humans are able to mindread to at least seven levels of embedding, both explicitly, through linguistic description, and implicitly, through observing social interactions. However, our data suggest that mindreading may be easier when stimuli are presented implicitly rather than explicitly. We argue that advanced mindreading abilities are to be expected in an extremely social species such as our own, where the ability to reason about others' mental states is an essential, ubiquitous and adaptive component of everyday life.

"Humans have the ability to represent representations… This metarepresentational ability is as distinctive of humans, and as important in understanding their behaviour, as is echolocation for bats" (Sperber, 1997, p.69).

1. Introduction

Mindreading is the ability to mentally represent others' mental representations. It is also known as mental metarepresentation, or theory of mind. Recursive mindreading is the ability to embed further levels of mental representation inside existing mental representations (e.g. I think$_0$ that you believe$_1$ that he thinks$_2$ that she wants$_3$… and so on; subscripts count the number of metarepresentations[3]). An intuitive and commonly held view is that high-level recursive mindreading (i.e. beyond first or second level) is cognitively demanding, and perhaps beyond normal human abilities (e.g. Clark, 1996; Gómez, 1994). Yet theoretical explanations of many important human behaviours and institutions, such as communication, religion, story-telling, and culture itself either argue or assume that humans can and do process high levels of recursive mindreading routinely and without difficulty (Dunbar, 2003, 2008; Sperber, 2000a; Tomasello, 2008). Furthermore, our natural ecology is a social one, in which both collaboration and competition are everyday activities (Byrne & Whiten, 1989; Dunbar, 2003; Humphrey, 1976). In such an environment, the ability to monitor and manage one's social environment, by reasoning about the motives and intentions of others, keeping track of others' relationships, deciding who to trust, and so on, is of critical importance. From this perspective, we should expect humans to be able to process mental (meta)representations with relative ease, at least when those representations are encountered within this social ecology.

There is a large literature on various aspects of first-level mindreading (e.g. the ability to reason about the mental state of another: I think$_0$ that you believe$_1$ some proposition). This includes, most prominently, its development in children (Baillargeon, Scott, & He, 2010; Wellman, Cross, & Watson, 2003), its role in some social cognitive disorders (Baron-Cohen, 1995; Chevallier, Kohls, Troiani, Brodkin, & Schultz, 2012), and its presence or absence in non-human primates (Call & Tomasello, 2008; Premack & Woodruff, 1978). In contrast, there is far less research dealing specifically with recursive

* Corresponding author. Department of Anthropology, Dawson Building, South Road, Durham, CH1 3LE.
E-mail address: t.c.scott-phillips@durham.ac.uk (T.C. Scott-Phillips).
[1] Joint first authors.
[2] @tscottphillips (twitter).
[3] There is inconsistency in the mindreading literature regarding how to count the levels. Some studies include the focal individual's perspective; others exclude it. This is the difference between counting the number of representations (I think$_1$ that Mary thinks$_2$…) or only the number of metarepresentations (I think$_0$ that Mary thinks$_1$…). Most adult studies use the former practice, whereas the developmental literature uses the latter. We follow this latter practice in this paper.

http://dx.doi.org/10.1016/j.evolhumbehav.2015.01.004
1090-5138/

mindreading, despite its theoretical importance for human social life. The handful of studies in adults that do exist report a prominent drop in performance after four levels of recursive mindreading (Kinderman, Dunbar, & Bentall, 1998; Lyons, Caldwell, & Shultz, 2010; Stiller & Dunbar, 2007). There is also a small literature on second- and third-level tasks in children's development, which finds that the ability to perform these higher-level tasks emerges later in development than competence in first-level tasks (see Miller, 2009 for a review).

However, previous research on high-order recursive mindreading may have significantly underestimated the extent of human recursive mindreading abilities, for at least two reasons. First, the stimuli used in previous studies have a number of shortcomings serious enough to raise doubts about their validity. We detail these issues in the next section. Second, previous studies tested recursive mindreading ability only explicitly, by presenting stimuli either as text to be read, or narration to be heard, and by testing understanding with direct questions, and not implicitly, by presenting stimuli as social events to be observed, and testing understanding by measuring reactions to those events. It may be the case, especially given the ecological perspective outlined above, that human mindreading abilities are fully expressed only when they are employed within social contexts i.e. when encountered implicitly (as opposed to being encountered as explicit, disembodied descriptions of those same contexts). This possibility is supported by findings in the developmental literature which shows that children pass implicit first-level mindreading tasks (false belief tasks) far earlier than they do equivalent explicit tasks: around the first birthday vs. around the fourth birthday (see Baillargeon et al., 2010 for a review of implicit false-belief tasks). Precisely what causes this dramatic difference is an unresolved issue in developmental psychology, but whatever the explanation, these results show that the mode of presentation can make a dramatic difference to performance, at least in children. Based on this finding, we might expect that adult performance on recursive mindreading tasks could also be facilitated by implicit presentation.

In sum, recursive mindreading plays an important role in explanations of many major human behaviours, yet there are reasons to think that we may not currently know or appreciate the full extent of this ability in adult humans. In this paper, we present a new study of recursive mindreading, which has two major advances on previous research. First, we use new stimuli designed to avoid the various methodological issues we have identified in previous studies (detailed below). Second, we use a 2×2 design of implicitly and explicitly presented stories, crossed with implicitly and explicitly presented questions. As such, the key novelty here is the use of implicit stimuli, which have not previously been used in the study of recursive mindreading in adults. We expected that, at least in conditions featuring implicit presentation, participants would succeed at recursive mindreading tasks at levels higher than those reported in previous studies. Correspondingly, our design includes questions of up to seven levels of mental metarepresentation, three levels higher than the typical level of successful performance in previous tasks (e.g. Kinderman et al., 1998; Lyons et al., 2010; Stiller & Dunbar, 2007).

2. Problems with previous research

Previous studies of recursive mindreading ability used versions of the Imposing Memory Task (IMT)[4] (e.g. Kinderman et al., 1998; Lyons et al., 2010; Stiller & Dunbar, 2007). The IMT has also been widely used as a measure of mindreading ability in studies designed to identify brain regions involved in mindreading (e.g. Lewis, Rezaie, Brown, Roberts, & Dunbar, 2011; Powell, Lewis, Dunbar, García-Fiñana, & Roberts, 2010), in studies designed to investigate the relationship between mindreading and various cognitive disorders (e.g. Frith & Corcoran, 1996; Kerr, Dunbar, & Bentall, 2003), and in studies designed to investigate the relationship between mindreading ability and other aspects of social psychology, in both adults and children (e.g. Henzi et al., 2007; Liddle & Nettle, 2006; Nettle & Liddle, 2008; Paal & Bereczkei, 2007; Sylwester, Lyons, Buchanan, Nettle, & Roberts, 2012). The IMT involves stories which are read aloud to participants, followed by a series of true-or-false or forced-choice mentalising questions, designed to test participants' understanding of the levels of recursive mindreading involved. Control questions are designed to test participants' ability to remember details of the stories that are unrelated to mental states, but which contain a matched number of elements to be remembered. We analysed the stimuli used in the IMT[5] and identified five main problems that, collectively, are significant enough to cast doubt on the conclusions drawn in these studies regarding the extent of human recursive mindreading ability. We describe these issues in the following paragraphs. A full breakdown of which of these criticisms apply to which questions in the IMT is provided as supplementary information (available on the journal's Website at www.ehbonline.org).

2.1. Broken conceptual chains

In some cases, mental questions are constructed in a way that allowed them to be processed in 'chunks', rather than as a single metarepresentational unit. For example, the following sentence is intended to test fourth level mindreading: 'Simon imagined$_1$ that Betty wanted$_2$ to marry Edward but that Edward really wanted$_3$ to marry Susan, whom Jim would like$_4$ to have married' (here and elsewhere in this paragraph, we have omitted the participant's own mental state, which, if we had included it, would have had the subscript 0 (i.e. 'The participant believes$_0$ that Simon imagined$_1$…')). However, this sentence does not contain one continuous chain of mental representations. Rather, it consists of three statements, joined by logical relationships: (i) Simon imagined$_1$ that Betty wanted$_2$ to marry Edward; (ii) Simon imagined$_1$ that Edward really wanted$_2$ to marry Susan; and (iii) Simon imagined$_1$ that Jim would like$_1$ to have married Susan. Consequently, constructions of this sort do not test 4th level mindreading; they test the conjunction of multiple cases of 2nd level mindreading. 13 of 50 mental questions in the IMT are constructed like this.

2.2. Simple substitution

Some stimuli are constructed in such a way that the entire sentence did not need to be processed in order to be answered correctly. An example is the forced choice between 'The girl whose car Simon works on practices dance with the person who is a loan officer in Edward's bank' and 'The girl whose car Simon works on practices dance with the person who is a computer consultant in Edward's bank' (from Rutherford, 2004). The only difference here is between 'loan officer' and 'computer consultant'. In many cases, one of these alternatives simply did not appear in the story at all. As such, the question can be answered by simply spotting the unfamiliar item: the full sentences, and the complex propositions they convey, do not need to be understood. This occurs in 6 of 50 control questions, and 1 of 50 mental questions.

[4] Although based on the IMT, some of the later studies do not use the name IMT. Here, we use IMT to refer to all studies based on the same general idea, and set of questions used.

[5] With one exception, none of the currently published studies that we are aware of provide a complete list of the specific questions used. The exception (Liddle & Nettle, 2006) used a version modified for children. We therefore analysed the complete set of IMT questions sent to us by R. Dunbar. This set of questions is an updated version of the materials used in the earliest IMT studies, and forms the basis for the materials used in later studies. As such, the items we analysed are representative of the stimuli used in this literature.

2.3. Impossible choices

Some questions in the IMT cannot be answered based on the information included in the story, or based on reasonable inference from the story. For example, in one story we are told that: Henry gave Sam faulty information; that Henry is a prankster, and Sam suspects him of playing a trick; and that their colleague Pete does not think Henry was trying to trick Sam. Crucially, however, Henry's actual motivations (rather than just Sam's suspicions about them) are never mentioned, and cannot be reasonably inferred. Nevertheless, one true/false question was the statement 'Henry wanted to play a trick'. 7 of 50 mental questions and 2 of 50 control questions were impossible to answer in this way.

2.4. Syntactic complexity

The mentalising questions in the IMT are more syntactically recursive than the corresponding control questions. To measure this, we counted the number of embedded clauses in each of the questions (e.g. subordinate clauses such as "Susan wants to marry Edward" in the sentence "Jim thinks that Susan wants to marry Edward"). We found that the average number of embedded clauses was significantly higher in the mentalising questions than the control questions (Median level of embedding for control questions: 0; median for mental questions: 2; Mann–Whitney U Test, $p < .001$). This difference in syntactic complexity is not an issue for studies that use control questions only as a way to test participants' comprehension of the stories. However it is an issue if these controls are used as an experimental contrast with mental questions, as some IMT studies have done. For instance, neuroimaging studies use the control questions as a baseline task (e.g. Lewis et al., 2011; Powell et al., 2010). Consequently, it is possible that any differences in the brain regions associated with the two conditions may be due to the additional syntactic demands of the mental questions, rather than to mindreading specifically.

2.5. Inappropriate control questions

Finally, the control questions used in the IMT are arguably not appropriate controls in the first place, since they do not involve conceptual embedding. Recursion is the repetition of a given feature, with each repetition embedded inside a previous instance of that feature (Karlsson, 2009). In a linguistic context, this would involve the embedding of a phrase within a similar phrase, for example:

[$_{NP}$ The book [$_{PP}$ on [$_{NP}$ the desk [$_{PP}$ in [$_{NP}$ the corner [$_{PP}$ of [$_{NP}$ the room]]]]]]]

This sentence involves a noun phrase (NP) 'the room' embedded within a prepositional phrase (PP) 'of the room' embedded within a larger NP 'the corner of the room', embedded within a larger PP, and so on. The syntax is recursive, and the concept itself is also recursive: each location is contained within another location. However, although syntactic recursion is often used to express conceptually recursive concepts, such as possession, recursive locations, or mental metarepresentations (e.g. "Jake thinks that I believe that Mary feels sad"), it is possible to express conceptual recursion without heavy syntactic recursion, through parataxis. For instance, the syntactically and conceptually recursive "Portia's dog Fido's ball" can also be expressed as "Portia has a dog. That dog is called Fido. Fido has a ball", which is conceptually but not syntactically recursive. As such, an appropriate control for recursive mindreading, which is conceptually recursive, would be to use control questions that are also conceptually recursive, while controlling for syntactic recursion across both types of question. However, control questions in the IMT do not contain the same level of syntactic recursion as mental questions (see 'Syntactic complexity', above). Mental questions also do not use recursive concepts in a controlled manner (see 'Broken conceptual chains', above). Instead, the mental questions in the IMT are more recursive than the control questions, both syntactically and conceptually.

2.6. Summary

Although there have been consistent findings across different studies that use the IMT, pointing perhaps to its internal consistency, when taken as a whole the problems discussed above cast some doubt on its validity as a measure of recursive mindreading ability. Partly in light of this, and partly because we wished to use implicit as well as explicit tasks (as detailed above, the IMT uses explicit questions only), we developed new stimuli for our study.

3. The current study

Our new study was designed to address the following questions: (i) Does the step-change in performance after four levels of metarepresentation reported in previous studies still occur with different stimuli, which avoid the issues identified above?; (ii) If not, does performance decline gradually, or not at all?; and (iii) Is any effect specific to implicit mindreading tasks, or does it generalize to non-mindreading tasks and/or explicit tasks, which have less ecological validity than implicit tasks?

4. Methods

4.1. Participants

We recruited 66 participants (41 F, 25 M; average age 22y 10 m). All participants were recruited from the student population of the University of Edinburgh, through the University Careers Service Student and Graduate Employment online database, and paid £7 for their participation. Participants were screened to ensure that they (1) were native English speakers, and (2) did not know any of the actors in the stories.

4.2. Materials

We wrote four original stories, each of which had a plot involving seven levels of recursively-embedded mental representation, and seven levels of a non-mental recursive concept, such as possession. The stories were written in two different formats: as scripts, to be acted; and as a narrative, to be read by a single story-teller. All details of plot, character, and so on were identical in both formats; only the manner of presentation differed. The scripts were then performed by actors (implicit story presentation) and filmed; the narrative was read out by a single story-teller and filmed (explicit story presentation). We also created, for each of the levels of mental and non-mental recursion, two scenes to follow the main story. One of these scenes was consistent with the relevant mental/non-mental aspect of the story, the other not. Again, these additional scenes were filmed both as scripts performed by actors (implicit test question presentation), and as narratives read by a single story-teller (explicit test question presentation). This gave a total of 14 test questions for each story (7 levels of recursion, each in mental and control conditions).

These stimuli were designed to avoid each of the problems identified in the previous section. First, we did not use any broken conceptual chains. Second, we took a number of steps to ensure that simple substitution was not possible, despite minimal differences between correct and incorrect answers in our explicit stimuli (see below). Third, the only differences between correct and incorrect stimuli involved differences in mental state attribution (e.g. 'wants' vs. 'doesn't want'), rather than other aspects of the scenario (e.g. 'loan officer' vs. 'computer consultant'), thus ensuring that there was no confound of syntactic complexity. Fourth, we ensured that there were no impossible choices. Fifth, we used control questions that had the same levels of conceptual and syntactic recursion as our experimental questions.

Collectively, these aspects of our stimuli ensured that the two options could only be differentiated if participants had correctly understood the embedded levels of recursive mindreading involved. In particular, non-mindreading strategies based on forms of simple substitutions should not succeed at better than chance levels, and in any case are not possible for implicit questions. More specifically, while the correct and incorrect choices in the explicit stimuli were very similar to one another, the differences they did have were precise and only those that involved recursive mindreading. For example, one 4th-level mental question involved the choice between 'Stephen knows that Elaine knows that Bernard feels she doesn't know him well enough to date' and 'Stephen doesn't know that Elaine knows that Bernard feels she doesn't know him well enough to date'. The only difference here is in the first intentional verb ('Stephen knows…' vs. 'Stephen doesn't know…'). This means that in order to answer the question correctly, participants must parse everything that follows this verb — which in itself involves interpreting three levels of mental metarepresentation — and then determine whether this is something that Stephen knows or not (a fourth level mental metarepresentation). Similarly, if correct and incorrect stimuli were also different at other points in the construction (e.g. 'Stephen knows that Elaine knows that Bernard feels she doesn't know him well enough to date' and 'Stephen doesn't know that Elaine doesn't know that Bernard feels she doesn't know him well enough to date'), then participants need only parse what follows those differences (i.e. they need only know if Elaine does or doesn't know that Bernard feels she doesn't know him well enough to date), and nothing before that. In short, our stimuli were designed to ensure that participants had to parse and comprehend the whole scenario in order to perform at above chance levels.

One of our stories is provided in the Appendix A, in both narrative and script forms, with a complete set of questions for each. The full set of all four stories is provided as supplementary information (available on the journal's Website at www.ehbonline.org), and the videos themselves are available at http://hdl.handle.net/10283/609.

4.3. Design

Each participant viewed and was tested on all four stories, with manner of story presentation and test question presentation fully crossed within subjects (i.e. participants saw one story with implicit presentation of the story and implicit testing, one with explicit story presentation and implicit testing, and so on), in a fully counterbalanced design, such that each story appeared in each position (1st viewed, 2nd viewed, etc.) an equal number of times, and each story appeared before and after each other story an equal number of times.

4.4. Procedure

Participants were first presented with the story video. They were able to watch this as many times as they liked before proceeding to the questions, after which they could not watch the story again (this mimics, in a different modality, the procedure used in several previous IMT tasks, in which participants read stories on paper). The fourteen test questions for each story (7 mental, 7 control) were presented in random order. For each question, two video frames were shown on the screen simultaneously, presenting the two forced choice options for that question, with left/right presentation randomized. Participants were able to watch both videos as many times as they liked (again, this mimics procedures from previous research), until they wished to identify, by mouse click, the video which they thought was consistent with the story. After each selection, participants were asked to rate their confidence, on a scale of 1 (lowest) to 10 (highest), that they had chosen the correct answer. After making a selection, participants were not able to return to that question.

4.5. Data analysis

Data for a single story from three participants could not be used due to computer error. We analysed three dependent variables: success (i.e. identifying the correct video) on test; the confidence ratings that participants attached to those responses; and the number of times that participants viewed each question video before answering (which may index whether participants found certain types of questions more difficult than others, and had re-watch them in order to answer). For the statistical analysis, multilevel models with random effects were employed: responses were fitted to a binomial distribution for the binary success DV, and a Poisson distribution for the confidence rating and number of additional video views DVs.[6] Analyses were conducted in the R programming environment (version 3.1.1, R Core Team, 2014) using the lme4 package (version 1.1–7; Bates, Maechler, Bolker, & Walker, 2014). Lme4 was used at default settings, except for the usage of the `bobyqa'-optimizer and increasing the maximum iterations to 100,000. The theoretically-motivated factors level of question complexity (1–7, indicating required level of metarepresentation), condition (mental or control), story presentation (implicit or explicit) and question presentation (implicit or explicit) were included as fixed effects for all DVs. Additionally, we included number of additional story views (i.e. beyond the first, obligatory viewing of the story) as a predictor, in order to control for the effects this might have on performance. Following Barr, Levy, Scheepers, and Tily (2013) we used a maximum random effects model, and consequently included subject (64 levels) and story (4 levels) as random effects, with by-subject and by-story random intercepts and random slopes for level, question type, story presentation and question presentation, which represented the most complex converging model. For further analyses, we reduced the random effects structure further, if the models did not converge. The validity of the mixed effects analyses were assessed by computing likelihood ratio tests comparing models containing effects with null models that contained the intercept and the random effect structure only (c.f. Mundry, 2011). Due to the nature of multilevel analysis, F-ratios will not be reported.

5. Results

5.1. Success rates

The model predicting success did significantly better than the null model ($\chi 2(15) = 28.244$, $p = .02$), and showed slight underdispersion (dispersion parameter of 0.771). Overall success rates for all questions were well above the level expected under chance performance (intercept in log-odds space of 2.356, corresponding to odds of greater than 10:1 of answering correctly, i.e. answering around 90% of all questions correctly). Moreover, there was very little evidence of any effect of level or condition on success (level: $\beta = -0.037$, SE $= 0.055$, $p = .494$; condition: $\beta = 0.092$, SE $= 0.329$, $p = .779$; see Fig. 1), nor of any interaction between these two factors ($\beta = -0.073$, SE $= 0.066$, $p = .266$).

There were two significant interactions between predictors of accuracy (Fig. 2). First, there was an interaction between condition and story presentation ($\beta = 0.727$, SE $= 0.282$, $p = .01$). Post-hoc tests using dummy-coded data and taking control-explicit story as the baseline showed that only the mental-explicit combination (i.e. the combination investigated in the IMT) performed worse than baseline (mental-explicit: $\beta = -0.366$, SE $= 0.173$, $p = .034$; control-implicit: $\beta = -0.526$, SE $= 0.391$, $p = .179$; mental-implicit: $\beta = -0.067$, SE $= 0.398$, $p = .866$).

Second, there was an interaction between story presentation and question presentation ($\beta = 0.885$, SE $= 0.304$, $p < .001$). Post-hoc

[6] As confidence ratings were generally towards the upper end of the 1 to 10 scale, confidence measures were subtracted from 10, in order to avoid predictions outside the possible range. Note that this reverses the expected sign of the coefficient estimates in the models for confidence.

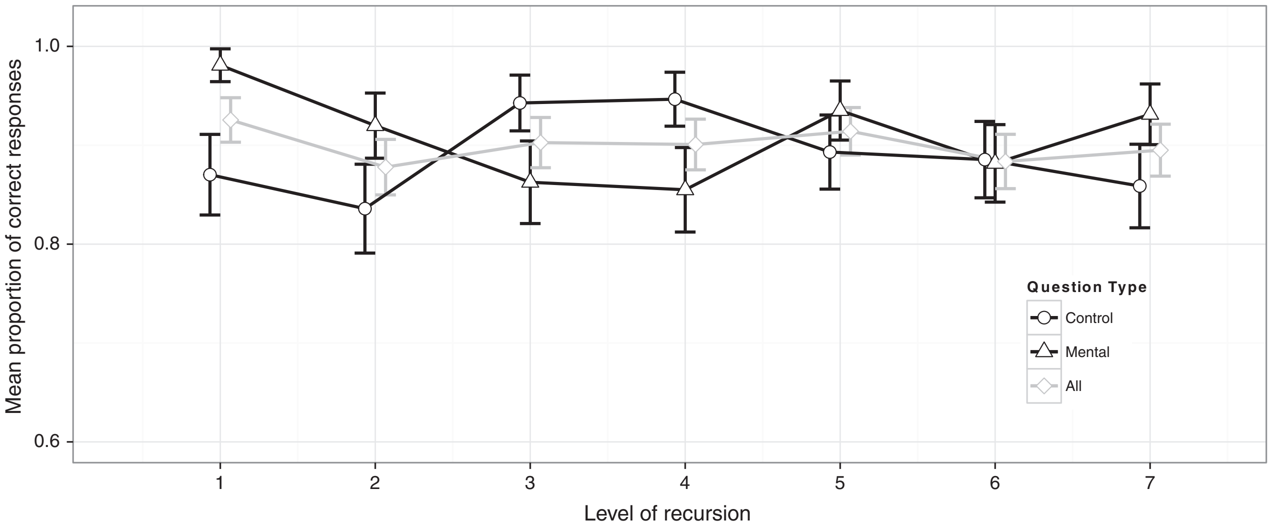

Fig. 1. Mean proportion correct (by participant) at each level of mindreading. Error bars give 95% CIs. These data show that performance on both mental and control tasks does not decrease as the level of embedding increases.

tests revealed no difference between the two modes of question presentation for explicitly-presented stories ($\beta = 0.089$, SE = 0.603, $p = .883$), but for implicitly-presented stories performance was significantly worse if the questions were presented explicitly ($\beta = 0.767$, SE = 0.364, $p = .035$). In other words, an implicit story followed by an explicit question was the most difficult combination of story presentation and question presentations. All other main effects and interactions were not significant ($p > .204$).

Finally, additional views of the story videos lead to a significant improvement in success (significant effect of number of additional story views: $\beta = 0.457$, SE = 0.138, $p < .001$). While the inclusion of number of additional story views as a factor does improve model fit significantly ($\chi2(1) = 11.312$, $p < .001$), a model lacking number of additional story views as a factor produced qualitatively similar results to those outlined above.

In summary, participants were able to successfully process recursive mental concepts even at high levels of recursion, and this was no more difficult than other, non-mental recursive concepts.

5.2. Judgements of confidence

For the confidence data the fitted model did significantly better than the null model ($\chi2(15) = 275.87$, $p < .001$), and had a dispersion parameter of 1.356 suggesting overdispersion within the acceptable range. Further analysis of confidence ratings revealed several effects of our manipulations. There was a small but significant effect of level ($\beta = 0.106$, SE = 0.022, $p < .001$; see Fig. 3): participants' confidence dropped as the levels increased even though, as noted above, their actual level of accuracy remained high. There were also a number of significant two- and three-way interactions involving level. Confidence ratings for mental but not control questions decrease with level (main model reveals a significant level × condition interaction, $\beta = 0.058$, SE = 0.012, $p < .001$; post-hoc tests using multilevel models on subsets of the data show a significant effect of level for mental questions, $\beta = 0.070$, SE = 0.0546, $p = .009$, but not for control questions, $\beta = 0.045$, SE = 0.030, $p = .134$). This interaction is further modulated by both story presentation and question presentation (level × condition

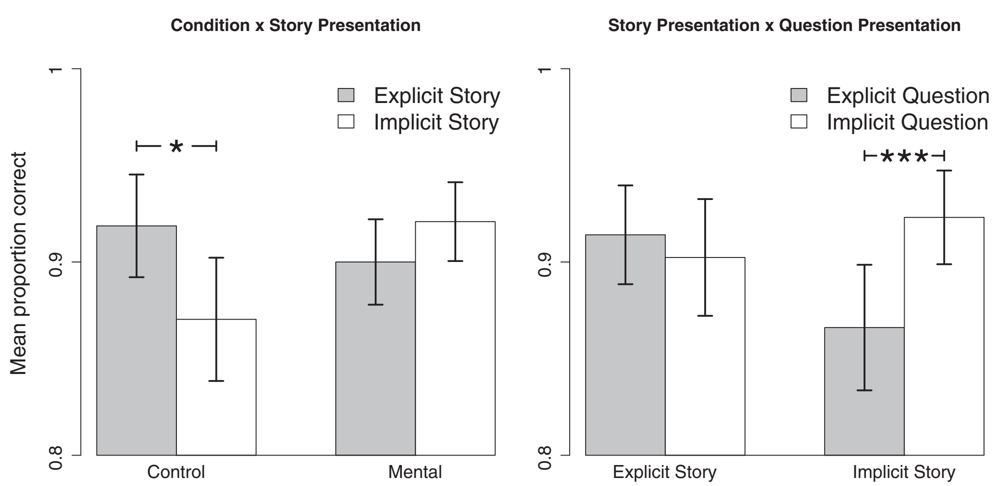

Fig. 2. The condition × story presentation (left) and story presentation × question presentation (right) interactions for accuracy. Bars give means of the by-participant mean success rates; error bars indicate 95% CIs. These results show that for control questions, performance was lower when the stories were presented implicitly rather than explicitly (left-hand side), and that performance was reduced when stories were presented implicitly and questions explicitly, in comparison to all other possible combinations of story presentation and question presentation (right-hand side). These were the only significant interactions for accuracy.

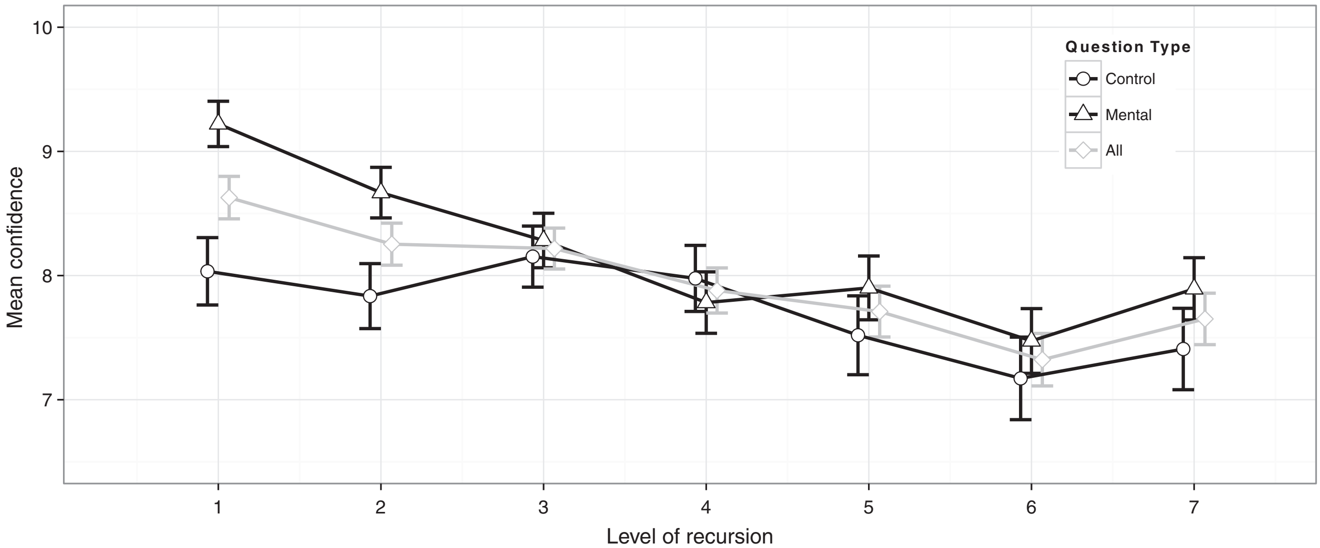

Fig. 3. Mean confidence (by participant). Error bars give 95% CIs. These data show that although actual performance (i.e. accuracy) does not decrease as the level of embedding increases (see Fig. 1), our participants' confidence in their answers does decrease.

× story presentation interaction: $\beta = 0.076$, SE $= 0.024$, $p = .001$; level × condition × question presentation interaction: $\beta = -0.136$, SE $= 0.024$, $p < .001$). Post-hoc tests on control questions reveal no two-way interactions between level and story presentation (($\beta = -0.016$, SE $= 0.017$, $p = .325$) or level and question presentation ($\beta = 0.020$, SE $= 0.016$, $p = .206$): however, these two-way interactions are significant for mental questions (level × story presentation: $\beta = 0.082$, SE $= 0.019$, $p < .001$; level × question presentation: $\beta = -0.116$, SE $= 0.018$, $p < .001$): confidence on mental questions decreases faster with increasing level when the story is presented implicitly, or when the questions are presented explicitly. Finally, there was also a three-way interaction between condition, story presentation and question presentation ($\beta = 0.358$, SE $= 0.098$, $p < .001$). Post-hoc tests using dummy-coded data and taking explicit story–explicit question as the baseline showed that for control questions, participants had significantly lower confidence for the implicit story–explicit question combination ($\beta = 0.464$, SE $= 0.164$, $p = .005$), with implicit story–implicit question ($\beta = 0.104$, SE $= 0.270$, $p = .700$) and explicit story–implicit question ($\beta = 0.281$, SE $= 0.201$, $p = .125$) not significantly different from the explicit story–explicit question intercept ($\beta = 0.438$, SE $= 0.146$, $p < .001$); in contrast, for mental questions, participants had higher confidence for the implicit story–implicit question combination than for explicit story–explicit question ($\beta = -0.537$, SE $= 0.179$, $p = .003$; other comparisons n.s., $p > .137$).

Finally, while there was a tendency for participants' confidence rating to increase with number of additional story views, this effect was not significant ($\beta = -0.086$, SE $= 0.049$, $p = 0.079$). Furthermore, including this factor did not significantly improve model fit ($\chi2(1) = 2.253$, $p = .133$) over an equivalent model lacking this factor, and the simpler model yielded qualitatively similar results to those discussed above.

5.3. Number of additional video views

The third dependent variable was the number of times each question video was viewed. The fitted model did significantly better than the null model ($\chi2(15) = 172.52$, $p < .001$) and had a dispersion parameter of 1.095, suggesting low overdispersion within the acceptable range.

This model revealed a small but significant effect of level ($\beta = 0.169$, SE $= 0.016$, $p < .001$): participants' required additional views of the question videos as levels increased. There were also significant effects of condition ($\beta = -0.272$, SE $= 0.075$, $p < .001$: mental questions required fewer additional plays) and question presentation ($\beta = -0.473$, SE $= 0.086$, $p < .001$: implicitly-presented questions required fewer additional plays), and various two-way interactions involving level, condition, and question presentation. Ultimately these are best explained by considering the significant three-way interaction between level, question presentation and condition ($\beta = -0.171$, SE $= 0.054$, $p = .001$; see Fig. 4). Post-hoc tests on subsets of the data show that the number of times the question videos were viewed increased with level for explicit questions (there were strong effects of level in explicit–control and explicit–mental conditions: explicit–control, $\beta = 0.207$, SE $= 0.030$, $p < .001$; explicit–mental, $\beta = 0.278$, SE $= 0.038$, $p < .001$) and for implicit–control questions ($\beta = 0.131$, SE $= 0.057$, $p = 0.022$), but not for implicit–mental questions ($\beta = 0.056$, SE $= 0.100$, $p = .577$), which showed no significant increase in additional question viewings with increasing complexity.

There was also a significant interaction between story presentation and question presentation ($\beta = -0.404$, SE $= 0.121$, $p < .001$). Follow-up analyses using dummy coded data with explicit story–explicit questions as the baseline revealed that, compared to this baseline, combinations of implicit story–implicit questions required significantly fewer views before answering ($\beta = -0.539$, SE $= 0.129$, $p < .001$), as did explicit stories–implicit question combinations ($\beta = -0.397$, SE $= 0.122$, $p = .001$). Implicit story–explicit question combinations required more views than baseline before answering, albeit not significantly so ($\beta = 0.170$, SE $= 0.113$, $p = .131$).

Finally, while there was a tendency for participants' to require fewer views of the question videos if they had watched the stories more often, this effect was not significant ($\beta = -0.042$, SE $= 0.075$, $p = 0.574$). Furthermore, including this factor did not significantly improve model fit ($\chi2(1) = 0.306$, $p = .580$) over an equivalent model lacking this factor, and the simpler model yielded qualitatively similar results to those discussed above.

6. Discussion

Our study produced several results worthy of note, of which two are of particular importance.

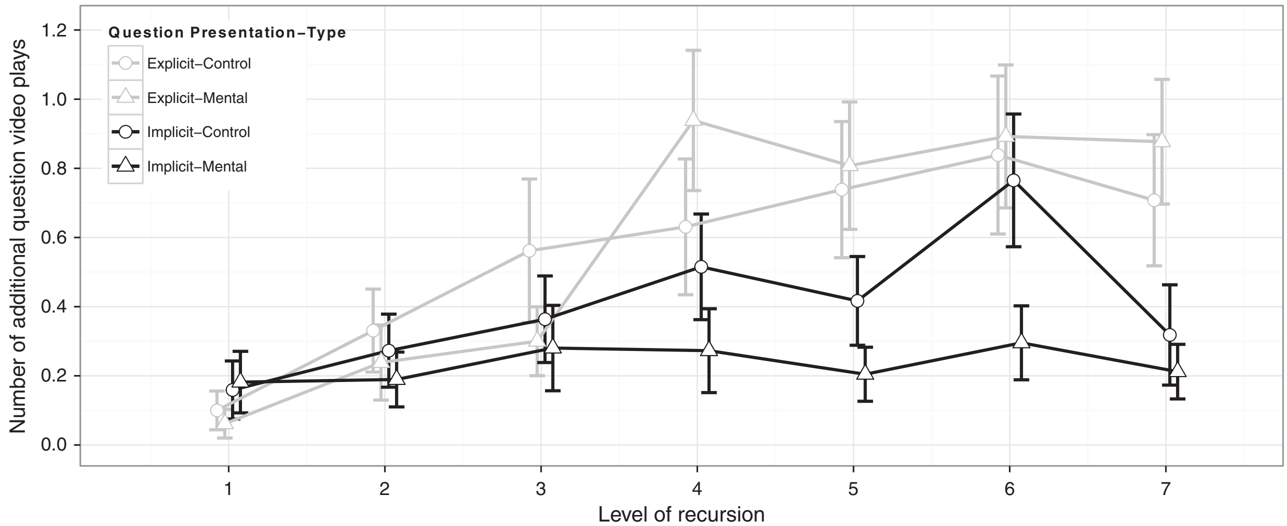

Fig. 4. Number of additional video views for each combination of condition and question presentation. Error bars give 95% CIs. These results show that for all control questions (i.e. both implicit and explicit), and for mental questions that were presented explicitly, participants chose to watch the videos increasingly often as level increased. This was, however, not true for mental questions presented implicitly, for which there was no corresponding increase with level. Given that this reduction in the number of views did not lead to any reduction in actual performance (see Figs. 1 and 2), this is tentative evidence that the implicit mental questions were processed more easily than all other types of question. In the Discussion we suggest that this may be because recursive mindreading is actually easy, when presented within is natural social context.

First, performance on mindreading tasks was high throughout (see Fig. 1). The design of our stimuli ensured that this level of performance is unlikely to be due to guesswork or any other strategy that did not involve recursive mindreading (see *Methods*). These findings run counter to the intuition that high level recursive mindreading tasks are cognitively demanding, and counter to results obtained in previous research that suggest that performance on mental questions decrease markedly after level 5. One way to understand this result is by analogy with our perceptual skills: a formal description of what is involved in, say, vision, is complex, but this does not mean that seeing is a cognitively demanding activity, beyond the ken of typical human abilities. Our results suggest that the same may be true of recursive mindreading, even at high levels of recursion. Interestingly, the intuition that high levels of recursive mindreading are particularly cognitively demanding extends even to the individuals involved: while actual performance remained high across all levels, confidence levels declined as level of embedding increased, for mental questions but not control questions (see Fig. 3).

Second, we found that participants viewed the videos more often as the level of complexity increased — except for implicitly presented mental questions i.e. except in those contexts that are most ecologically valid (see Fig. 4). This tentatively suggests that recursive mindreading is especially easy when employed within its natural environment, and that it is otherwise no more or less easy than recursive tasks in general. As we emphasised in the Introduction, humans' natural ecology is social. Correspondingly, prominent accounts of the evolution of human cognition emphasise the importance of specifically social cognition, including mental state attribution (Sterelny, 2003; Tomasello, 2014; Tooby & Cosmides, 1992). Our results make sense from this ecological perspective: recursive mindreading is an essential, ubiquitous, and adaptive component of everyday life, and as such, we should expect that we are good at it. A natural extension of our study, which would increase the ecological validity, would be to limit the participants to single views of both the story and the question videos (rather than the multiple views that, following previous research, we allowed for in the current design).

These results are consistent with the picture emerging from the literature on adult first-level mindreading, which shows that mindreading may be less like thinking, and more like perception i.e. something that we do unconsciously, as part of the background cognition that manages much of our daily lives (Apperly, 2011). Several experiments have now shown that in implicit contexts, we track the beliefs of others automatically, as part of our intuitive monitoring of the world around us, and that like our perceptual experiences, these representations of others' mental states fade quickly if we do not focus on them (e.g. Kovács, Téglás, & Endress, 2010; Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010; van der Wel, Sebanz, & Knoblich, 2014). Our results tentatively suggest that the same may be true of recursive mindreading. In particular, we found that although participants in general increased their number of views of the question videos as level increased, suggesting an increase in the level of difficulty, this was not true of implicit mental questions.

It is also instructive to compare our results with the developmental literature. First, note that the classic Sally-Anne false belief task uses an implicit story (acted out using dolls, albeit with explicit commentary attached), followed by an explicit question. Our results show that this implicit–explicit combination is the most challenging combination for adults in recursive mindreading tasks (see Fig. 2). The robustness of this finding, and the exact reasons for it, are topics for future research, but it does raise the possibility that the classic false belief task involves the most cognitively demanding combination of story and question presentation possible. More generally, our results suggest two possible lines of future research on the development of mindreading abilities: (i) the use of explicit–explicit and/or explicit–implicit methods, in order to make comparisons with the existing implicit–explicit and implicit–implicit approaches; and (ii) the investigation of higher-level, recursive mindreading abilities in children, using implicit–implicit methods. Implicit–Implicit methods have dramatically re-shaped our understanding of the development of simple mindreading abilities, but this advance has not yet been extended to the development of recursive mindreading abilities.

These results differ somewhat from previous research on adult recursive mindreading, which found a prominent drop in performance after four levels of metarepresentation. We suggest two possible reasons for this. The ecological validity of our implicit tasks cannot fully explain this, since we also find high levels of performance on explicit tasks. A more likely explanation is the various methodological problems we have identified with the IMT, which previous studies used as a measure of mindreading ability (see *Problems with previous research*, above). These methodological problems raise the possibility that previous results may not accurately reflect human mindreading abilities, and their relationship to other aspects of social psychology, as accurately as possible. Further investigation into the exact reasons for the differences between our study and previous research may be warranted.

The broader implications of our findings are several. In particular, our results should reduce concerns that some theoretical explanations of many important human behaviours and institutions are implausible precisely because they invoke high level recursive mindreading abilities. For example, the most prominent theoretical accounts of human communication argue that it involves the expression, on the part of the speaker, of an intention that the audience recognises that the speaker has an intention to inform the audience — and that the audience must recognise these embedded intentions (Csibra, 2010; Grice, 1969; Sperber, 2000b; Tomasello, 2008). Several researchers have argued that, while theoretically cogent, this analysis is empirically implausible, precisely because it depends upon high levels of recursive mindreading, which are assumed to be cognitively demanding (e.g. Breheny, 2006; Clark, 1996; Glüer & Pagin, 2003; Gómez, 1994). Our results suggest that these concerns are likely unfounded: at least in the contexts we explored here, recursive mindreading poses no particular challenges for adult humans, even at high levels of embedding. The same point applies to numerous other activities that have been argued to depend upon recursive mindreading, such as language, story-telling, culture, and even consciousness (Corballis, 2011, Dunbar, 2003, 2005, 2008; Graziano, 2013; Sperber, 2000a).

Supplementary materials

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.evolhumbehav.2015.01.004.

Acknowledgements

COG gratefully acknowledges financial support from the Skye Foundation. TSP gratefully acknowledges financial support from the Leverhulme Trust and the ESRC. We would like to thank Tamsin Saxton for valuable discussion of the methods, Robin Dunbar for sending us the IMT used in previous studies, and Mateo Obregón and Roger Mundry for statistical advice.

Appendix A. Example script & narrative (story: 'Bernard')

Narrated version

One evening, Megan finds out that her sister Lauren wants to go out with a boy in her biology class, Stephen. Megan tells Lauren that Stephen used to be best friends with a boy called Chris, who is now Megan's best friend. Lauren tells Megan that she saw Stephen smiling and flirting with their cousin, Elaine, and so she thinks Stephen might want to go out with Elaine. Because Lauren thinks Stephen likes someone else, she is too nervous to ask him out.

Megan talks to Elaine at school and finds out that Elaine actually wants to go out with Bernard, whom Megan knows from the school play. Megan learns that Elaine and Bernard are next-door neighbours, and that Bernard thinks that Elaine doesn't know him well enough to date. Elaine tells Megan that Stephen knows how Elaine feels about Bernard and how Bernard feels about Elaine.

Megan later talks to her friend Chris about the situation, realising that if Lauren knew about Elaine's situation, and knew that Stephen knows about it too, Lauren would realise that Stephen doesn't want to go out with Elaine, and might work up the courage to ask him out. Megan plans to tell Lauren about everything that evening.

Mental questions for narrated version

1.	A.	Elaine likes Bernard
	B.	Elaine likes Stephen
2.	A.	Megan knows that Lauren wants to ask Stephen out
	B.	Megan doesn't know that Lauren wants to ask Stephen out
3.	A.	Elaine doesn't know that Bernard feels that she doesn't know him well enough to date
	B.	Elaine knows that Bernard feels that she doesn't know him well enough to date
4.	A.	Stephen knows that Elaine knows that Bernard feels she doesn't know him well enough to date
	B.	Stephen doesn't know that Elaine knows that Bernard feels she doesn't know him well enough to date
5.	A.	Megan knows that Stephen doesn't know that Elaine knows that Bernard feels that she doesn't know him well enough to date
	B.	Megan knows that Stephen knows that Elaine knows that Bernard feels that she doesn't know him well enough to date
6.	A.	Chris knows that Megan knows that Stephen knows that Elaine knows that Bernard feels that she doesn't know him well enough to date
	B.	Chris doesn't know that Megan knows that Stephen knows that Elaine knows that Bernard feels that she doesn't know him well enough to date
7.	A.	Megan wants Lauren to know that she, Megan, knows that Stephen knows that Elaine knows that Bernard feels she doesn't know him well enough to date, so that Lauren asks Stephen out
	B.	Megan doesn't want Lauren to know that she, Megan, knows that Stephen knows that Elaine knows that Bernard feels she doesn't know him well enough to date, so that Lauren doesn't ask Stephen out

Control questions for narrated version

1.	A.	Lauren is Megan's sister
	B.	Lauren is Megan's friend
2.	A.	Stephen has Biology with Megan's sister Lauren
	B.	Stephen was in the school play with Megan's friend Chris
3.	A.	Chris used to be best friends with Stephen, who has Biology with Megan's sister
	B.	Chris used to be best friends with Bernard, who acted in the school play with Megan's sister
4.	A.	Megan is best friends with Chris, who used to be best friends with Stephen, who has Biology with Megan's sister Lauren
	B.	Megan's sister Lauren used to go out with Chris, who is best friends with Elaine, who has Biology with Stephen
5.	A.	Bernard acted in the play with Megan, who is the best friend of Chris, who used to be best friends with Stephen, who takes Biology with Lauren who is Megan's sister
	B.	Bernard acted in the play with Lauren, who used to go out with Chris, who used to be best friends with Stephen, who lives next door to Elaine who is Lauren's cousin
6.	A.	Elaine lives next door to Bernard, who acted in the play with Megan, who is the best friend of Chris, who used to be best friends with Stephen, who takes Biology with Lauren, who is Megan's sister
	B.	Elaine lives next door to Bernard, who is the best friend of Chris, who used to be best friends with Stephen, who was in the play with Lauren's sister Megan, who has Biology with Elaine
7.	A.	Megan's cousin is Elaine, who lives next door to Bernard, who acted in the play with Megan, who is the best friend of Chris, who used to be best friends with Stephen, who has Biology with Megan's sister Lauren
	B.	Megan's cousin is Elaine, who has Biology with Bernard, who acted in the play with Stephen, who used to go out with Megan, who is the best friend of Chris, who has Biology with Megan's sister Lauren

Scripted version

Scene 1

Megan: Lauren, your phone just buzzed.

Lauren snatches up her phone

Megan: Ooooh, is that a booooy?

Lauren: Maybe…OK, OK, it's Stephen. You know, the one in my biology class.

Megan: Oh, I didn't know you had a thing for him. Did you know he used to be best friends with Chris?

Lauren: Which Chris?

Megan: The Chris who's my best friend. Yeah, they were friends in primary school.
Lauren: That's weird. Anyway…I was actually thinking about working up the guts to ask him out, but I'm nervous.
Megan: Haha, you'd better not let our Dad find out…Anyway, why are you nervous? I'm sure he'll say yes.
Lauren: Well, I saw him chatting to Elaine at lunchtime the other day, and he was all smiley and flirty, so I think he's planning to ask her out.

Scene 2
Megan: Hey Elaine!
Elaine: Oh, hey cousin, could you hold this for a second?
Megan: So, I saw you flirting with Stephen the other day…he's cute!
Stephen walks past them, they pause awkwardly and giggle
Elaine: Haha, yup, he is, but there's nothing going on.
Megan: Why not?
Elaine: I kind of have a thing for Bernard…
Megan: Bernard? The one who was in the school play with me?
Elaine: Yeah, that's the one.
Megan: Oh, I didn't know you guys knew each other.
Elaine: I live next door to him. Anyway, he doesn't like me back.
Megan: Why on earth not?
Elaine: Well, I dunno if it's that he doesn't like me, it's more just that he thinks I don't know him well enough to really want to date him. He says if I knew him better I wouldn't want to… really weird.
Megan: Ugh, he's just being melodramatic. I'm sure he'll come round. Does Stephen know? He looked pretty into you when you were chatting.
Elaine: Yup, I told him about it ages ago, we were just laughing about Mr Murray's new haircut.
Megan: Ah, OK. Anyway, I'd better run – hope Bernard realises what he's missing!

Scene 3
Megan: …so it turns out that Elaine and Stephen aren't into each other at all. Elaine is actually really into Bernard.
Chris: The acting one?
Megan: Yeah. Bernard's super melodramatic, too – he doesn't want to date Elaine because he reckons Elaine doesn't know him well enough or something.
Chris: That's weird. And Stephen knows that Elaine's not into him?
Megan: Mm, Elaine told him about the whole Bernard thing ages ago.
Chris: So Lauren…
Megan: Right! So, Lauren doesn't want to ask Stephen out because she thinks he's into Elaine – but if she knew that Stephen knows that Elaine likes Bernard, and that Stephen knows that Elaine's not into him, she might work up the guts to ask Stephen out.
Chris: I guess…so are you going to tell her?
Megan: Yeah, I'm going to tell her the whole thing tonight.

Mental questions for scripted version

1.	A.	Elaine:	Ahhh, there goes Bernard, he's so cute…can't remember the last time I liked someone so much.
	B.	Elaine:	Ahhh, there goes Stephen, he's so cute…can't remember the last time I liked someone so much.
2.	A.	Chris:	I heard your sister wants to ask Stephen out!
		Megan:	What? She hasn't told me that!
	B.	Chris:	Lauren told me today that she wants to ask Stephen out.
		Megan:	Oh, yeah, she told me about it last night
3.	A.	Elaine:	I worked up the guts to ask Bernard out today. He said no, but I have no idea why he won't go out with me
	B.	Elaine:	I worked up the guts to ask Bernard out today. He said no. He reckons I don't know him well enough
4.	A.	Megan:	Hey, Stephen, I thought you should know something.
		Stephen:	Yeah?
		Megan:	It's about Elaine – you know she has a thing for Bernard?
		Stephen:	Yeah, and he doesn't want to go out with her because he thinks she doesn't know him enough? Poor girl, she was crushed when she found out
	B.	Megan:	Oh, I have some hot gossip – did you hear about Elaine?
		Stephen:	No – what about her?
		Megan:	She's really into Bernard, but Bernard doesn't like her back because he thinks Elaine doesn't know him well enough.
		Stephen:	Seriously? That's a weird reason not to go out with someone. Why didn't Elaine tell me about this?
5.	A.	Megan:	So Elaine told me today that she has a thing for Bernard, but Bernard doesn't like her back because he thinks Elaine doesn't know him well enough.
		Chris:	Elaine has a thing for Bernard? But I saw her flirting with Stephen!
		Megan:	Mm-hmm, and Stephen has no idea, either that Elaine likes Bernard, or that Bernard doesn't like her back.
		Chris:	Yikes, that's not cool – she needs to tell him.
	B.	Megan:	So Elaine told me today that she has a thing for Bernard, but Bernard doesn't like her back because he thinks Elaine doesn't know him well enough.
		Chris:	Elaine has a thing for Bernard? But I saw her flirting with Stephen!
		Megan:	Oh, yeah, he knows, she told him.
		Chris:	Knows which bit?
		Megan:	Knows that Elaine likes Bernard and that Bernard told Elaine that he won't go out with her.
6.	A.	Chris:	Yeah, so Megan heard some juicy gossip today…Stephen too? Haha, I bet it's the same gossip – they both know about this, anyway… So you know Elaine? She has a mega thing for Bernard… uh-huh… and Bernard told her – yeah, told Elaine – that he doesn't want to go out with her because Elaine doesn't know him well enough. I know, it's weird, right? So, wait, you spoke to Megan earlier and asked her about Elaine and Stephen, and Megan said she didn't know anything about it? That's weird, she definitely knows.
	B.	Chris:	Yeah, so what's this hot gossip? Elaine and Bernard? Why not? Because she doesn't know him well enough? That's weird…does Stephen know? Ok, that's good, it would be really awful of Elaine if he didn't. And Megan knows too? What? Has she heard that Stephen knows about all of it? That's weird, Megan didn't tell me anything.
7.	A.	Megan:	So, I'm thinking that Lauren needs to know what I heard. Right? Because if she knows what I know right now, about Elaine's crush, and Bernard's rejection, and that Stephen knows the whole thing…she'll work up the guts to ask him out! So I'm going to tell her tonight.
	B.	Megan:	Well, if you think about, if Lauren knew what I heard today – and if she knew that Stephen knew all about it too, about Elaine's crush and Bernard's weird reason for rejection and everything – she'd ask Stephen out. But I don't want her to do that, so I'm not going to tell her.

Control questions for scripted version

1.	A.	Megan:	So I was chatting to my friend, Lauren…
	B.	Megan:	So I was chatting to my sister, Lauren…
2.	A.	Megan:	You know Stephen? The one who has Biology with my sister Lauren?
	B.	Megan:	You know Stephen? The one who was in the school play with my friend Chris?
3.	A.	Megan:	Did you know that Chris used to be best friends with Stephen?
		Elaine:	Which Stephen?
		Megan:	The one who has Biology with my sister Lauren.
	B.	Megan:	Did you know that Chris used to be best friends with Bernard?
		Elaine:	Which Bernard?
		Megan:	The one who was in the school play with my sister Lauren.
4.	A.	Megan:	You know Stephen, who has Biology with my sister? He used to be best friends with my best friend, Chris.
	B.	Megan:	You know my sister Lauren used to go out with Chris, who's best friends with Elaine? She has Biology with Stephen.
5.	A.	Elaine:	No, the other Megan, the one whose best friend Chris used to be best friends with this guy Stephen who has Biology with Lauren, Megan's sister…yeah, she knows Bernard from the school play.
	B.	Elaine:	No, the other Lauren, the one who used to go out with that Chris guy who used to be best friends with Stephen…yeah, Stephen who lives next door to me. Yup, that's the Lauren who's my cousin.

(continued on next page)

6. A. Elaine: You know Bernard, who was in the play with Megan? No, the Megan whose best friend is that Chris guy, the one who used to be best friends with Stephen…nuh-uh, the other Stephen, the one who has Biology with Megan's sister Lauren. Yeah, that Megan. Remember she co-starred with Bernard? That's my next-door neighbour.
   B. Elaine: You know Bernard, Chris's best friend? No, the Chris who used to be best friends with Stephen, that guy who starred in the play with Lauren's sister Megan, who's in Biology with us. Yeah, that Bernard is my next-door neighbour.
7. A. Megan: My cousin, Elaine, the one who lives next door to Bernard? Argh, you know Bernard. He was in the play with me and my best friend Chris…Chris who used to be best friends with Stephen, the guy in Biology with my sister Lauren.
   B. Megan: You know my cousin Elaine, who has Biology with Bernard? Argh, you know Bernard…he acted in the play with my ex, Stephen, and my best friend Chris…Chris who has lives next door to me and my sister Lauren.

References

Apperly, I. A. (2011). *Mindreaders: The cognitive basis of 'Theory of Mind'.* New York, NY: Psychology Press.

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, *14*(3), 110–118.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind.* Cambridge, MA: MIT press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–7. URL http://CRAN.R-project.org/package=lme4

Breheny, R. (2006). Communication and folk psychology. *Mind & Language*, *21*(1), 74–107.

Byrne, R. W., & Whiten, A. (1989). *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes, and humans.* Oxford: Oxford University Press.

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, *12*(5), 187–192.

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.

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

Corballis, M. C. (2011). *The recursive mind: The origins of human language, thought, and civilization.* Princeton, NJ: Princeton University Press.

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language*, *25*(2), 141–168.

Dunbar, R. I. M. (2003). The social brain: Mind, language, and society in evolutionary perspective. *Annual Review of Anthropology*, *32*, 163–181.

Dunbar, R. I. M. (2005). Why are good writers so rare? An evolutionary perspective on literature. *Journal of Cultural and Evolutionary Psychology*, *3*(1), 7–21.

Dunbar, R. I. M. (2008). Mind the gap: Or why humans aren't just great apes. *Proceedings of the British Academy*, *154*, 403–423.

Frith, C. D., & Corcoran, R. (1996). Exploring 'theory of mind' in people with schizophrenia. *Psychological Medicine*, *26*(03), 521–530.

Glüer, K., & Pagin, P. (2003). Meaning theory and autistic speakers. *Mind & Language*, *18*(1), 23–51.

Gómez, J. C. (1994). Mutual awareness in primate communication: A Gricean approach. In S. T. Parker, R. W. Mitchell, & M. L. Boccia (Eds.), *Self-Awareness in Animals and Humans* (pp. 61–80). Cambridge: Cambridge University Press.

Graziano, M. S. (2013). *Consciousness and the social brain.* Oxford University Press.

Grice, H. P. (1969). Utterer's meaning and intention. *The Philosophical Review*, *78*(2), 147–177.

Henzi, S. P., de Sousa Pereira, L. F., Hawker-Bond, D., Stiller, J., Dunbar, R. I. M., & Barrett, L. (2007). Look who's talking: Developmental trends in the size of conversational cliques. *Evolution and Human Behavior*, *28*(1), 66–74.

Humphrey, N. K. (1976). The social function of intellect. In P. P. G. Bateson, & R. Hinde (Eds.), *Growing Points in Ethology* (pp. 303–317). Cambridge: Cambridge University Press.

Karlsson, F. (2009). Syntactic recursion and iteration. In H. van der Hulst (Ed.), *Recursion and Human Language: Studies in Generative Grammar* (pp. 43–67). Berlin: Mouton de Gruyter.

Kerr, N., Dunbar, R. I., & Bentall, R. P. (2003). Theory of mind deficits in bipolar affective disorder. *Journal of Affective Disorders*, *73*(3), 253–259.

Kinderman, P., Dunbar, R., & Bentall, R. P. (1998). Theory-of-mind deficits and causal attributions. *British Journal of Psychology*, *89*(2), 191–204.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830–1834.

Lewis, P. A., Rezaie, R., Brown, R., Roberts, N., & Dunbar, R. I. M. (2011). Ventromedial prefrontal volume predicts understanding of others and social network size. *NeuroImage*, *57*(4), 1624–1629.

Liddle, B., & Nettle, D. (2006). Higher-order theory of mind and social competence in school-age children. *Journal of Cultural and Evolutionary Psychology*, *4*, 231–246.

Lyons, M., Caldwell, T., & Shultz, S. (2010). Mind-reading and manipulation - Is Machiavellianism related to theory of mind? *Journal of Evolutionary Psychology*, *8*(3), 261–274.

Miller, S. A. (2009). Children's understanding of second-order mental states. *Psychological Bulletin*, *135*(5), 749–773.

Mundry, R. (2011). Issues in information theory-based statistical inference – A commentary from a frequentist's perspective. *Behavioral Ecology and Sociobiology*, *65*(1), 57–68.

Nettle, D., & Liddle, B. (2008). Agreeableness is related to social-cognitive, but not social-perceptual, theory of mind. *European Journal of Personality*, *22*, 323–335.

Paal, T., & Bereczkei, T. (2007). Adult theory of mind, cooperation, Machiavellianism: The effect of mindreading on social relations. *Personality and Individual Differences*, *43*, 541–551.

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

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

R Core Team (2014). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing (http://www.R-project.org).

Rutherford, M. D. (2004). The effect of social role on theory of mind reasoning. *British Journal of Psychology*, *95*(1), 91–103.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1255.

Sperber, D. (1997). Intuitive and reflective beliefs. *Mind & Language*, *12*(1), 67–83.

Sperber, D. (Ed.). (2000). *Metarepresentations: An interdisciplinary perspective*. Oxford: Oxford University Press.

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

Sterelny, K. (2003). *Thought in a hostile world: The evolution of human cognition.* Oxford: Wiley-Blackwell.

Stiller, J., & Dunbar, R. I. M. (2007). Perspective-taking and memory capacity predict social network size. *Social Networks*, *29*(1), 93–104.

Sylwester, K., Lyons, M., Buchanan, C., Nettle, D., & Roberts, G. (2012). The role of theory of mind in assessing cooperative intentions. *Personality and Individual Differences*, *52*, 113–117.

Tomasello, M. (2008). *Origins of human communication.* Cambridge, MA: MIT Press.

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

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (pp. 19–136). Oxford: Oxford University Press.

van der Wel, R. P. R. D., Sebanz, N., & Knoblich, G. (2014). Do people automatically track others' beliefs? Evidence from a continuous measure. *Cognition*, *130*(1), 128–133.

Wellman, H. M., Cross, D., & Watson, J. (2003). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*(3), 655–684.