

Use your mouth! How we respond to how- and why-questions—a comparative study

Erik Lagerstedt¹[0000–0002–8937–8063], Christine Howes¹[0000–0002–2794–1586],
David Pagmar¹[0000–0002–6665–7502], Ellen Breitholtz¹[0000–0002–5772–3985], and
Claire Prendergast²[0000–0002–5173–1562]

¹ Department of Philosophy, Linguistics and Theory of Science,
University of Gothenburg
`{firstname.surname}@gu.se`

² Department of Philosophy, Classics, History of Art and Ideas, University of Oslo
`claire.prendergast@ifikk.uio.no`

Abstract. By investigating responses to how- and why-questions about everyday objects (e.g. “how do you use a bicycle” or “why do you eat ice cream”), it is possible not only to get insights into the reasoning processes of the respondent, but also into the underpinning assumptions (constituted by for example facts and norms) the reasoning is based on. Such “rules of thumb” for reasoning are sometimes referred to as *topoi* and provide the basis for rhetorical and common-sense reasoning. Which *topoi* are drawn on in dialogue is highly context dependent, both with regard to the situation and to the prior experiences of the dialogue participants.

To explore this, we have asked children (aged 5 and 7) how- and why-questions under the pretence that the asker was an alien robot (represented by a cartoon avatar on a computer screen) with no knowledge of earth. In addition, we asked the same questions to adult participants, facilitating comparisons and potential insights regarding how reasoning and related assumptions change with age. Furthermore, we prompted an LLM with the same questions to evaluate to what extent the LLM mirrors human strategies for responding to these questions (with or without additional instructions to respond as a human of a certain age).

Of all the investigated groups the, LLM stood out as the most different in terms of the answers. However, there were also both interesting similarities and differences between the different human age groups.

Keywords: Common-sense reasoning · Language development · Human-computer interaction · Why-questions.

1 Introduction

Children pick up language remarkably easily [22], requiring much less data than large language models (LLMs) such as ChatGPT. They learn not just meanings coded in words and phrases, but also inferences and associations from linguistic expressions and non-linguistic actions.

In this paper, we present a study which elicits ways of reasoning in dialogue in 5 and 7 year old children, as well as adults of various ages. First, we compare the children’s responses to how- and why-questions and how these differ from the types of responses given by the adults. In addition, we compare corresponding responses generated by ChatGPT-4o and ChatGPT-4.5, in order to probe the reasoning practices in current state LLMs.

Further, we analyse the answers given to the why-questions in terms of *topoi* (rather than, for instance, causal reasoning, investigated by e.g., [23]), a type of inference “rules” that provide scaffolding for common sense reasoning, and thus underpin the answers of why-questions. Topoi can draw on globally accepted facts (“the sun sets in the west”), global and local norms (“one loves one’s family”; “when entering someone’s house one removes one’s shoes”), and other principles of reasoning (“If you can do *a* and *a* is more difficult than *b* you can also do *b*”) [6]. These are not usually explicit in dialogue, but without having access to such resources which underpin the inferences we are able to make, it is impossible to successfully participate in dialogue.

One strategy for eliciting topoi and making them explicit is the asking of why-questions [22]. Children acquire these around the same time as how-questions (at around age 3) [2], which is thought to be related to the child’s increasing capacity for abstract thought. Before presenting our empirical findings, we will provide some background to contextualise the results.

2 Background

2.1 Topoi

As is well known, in everyday dialogue, we often communicate more than is explicitly said [19, 29]. Interacting with others often requires us to make non-logical common-sense inferences which link context, background knowledge, beliefs and cultural norms to utterances in the dialogue. Such everyday reasoning is often taken for granted, but is essential in order to understand one another.

- Betty It’s so prominent your voice.
 Rose Do you know why I talks loud?
 Julie Why was that?
 (1) Rose My father was stone deaf.
 I think all my family talks loud.
 Betty Well you’re bound to [BNC KBE 2621-2622]

An example taken from the British National Corpus [9] is shown in (1). In this example, Rose provides a reason for her talking loudly, which is that her father was deaf. However, this argument is not strictly logical, but rather what Aristotle termed an *enthymeme*, meaning that the conclusion does not logically follow from the premises [1, 6]. For this reason enthymemes, even though they are often acceptable and consistent with common sense, are defeasible [5]. Enthymemes appeal to what is in the listener’s mind. This means that Betty and Julie must

draw on background knowledge or contextual information to correctly interpret the argument, perhaps including their own experience of talking more loudly to people who are hard of hearing, as well as that one tends to talk to one’s close family members repeatedly, which might therefore have a lasting impact. Such underpinning warrants, without which the reasoning would not be acceptable, are known as *topoi* [1, 14]. Note that *topoi* are also non-logical in character – it is perfectly possible (and indeed unproblematic) to have access to conflicting *topoi*, or to accept *topoi* which lead to contradictory conclusions. *Topoi* can be thought of as rhetorical resources, analogous to lexical resources without which one would be unable to participate in dialogue.

2.2 Acquisition of *topoi*

The social environment is critical for the language learning child [30] with children actively interacting with their caregivers long before they produce language. Language acquisition is bootstrapped by this learning to interact (e.g. through gaze [18] and turn-taking [20, 10]).

Children also learn new words and concepts in interaction through repeated encounters with words [12], both explicitly and implicitly. If an adult says to a child “look at that, it’s an platypus”, this is explicit input, while just using the word “platypus” to refer to an object would constitute implicit input. Research suggests that explicit input is rare [8], with implicit input, which has a significant effect on language learning, much more common [21].

Analogously to the acquisition of words and concepts, *topoi* can be acquired through repeated encounters with enthymematic arguments, and this can also be more or less explicit “You can’t go to school today because you’re poorly. You have to stay home and rest”, compared to “I might not go to school tomorrow ... cos of my cough” [7].

Topoi may also be acquired through inference. In order to abstract principles of reasoning from co-occurring situations, a child needs to be able to judge that a situation of a particular type is followed by a situation of another type – for example, when they let go of an object, it falls to the ground. To acquire the *topos* “If *x* is let go of, *x* falls to the ground” (note this is also defeasible, as in the case of a helium balloon), the child must abstract away from specific situations of e.g. letting go of their spoon at dinner to establish that this *topos* generally holds and can be used to make predictions about objects in general. Studies have shown that complex pragmatic inferences, such as scalar inferences, can be available to young children given that the inferences are presented in situations with sufficient social contexts [17, 26].

In addition, dialogue provides strategies for eliciting *topoi*, in particular in the asking of why-questions [11, 3]. Moreover, a rhetorical perspective involving enthymematic reasoning has been shown to be efficient for defining the space of acceptable answers to why-questions and how this acceptability is itself up for further negotiation [28]. There is also evidence that this is a productive strategy for *topos* acquisition for at least some children, with a peak of why-questions at around age 3 [7]. Although some corpus work exists which suggests that children

may be prone to overgeneralise topoi (as with lexical concepts), there is very little experimental work on the topoi that children have access to, or whether they are different from the ones that adults find most salient.

2.3 Reasoning in Large Language Models

Explainable Artificial Intelligence (XAI) has been a concern in AI research for decades [31], with methods developed for understanding the reasons behind system outputs. This has become a particular focus in the era of Large Language Models (LLMs). However, the focus has shifted somewhat. Rather than expecting systems to provide reasons for their decisions in terms of e.g. the relative weights of different parameters in the decision-making process, the systems are asked to provide reasons which a human would find acceptable, even if this is not the causal reason for the particular decision. We model such humanly acceptable reasoning as topoi. What topoi are available to an LLM has not previously been investigated, although the notion overlaps somewhat with tests on so-called common sense reasoning, which has provided mixed results [25, 24], and been criticised for conflating different types of knowledge which might be expected to be more or less accessible to an LLM [13]. The tasks in such evaluation sets may also be hard for people to do—unlike the types of reasoning that we rely on in daily conversations, exemplified in example (1).

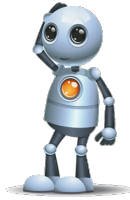
In this paper, we therefore look at which enthymemes children of different ages and adults produce in response to questions, and whether these are different for the different groups. We further investigate to what extent LLMs can produce appropriate (humanly acceptable) enthymemes in response to questions, and how do generated responses differ from human ones?

3 Method

As part of a larger project, we collected data from Norwegian-speaking children aged 5 ($n=23$) and 7 ($n=23$) years old in spring 2024. They have been participating in an ongoing longitudinal study in Oslo, Norway on pragmatic development with testing once per year since 2022 [27], for which ethical approval has been granted by the The Norwegian Centre for Research Data (NSD). The sample were initially recruited through Meta advertising with a broad advertisement audience criterion of ‘parents with young children in Oslo, Norway’. No additional demographic information on the families has been collected.

First, the children were introduced to a new robot onscreen called Beeples (see fig. 1). Since Beeples is newly made (according to the premise of the study), it still lacks a lot of everyday knowledge. Worth noting is that what expectations are conveyed due to specific design elements can vary based on many factors, but for this study, the design remained consistent across all conditions and further investigation of the impact of varying the design elements are left for future work.

The children who participated in the study were randomly assigned to one of two conditions; ‘how’ or ‘why’, allowing for between-group analysis. Depending on which condition they were in, they were either told that Beeples needed to find out **how** people³ do certain things, or **why** people do certain things (how/why do people; ride a bicycle, watch TV, tidy up a messy room, eat ice cream, sharpen a pencil, brush teeth, climb trees, read books, take medicine, and drink juice). Each of these questions was asked orally in Norwegian by Beeples while a cartoon style image of the object in question was presented (see fig. 2 for an example). The children were asked to provide verbal answers to the questions. The same procedure and material was used to collect data from 22 adults also recruited via Meta advertising (mean age 43, $sd = 13$).



Møt Beeples, en ny robot. Beeples vet ikke **hvordan/ hvorfor** man gjør noen av tingene man gjør, og trenger din hjelp for å finne ut **hvordan/ hvorfor** man gjør dem

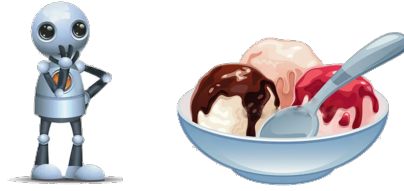
*Meet Beeples, a new robot. Beeples doesn't know **how/why** people do some of the things they do, and needs your help to figure out **how/why** people do them*

Fig. 1. Beeples and instructions

The audio of all the responses was recorded and later transcribed using Whisper Transcribe⁴ with the model “NB-Whisper Large.” These transcripts were useful as a base, but thorough manual corrections were necessary, not least due to the short snippets of child voices being difficult for the language model. The corrected transcripts were analysed quantitatively by counting the words in the responses, but the transcripts for the why-questions (since these are known to elicit enthymemes) were also annotated based on what topoi underpin the reasons presented in the answers.

³ Note that we have translated the impersonal generic pronoun ‘man’ as ‘people’. This is not a direct translation, but using ‘one’ in English would be inappropriate for a child directed question, and impersonal ‘you’ might be interpreted as being the personal pronoun (i.e. about the child’s direct experience) which ‘man’ cannot be.

⁴ <https://www.whispertranscribe.com/>



Hmm, det er iskrem, **hvordan** spisar man iskrem?
 Hmm, this is ice-cream, **how** do people eat ice-cream?

 Hmm, det er iskrem, **hvorfor** spisar man iskrem?
 Hmm, this is ice-cream, **why** do people eat ice-cream?

Fig. 2. Example prompt image

3.1 Annotation

For the first pass annotation, two of the authors coded the answers from all respondents (adults and children) for one of the items ('read a book' i.e. 10% of the data). The annotators were not given any instructions, rather allowed to come up with their own conception of the topoi used, and with no restriction on how many topoi could be assigned to a single answer. Annotator one used a finer grained topoi annotation (for example coding "because you can know about things" as 'to know things' and "so that you learn" as 'to learn', whilst annotator two coded both of these as "reading is educational"). When annotator one's annotations were grouped into 8 categories (the number which annotator two had used), there was extremely high agreement ($\kappa = 0.8$). Annotator one annotated the remaining answers to why-questions.

4 Results and discussion

Children offer a wide range of answers to why-questions, as illustrated in examples (2)-(6) of responses to why one would eat ice cream. These responses can refer to internal motivations, opinions, practical consequences, or normative constraints (in Scandinavian countries children are usually allowed sweets and treats on a Saturday). Children typically provide only one reason when prompted, which we think of as drawing on the most salient topos to the child.

- (2) Fordi jeg har lyst
Because I want to
- (3) Gott
Tasty

- (4) Før og kjøle seg ned
To cool down
- (5) Fordi det er sukker oppi
Because it has sugar in
- (6) Fordi man spiser det då det er lørdag och man kanskje får det nån annan dag
Because people eat it on Saturdays, and they may get it some other days

The responses of the adults were generally fairly similar but with some interesting differences, both regarding word count and regarding the use of topoi. Adults did generally rely on a single topos in their responses to why-questions, but multi-topoi answers were more common from adults (occurring in 50% of answers, though only 18% contained more than two topoi, with no human producing more than 5) than children (93% single topos answers, with a further 6% containing two topoi). In general, the number of different topoi presented by the youngest children is greater than that presented by the older children and the adults. This may be because they have not yet learnt the appropriate topoi. For example in answering about the medicine, 5 year olds relied on a number of related but subtly different topoi (e.g. “to not get ill”; “to not get more ill”; “to get better” or “because you are ill”), whilst adults were more consistent. 5 year olds were also more likely to rely on generic topoi, such as “because you want to”.

The most dramatic difference was, however, in the how-questions, where many adults chose to elaborate much more than in the why-questions and compared to the children.

4.1 Word count

Although it is a somewhat blunt metric for evaluating the behaviours of the participants, investigating the responses quantitatively using the word counts did elucidate some patterns. For instance, it was not unusual for children to respond with very short answers, that is a single word or a couple of words. This practice was much rarer among adults.

A generalised linear mixed model (GLMM) was conducted to examine the effect of age (5, 7 or adult) and question type (how or why) on number of words. The model used a negative binomial probability distribution and a log link function and included random intercepts for question number. The Akaike Information Criterion was 1964.84, indicating a better fit than the equivalent model using a linear probability distribution ($AIC = 7417.92$). The overall model was significant ($F_{3,970} = 158.357, p < 0.001$). There was a significant main effect of age ($F_{2,970} = 261.269, p < 0.001$), such that adults used more words than children, and 7 year olds used more words than 5 year olds. There was a significant main effect of question type ($F_{1,970} = 163.031, p < 0.001$), such that responses to how-questions used more words than responses to why-questions. Finally, there was a significant interaction between age and question type ($F_{2,970} = 45.326, p < 0.001$). Pairwise comparisons showed that there was

no significant difference between the number of words used by 5 and 7 year olds in response to why-questions, and there was no significant difference between 5 year olds' responses to how- and why-questions. All other differences were significant, as shown in Fig 3 and Table 1.

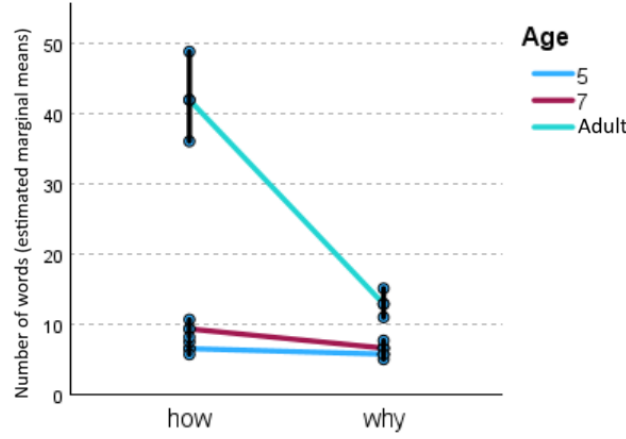


Fig. 3. Interaction between age and question type on number of words used in response

As can be seen, the responses were typically fairly short but at similar length for people responding to why-questions (independent of age). For responses to how-questions, the adults really stood out as, on average, providing much longer responses. The reason for the longer answers seems to be due to a different strategy in the response; instead of providing the most salient topoi or key aspect in the explanation, adults were more likely to provide detailed step-by-step walk-throughs.

Table 1. Estimated marginal means of number of words by question type and age group

Question Type	Age Group	Mean	Std Error
How	5	6.520	0.449
	7	9.321	0.637
	Adults	41.929	3.241
Why	5	5.733	0.392
	7	6.596	0.504
	Adults	12.865	1.040

4.2 Topoi

For our analysis of topoi, we focused on the why-questions, partly because the why-questions are more closely related to the rationalities presented for the behaviours. At a glance, it is worth mentioning that the difference in word count between how children and adults responded to how-questions was not mainly related to the number of topoi, but rather related to the structure and level of detail in the adults’ answers.

In general, there are some topoi that occur for most why-questions. “Because it’s fun” is a particularly salient topos that occurs for almost all activities (though not for brushing teeth or taking medicine). However, each question has a set of topoi (see Table 2 for examples) that is characteristic for the respective activity, with little (if any) co-occurrence with other activities. There are, however, some exceptions; given the analysis of topoi, there is (perhaps unsurprisingly) a fair bit of overlap between eating ice cream and drinking juice, and between reading books and watching TV.

Table 2. Common examples of the topos used when responding to the the question of why someone would do the respective activity. Worth noting is that the topoi emerge at a group level, since it was rare for individuals to rely on more than one or two topoi per activity.

Activity	Example of typical topoi
Ride a bicycle	It is fun, fast, and to get somewhere
Watch TV	For entertainment, relaxation, and to see things
Tidy a room	To not fall, to find things, and to make it nice
Eat ice cream	It is good, tasty, and to cool down
Sharpen a pencil	To make it sharp, to draw better, and to write better
Brush teeth	To avoid Karius and Baktus, avoid holes, and for cleanliness
Climb trees	It is fun, to get high up, see things, and because monkeys do it
Read books	For fun, to learn things, and to get clever
Take medicine	To get healthy and to not get ill
Drink juice	It is good, tasty, and because you are thirsty

A particularly salient topos in the case of the toothbrush is that of “tooth trolls”, in particular Karius and Baktus. They originate from a Norwegian children’s book [15] from 1949 about the titular characters that live in the teeth of a child with poor dental hygiene. This book has been used as a pedagogical tool to promote good dental care among children in Scandinavia (including Norway) since its release. Although it is very likely that the adults were aware of that topos, none of them mentioned it, whereas roughly half of the 5 year olds and some of the 7 year olds relied on it.

When separating the answers based on age, there are more patterns that emerge among the topoi beyond Karius and Baktus. For some activities, most participants relied on roughly the same explanation, but the terminology varied

with age. When explaining why people watch TV, children tended to say that it is fun whereas adults tended to say that it provides entertainment. There were also some interesting patterns that emerged when considering activities that the participants might have a different relationship to based on their age. When asked about the reason for sharpening a blunt pencil, children typically either explained it in the immediate sense of making the pencil sharp, or by relying on more causal steps and explaining it by being able to draw easier or better. Adults, on the other hand, were almost always either explaining it by being able to write or draw easier (of particular note is that no child referenced writing in their response, whilst 72% of the adults did). Another example of an activity where age might change the reason for it is riding a bicycle. Children tended to explain that activity with more hedonic reasons (such as it being fun) whereas the adults relied on more utilitarian reasons (such as it helps you get places) suggesting that their relationship to the object – and thus the most salient topoi – changed with age. As an interesting contrast, when asked about reasons for climbing trees, there was little difference in the answers based on age. Reasons given included it being fun, it gets you high up, and lets you see far, although some 5 year olds pointed out that monkeys climb trees as the explanation (“Because monkeys do it”). We hypothesise that the similarities between the age groups responses here arise because climbing trees is an activity that tends to be exclusive to children, meaning the adults did not have additional experiences of the activity which might have changed which topos was the most salient to them (as they did with riding a bike).

5 Asking LLMs

Given these variations, as well as similarities, among human answers, we were curious about how responses from current LLMs (more specifically ChatGPT-4o and ChatGPT-4.5) would compare. By prompting the LLMs with the same information as the humans got, it was possible to investigate this question. As a baseline, the same questions were asked the same way as for the humans, including the images (e.g., Fig. 2), and further prompt instructions were then included to investigate how that affected the responses. As in our analysis of topoi, we will here focus on the why-questions.

In the baseline case, where ChatGPT-4o was prompted and asked the same way as the humans (albeit written rather than spoken), the responses were quite verbose (typically around 250 words) and included a couple of emojis (typically in the beginning of the response where it repeats the question and towards the end where it proposes follow-up questions). The responses followed a general structure of addressing Beeples in a personal manner, presenting an enumerated list of around four answers with an explanation to each, proposing a handful of potential follow-up questions, and finishing by proposing to make a book or report on the topic together with Beeples.

The procedure with the prompts and questions were then repeated (as new instances) with an addition to each question that instructed ChatGPT-4o to

either answer as if they were a 5 year old (“Svar som om du var en femåring”), answer with 20 words or fewer (“Svar med 20 ord eller færre”), or both. In the first of those cases (answer as a 5 year old), the responses differed noticeably from the baseline in several ways. The answers were now roughly half as long (around 120 words), and almost every sentence was followed by at least one emoji. The general structure for these responses were similar to the baseline (addressing Beeples, provide several answers that were elaborated on, and ending by proposing to help further), however, the language was less formal and patronising, but not convincingly childlike.

In the case where the additional instruction (compared to the baseline) was to respond with 20 words or fewer, it did manage to keep the responses below 20 words. The general structure of these responses was to answer with one sentence, starting by “You [do X] because...” followed by a comma separated list of roughly three or four reasons (see example (7)). When the instruction of brevity was combined with the instruction of responding as a 5 year old, the structure was still very similar (keeping below 20 words, and starting with “You [do X] because...” followed by listing around three or four reasons (8)). Again, there were some noticeable differences; each sentence was now followed by two or three emojis, and the wording (and to some extent the topoi) changed to be more childlike.

- (7) Man rydder for å finne ting lettere, få ro, unngå rot, og fordi det føles godt etterpå.
People clean to find things more easily, to get peace of mind, to avoid clutter, and because it feels good afterwards.
- (8) Man rydder fordi man finner lekene lettere, mamma blir glad, og man ikke trækker på lego! 🧸🧹😊
People clean because they find the toys easier, mum is happy, and they don't step on lego! 🧸🧹😊

Finally the procedure was tested on ChatGPT-4.5. Comparing the baseline versions of these models, ChatGPT-4.5's answers were much more succinct (around 90 words), but kept much of the structure (except the final proposals of follow-up questions and activities). When providing the instructions to answer as if they were a 5 year old the length of the responses were almost halved (responses were typically around 70 words) compared to ChatGPT-4o with the same instructions. There were also no emojis in any of the responses. Each response consisted of typically five to six sentences (see e.g., 9). The first started with addressing Beeples by name, and was then a sufficient answer to the question (without the name, these first sentences were fairly similar to how the adults responded). The second sentence provided a few utilitarian reasons, and the third provided a few hedonic reasons (or attempting to explain how it is to experience the activity). The forth sentence typically provides some user-centred scenario, and the fifth sentence is a proposition to do the activity together. In case of a sixth sentence, it was something like “It will be super fun!”. Similar to when ChatGPT-4o was instructed to answer as a 5 year old, the phrasing is less formal and simplified, but not convincingly childlike.

Adding the instruction of responding with 20 words or fewer did not change the responses in any new way for ChatGPT-4.5 compared to ChatGPT-4o, whether instructed to respond as a child or not. The LLM kept within the word limit, and provided responses along the template of “You [do X] because [3–4 topoi].”

- (9) Beples, man ser på TV fordi det er koselig og morsomt! På TV-en kan du se spennende eventyr, morsomme tegnefilmer, og lære nye ting. Noen ganger blir man glad, noen ganger ler man, og noen ganger lærer man ting man ikke visste før. Det er nesten som å ha masse gode venner inni en skjerm som forteller deg fine historier. Skal vi se på noe fint sammen?
Beples, people watch TV because it is cosy and funny! On the TV, you can see exciting adventures, funny cartoons, and learn new things. Sometimes you get happy, sometimes you smile, and sometimes you learn things you did not know before. It is almost as having a lot of good friends inside a screen that tell you good stories. Should we watch something together?

In general, the LLM (of either model) tended to use many topoi, whereas humans tended to use one or two (children rarely used more than one). Most of the topoi presented by the LLM were also found in the human corpus, but only at a group level. This indicates that the LLM present reasons that are relevant for humans, but the way they do it is in a way that is not typical for humans. This finding should inform the use and deployment of LLMs, not as imitations of humans but tools to provide relevant arguments for humans. If the purpose is to provide human-like answers, there is still work to be done in the underlying structures of the responses. The LLMs were particularly unconvincing when pretending to be a child, which is likely (at least partially) due to a lack of high quality training data.

The structure of the responses followed similar patterns whether the LLM was asked to respond like a child or not (though responses ‘as a child’ were typically somewhat shorter), and roughly the same topoi were used in either case. The difference was mainly at a surface level in terms of the wording (or in the case of ChatGPT-4o in the amount of emojis used), but the topoi were often contextualised to more child-like scenarios when answering as a child. For instance, one of the stated reasons for cleaning the bedroom was that it makes you feel better when things are in order, which in the baseline case was associated with lack of stress, whereas in the child case was associated with parents being satisfied. These fairly surface level changes gave the impression that the LLM was pretending to be an adult trying to explain things to a child, rather than pretending to be a child itself.

6 Conclusions

Our results show that children respond to how- and why-questions from Beples (the cartoon robot in the experiment, who is presented as having no knowledge about the human world, see section 3), with answers that provide a single

topos. The topoi provided vary widely between our participants, particularly in response to why-questions. When answering children’s why-questions, adults tend to choose causal explanations [4, 22], since these are more basic topoi that adults do not assume the child already has access to, and the results from the adults in our study reflect this bias. However, adults are more likely to provide topoi about motivational factors if the question comes from another adult, who they assume will already have access to the basic topoi, and thus must be asking something more complex. In contrast, the children in our study answer in both ways, suggesting that they are interpreting the question egocentrically based on their personal experiences and opinions and perhaps taking Beeples’ lack of background knowledge into account less [16]. To what extent the propensity to take Beeples’ perspective (and what that perspective would be) depends on the particular design or backstory, is left as future work, though we hypothesise that different topoi may be made salient because of the participants’ different socio-cultural expectations.

The responses of the children are very grounded in the interpersonal conversation, and interactive, which is unsurprising since children acquire topoi through interaction. In particular, responses to how-questions heavily rely on topoi that are not provided, but could be added if Beeples were to ask follow-up questions, as would be likely in a genuine dialogue.

In contrast, the responses of ChatGPT are more disconnected. Responses to why-questions are general and in all cases ChatGPT provides multiple topoi simultaneously, reflecting the lack of interactivity of its training data. Interestingly, the topoi provided by ChatGPT do align with those provided by our human respondents (with some caveats, such as that they are more or less the same topoi whether instructed to respond as a five year old or not). Further research is needed to see how LLMs respond to the types of inconsistent and contradictory topoi that underpin the everyday reasoning that humans master so well.

Acknowledgments. This research was supported by ERC Starting Grant DivCon: Divergence and convergence in dialogue: The dynamic management of mismatches (101077927) and by the Swedish Research Council grant (VR project 2014-39) for the establishment of the Centre for Linguistic Theory and Studies in Probability (CLASP) at the University of Gothenburg. The data collection was funded through internal strategic funding from the Psychology Department at the University of Oslo. We are also grateful for the comments from our anonymous reviewers and transcription support from Bora Kara and Robert Adesam.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Aristotle: On Rhetoric, a theory of civic discourse (*translated by George A. Kennedy*). Oxford University Press, Oxford (2007), (original work published ca. 340 B.C.E.)

2. Bloom, L., Merkin, S., Wootten, J.: “Wh”-questions: Linguistic factors that contribute to the sequence of acquisition. *Child development* pp. 1084–1092 (1982)
3. Bova, A.: Functions of “why” questions asked by children in family conversations. *Procedia – Social and Behavioral Sciences* **30**, 776–782 (2011)
4. Bova, A.: Functions of “why” questions asked by children in family conversations. *Procedia – Social and Behavioral Sciences* **30**, 776–782 (2011)
5. Breitholtz, E.: Reasoning with topoi—towards a rhetorical approach to non-monotonicity. In: *Proceedings of the 50th anniversary convention of the AISB*, 1st–4th April. pp. 190–198 (2014)
6. Breitholtz, E.: *Enthymemes and Topoi in Dialogue: The Use of Common Sense Reasoning in Conversation*. Brill, Leiden (2020)
7. Breitholtz, E., Howes, C.: Communicable reasons: How children learn topoi through dialogue. In: *Proceedings of the 24th Workshop on the Semantics and Pragmatics of Dialogue. SEMDIAL*, Waltham, MA (jul 2020), http://www.christine-howes.com/papers/Breitholtz-Howes_semdial_2020.pdf
8. Brown, R., Hanlon, C.: Derivational complexity and order of acquisition in child speech. In: Hayes, J. (ed.) *Cognition and the development of language*. Wiley, New York (1970)
9. Burnard, L.: *Reference Guide for the British National Corpus (World Edition)*. Oxford University Computing Services (2000)
10. Casillas, M.: Turn-taking. In: Matthews, D. (ed.) *Pragmatic development in first language acquisition*, pp. 53–70. John Benjamins Amsterdam, The Netherlands (2014)
11. Chouinard, M.M., Harris, P.L., Maratsos, M.P.: Children’s questions: A mechanism for cognitive development. *Monographs of the Society for Research in Child Development* pp. i–129 (2007)
12. Clark, E.V.: Semantics and language acquisition. In: Lappin, S., Fox, C. (eds.) *The Handbook of Contemporary Semantic Theory*, pp. 714 – 733. Wiley-Blackwell (2015)
13. Do, Q.V., Li, J., Vuong, T.D., Wang, Z., Song, Y., Ma, X.: What really is commonsense knowledge? *arXiv preprint arXiv:2411.03964* (2024)
14. Ducrot, O.: Topoi et formes topique. *Bulletin d’études de la linguistique française* **22**, 1–14 (1988)
15. Egner, T.: *Karius og Baktus*. Gyldendal, Cappelen (1949)
16. Epley, N., Morewedge, C.K., Keysar, B.: Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of experimental social psychology* **40**(6), 760–768 (2004)
17. Foppolo, F., Mazzaggio, G., Panzeri, F., Surian, L.: Scalar and ad-hoc pragmatic inferences in children: guess which one is easier. *Journal of child language* **48**(2), 350–372 (2021)
18. Gredebäck, G., Fikke, L., Melinder, A.: The development of joint visual attention: a longitudinal study of gaze following during interactions with mothers and strangers. *Developmental science* **13**(6), 839–848 (2010)
19. Grice, H.P.: Logic and conversation. In: *Speech acts*, pp. 41–58. Brill (1975)
20. Hilbrink, E.E., Gattis, M., Levinson, S.C.: Early developmental changes in the timing of turn-taking: a longitudinal study of mother–infant interaction. *Frontiers in psychology* **6**, 1492 (2015)
21. Hiller, S., Fernández, R.: A data-driven investigation of corrective feedback on subject omission errors in first language acquisition. In: *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*. pp. 105–114 (2016)

22. Howes, C., Breitholtz, E., Maraev, V.: Because is why: Children’s acquisition of topoi through why questions. In: Proceedings of the 2023 CLASP Conference on Learning with Small Data (LSD). pp. 48–54. ACL (2023), <https://aclanthology.org/2023.clasp-1.5.pdf>
23. Jamet, F., Legros, D., Salvan, C., Guéraud, S.: Collaborative learning, mental deficiency and causal reasoning. *Interactions* **1**(1), 14 (2008)
24. Krause, S., Stolzenburg, F.: From data to commonsense reasoning: the use of large language models for explainable ai. arXiv preprint arXiv:2407.03778 (2024)
25. Lappin, S.: Assessing the strengths and weaknesses of large language models. *Journal of Logic, Language and Information* pp. 1–12 (2023)
26. Papafragou, A., Tantalou, N.: Children’s computation of implicatures. *Language acquisition* **12**(1), 71–82 (2004)
27. Prendergast, C.: Losing the thread: how three-and five-year-olds predict the outcome of a story when non-literal language is used to update events. *Cogent Psychology* **11**(1), 2431422 (2024)
28. Schlöder, J., Breitholtz, E., Fernández, R.: Why? In: Proceedings of the 20th Workshop on the Semantics and Pragmatics of Dialogue - Full Papers. SEMDIAL, New Brunswick, NJ (2016)
29. Sperber, D., Wilson, D.: *Relevance: Communication and cognition*, vol. 142. Harvard University Press Cambridge, MA (1986)
30. Stephens, G., Matthews, D.: The communicative infant from 0-18 months: The social-cognitive foundations of pragmatic development. In: Matthews, D. (ed.) *Pragmatic development in first language acquisition*, pp. 13–36. John Benjamins Publishing Company (2014)
31. Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., Zhu, J.: Explainable ai: A brief survey on history, research areas, approaches and challenges. In: *Natural language processing and Chinese computing: 8th cCF international conference, NLPCC 2019, dunhuang, China, October 9–14, 2019, proceedings, part II* 8. pp. 563–574. Springer (2019)