Language Acquisition

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hlac20

Cognitive inhibition explains children's production of medial *wh*-phrases

Adam Liter, Elaine Grolla & Jeffrey Lidz

To cite this article: Adam Liter, Elaine Grolla & Jeffrey Lidz (2022): Cognitive inhibition explains children's production of medial *wh*-phrases, Language Acquisition, DOI: 10.1080/10489223.2021.2023813

To link to this article: https://doi.org/10.1080/10489223.2021.2023813

Published online: 11 May 2022.

LANGUAGE ACQUISITION
https://doi.org/10.1080/10489223.2021.2023813

Routledge
Taylor & Francis Group

Check for updates

Cognitive inhibition explains children's production of medial *wh*-phrases

Adam Liter[a], Elaine Grolla[b], and Jeffrey Lidz[a]

[a]University of Maryland at College Park; [b]University of São Paulo

ABSTRACT

Non-adult-like linguistic behavior in children is sometimes taken as evidence for endogenous factors that drive selection of grammatical features from the child's hypothesis space of possible grammars. Analyses of English-acquiring children's productions of medial *wh*-phrases exemplify this trend in particular. We provide an alternative account of these productions as performance errors arising from under-developed cognitive inhibition. We offer experimental evidence in favor of our failure of inhibition account. The results argue against treating these errors as reflecting incomplete or non-target acquisition of grammatical features. Instead, the results support a theory of how these errors arise and are subsequently purged from children's productions that reduces to a theory of how cognitive inhibition develops during childhood.

ARTICLE HISTORY
Received 13 November 2020
Accepted 23 November 2021

KEYWORDS
Syntax; movement phenomena; *wh*-movement; executive function; inhibition control

1. Introduction

Children often exhibit non-adult-like linguistic behavior. When children do behave differently from adults in some linguistic domain, there are generally three types of explanations that are offered to account for such differences. One type of explanation accounts for their non-adult-like linguistic behavior as the result of an immature knowledge state, with representations that are fundamentally different in kind from those of adults (see, among others, Roeper & Matthei 1975, Rizzi 1993, Tomasello 2000). A second type of explanation accounts for children's non-adult-like linguistic behavior as the result of an interaction of target grammatical knowledge and their extragrammatical performance systems (see, among others, Hamburger & Crain 1982, Bloom 1990, Mazuka, Nobuyuki & Oishi 2009). A third type of explanation accounts for children's non-adult-like linguistic behavior as the result of grammatical knowledge that traffics in the same sorts of representations that adults have, but this knowledge happens to be appropriate for some other language that is not the target language of acquisition; the behavior reflects non-target grammatical knowledge that is possible in the world's languages (see, among others, Hyams 1983, 1986; Pierce 1992, Guasti, Thornton & Wexler 1995).

This third kind of analysis is potentially highly informative, as it entails a learning system in which endogenous factors drive the learner's selection of grammatical features. Because of these implications, it is important to hold such accounts to careful scrutiny.

CONTACT Adam Liter io@adamliter.org University of Maryland at College Park, Linguistics, 1401 Marie Mount Hall, 7814 Regents Dr., College Park, MD 20742-7505.

One highly prominent case where such an analysis has been offered is in the case of children's production of medial *wh*-phrases. For example, English-acquiring children are known to sometimes produce questions with medial *wh*-phrases in elicited production tasks (e.g., Thornton 1990).[1] Examples of such questions are given in (1).[2]

(1) a. [$_{CP}$ Who do you think [$_{CP}$ **who** Grover wants to hug?]] (Tiffany 4;09) (Thornton 1990:87)
b. [$_{CP}$ What do you think [$_{CP}$ **who** ate this?]] (Kelly 3;11) (Thornton 1990:232)

This finding has been noteworthy for at least two reasons. First, questions such as (1) are not adult-like in the target language of acquisition; that is, they are ungrammatical in adult English. Second, such productions look very much like structures that are grammatical in other adult languages.

For example, some German dialects (among other languages) exemplify *wh*-copying, where the same *wh*-phrase occurs at the edge of the matrix clause and at the edge of the embedded clause, such as in (2).

(2) Wen glaubst du [$_{CP}$ wen sie getroffen hat]?
who think you who she met has
'Who do you think she met?'
(Felser 2004:544)

This grammatical (in adult German) question looks remarkably similar to the non-adult-like question of the English-acquiring child in (1a).

Similarly, languages such as Bangla (Bayer 1996), Romani and German (McDaniel 1986, 1989), Afrikaans (du Plessis 1977), Iraqi Arabic (Wahba 1991), and Hungarian (Horvath 1997) exhibit scope marking (sometimes also called partial *wh*-movement), where the *wh*-phrase in the embedded clause takes matrix scope, and the *wh*-phrase in the matrix clause just serves to mark the position from which the intermediate *wh*-phrase takes scope. An example from Bangla is given in (3).

(3) Tumi ki bhebe-cho [$_{CP}$ ke baRi kore-che]?
you what think who house make
'Who do you think has built a house?'
(Bayer 1996:293)

Again, this grammatical (in adult Bangla) question looks remarkably similar to the non-adult-like question of the English-acquiring child in (1b).

Given that these structures are possible in the world's languages (i.e., given that they are allowed by Universal Grammar) and given the similarities between these structures and the utterances of some English-acquiring children, such as those in (1), some researchers have proposed that English-acquiring children have temporarily adopted a non-target (but adult-like) grammar (e.g., Thornton 1990, Thornton & Crain 1994, McDaniel, Chiu & Maxfield 1995, Roeper & de Villiers 2011, de Villiers, de Villiers & Roeper 2011). We will call such accounts non-target grammar accounts. Again, such accounts make strong commitments to certain types of endogenous factors that guide language acquisition, and so they should be critically evaluated against alternative accounts.

[1]Similar findings are reported for Dutch learners (van Kampen 1997, Jakubowicz & Strik 2008), Spanish learners (Gutiérrez Mangado 2006), and French learners (Oiry 2006, Demirdache & Oiry 2008, Jakubowicz & Strik 2008). In this article, we focus on English-acquiring children, but we expect that our account of English-acquiring children's production of medial *wh*-phrases would apply to children acquiring other languages where these questions are also ungrammatical in the target language of acquisition.

[2]Thornton (1990:ii) described only the productions like (1a), with the same *wh*-word in both positions, as "medial-Wh questions"; she described productions like (1b) as partial movement questions. Throughout this article, we refer to both types of questions, at least as produced by English-acquiring children, as medial *wh*-questions. While this conflicts with the terminology originally put forward by Thornton (1990), it more accurately reflects our account of these non-target productions, which holds that they arise via the same mechanism.

Indeed, we contrast such non-target grammar accounts of these medial *wh*-phrase productions with an account that exemplifies the second kind of analysis mentioned at the outset—namely, we offer an analysis where the medial *wh*-phrase productions of English-acquiring children arise as an interaction of target grammatical knowledge and an immature extragrammatical performance system. The idea of our account, which we originally proposed in Grolla & Lidz (2018), is that English-acquiring children produce sentences like those in (1) because they may have trouble inhibiting the pronunciation of the copy of the *wh*-phrase at the edge of the embedded clause. Because *wh*-movement is successive cyclic (see, among others, Chomsky 1973, Torrego 1983), there is a copy of the *wh*-phrase in that position, which the child may accidentally pronounce, especially if they have underdeveloped inhibition control, despite having a grammar that forbids its pronunciation.[3] This account will be discussed in more detail below.

In this article, we report experimental evidence that favors the failure of inhibition account over the non-target grammar accounts. Non-target grammar accounts have taken children's non-target productions of medial *wh*-phrases as evidence of two things: (i) that there is an innately specified and restricted hypothesis space, and (ii) that the learning mechanism the child employs to move around in this hypothesis space can be driven by endogenous factors that are not present in the input (since there is no evidence for either *wh*-copying or scope marking questions in adult English). Although both of these points very well may be true (for the first point, see, for example Berwick et al. 2011, for the second point, see, for example Han, Musolino & Lidz 2016), our results suggest that English-acquiring children's medial wh-questions do not constitute evidence in favor of either of these points. The article proceeds as follows. In § 2, we discuss our failure of inhibition account and non-target grammar accounts in more detail. In § 3, we discuss the experiment that we conducted to tease apart non-target grammar accounts and our inhibition account. In § 4, we conclude.

2. Different developmental accounts

In this section, we contrast our own failure of inhibition account (§2.1) with the various non-target grammar accounts that have been proposed in the literature (§2.2).

2.1. *A failure of inhibition account*

We propose a failure of inhibition account of English-acquiring child productions like those in (1). Inhibition is one of a variety of different executive functions. Executive functions are functions that allow an individual to guide and direct their thoughts and actions (e.g., Lezak 1995:42); these functions are thought to be associated with the prefrontal cortex (cf. Miller & Cohen 2001), which undergoes development well into adolescence (cf. Huttenlocher & Dabholkar 1997).

A variety of different executive functions have been proposed in the literature including, for example, shifting (the executive function that coordinates moving from one task to another or one mental set to another), updating (the executive function that coordinates the updating and monitoring of working memory), and inhibition (the executive function that suppresses responses that may be prepotent but inappropriate) (cf. Miyake et al. 2000:54–58). An area of active research in cognitive psychology investigates the extent to which different proposed executive functions truly constitute separate cognitive abilities; for example, Miyake et al. (2000) provide evidence that the three aforementioned executive functions—shifting, updating, and inhibition—are all in fact distinct executive functions. Likewise, Harnishfeger (1995) argues that inhibition should be further distinguished between motor response inhibition and cognitive inhibition (see also Dempster 1991, 1993).

[3]For the sake of expository convenience, we will talk in terms of copies of *wh*-phrases throughout this article. However, our account does not depend on the copy theory of movement being true. In fact, all that matters for our purposes is that some form of the *wh*-phrase is reactivated in the mind of the producer at places in the structure where copies are posited to be, due to the successive cyclic nature of *wh*-movement. There is a good amount of psycholinguistic evidence that supports this view, which we will discuss in § 2.1. As long as one believes the psycholinguistic findings, our account goes through, and it doesn't particularly matter which syntactic theory underpins these psycholinguistic facts.

Importantly, executive functions are relevant to the real-time deployment of grammatical knowledge (for parsing/comprehension, see, e.g., Engelhardt, Nigg & Ferreira 2013, Novick et al. 2014, Hsu & Novick 2016, Thothathiri et al. 2018 and for production/generation, see, e.g., Shao, Meyer & Roelofs 2013, Freund, Gordon & Nozari 2016, Trude & Nozari 2018, Nozari & Novick 2017, Nozari & Omaki 2018). For example, Novick et al. (2014) show that cognitive control training that enhances conflict resolution, which plausibly involves inhibition control, improves adults' ability to comprehend garden path sentences. Similarly, Hsu & Novick (2016) show that having good cognitive control (as measured by successfully detecting a conflict in an incongruent Stroop trial) leads to more quickly discarding the incorrect parse when processing a sentence that is temporarily ambiguous. And on the production side of things, Nozari & Omaki (2018) show (i) that individuals with better inhibitory control produce fewer agreement attraction errors (e.g., producing *are* in a sentence like *The snake next to the elephants are green*; cf. Bock & Miller 1991) and (ii) that taxing inhibitory control resources on a trial-by-trial basis leads to more agreement attraction errors. With children specifically, both Woodard, Gleitman & Trueswell (2016) and Qi et al. (2020) report evidence suggesting that children's ability to recover from an early but incorrect parse is predicted by their executive functioning (although see Huang & Hollister (2019) for an alternative interpretation of the data).

Equally importantly, executive functions are known to not reach maturity until sometime in late adolescence (see, among others, Anderson et al. 2001, Anderson 2002, Diamond 2002, Luciana et al. 2005, Davidson et al. 2006). For example, Davidson et al. (2006) showed that children's performance on a variety of tasks that measure different executive functions increases from ages 4 to 13. Children's performance on these tasks at an early age is often quite poor, indicating that the executive function being probed is still developing. This is consistent with brain development in the region that is thought to underlie executive functions, the prefrontal cortex, still undergoing development into mid-to-late adolescence (see, e.g., Huttenlocher & Dabholkar 1997:170). Nonetheless, development of the various executive functions is not monolithic. For example, Dempster (1993) suggests that there may be an ordering to the development of different types of inhibition, with the development of motor response inhibition preceding the development of cognitive inhibition.

Now, the developmental account of English-acquiring children's productions of medial *wh*-phrases that we propose assumes that children have acquired the target grammar; that is to say, children know that the *wh*-phrase can only be pronounced once at the left edge of the matrix clause in a multiclausal *wh*-question. However, because children's executive function of cognitive inhibition is not fully developed, children may fail to inhibit the pronunciation of the copy of the *wh*-phrase that is at the edge of the embedded clause, leading to a grammatically illicit medial *wh*-phrase production. Such productions are expected when cognitive inhibition is poor if we adopt the model of sentence production in Dell (1986) and if we adopt the hypothesized explanations of some interesting psycholinguistic facts about how fillers (e.g., *wh*-phrases) are related to their gap sites.

Following Dell (1986:289–290), we assume that, in the course of sentence production, elements with high activation are what get pronounced and that speech errors arise from incorrect items having higher activation than the correct item. Dell (1986:286–287) posits that activation is a real-valued property. Furthermore, whenever an item has an activation greater than 0, it spreads some of its activation to nearby items in the lexicon. Dell additionally assumes that activation decays exponentially over time towards 0. Dell's model of spreading and decaying activation, coupled with possible small random fluctuations in activation as well as the potential for differing initial background activations of items due to what the speaker is thinking about, what the discourse makes salient, etc., provides an account of a variety of different types of speech errors, including substitution errors, such as *pass the pepper* when what was really meant was *pass the salt*, and perseveration errors, such as *class will be about discussing the class* when what was really meant was *class will be about discussing the test* (Dell 1986:285).

In effect, we take English-acquiring children's productions of medial *wh*-phrases to be speech errors. To see how this is spelled out, we must consider the mechanisms involved in relating fillers to gaps. First, there is a variety of empirical evidence that fillers are actively maintained in memory until they are associated with their gap site. There is evidence to this effect for both comprehension and production. One piece of evidence in favor of this theory comes from Gibson & Warren (2004). They show that reading times of the predicate from which a *wh*-phrase receives its theta role are faster when there is an intermediate copy of the *wh*-phrase in between the *wh*-phrase and the gap site than when there is not an intermediate copy of the *wh*-phrase, with the length of the sentences held constant. They explain this by assuming that the *wh*-phrase is maintained in memory after being read and is reactivated at the position with the copy, leading to higher activation and thus faster reading times than if there were no intermediate position with a copy.

Similarly, there are also a variety of complexity effects in production that support the idea that *wh*-phrases are maintained in memory until integration with the predicate that they either receive their theta role from or modify. For example, Scontras et al. (2015) and Scontras, Badecker & Fedorenko (2017) show that shorter subject dependencies are produced quicker than longer object dependencies. An explanation for this finding is that, in the longer dependency, the memory representation of the filler has decayed more, and so it may take longer to reactivate the filler when the gap position is reached, leading to overall production times that are slower when the dependency spans a greater distance.[4] A similar finding is that passives are often produced inside of relative clauses, which is hypothesized to result from attempts to minimize the dependency length by making the head of the relative clause the subject of the relative clause, thereby reducing the overall cost of the production (see, e.g., Fadlon et al. 2019). For example, people will prefer to produce (4b) over (4a).[5]

(4) a. The day worker that the farmer praised
 b. The day worker that was praised by the farmer
 (Fadlon et al. 2019:46)

A second relevant fact about the psycholinguistic mechanisms involved in filler-gap constructions is that it is perhaps only syntactic category information and an abstract *wh*-feature that are actively maintained in memory. Wagers & Phillips (2014) show this to be the case at least in the comprehension of long-distance dependencies. They conduct a reading-time study and find that participants are slower at reading the region around the gap site if the lexical content of the *wh*-phrase is semantically odd with respect to the predicate at the gap site, but only when the dependency is a short dependency. In other words, if the dependency is long (e.g., spans a clause), the lexical content of the *wh*-phrase will not be actively maintained in memory; only the syntactic category and a *wh*-feature are maintained in memory, and so no slowdown due to semantic anomaly results when the reader reaches the gap site. Importantly, however, there is a slowdown that occurs after the gap site, suggesting that this lexical content is retrieved from memory, and so the slowdown effect from the semantic oddity is seen after this retrieval process takes place.

If we adopt the working assumption that production uses the same mechanisms as comprehension (cf. Momma & Phillips 2018), then a producer will also only maintain a *wh*-feature in memory after uttering the *wh*-phrase until associating it with a gap site that is far away (in another clause). We now have the ingredients needed to state our account precisely in terms of independently motivated performance mechanisms. We describe how our failure of inhibition account provides an explanation for the utterances in (1), repeated here as (5) for convenience, even though, by hypothesis, the speaker has acquired the target, adult-like English grammar.

(5) a. [$_{CP}$ Who do you think [$_{CP}$ **who** Grover wants to hug?]] (Tiffany 4;09)
 (Thornton 1990:87)
 b. [$_{CP}$ What do you think [$_{CP}$ **who** ate this?]] (Kelly 3;11)
 (Thornton 1990:232)

[4]Scontras et al. (2015) also show that it takes longer to begin a sentence with a longer dependency, suggesting that there is a planning cost as well.

[5]For additional psycholinguistic evidence that the production of linguistic dependencies, not necessarily just A-bar- (i.e., *wh*-) dependencies, involves active maintenance of the relevant element in question in memory, see also Badecker & Kuminiak (2007) and Franck et al. (2010).

For (5a), the speaker starts by uttering *who*. Because this is a long-distance dependency, only a *wh*-feature and the syntactic category information is maintained in memory (Wagers & Phillips 2014). However, when the speaker reaches the beginning of the embedded clause, the *wh*-feature receives a boost in activation (Gibson & Warren 2004). Even though the child's grammar forbids the pronunciation of the element in this position, the child may nonetheless accidentally utter the *wh*-phrase, *who*, because sentence production involves producing those things that are highly active (Dell 1986), and the *wh*-phrase just received a boost in activation by way of reactivation of the *wh*-feature. The child may be particularly likely to make this sort of speech error if their cognitive inhibition is poor. Recall that cognitive inhibition is the executive function that suppresses responses that may be prepotent but inappropriate (see, e.g., Dempster 1991, 1993). That is to say, the role that cognitive inhibition plays in our account is in suppressing the pronunciation of a reactivated *wh*-phrase in accordance with the grammar; however, if an individual's cognitive inhibition is underdeveloped, they might fail to inhibit the pronunciation, effectively resulting in a speech error, since what they said does not accord with their grammar.[6]

As for the production in (5b), the story is very similar. Recall that Dell (1986) allows for differing initial background activations of lexical items. Initial activation levels could be influenced by a variety of things, including things like what the speaker happens to be thinking about, what the discourse makes salient, what another speaker just uttered, etc. Indeed, as discussed previously, the model in Dell (1986) accounts for speech errors involving substitutions, such as substituting *pepper* for *salt*, perhaps because of different initial levels of activation and/or spreading activation. So, even though the speaker intends to ask a *who* question, *what* may, in this particular instance, initially have a higher level of activation than *who* for whatever reason, and so the speaker may accidentally utter *what*. Sentence production proceeds, and the *wh*-feature is then maintained in memory and reactivated at the edge of the embedded clause because of the copy of the *wh*-phrase in this position. Activation spreads, including to conceptually related words (Dell 1986:287–289), and so the activation can spread to all *wh*-words. Moreover, activation decays over time as well (Dell 1986:287), so the activation of *what* might have decayed towards 0, and *who* is now more highly active than *what*, especially since the speaker intended to ask a *who* question anyway. Finally, again, even though the child's grammar disallows the pronunciation of *who* in this position, they may accidentally pronounce it because it has a high level of activation, especially if their cognitive inhibition is poor.[7]

A failure of inhibition account therefore straightforwardly predicts that a measure of cognitive inhibition should correlate with productions of medial *wh*-phrases, with children who exhibit poor cognitive inhibition being more likely to produce medial *wh*-phrases than children who exhibit good

[6]A reviewer asks whether our account also predicts that *wh*-phrases should be pronounced in other locations along the movement path. We return to this in more detail in § 3.5.3. However, in brief, such productions are, in principle, expected given our account. That being said, we expect them to be extremely rare. Even when cognitive inhibition fails, we would expect the failure to still be constrained by aspects of the grammar, including the statistical distribution of the utterances that the grammar gives rise to. Specifically, *wh*-phrases are only ever pronounced in base positions in very highly constrained contexts in English (echo questions and multiple questions), while, on the other hand, *wh*-phrases occur more frequently at the clause boundary in English, including in embedded interrogatives (*I wonder what time it is*) and embedded polar interrogatives (*do you know what time it is*). So, because the child would only hear *wh*-phrases in base positions on an extremely rare basis, we would expect that children would exhibit many fewer instances of failed cognitive inhibition in such cases, compared to at the clausal boundary. If anything, children might instead be more likely to produce the corresponding DP that gives the answer in the base position, especially since they know the answer given the setup of the task and since it is licit (and common) to pronounce DPs in the complement to V position in the adult target language. This is similar to an aspect of Dell's (1986:291–292) theory which holds that speech errors will follow the categorical constraints of the grammar.

[7]Lutken, Legendre & Omaki (2020) propose an account of medial *wh*-phrase productions that is similar to our account. We will return to this in § 3.5.3. However, it is worth noting that they claim our account, as proposed in Grolla & Lidz (2018), cannot account for medial *wh*-phrase productions where the *wh*-phrases are distinct (Lutken, Legendre & Omaki 2020:43, note 3). As we've just discussed, this is not the case. Our account does offer an explanation of medial *wh*-phrase productions with distinct *wh*-phrases. And, in fact, they themselves adopt two distinct accounts of the two different types of productions, with the productions involving distinct *wh*-phrases hypothesized to be restart errors (Lutken, Legendre & Omaki 2020:38–39).

cognitive inhibition. On the other hand, non-target grammar accounts, which we briefly discuss next, make no such prediction; in fact, they most straightforwardly predict that there should be no relation between measures of cognitive inhibition and medial *wh*-phrases production.

2.2. Non-target grammar accounts

As noted, the observation that productions like (5) look like grammatical sentences in adult grammars of other languages is a common departure point for the various non-target grammar accounts (see, e.g., Thornton 1990:209 ff., McDaniel, Chiu & Maxfield 1995:710 ff., de Villiers, de Villiers & Roeper 2011:353–354). Despite this, the various non-target grammar accounts of English-acquiring children's productions of questions like (5) that have been proposed all differ somewhat. In our discussion here, we will abstract away from these details and instead raise some questions about the evidence that has been cited in favor of these non-target grammar accounts and raise some general issues for this class of similar accounts.

The details of the various accounts are largely immaterial to the experiment that we conduct, which is aimed at teasing apart the non-target grammar accounts from our own failure of inhibition account. As noted previously, the non-target grammar accounts most straightforwardly predict that there should be no relation between a measure of cognitive inhibition and production of medial *wh*-phrases. These accounts most straightforwardly make this prediction because there is no apparent reason that someone with better cognitive inhibition should have arrived at the target grammar sooner than someone with worse cognitive inhibition, since cognitive inhibition is just the ability to suppress a prepotent but inappropriate response. In other words, under the non-target grammar accounts, one would expect children with differing cognitive inhibition abilities in both groups, those with the target grammar and those with the non-target grammar.[8] Since our experiment is designed to look at the relation between cognitive inhibition and medial *wh*-phrase production, we abstract away from the details of the various non-target grammar accounts. We will instead first discuss some of the evidence that has been argued to favor these accounts, and then we will discuss some issues for these accounts.

2.2.1. Evidence for non-target grammar accounts and its issues

There are generally three pieces of putatively converging evidence that are adduced in favor of the hypothesis that English-acquiring children who produce sentences like (5) have a non-target grammar. The first piece of evidence is just the fact that children between the ages of 3 and 6 produce such sentences and that these sentences look remarkably similar to grammatical adult-like utterances in other non-target languages, as originally reported in Thornton (1990). A second piece of evidence comes from children's answers to certain kinds of questions. For example, de Villiers, Roeper & Vainikka (1990) found that younger children in their sample of 3- to 6-year-olds would respond to questions such as (6) with an answer to the *who* question (which is an ungrammatical interpretation of this question in the target language), not an answer to the *how* question.

(6) How did Big Bird say who to paint?
(de Villiers, Roeper & Vainikka 1990:282)

One possible analysis of this is that children treat sentences like (6) as scope marking constructions, with *how* just marking the matrix scope of *who* (see also Thornton & Crain 1994, Thornton 1995, de Villiers & Roeper 1995, de Villiers et al. 2008, de Villiers, Kotfila & Klein 2019). Finally, a third piece of

[8]It is of course possible that executive functions impact the speed of language acquisition, but, absent a detailed and testable proposal to this effect, non-target grammar accounts predict there to be no relation between cognitive inhibition and medial *wh*-phrase production. Moreover, given the issues for non-target grammar accounts that we discuss immediately following as well as converging evidence from other experiments like the ones reported in Lutken, Legendre & Omaki (2020), we think that non-target grammar accounts, even if they were supplemented with a developmental proposal in the vein suggested here, are likely wrong.

evidence comes from the fact that some children between the ages of 3 and 5 report questions with medial *wh*-phrases as being acceptable, despite being ungrammatical in their target language of acquisition (McDaniel, Chiu & Maxfield 1995).

With regard to the second piece of evidence, it is likely that these sorts of interpretations are driven by extragrammatical factors and do not constitute evidence that English-acquiring children have a non-target grammar. Lutken, Legendre & Omaki (2020) make a convincing empirical case for this interpretation of these findings. They conducted similar question-after-story tasks, with sentences involving *what* and *that* and sentences involving *how* and *what*, as in (7a) and (7b).

```
(7)  a.  What did Evil Steve tell Detective Sherry that he was gonna steal?
     b.  How did Evil Steve tell Detective Sherry what he was gonna steal?
```

These sentences were asked after a story in which Evil Steve told Detective Sherry that he was going to steal the Queen's ring, not the Queen's crown, in order to try to throw Detective Sherry off his trail. The story is constructed in such a way so that the manner of telling is at issue, and the story is also constructed such that what Evil Steve actually ends up stealing (the crown) is different from what he tells Detective Sherry that he is going to steal (the ring). Lutken, Legendre & Omaki (2020:26–28) find that children correctly answer (7b) with the manner of telling 67.7% of the time. Moreover, true object answers (the crown) and false object answers (the ring) only occur 8.3% of the time each in response to (7b). Only the false object answers would constitute evidence in favor of the idea that English-acquiring children have adopted a non-target grammar because it is only this kind of an interpretation where the *wh*-phrase also takes scope over *tell*. Notably, these sorts of answers were less frequent than non-adult-like true object answers (the crown) to (7a), which occurred 16.7% of the time; thus, it is likely that these false object answers (the ring) to (7b) are just noise in the data. In other words, these sorts of answers should be understood as a fact about performance, not a fact about grammatical competence. And to the extent that these sorts of answers were more prevalent in previous studies (some of which were investigating other issues), it is likely that they were pragmatic artifacts (cf. Hamburger & Crain 1982, Conroy et al. 2009).

Notably, Lutken, Legendre & Omaki (2020:30–35) also find no correlation between false object answers and productions with medial *wh*-phrases when they conducted a within-subjects experiment to test this. Yet this correlation would be expected under non-target grammar accounts.

With respect to the third piece of evidence (the fact that children report some sentences with medial *wh*-phrases as being acceptable), we just want to note that performance measures have always been understood to only constitute defeasible evidence as to the grammatical status of the sentences in question. For example, we know from center embedding sentences and garden-path sentences on the one hand (see, e.g., Frazier & Fodor 1978) and Negative Polarity Item (NPI) illusions on the other hand (see, e.g., Drenhaus, Frisch & Saddy 2005) that acceptability and grammaticality can come apart, in both directions. Thus, it is possible that the acceptability judgments reported in McDaniel, Chiu & Maxfield (1995) do not reflect the grammatical competence of the children that took part in the study. That is to say, it's possible that the sentences that children reported as acceptable were nonetheless ungrammatical with respect to their own mental grammars. There were 32 children who took part in their acceptability judgment task, ranging in age from 2;11 to 5;07; the majority were 3- and 4-year-old children. Acceptability judgments are quite hard to elicit from young children, for a variety of reasons, so it is quite possible that their acceptability judgments did not reflect their own mental grammars. Moreover, overall acceptance rates for the sentences with medial *wh*-phrases were generally low (around 20%), compared to acceptance rates for sentences that are grammatical in adult English (cf. McDaniel, Chiu & Maxfield 1995:724, Table 2). Given these considerations and given the converging evidence from studies like Lutken, Legendre & Omaki (2020) and the following experiment suggesting these aspects of child English reflect performance factors, we think this is the correct interpretation of these judgments.

Next, we briefly discuss some general issues that the non-target grammar accounts face.

2.2.2. *General issues with non-target grammar accounts*

Perhaps the biggest issue that non-target grammar accounts face is the fact that languages with scope marking and *wh*-copying are not monolithic. As just one example of this fact, there are some dialects of German and Romani that allow scope marking but not *wh*-copying (see, e.g., McDaniel 1986:93–185). Nonetheless, almost all of the non-target grammar accounts treat children's productions of both sorts of constructions as arising from the same piece of hypothesized (non-target) grammatical knowledge. Thus, on these accounts, it is actually quite unexpected that there would be German and Romani dialects with scope marking but not *wh*-copying. These accounts only expect there to be dialects with both constructions or neither.[9]

Relatedly, children do actually produce some utterances that are not possible in the world's languages. For example, Thornton (1995:151) reports some utterances that involve a *wh*-phrase with *which* in the matrix clause and some other *wh*-phrase at the edge of the embedded clause, as in (8).

```
(8) Tiffany 4;9
    a.  Which Smurf do you think who has roller skates on?
    b.  Which animal do you think what really says "woof woof"?
```

Such utterances are, as far as we know, not possible in the world's languages. Although it is possible to have a *wh*-phrase with *which* appear at the edge of the embedded clause in a scope marking construction, it is not possible to use a *wh*-phrase with *which* in a *wh*-copying construction (see, e.g., McDaniel, Chiu & Maxfield 1995:712). But the utterances in (8) are neither *wh*-copying nor scope marking constructions, since the *wh*-phrase with *which* appears in the matrix clause (not at the edge of the embedded clause), and the *wh*-phrase at the edge of the embedded clause is a different *wh*-phrase.

These utterances are problems for non-target grammar accounts, but they can be straightforwardly accounted for with our failure of inhibition account, as discussed in § 2.1. Because it is just the *wh*-feature that is maintained in memory, when the speaker starts the embedded clause, they reactivate this *wh*-feature, which causes spreading activation to all *wh*-words, and so the speaker might accidentally produce a different *wh*-phrase depending on the prior activations of the different *wh*-words, especially if they are not good at inhibiting prepotent but inappropriate responses (they have poor cognitive inhibition).

We conclude that some of the evidence in favor of non-target grammar accounts is not as definitive as it might have appeared. Moreover, there are some non-trivial issues that non-target grammar accounts face. We next turn to experimental evidence that supports our inhibition account above and beyond any non-target grammar account.

3. The experiment

We conducted four different tasks with each child in order to tease apart our failure of inhibition account from non-target grammar accounts. Specifically, we conducted (i) an elicited production task, (ii) a task that measures cognitive inhibition, (iii) a task that measures motor inhibition, and (iv) a task that measures motor ability. The tasks themselves are described in more detail in § 3.2. We next describe the predictions of the various accounts.

3.1. *Predictions*

As noted in § 2, the non-target grammar accounts predict there to be no relation between measures of inhibition control and production of medial *wh*-phrases. On the other hand, our failure of inhibition account predicts that children who exhibit poor inhibition control will be more likely to produce

[9]One exception to this is the non-target grammar account in Thornton (1990), where children's productions with non-identical *wh*-phrases are treated as distinct from children's productions with identical *wh*-phrases; however, no concrete analysis of the former sorts of productions is given. McDaniel, Chiu & Maxfield (1995:740) do also briefly acknowledge this problem, and they propose, in passing, an *ad hoc* solution to the issue, which is particular to the details of their account that we've glossed over.

medial *wh*-phrases than children who exhibit good inhibition control. In fact, as discussed previously, both Harnishfeger (1995) and Dempster (1991, 1993) argue that cognitive and motor response inhibition are distinct executive functions and that the development of motor response inhibition precedes the development of cognitive inhibition. Cognitive inhibition, unlike motor inhibition, involves the ability to suppress cognitive processes, such as the suppression of stray thoughts, suppression of memory processes, etc. (cf. Harnishfeger 1995:184). So our account therefore more specifically predicts that the children who exhibit poor *cognitive* inhibition will be more likely to produce medial *wh*-phrases than children who exhibit good *cognitive* inhibition.

Our account does not predict there to be any correlation between likelihood of medial *wh*-phrase production and motor inhibition or motor ability. We included these two additional tasks as control tasks, in effect; if we do not see a correlation between these two measurements and the likelihood of medial *wh*-phrase production, then this is at least suggestive that medial *wh*-phrase production is not merely predicted by general measurements of development, modulo, of course, the necessary caveats about the interpretation of null effects.

3.2. *Methods and materials*

Next, we describe each of the four tasks in more detail.

3.2.1. *Elicited production*

For the elicited production task, children watched short animated videos while a puppet, Snuggles, was hiding. The videos were short clips from a Brazilian animated cartoon show, *Turma da Mônica* ('Monica's Gang'), that lacked dialogue.[10] There were six different clips that averaged 28.167 seconds in length; the range of clip length was 23–30 seconds.

After each video, Snuggles came out of hiding, and the child was encouraged to ask the puppet between two and four different questions per video. For example, in one video Mônica chased two other children (who've stolen her stuffed animal) across a tightrope. The experimenter would then prompt the participant to ask Snuggles a question by whispering something like (9) to the child. (The prompt was whispered so that Snuggles couldn't hear the answer to the question.)

(9) We know that it was the girl that was chasing the boys, but let's ask Snuggles who he thinks.

The participant was expected to ask Snuggles the question in (10).

(10) Who do you think was chasing the boys?

As noted, there were between two and four questions that we elicited per video clip, for a total of 21 questions. We elicited eight subject questions, seven object questions, and six adjunct questions. A full list of the prompts and the intended elicited questions can be found in the appendix.

The first video clip had three target questions (two subject questions and one object question). During the trials for this first video, if the participant failed to ask a multiclausal question, the experimenter reminded the participant that we want to know what Snuggles *thinks* and then prompted the participant to ask again. The hope was that the participant would then ask Snuggles a multiclausal question where the first clause was *what/who/how/which game/etc. do you think*. This was not done for subsequent video clips. In our following statistical analysis, we only analyze multiclausal utterances where there could have been a medial *wh*-phrase.

[10]The YouTube channel for *Turma da Mônica* can be found at the following link: https://www.youtube.com/channel/UCV4XcEqBswMCryorV_gNENw

3.2.2. *Cognitive inhibition task*

To measure cognitive inhibition, we used a picture-naming task from the literature on cognitive inhibition, following Kipp & Pope (1997). Participants were asked to name items as quickly as possible in a picture book, *Anno's Journey*, as the experimenter pointed to them. There were two 2-minute blocks for this task. In the first 2-minute block, the participant was told that they were not to name the item the experimenter pointed to if it was an animal (a "distractor" item). In the second 2-minute block, the participant was told to name every item, including the animals. The measure of cognitive inhibition from this task is whether or not a participant erroneously named a distractor in the first block since erroneously naming a distractor is indicative of not being able to inhibit the prepotent response to name the item. The second block was included to ensure that all participants were capable of naming animals, which they were.

Before the first 2-minute block, there was a brief 30-second training period where the experimenter pointed to a few items and the participant named them, just so that the participant could get comfortable with the task.

3.2.3. *Motor inhibition task*

To measure motor response inhibition, we implemented a Simon task (cf. Simon & Rudell 1967, Simon 1969) in PsychoPy (Peirce 2007, 2009; Peirce et al. 2019), following a spatial version of the Simon task as described in Davidson et al. (2006). Specifically, there were three conditions: a congruent, an incongruent, and a mixed condition. In the congruent condition, participants saw a heart appear on the screen and had to press a key on the keyboard that is on the same side of the keyboard as the heart (either the *z* or *m* key). In the incongruent condition, the participant saw a flower and had to press the key on the opposite side of the flower. In the mixed condition, the participant saw both hearts and flowers and had to press the relevant key depending on the side of the screen that the image appeared on and what the image was.

The trials were blocked such that the participant always did the congruent trials, then the incongruent trials, and finally the mixed trials. There were 12 congruent trials, 12 incongruent trials, and 16 mixed trials. Half of the stimuli appeared on the left, and half appeared on the right. Each participant was presented the stimuli in the same pseudorandom order. In between each stimulus, there was a 1,000-millisecond interstimulus interval during which a fixation cross appeared in the middle of the screen. The participant had 1,500 milliseconds to press either *z* or *m* after the stimulus appeared. If the child did not press a key before these 1,500 milliseconds elapsed, the next trial began. Prior to the congruent and incongruent trials, there was a brief training period with a few trials of each type. There was no training prior to the mixed trials. Screenshots of the mixed block instructions, a heart trial, and a flower trial from the PsychoPy experiment are included in Figure 1.

The incongruent trials and the mixed trials measure motor response inhibition. Because there is a preference to respond on the same side as the stimulus, the participant must first inhibit this response in an incongruent trial and then respond on the opposite side (cf. Davidson et al. 2006). Incongruent trials occur in the mixed block, so results from the mixed block could be used as a measure of motor inhibition, too. However, we expected all children to generally perform poorly in the mixed block because of how difficult it is. Moreover, the mixed block measures both motor inhibition and the executive function of shifting, since the participant has to switch between the appropriate response for the heart and the appropriate response for the flower. We therefore chose to use mean error rate in the incongruent trials as our measure of motor inhibition in the following statistical analysis.

3.2.4. *Motor ability task*

Finally, the task that we used to measure general motor ability was a motor sequencing task from Carlson & Moses (2001), which they adapted from Welsh, Pennington & Groisser (1991). This task used a toy piano with four different keys. The participant was asked to play each key once in a sequence as many times as possible in 10 seconds, without skipping a key or pressing a key twice. The number of successfully completed scales was used as the measure of motor ability in our following statistical analysis.

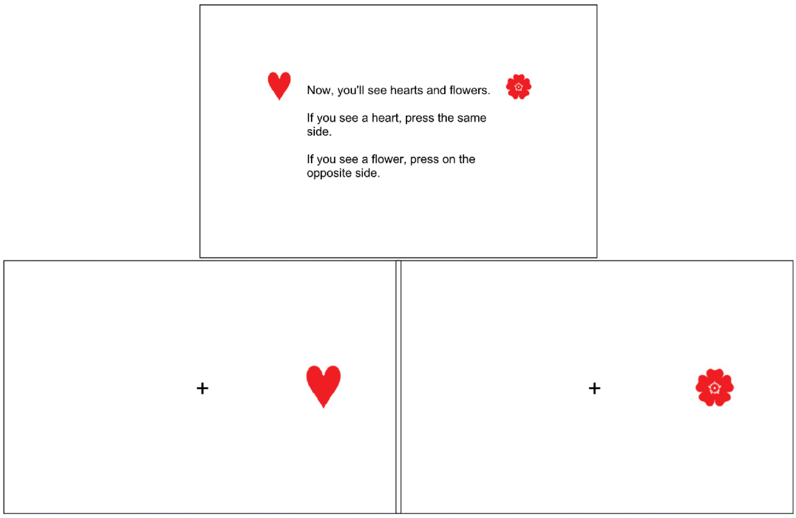


Figure 1. Mixed instructions and example trials from Simon task.

3.3. *Participants*

We collected data from 100 participants. This data set is a superset of the preliminary data from 32 participants that was reported in Grolla & Lidz (2018).

Of these 100 participants, one participant's sessions were not recorded, and so we could not transcribe the data from this participant. Thirteen participants did not complete all four tasks, and one participant's parent/guardian reported that they did not hear English at least 80% of the time, a threshold for inclusion that was determined before data collection began. This leaves 85 participants. Of these 85 participants, 79 produced at least one multiclausal *wh*-question such that there could have been a medial *wh*-phrase. Thus, in what follows, we analyze the data from these 79 participants.[11]

Participants were recruited from the University of Maryland's Center for Young Children (an on-campus school for children) or via the University of Maryland's Infant and Child Studies Consortium. Testing happened in one of two ways. Either participants completed all four tasks in a single session, with a 5–10 minute break in between the elicited production task and the other three tasks, or participants completed the tasks in two different sessions, one with the elicited production task, and one with the other three tasks. The elicited production task was usually done first; the mean number of days between the elicited production task and the other three tasks was 7.58 days, with the range being from −7 to 68. The average age of participants during the session with the elicited production task was 4;08,18, with the range being from 3;06,27 to 6;02,29. The average age of participants during the session with the other three tasks was 4;08,20, with the range being from 3;06,28 to 6;05,1.

3.4. *Results*

We begin discussion of the results from our experiment with some descriptive statistics and visualizations of the measures from each of the four tasks.

[11]The data are available at https://osf.io/vp6mg/

3.4.1. *Elicited production results*

The elicited production results were transcribed and then coded for various linguistic properties of interest. Each utterance was coded for (i) the type of question it was (subject, object, or adjunct), (ii) whether it was a *wh*-question with two clauses (yes or no; i.e., embedded polar interrogatives such as *Do you know what was under the top hat?* were coded as no), what matrix *wh*-word was used (e.g., *who, what, which NP*, etc.), (iii) whether there was a medial *wh*-phrase (yes or no), (iv) what medial *wh*-word was used (if any), (v) whether the matrix *wh*-word was the same as the medial *wh*-word (yes, no, or N/A), (vi) whether there was a *that*-trace violation (yes or no), and (vii) whether there was resumption of the argument in the base position (yes or no). Additionally, following Lutken (2021:171), we coded for several different types of disfluencies. We coded (i) whether a filler was used, such as *uh* or *um* (yes or no), (ii) whether (at most) one word was repeated (yes or no), (iii) whether more than one word was repeated and all repeated words were the same (yes or no), and (iv) whether one or more words were repeated with at least one of the words being changed in the process (yes or no). For each of these four disfluencies, we additionally coded whether the disfluency occurred at the clause boundary (yes, no, or N/A in the case of single-clause utterances), where, for object and adjunct questions, the clause boundary was defined as after the matrix verb but before the subject of the embedded clause and where, for subject questions, the clause boundary was defined as after the matrix verb but before the embedded verb. Disfluencies occurring more frequently at the clause boundary have been claimed to be indicative of greater processing load in that region (cf. McDaniel, McKee & Garrett 2010).

Out of the 21 questions that we attempted to elicit with each participant, participants asked on average 17.88 multiclausal *wh*-questions that could have had a medial *wh*-phrase (range: 3–21; *SD*: 4.91). A breakdown of this average across the three different question types is given in Table 1.

Thirty-seven participants did not ask questions with medial *wh*-phrases, and 42 participants asked at least one question with a medial *wh*-phrase. For these 42 participants, on average, 17.80% of their questions where medial *wh*-phrases were possible did in fact contain a medial *wh*-phrase (range 4.76%–53.33%; *SD*: 14.51%). For example, let's consider the data from one of our participants; in (11), we give the questions from this participant that contained medial *wh*-phrases.

```
(11) Samantha 4;07,23
     a. Adjunct questions with medials (3 out of 4 = 75%)
          i. What do you think how he got lift up?
         ii. Where do you think where they were walking?
        iii. Where do you think where he was hiding?
     b. Object questions (1 out of 5 = 20%)
          i. What do you think what the girls were eating?
     c. Subject questions (7 out of 12 = 58.33%)
          i. Who do you think who kissed the boy?
         ii. Who do you think what popped the balloons?
        iii. What do you think what cut the rope?
         iv. What do you think who fell?
          v. Who do you think what who which kid flew in the sky?
         vi. What do you think which kid jumped into the water?
        vii. Who do you think that who's great at soccer?
```

Table 1. Average number of produced questions where medial *wh*-phrase was possible, by question type, with standard deviations in parentheses.

	Adjunct	Object	Subject	Total
Produced	4.98 (1.58)	5.43 (1.93)	7.48 (2.58)	17.88 (4.91)
Elicited	6	7	8	21

This participant asked 21 multiclausal questions (21 questions where medial *wh*-phrases were possible), and 11 of these did in fact contain medial *wh*-phrases.[12] This particular participant was near the upper end of the range of medial *wh*-phrase production (53.33%), with 52.38% of their multiclausal *wh*-questions containing medial *wh*-phrases. No child produced primarily or exclusively questions with medial *wh*-phrases. A reviewer asks whether children produced one specific type of medial *wh*-structure or whether there was a mix. To address this, we looked at the nine children in our data set who produced more than three medial *wh*-structures (range: 3;07,20–5;05,4; mean: 4;05,17). Of these nine children, five exclusively produced medial *wh*-phrase questions with identical matrix and medial *wh*-words, and four participants produced medial *wh*-phrase questions with a mix of identical and non-identical matrix and medial *wh*-words.

In addition to breaking things down by individuals, we also broke things down by question type. We calculated the rates of medial *wh*-phrase by question type; the proportions are plotted in Figure 2.

As can be seen in Figure 2, medial *wh*-phrases were not all that frequent, occurring in 4.52% of the adjunct questions in our sample, 5.70% of the object questions, and 14.95% of the subject questions. Overall, collapsing across question type, out of the 1,352 multiclausal *wh*-questions that we analyzed in our final sample of 79 participants, 124 (9.17%) of them had medial *wh*-phrases.[13]

Of these 124 questions with medial *wh*-phrases, 81 (65.32%) had matching, 19 (13.52%) had *what* in the matrix clause and the target *wh*-phrase in medial position, and 24 (19.35%) involved some other combination (predominantly (15 of 24) a complex *wh*-phrase in the matrix clause, such as *which kid*, and a related but simplex *wh*-phrase in the medial position, such as *who*). This information is summarized in Table 2.[14]

It is also worth noting that there were many more subject questions with medial *wh*-phrases than there were object or adjunct questions with medial *wh*-phrases in our data. This is reminiscent of the observation in Thornton & Crain (1994:220) that children's production of medial *wh*-phrases persisted longer in subject questions. The evidence that Thornton & Crain (1994) report for this is longitudinal, although they do not give concrete summary statistics. Moreover, it is worth noting that Lutken, Legendre & Omaki (2020) do not see more medial *wh*-phrases with subject questions than with object questions. We will return to this point more in § 3.5.3.

Before turning to the results from the other tasks, we also briefly address the presence of that-trace violations and general disfluencies in our data set. There were 60 instances of that-trace violations in our data set, from 22 different participants (range: 3;07,20–5;11,10; mean: 4;08,26). The baseline

[12]Note that this participant asked 12 subject questions even though we only tried to elicit eight; similarly, this participant asked five object questions, but we tried to elicit seven, and this participant asked four adjunct questions, but we tried to elicit six. Participants did not always ask the type of question that we tried to elicit. We return to this in § 3.4.6 and § 3.5.3.

[13]For comparison, 22% of the utterances in Experiment 1 (30 children, 354 child utterances) from Lutken, Legendre & Omaki (2020) included medial, and 15.4% from their Experiment 3 (20 children, number of child utterances not reported) included medial *wh*-phrases.

[14]A reviewer notes that our account would predict there to be all three of these types of medial *wh*-phrase productions, but the reviewer points out that the data sets from Thornton (1990) and Lutken, Legendre & Omaki (2020) predominantly include productions with matching *wh*-phrases and productions with *what* + target *wh*-phrase. The reviewer is correct to point out that our theory predicts a mix of production types, including non-matching medial *wh*-phrase productions that aren't solely of the *what* + target *wh*-phrase type. This is indeed what we observe in our data set, as can be seen in Table 2.

Moreover, it's not clear to us that these productions actually are absent from the data sets of Thornton (1990) and Lutken, Legendre & Omaki (2020). It is hard to ascertain the extent to which they were observed in Thornton (1990) since few summary statistics are given. That being said, Thornton (1990:237) does report some such productions, including *Which Smurf do you think who has roller skates on?, Which animal do you think what really says "woof woof"?, Which one do the bear and the squirrel think who has two bears?*, and *Which guy did they say which had the orange marble?*. On the other hand, Lutken, Legendre & Omaki (2020:16) do give the relevant summary statistics for their Experiment 1. In this experiment, a majority of the utterances with non-matching *wh*-phrases, 11 of 14, are *not* of the *what* + target *wh*-phrase type. This is consistent with our account. Furthermore, if these types of medial *wh*-phrase productions are more absent from other data sets than one would expect given our account, we suspect this is due to other studies involving fewer participants and fewer utterances than is the case for our study; that is to say, these other studies might just have happened to fail to observe a preponderance of various other non-matching types of medial *wh*-structures in virtue of smaller sample sizes.

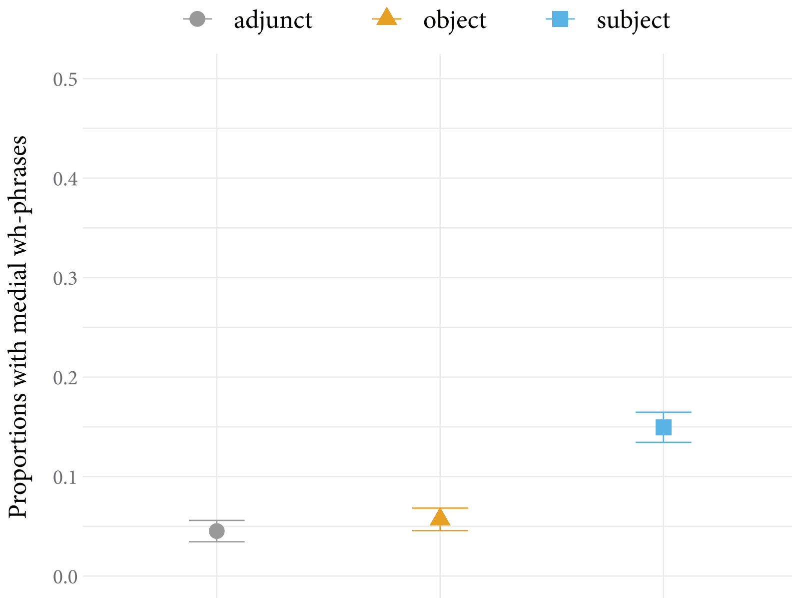


Figure 2. Proportions of questions with medial *wh*-phrases, by question type; error bars indicate one standard error of the proportion.

Table 2. Number of different types of medial *wh*-phrase productions in our data set.

Matching *wh*-phrases	what + target *wh*-phrase	Other
81 (65.32%)	19 (13.52%)	24 (19.35%)

average rate of *that* usage in grammatical adult-like contexts (adjunct and object questions) was 9.07%. This is comparable to an average rate of *that* usage in subject questions (which is ungrammatical in the adult target language), which was 11.3%.

For disfluencies, we looked at all 1,612 utterances from the 79 participants whose data we analyzed. This therefore includes the 1,352 multiclausal *wh*-questions that are of main interest, as well as the 260 utterances that were not multiclausal *wh*-questions. Of these 1,612 utterances, 46 included fillers. Of these 46, 41 occurred in utterances with two clauses, and 23 of these 41 occurred at the clause boundary. Fifty-three of the 1,612 utterances contained the repetition of at most a single word, and of these, 42 were multiclausal utterances, with 10 of the 42 occurring at the clause boundary. Twenty of the 1,612 utterances included the repetition of more than just a single word. Of these, 18 were in multiclausal utterances, and two of the 18 occurred at the clause boundary. Finally, there were 68 out of 1,612 utterances with a repetition of one or more words where at least one word was changed. Of these, 64 occurred in multiclausal utterances, and 12 of the 64 were at the clause boundary. This information is summarized in Table 3.

3.4.2. *Cognitive inhibition results*

Next, we present some descriptive results from the cognitive inhibition task. In Figure 3, we plot the proportion of participants that erroneously named a distractor in the first 2-minute block of the cognitive inhibition task, split by whether the participant produced any medial *wh*-phrase throughout the course of the elicited production task. This figure is intended solely as a descriptive summary of the results from this task, but splitting the data in this way anticipates the statistical analysis we report in § 3.4.6 and is a useful way to visualize the descriptive results in light of the predictions that the different accounts make.

Table 3. Number of occurrences of each disfluency type and number of each type that occurred at a clause boundary.

Disfluency type	Number occurrences	Number at clause boundary
Filler	46 (2.9%)	23 (56.1%)
Repetition	53 (3.3%)	10 (23.8%)
Multi-word repetition	20 (1.2%)	2 (11.1%)
Repetition w/correction	68 (4.2%)	12 (18.8%)
Total	187	47

Note: Percentages are given in parentheses for the proportion of these disfluencies relative to the total number of utterances and relative to the number of utterances with that disfluency type and with multiple clauses respectively. Percentages are not given for the totals because these disfluencies were not mutually exclusive (i.e., they can, and occasionally did, occur in the same utterance).

As can be seen in Figure 3, more participants who produced at least one medial *wh*-phrase erroneously named a distractor than participants who did not produce any medial *wh*-phrases.

3.4.3. *Motor inhibition results*

Moving to the motor inhibition results, in Figure 4, we plot the mean error rate by group and by the three different conditions in this task. As can be seen in the figure, the mean error rates for the congruent and mixed conditions were similar between the two groups.

For the incongruent condition, the group who produced at least one medial made more errors than the group who did not produce any medials.

3.4.4. *Motor ability results*

Next, we present descriptive results of the motor ability task. In Figure 5, we plot the average number of successful scales that participants completed, by group. As can be seen in the figure, both groups successfully completed, on average, approximately the same number of scales in this motor sequencing task.

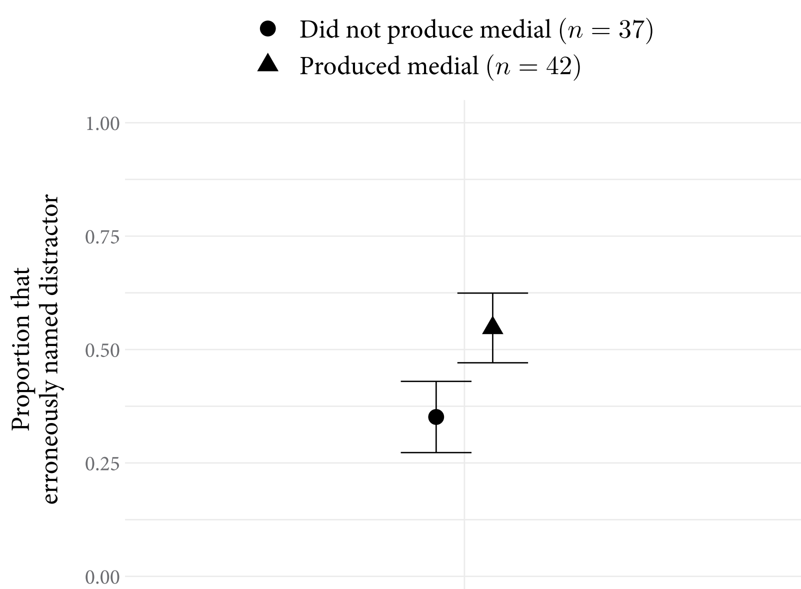


Figure 3. Proportion of participants who erroneously named distractor, by group; error bars indicate one standard error of the proportion.

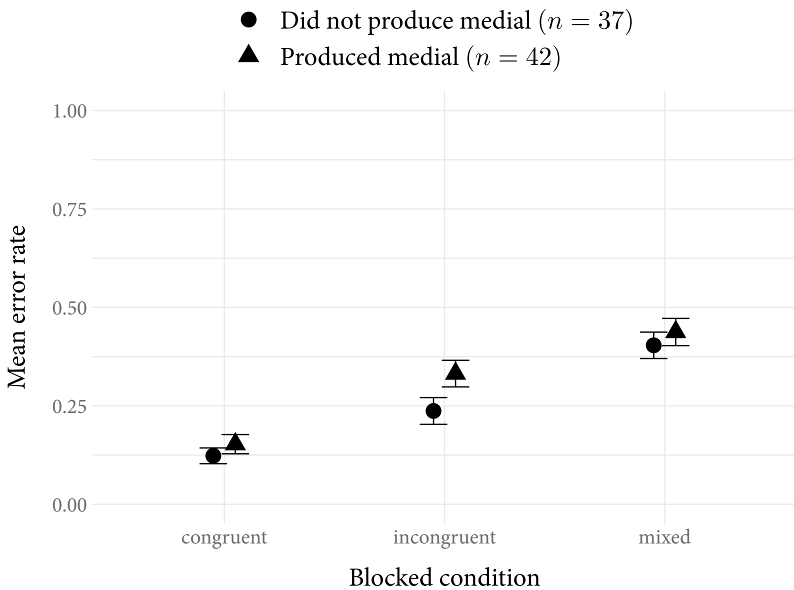


Figure 4. Mean error rate, by group and by condition; error bars indicate one standard error of the mean.

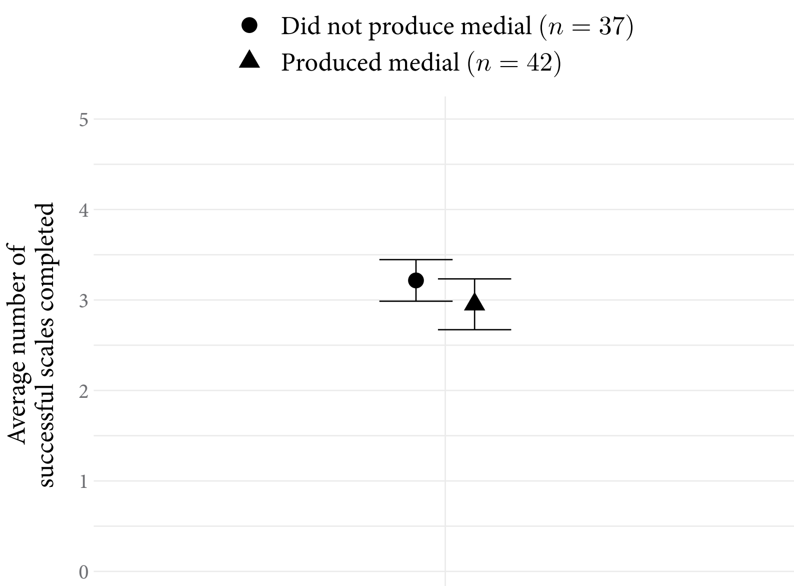


Figure 5. Average number of successful scales completed, by group; error bars indicate one standard error of the mean.

3.4.5. *Intertask correlations*

Finally, before turning to the statistical analysis where we model the presence of a medial *wh*-phrase in a given production as a function of these various measures, we report Pearson's correlation coefficients between the different task measures that we just discussed in Table 4.

3.4.6. *Statistical analysis*

We next present a statistical analysis of these results aimed at teasing apart the non-target grammar accounts from our failure of inhibition account on the basis of the predictions discussed previously. To reiterate, our failure of inhibition account predicts that cognitive inhibition should be positively correlated with the production of medial *wh*-phrases, whereas the non-target grammar accounts predict no such relation. To test these predictions, we fit several different logistic mixed-effects models to the data. The models were fit using R (R Core Team 2017) and the package lme4′s glmer function (Bates et al. 2015).

Each model was fit to predict whether or not a given trial where a medial *wh*-phrase was possible in fact had a medial *wh*-phrase. We fit six different models, and we compared the model using Akaike Information Criterion (AIC) (cf. Akaike 1973) and χ^2 tests of the log likelihood ratios. The fixed effects in our six different models included an intercept term, a categorical variable indicating the type of question that the participant actually asked, adjunct, object, or subject (*not* the type of question that we tried to elicit; this variable was deviation coded with adjunct as the reference level), a categorical variable indicating whether or not the participant switched to asking a type of question other than the type that we intended to elicit on that particular trial (treatment coded as yes or no), a categorical variable indicating whether or not the participant erroneously named a distractor in the cognitive inhibition task (treatment coded as yes or no), a variable coding the participant's mean error rate in the incongruent condition of the motor inhibition task, and a variable coding the number of successful scales that the participant completed in the motor ability task.

By modeling our results with generalized linear models, we can control for the effects of other potential covariates in the statistical analysis by including them as effects in the model (see, e.g., Baayen, Davidson & Bates 2008:399–400). Particularly given the observation in Thornton & Crain (1994) that medials persisted longer with subject questions, we planned to include a categorical fixed effect predictor for the type of question that the participant actually asked. As discussed in fn. 1, this was not always the type of question that we tried to elicit, so we also included a categorical variable indicating whether or not the participant switched to asking a question of a different type. Beyond this, our fixed effects included the measures from our three other tasks, as discussed in § 3.2. The fixed effects of the six different models are summarized in Table 5. This allowed us to compare models in a way that is consistent with the predictions of our account. To reiterate, our account does not predict there to be any relation between the number of successful scales completed and the production of medial *wh*-phrases. Similarly, our account does not predict there to be a relation between motor inhibition and the production of medial *wh*-phrases. Our account does, however, predict there to be a correlation between cognitive inhibition and production of medial *wh*-phrases. We therefore constructed models that successively eliminated the predictors that we did not expect to correlate with production of medial *wh*-phrases, as specified in the table.

In order to be able to do model comparison in this way, the random effects structures of the models need to be the same. We used the maximal random effects structure justified by design of the experiment that led to model convergence (cf. Barr et al. 2013). Specifically, our models were fit with random intercepts for participant and for trial. The results of comparing the models in this way are given in Table 6. We report values for AIC and χ^2 tests of the log likelihood ratios.

Table 4. Pearson's correlation coefficients for various task measures.

	Congruent error	Incongruent error	Mixed error	Number scales
Erroneously named distractor	0.12	0.18	0.16	−0.22
Congruent error		0.11	0.26	−0.36
Incongruent error			0.42	−0.26
Mixed error				−0.20

Table 5. The fixed effects of our six different logistic mixed-effects models.

	Models					
	m1	m2	m3	m4	m5	m6
Intercept	✓	✓	✓	✓	✓	✓
QuestionType		✓	✓	✓	✓	✓
SwitchedType			✓	✓	✓	✓
NamedDistractor				✓	✓	✓
IncongruentError					✓	✓
NumberOfScales						✓

AIC is a way of measuring how well a model fits the data while also penalizing a model for having lots of predictors, since a model with a large enough number of predictors will be able to fit the data perfectly. Lower values for AIC are better, indicating good model fit without having too many predictors. χ^2 tests of the log likelihood ratios of the various models yield p values, which can be interpreted just like other p values. As can be seen in the table, both of these measures indicate that the best model is the fourth model.[15,16]

We can then look details of the best fitting model, as in Table 7. The fixed effects that were significant predictors of whether or not the utterance contained a medial *wh*-phrase included being a subject question, being uttered on a trial when the participant switched the question type, and being uttered by a participant who had erroneously named a distractor.

Because the generalized linear model is a logistic model, it is not straightforward to interpret the numeric value of the estimates without transforming them. The exponentiated estimates yield odds ratios indicating the change in the odds of being an utterance with a medial *wh*-phrase, for every one-unit change in the value of the predictor. All of the predictors in our best fitting model are categorical

Table 6. Comparing the different models.

	AIC	Log likelihood	χ^2	p value	
m1	729.97	−361.99			
m2	700.79	−345.40	33.18	6.24e−8	***
m3	691.03	−339.51	11.76	6.036e−4	***
m4	687.94	−336.97	5.09	0.0241	*
m5	688.09	−336.05	1.85	0.1741	
m6	688.92	−335.46	1.17	0.2789	

Table 7. Details of the best fitting model.

Fixed effect	Estimate	Std. Error	z value	p value	
Intercept	−3.952	0.382	−10.340	< 2e−16	***
QuestionType – object	−0.249	0.211	−1.178	0.239	
QuestionType – subject	0.833	0.187	4.463	8.08e−6	***
SwitchedType	1.228	0.353	3.480	5.01e−4	***
NamedDistractor	0.992	0.442	2.246	0.0247	*

[15]Note that the fixed effect coding the error rate in the incongruent condition of the motor inhibition task was not in the best fitting model, despite the seeming difference between groups that can be seen in Figure 4, where we simply plotted some descriptive statistics about the measures from this task. This is consistent with the predictions of our account, as discussed in § 3.1.

[16]A reviewer points out that there is very little variability in our motor ability task measure and so, because of this, the measure may not have been predictive of medial *wh*-phrase production. This is indeed possible. As Hedge, Powell & Sumner (2018) point out, when between-subject variability in a measure is low, detecting a correlation with that measure becomes mathematically more unlikely. That being said, neither our account nor any other account predicted there to be a correlation between medial *wh*-phrase production and motor ability. We only included this task as a control task. If we had a motor ability measure with more between-subject variability, our account (and all other accounts) would still predict there to be no correlation with medial *wh*-phrase production.

predictors, so, for example, the value $e^{0.992} \approx 2.70$ can be interpreted as indicating that an utterance is 2.70 times more likely to have a medial *wh*-phrase if it was uttered by a participant who named a distractor in the cognitive inhibition task than if it was uttered by a participant who did not name a distractor in that task. To facilitate interpretation of the best fitting model, we plot these odds ratios in Figure 6.

In addition to being 2.70 times significantly more likely on trials uttered by participants who named a distractor, medial *wh*-phrases were also 2.30 times significantly more likely in subject questions than the average question and 3.41 times significantly more likely on trials where the participant asked a different question type other than the type that we tried to elicit.

3.5. *Discussion*

As predicted by our failure of inhibition account, cognitive inhibition is a significant predictor of medial *wh*-phrase production by English-acquiring children. This relationship is unexpected (and unexplained) on all of the non-target grammar accounts. In addition to the various issues with the different non-target grammar accounts that we discussed in § 2.2, this experimental result lends support to our failure of inhibition account over the non-target grammar accounts.

Before turning to some general discussion about our data set, our best-fitting model, our account, and a comparison with the performance account from Lutken, Legendre & Omaki (2020), we wish to further address the relation between medial *wh*-phrase productions with both *that*-trace violations and disfluencies, as helpfully suggested by a reviewer.

3.5.1. *Medial wh-phrases and that-trace violations*

A reviewer asked about the presence of *that*-trace violations in our data set and whether such productions could be accounted for under our account. Although studies that discuss medial *wh*-phrase production often also report instances of *that*-trace violations (see, e.g., Thornton 1990,

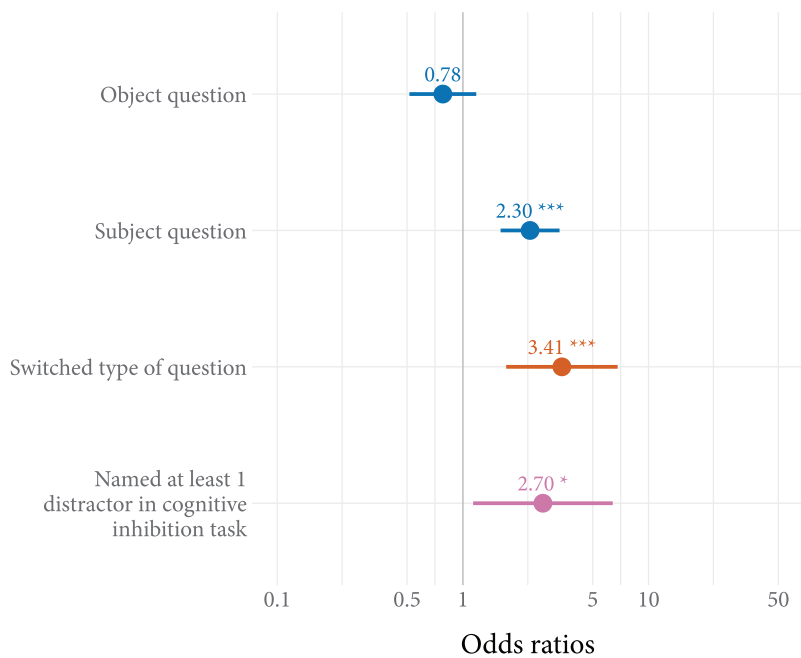


Figure 6. Odds ratios of the best fitting model.

Lutken, Legendre & Omaki 2020), we are not aware of any study that reports statistical evidence for a relation between these two types of productions. As such, it is not clear whether a theory should be made to account for both types of productions.

Looking at our data set specifically, we tested whether there is a correlation between a participant's rate of medial *wh*-phrase production and a participant's rate of *that*-trace violation productions using a Pearson's correlation test and found no evidence for a significant correlation, $r = 0.049, t(77) = 0.432, p = 0.6669$. That being said, we did have a small number of *that*-trace violations in our data set (60, from 22 participants), so the lack of correlation could be due to lack of statistical power, in addition to the normal caveats about interpreting null results. Indeed, it is perhaps suggestive that 17 of the 22 participants who produced *that*-trace violations in our data set also produced medial *wh*-phrases (although 25 of the 42 who produced medial *wh*-phrases did not produce any *that*-trace violations).

If it were to turn out that these things are indeed correlated, we do think our account could explain such a correlation. The story would be similar to the story we sketched for medial *wh*-phrases. Specifically, in subject questions, the grammar of the child says that you can only use the null complementizer ($[_C\ \emptyset]$), but when you activate $[_C\ \emptyset]$ you get spreading activation to other complementizers, such as $[_C$ *that*$]$. Thus, in some cases, $[_C$ *that*$]$ might end up being more active than the null complementizer, and so the child might pronounce $[_C$ *that*$]$ instead of $[_C\ \emptyset]$, even though it goes against the child's grammar. The only difference between this and the medial *wh*-phrase case is that there is no reactivation of an element you've already uttered. Instead, the triggering event is just the initial activation of the embedded complementizer, which can spread.

The implicational relation we just observed (if you're a *that*-trace violation producer, you're also likely a medial *wh*-phrase producer, but not the converse) is perhaps even suggestive that this account might be right. In the case of *that*-trace violations, you're only dealing with spreading activation from an initial activation of a lexical item, whereas in the case of medial *wh*-phrase productions, you have spreading activation from reactivation of a lexical item (or its features); that is, the spreading activation is likely weaker in the *that*-trace violation case, and so if you see failure of cognitive inhibition in these cases, you'll also see failure of cognitive inhibition in cases where you have stronger spreading activation (but not the converse).

All of this, however, is highly speculative since we do not have concrete evidence of a relation between *that*-trace violating productions and medial *wh*-phrase productions, nor are we aware of any such evidence from other studies. This possibility is worth further future investigation, but it will likely require a much larger data set than ours in order to avoid lacking sufficient statistical power to detect a possible correlation.

3.5.2. *Medial* wh-*phrases and disfluencies*

A reviewer also suggested looking at whether there was a relation between the production of medial *wh*-phrases and disfluencies. The reviewer noted that under performance accounts, but not non-target grammar accounts, one might expect productions with medial *wh*-phrases to have more disfluencies. Of the 124 out of 1,352 multiclausal utterances with a medial *wh*-phrase in our data set, 30 also contained a disfluency, and 109 of the 1,228 multiclausal utterances without a medial *wh*-phrase contained a disfluency. A χ^2 test of independence indeed suggests that the presence of a medial *wh*-phrase is not independent of the presence of a disfluency in a given utterance, $\chi^2(1, N = 1,352) = 27.011, p = 2.023\text{e–}7$. This is consistent with the reviewer's suggestion and lends support to both our account and the account in Lutken, Legendre & Omaki (2020) over and above the non-target grammar accounts, though we still think our account offers a better explanation than the performance account in Lutken, Legendre & Omaki (2020); see § 3.5.3 for more discussion.

3.5.3. *General discussion*

Furthermore, there is another aspect of the data from our participants that we think lends support to our account. Some of our participants produced questions where the *wh*-chain was resumed with an answer. All 20 instances of this from our data set are given in (12), with the relevant words in bold.[17]

(12) a. Natalie 4;10,17[18]
 i. **What** were the boys playing with **the ball**?
 b. Ina 4;05,13
 i. **Which kid** do you think that **it** was really right?
 c. Eva 4;07,00
 i. **Who** do you think **he** was really good at it?
 d. Ralli 3;08,28
 i. **Who** do you think **the boy** hide?
 ii. **Who** do you think **a kid** jumped in the water?
 iii. **Who** do you think the boy saw **the girl**?
 e. Meghan 4;11,06
 i. **What** do you think that the girl was holding **a bunny**?
 ii. **What** do you think the girl kissed **the boy**?
 f. Alex 5;04,28
 i. **What** did you think someone was walking on **the rope**?
 ii. **What** do you think **the boy** got hit?
 g. Samantha 4;07,23
 i. **Who** do you think **girl** was chasing the boys?
 h. Kaylee 4;08,16
 i. Snuggles, **where** do you think the girl was **behind the fence**?
 i. Trey 4;04,09
 i. Snuggles, how do you think the boy . . . the yellow green boy it was flying **with balloons**?
 j. Thomas 4;00,07
 i. **Where** do you think the green boy hided **under the water**?
 ii. **Where** do you think the girl was **behind the fence**?
 k. Ellie 3;07,00[19]
 i. **What** did the boy saw **the girl**?
 l. Annabelle 4;04,08
 i. **Who** you thinks **he**'s the good of soccer?
 m. James 4;02,11
 i. Snuggles, **how** do you think the girl popped . . . popped [inaudible] **watermelon seeds**?
 ii. **Where** . . . **where** you think the boys was hiding **in the water**?
 iii. Snuggles, **where** do you think the boy was hiding **over the fence**?

Such non-target productions can be accounted for with our failure of inhibition account in effectively the same way that non-target medial *wh*-phrase productions are accounted for. Specifically, upon encountering the base position of the *wh*-chain during the course of sentence production, the *wh*-feature is reactivated and this leads to spreading activation; this activation spreads to *wh*-words, but it could also spread to related elements, which would include the actual answer to the question, particularly given the design of the experiment in which the

[17]The question in (12m-i) clearly involves argument resumption in the context of the video, which involved a girl popping balloons by spitting watermelon seeds at them. The "[inaudible]" portion of the transcription presumably involved the preposition *with*.
[18]Note that this was not one of the 1,352 multiclausal utterances.
[19]Note that this participant did not complete all four tasks and so their data is not analyzed in our main results. This was also not a multiclausal utterance.

participant actually knows the answer to the question in the embedded clause. The activation of the answer may, on occasion, exceed the activation of the *wh*-phrase, depending on the prior background activations; moreover, even though the child's grammar forbids pronouncing anything in this position, the child's developing cognitive inhibition control may sometimes fail to prevent them from pronouncing the highly active lexical material. Our inhibition account therefore provides a unified account of these sorts of non-target productions, something the non-target grammar accounts cannot offer.

Moreover, as can be seen in (12), we did not see any instances of *wh*-phrase production in the base position. Although such productions are, in principle, expected under our account, we would expect them to be exceedingly rare, even more rare than cases in which the DP that answers the question is produced in the base position, as we discussed in fn. 6. Given that, even in our relatively large data set of 1,352 multiclausal utterances, we only had 20 cases of argument resumption and 124 cases of medial *wh*-phrases, we would only expect to see such productions in a much larger data set.

To reiterate some of the previous discussion, even in cases of cognitive inhibition failing, we would still expect productions to be constrained by properties of the grammar (see also Dell 1986:291–292), including the statistical distribution of utterances that the grammar licenses and which the child has been exposed to in the course of acquisition. Since *wh*-phrases only occur in base positions in very highly constrained contexts, we would expect them to be exceedingly rare in children's productions, and, if anything, we would expect to see more cases of argument resumption. This is exactly what we see in our data set.

Relatedly, we think this account could also be straightforwardly extended to provide an account of the fact that children sometimes produce resumptives in relative clauses, despite such productions being ungrammatical in the target language. (13) gives an example of this from Palestinian Arabic (see, among others, Labelle 1990, McKee & McDaniel 2001, Botwinik, Bshara & Armon-Lotem 2015).

(13) `iz-zara:fi illi l-walad ḥaẓan **`iz-zara:fi**
the-giraffe that the-boy hugged **the-giraffe**
'The giraffe that the boy hugged'
(Botwinik, Bshara & Armon-Lotem 2015:49, ex. (20c))

Our account would explain such productions in a very similar manner. Specifically, the idea is that some representation of the filler is maintained in memory until the gap position is reached. When the gap position is reached, the element receives a spike in activation, and if the children are not particularly good at preventing prepotent but inappropriate responses, they may accidentally pronounce the element even though it is ungrammatical to do so in their mental grammar.

Another similar non-adult-like phenomenon that we think our account could also in principle explain is children's productions of two auxiliary verbs, such as in (14) (see, among others, Stromswold 1990, Guasti, Thornton & Wexler 1995).

(14) Did I didn't mean to?
(Stromswold 1990:58)

Both of these phenomena are quite similar to children's non-adult-like productions of medial *wh*-phrases, and it would be straightforward to conduct the same sort of experiment that we report here but with these other two phenomena.[20]

[20]One author of this article has collected some preliminary data with Brazilian-Portuguese-acquiring children that tentatively suggests that the production of resumptive elements in relative clauses is also correlated with cognitive inhibition. Unfortunately, this data collection effort has been interrupted by the COVID-19 global pandemic.

Finally, we discuss the other two significant predictors of medial *wh*-phrase production in our best fitting model. Although we did not expect these predictors to be significantly correlated with medial *wh*-phrase production *a priori*, we think these correlations can be explained on our account. First, our best fitting model indicated that medial *wh*-phrases were significantly more likely on trials where the participant asked a type of question other than the type that we tried to elicit. We think this is consistent with our failure of inhibition account. Recall that the participant was prompted with (9), repeated here as (15) for convenience.

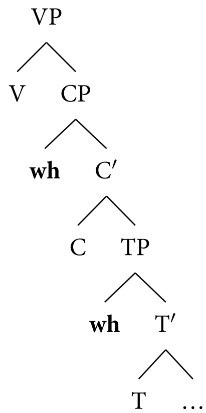
(15) We know that it was the girl that was chasing the boys, but let's ask Snuggles who he thinks

In order to do the task, the child must first recognize what has been clefted, realize that that thing is what they must ask the puppet about, and then form a multiclausal question that queries the puppet for what it thinks is the head of the cleft. This is a non-trivial task and presumably taxing on the child's executive functioning abilities generally. On trials where children's executive functioning abilities are particularly taxed, for whatever reason, we might see this reflected in the participants simply deciding to ask another question that is easier for them to formulate. If question switching is indeed indicative of taxed executive function, then it is not surprising that we would see more medial *wh*-phrases on trials where the participant has switched the question type. It should be noted, however, that on the 26 trials in our data set where this happened and where a medial *wh*-phrase was produced, 22 of these trials involved switching from a non-subject question to a subject question. Although the fixed effect for question switching was still a significant predictor in our best-fitting model alongside the fixed effect for subject questions, it could be that the significance of this predictor is really driven by the accidental fact that most of the switched-to questions were subject questions. Exploring this possibility further in future work could help to adjudicate between these possibilities.

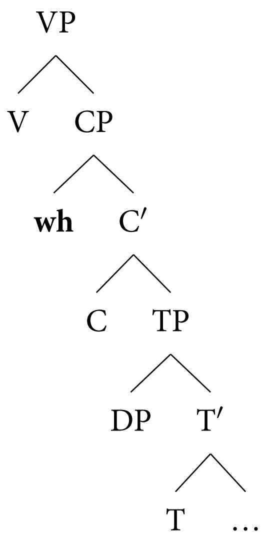
Second, our best fitting model also included a significant correlation between being a subject question and having a medial *wh*-phrase. As noted, this is reminiscent of the observation in Thornton & Crain (1994) that medial *wh*-phrases persisted longer in subject questions, although, again, Lutken, Legendre & Omaki (2020) did not find this to be the case in their results. It could be that the result in our data is spurious, or it could be that the effect is somewhat small, and Lutken, Legendre & Omaki (2020) failed to detect it because they had many fewer participants in their experiments than we have in ours. We suspect this subject effect is a genuine effect, and we think it is in fact consistent with our failure of inhibition account, though future experiments should investigate this in more detail.

We think medial *wh*-phrases are more likely to occur in subject questions on our account for the following reason. First, to reiterate, our account holds that medial *wh*-phrases occur in utterances by children because the *wh*-phrase becomes highly active due to spreading activation from the reactivated *wh*-feature at the site of the copy of the *wh*-phrase, and elements that are highly active are what get pronounced in production planning, as per Dell (1986); moreover, this happens *despite* the child's grammar forbidding it because the child's cognitive inhibition control is still developing. Now, when planning the production of a multiclausal subject *wh*-question, it is plausible that there is a unit of production planning that includes CP and TP of the embedded clause to the exclusion of the embedded VP. If so, then this unit of production planning would contain two copies of the *wh*-chain in the case of subject questions but only one copy of the *wh*-chain in adjunct and object questions, as can be seen in (16).

(16) a. Subject question



b. Object and adjunct questions



Therefore, when planning the beginning of the embedded clause of a multiclausal subject question, there would be two instances of reactivation of the *wh*-feature, both of which would lead to spikes of spreading activation to the *wh*-element. Thus, in a subject question, the spreading reactivation could be as much as twice as high than in adjunct or object questions, and it could therefore be all that more difficult to prevent oneself from pronouncing the highly active lexical material, particularly if one's cognitive inhibition is still developing. We therefore think that subject questions being significantly predictive of medial *wh*-phrase production is consistent with our failure of inhibition account.[21] It would be interesting for future work to investigate this relationship further.

[21]A reviewer asks whether the account from McDaniel et al. (2015) could explain the subject asymmetry that we see in our data set. McDaniel et al. posit that there is a general production pressure such that subject extraction across a clausal boundary is more difficult than other types of extraction across a clausal boundary. They furthermore argue that the effects of this production pressure are grammaticalized (2015:421, fn. 7), in an attempt to explain certain Empty Category Principle contrasts and *that*-trace effects in English.

We do not think this account could explain the subject asymmetry that we see in our data. The core of their account still turns on the status of the grammar, even if the nature of grammar has been in part shaped by certain production pressures. We see no evidence in our data set that leads us to posit different grammars for different children at different times. That being said, perhaps the production pressures that disfavor subject extraction across a clausal boundary, which McDaniel et al. (2015) posit as the basis for some grammatical knowledge, could still explain the subject asymmetry in our data set, without appeal to the grammatical knowledge itself. We still, however, fail to see how this would explain the preponderance of medial *wh*-phrases with subject questions. It's not clear why the general difficulty of subject extraction across a clausal boundary should result in a greater frequency of medial *wh*-phrases. Moreover, if their account is correct, it's actually surprising that in 22 of the 26 cases in our data set where children asked a question with a medial *wh*-phrase while also switching the type of question to one other than the one that we tried to elicit, they switched an intended non-subject question to a subject question.

Before concluding, we wish to briefly return to the alternative performance account offered by Lutken, Legendre & Omaki (2020). As noted in fn. 7, they offer a performance-based account of these utterances that is very similar to our own. Their account also relies on the idea that fillers are actively maintained in memory throughout production, but they propose that children struggle to maintain the filler in memory and may sometimes pronounce it at the clause boundary as an attempt to strengthen their memory representation of the filler (Lutken, Legendre & Omaki 2020:37–38). Although very similar in spirit, we think there is reason to favor our proposal. First, we find it a bit counterintuitive that their proposal is one where you pronounce the elements that you are forgetting.[22] This seems directly at odds with the model of sentence production proposed in Dell (1986), where you pronounce the elements that are highly active in memory. Second, the experimental results we present here favor our account over their account. If pronunciation of the filler at the clause boundary is an attempt to strengthen its representation in memory, there's no apparent reason why these productions should correlate with cognitive inhibition. That is to say, like the non-target grammar accounts, the most straightforward prediction that this account makes with respect to the experiment that we conducted is that there should be no correlation between cognitive inhibition and medial *wh*-phrase production. Moreover, it is unclear on their account why production of the *wh*-phrase should occur at the clause boundary. If pronouncing it just helps to strengthen the memory representation of it, one might expect to find pronunciations of the *wh*-phrase in other locations, but both in our own data and others' data, we seem to largely only see pronunciations of the *wh*-phrase in places where there is a syntactic copy. This is expected on our account because that syntactic copy leads to a spike in activation from the spreading activation of the reactivated *wh*-feature.

Lastly, they criticize our account for only being able to account for productions with identical *wh*-phrases; as we've argued, we think our account can indeed account for both sorts of productions. Moreover, their own account explains productions with non-identical *wh*-phrases as a separate sort of performance error—namely, restart errors. Part of the evidence they provide in favor of this separate account is that there is some subject-auxiliary inversion in the embedded clauses. It may very well be true that some of these productions are restart errors, but we do not think all of them are, and our account offers an explanation of the ones that are not. For example, (17) is an example from our data set where the *wh*-phrases are not identical and there is no subject-auxiliary inversion in the embedded clause; thus, it is unlikely that this is a restart error.

(17) What do you think where the girl was walking? (Sophia 4;04,04)

Of the 11 non-subject questions in our data set with medial *wh*-phrases, where the *wh*-phrases were not identical, only two of them involved subject-auxiliary inversion in the embedded clause.[23] Our account still provides an explanation for these nine other utterances, but it's not immediately clear that their account does.

[22]Lutken & Legendre (2020) and Lutken (2021:256–260) do find a negative correlation between a composite working memory score and the production of medial *wh*-phrases of $r = -0.2$, which could be taken as further evidence in support of the account in Lutken, Legendre & Omaki (2020). However, although this correlation is significant, it is a relatively small correlation. Moreover, it could be that they see a correlation between working memory and medial *wh*-phrase production because the development of working memory is correlated with the development of cognitive inhibition (see, e.g., Carlson, Moses & Breton 2002). This would explain both our finding as well as theirs.

One possibility would be to redo our study but also include the working memory measure from Lutken & Legendre (2020) and Lutken (2021). If one still observed a correlation between cognitive inhibition and the production of medial *wh*-phrases in such a followup where working memory is controlled for in the statistical analysis, this would suggest that our account is correct and that the correlation with working memory arises because of an underlying correlation with cognitive inhibition. Another possibility for teasing these things apart is to upregulate cognitive inhibition on a trial-by-trial basis (cf. Hsu & Novick 2016); if one were to see fewer medial *wh*-phrases on trials where cognitive inhibition has been upregulated and more on trials where cognitive inhibition has not been upregulated, this would be consistent with our account but unexplained on the account from Lutken, Legendre & Omaki (2020).

[23]We do not consider subject questions here, since there would be no observable subject-auxiliary inversion in these questions anyway.

4 Conclusion

In this article, we have offered an inhibition-based account of children's medial *wh*-phrase productions based on an independent theory of sentence production and speech errors and independently motivated mechanisms in the performance systems for dealing with filler-gap dependencies (§2.1). We argued that our inhibition-based account is a better account than any of the non-target grammar accounts that have been proposed (e.g., Thornton 1990, Thornton & Crain 1994, McDaniel, Chiu & Maxfield 1995, Roeper & de Villiers 2011, de Villiers, de Villiers & Roeper 2011). As noted at the outset of the article, the non-target grammar accounts are interesting because they make non-trivial commitments to specific types of endogenous factors that drive grammatical feature selection. As such, such accounts should be subjected to careful scrutiny. We've argued that, upon careful scrutiny, the evidence in favor of these non-target grammar accounts is not as compelling as it might have initially appeared (§2.2.1), there are some non-trivial issues with non-target grammar accounts (§2.2.2), and, in fact, there is empirical evidence that favors our failure of inhibition account over any of the non-target grammar accounts (§3).

Moreover, we think these results highlight how important it is for those studying language acquisition to have a good understanding of (i) linguistic representations, (ii) properties of the extralinguistic performance systems that deploy these representations in real time, and (iii) aspects of the development of these extralinguistic performance systems. By incorporating an understanding of how extralinguistic cognitive mechanisms are engaged during sentence production/comprehension, acquisitionists will be in a better position to diagnose the grammatical competence that underpins linguistic performance. As we have seen here, sometimes children's errors may be a reflection not of their grammars but of how they put those grammars to use (see also Shipley, Smith & Gleitman 1969, Ibbotson & Kearvell-White 2015, Yuile & Sabbagh 2020). Whenever non-adult-like linguistic behavior can be attributed to aspects of extragrammatical performance systems, the theoretical burden of having to identify the factors that would lead children to acquire grammatical features for which they have no evidence and the factors that would drive them to purge those features from their linguistic system vanishes. That is to say, by treating children's production errors as the interaction between grammar and extralinguistic performance systems, the theory of how such errors arise and are purged from children's productions will reduce to the theory of how those extralinguistic systems develop in childhood.

Acknowledgments

We would like to thank all of the children who participated in this study as well as their parents/guardians. For helpful suggestions and feedback, we are grateful to the UMD acquisition lab as well as audiences at the 42nd Boston University Conference of Language Development and the 2019 Michigan State Undergraduate Linguistics Conference.

Funding

The first author was partially supported by an NSF-NRT grant (NSF: #1449815) while working on this project, and the second author was partially funded by the Brazilian National Council for Scientific and Technological Development (CNPq), grant #306462/2020-6. We are grateful for the support of both of these agencies.

Disclosure Statement

No potential conflict of interest was reported by the authors.

ORCID

Adam Liter http://orcid.org/0000-0002-1517-2900
Elaine Grolla http://orcid.org/0000-0001-8126-0493
Jeffrey Lidz http://orcid.org/0000-0001-8829-1495

References

Akaike, Hirotugu. 1973. Information theory as an extension of the maximum likelihood principle. In Boris N. Petrov & Frigyes Csaki (eds.), *Second International Symposium on Information Theory*, 267–281. Budapest, Hungary: Akademiai Kiado.

Anderson, Vicki A. 2002. Executive function in children: Introduction. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence* 8(2). 69–70. doi:10.1076/chin.8.2.69.8725

Anderson, Vicki A., Peter Anderson, Elisabeth Northam, Rani Jacobs & Cathy Catroppa. 2001. Development of executive functions through late childhood and adolescence in an Australian sample. *Developmental Neuropsychology* 20(1). 385–406. doi:10.1207/S15326942DN2001_5

Baayen, R. Harald, Douglas J. Davidson & Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4). 390–412. doi:10.1016/j.jml.2007.12.005

Badecker, William & Frantisek Kuminiak. 2007. Morphology, agreement and working memory retrieval in sentence production: Evidence from gender and case in Slovak. *Journal of Memory and Language* 56(1). 65–85. doi:10.1016/j.jml.2006.08.004

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278. doi:10.1016/j.jml.2012.11.001

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixedeffects models using lme4. *Journal of Statistical Software* 67(1). 1–48. doi:10.18637/jss.v067.i01

Bayer, Josef. 1996. *Directionality and logical form: On the scope of focusing particles and* wh-*in-situ* (Studies in Natural Language and Linguistic Theory 34). Dordrecht, The Netherlands: Springer. doi:10.1007/978-94-017-1272-9

Berwick, Robert C., Paul Pietroski, Beracah Yankama & Noam Chomsky. 2011. Poverty of the stimulus revisited. *Cognitive Science* 35(7). 1207–1242. doi:10.1111/j.1551-6709.2011.01189.x

Bloom, Paul. 1990. Subjectless sentences in child language. *Linguistic Inquiry* 21(4). 491–504.

Bock, Kathryn & Carol A. Miller. 1991. Broken agreement. *Cognitive Psychology* 23(1). 45–93. doi:10.1016/0010-0285(91)90003-7

Botwinik, Irena, Reem Bshara & Sharon Armon-Lotem. 2015. Children's production of relative clauses in Palestinian Arabic: Unique errors and their movement account. *Lingua* 156. 40–56. doi:10.1016/j.lingua.2014.10.007

Carlson, Stephanie M. & Louis J. Moses. 2001. Individual differences in inhibitory control and children's theory of mind. *Child Development* 72(4). 1032–1053. doi:10.1111/1467-8624.00333

Carlson, Stephanie M., Louis J. Moses & Casey Breton. 2002. How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development* 11(2). 73–92. doi:10.1002/icd.298

Choe, Jinsun & Kamil Deen. 2016. Children's difficulty with raising: A performance account. *Language Acquisition* 23 (2). 112–141. doi:10.1080/10489223.2015.1047097

Chomsky, Noam. 1973. Conditions on transformations. In Stephen R. Anderson & Paul Kiparsky (eds.), *A festschrift for Morris Halle*, 232–286. The Hague, The Netherlands: Mouton.

Conroy, Anastasia, Eri Takahashi, Jeffrey Lidz & Colin Phillips. 2009. Equal treatment for all antecedents: How children succeed with Principle B. *Linguistic Inquiry* 40(3). 446–486. doi:10.1162/ling.2009.40.3.446

Davidson, Matthew C., Dima Amso, Loren Cruess Anderson & Adele Diamond. 2006. Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia* 44(11). 2037–2078. doi:10.1016/j.neuropsychologia.2006.02.006

Dell, Gary S. 1986. A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93(3). 283–321. doi:10.1037/0033-295X.93.3.283

Demirdache, Hamida & Magda Oiry. 2008. On the syntax and semantics of LD questions in child French. In Anna Gavarró & M. João Freitas (eds.), *Language acquisition and development: Proceedings of GALA 2007*, 177–188. Newcastle, UK: Cambridge Scholars Publishing.

Dempster, Frank N. 1991. Inhibitory processes: A neglected dimension of intelligence. *Intelligence* 15(2). 157–173. doi:10.1016/0160-2896(91)90028-C

Dempster, Frank N. 1993. Resistance to interference: Developmental changes in a basic processing mechanism. In Mark L. Howe & Robert Pasnak (eds.), *Emerging themes in cognitive development*, vol. 1, chap. 1, 3–27. New York, NY: Springer. doi:10.1007/978-1-4613-9220-0_1

Diamond, Adele. 2002. Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In Donald T. Stuss & Robert T. Knight (eds.), *Principles of frontal lobe function*, 466–503. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780195134971.003.0029

Drenhaus, Heiner, Stefan Frisch & Douglas Saddy. 2005. Processing negative polarity items: When negation comes through the backdoor. In Stephan Kepser & Marga Reis (eds.), *Linguistic evidence: Empirical, theoretical and computational perspectives* (Studies in Generative Grammar 85), 145–164. Berlin, Germany: Mouton de Gruyter. doi:10.1515/9783110197549.145

Engelhardt, Paul E., Joel T. Nigg & Fernanda Ferreira. 2013. Is the fluency of language outputs related to individual differences in intelligence and executive function? *Acta Psychologica* 144(2). 424–432. doi:10.1016/j.actpsy.2013.08.002

Fadlon, Julie, Adam M. Morgan, Aya Meltzer-Asscher & Victor S. Ferreira. 2019. It depends: Optionality in the production of filler-gap dependencies. *Journal of Memory and Language* 106. 40–76. doi:10.1016/j.jml.2019.02.005

Felser, Claudia. 2004. *Wh*-copying, phases, and successive cyclicity. *Lingua* 114(5). 543–574. doi:10.1016/S0024-3841(03)00054-8

Franck, Julie, Gabriela Soare, Ulrich H. Frauenfelder & Luigi Rizzi. 2010. Object interference in subject–verb agreement: The role of intermediate traces of movement. *Journal of Memory and Language* 62(2). 166–182. doi:10.1016/j.jml.2009.11.001

Frazier, Lyn & Janet Dean Fodor. 1978. The sausage machine: A new two-stage parsing model. *Cognition* 6(4). 291–325. doi:10.1016/0010-0277(78)90002-1

Freund, Michael, Barry Gordon & Nazbanou Nozari. 2016. Conflict-based regulation of control in language production. In Anna Papafragou, Daniel Grodner, Daniel Mirman & John Trueswell (eds.), *Proceedings of the 38th annual conference of the Cognitive Science Society*, 1625–1630. Austin, TX: Cognitive Science Society.

Gibson, Edward & Tessa Warren. 2004. Reading-time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax* 7(1). 55–78. doi:10.1111/j.1368-0005.2004.00065.x

Grolla, Elaine & Jeffrey Lidz. 2018. A performance account for medial *wh*-questions in child English. In Anne B. Bertolini & Maxwell J. Kaplan (eds.), *Proceedings of the 42nd annual Boston University Conference on Language Development [BUCLD 42]*, 289–302. Somerville, MA: Cascadilla Press.

Guasti, Maria Teresa, Rosalind Thornton & Kenneth Wexler. 1995. Negation in children's questions: The case of English. In Dawn MacLaughlin & Susan McEwen (eds.), *Proceedings of the 19th annual Boston University Conference on Language Development [BUCLD 19]*, vol. 1, 228–239. Somerville, MA: Cascadilla Press.

Gutiérrez Mangado, Mara Juncal. 2006. Acquiring long-distance*wh*-questions in L1 Spanish: A longitudinal investigation. In Vincent Torrens & Linda Escobar (eds.), *The acquisition of syntax in Romance languages* (Language Acquisition and Language Disorders 41), 251–287. Amsterdam, The Netherlands: John Benjamins. doi:10.1075/lald.41.13gut

Hamburger, Henry & Stephen Crain. 1982. Relative acquisition. In Stanley A., Kuczaj II (ed.), *Language development: Syntax and semantics*, vol. 1, chap. 7, 245–274. Hillsdale, NJ: Lawrence Erlbaum.

Han, Chung-hye, Julien Musolino & Jeffrey Lidz. 2016. Endogenous sources of variation in language acquisition. *Proceedings of the National Academy of Sciences* 113(4). 942–947. doi:10.1073/pnas.1517094113

Harnishfeger, Katherine Kipp. 1995. The development of cognitive inhibition: Theories, definitions, and research evidence. In Frank N. Dempster & Charles J. Brainerd (eds.), *Interference and inhibition in cognition*, chap. 6, 175–204. San Diego, CA: Academic Press. doi:10.1016/B978-012208930-5/50007-6

Hedge, Craig, Georgina Powell & Petroc Sumner. 2018. The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods* 50(3). 1166–1186. doi:10.3758/s13428-017-0935-1

Horvath, Julia. 1997. The status of "*wh*-expletives" and the partial *wh*-movement construction of Hungarian. *Natural Language & Linguistic Theory* 15(3). 509–572. doi:10.1023/A:1005842214213

Hsu, Nina S. & Jared M. Novick. 2016. Dynamic engagement of cognitive control modulates recovery from misinterpretation during real-time language processing. *Psychological Science* 27(4). 572–582. doi:10.1177/0956797615625223

Huang, Yi Ting & Erin Hollister. 2019. Developmental parsing and linguistic knowledge: Reexamining the role of cognitive control in the kindergarten path effect. *Journal of Experimental Child Psychology* 184. 210–219. doi:10.1016/j.jecp.2019.04.005

Huttenlocher, Peter R. & Arun S. Dabholkar. 1997. Regional differences in synaptogenesisin human cerebral cortex. *The Journal of Comparative Neurology* 387(2). 167–178. doi:10.1002/(sici)1096-9861(19971020)387:2<167::aid<167::aid cne1>3.0.co;20.co;2z

Hyams, Nina. 1983. *The acquisition of parameterized grammars*. New York: The City University of New York dissertation.

Hyams, Nina. 1986. *Language acquisition and the theory of parameters* (Studies in Theoretical Psycholinguistics 3). Dordrecht, The Netherlands: D. Reidel. doi:10.1007/978-94-009-4638-5

Ibbotson, Paul & Jennifer Kearvell-White. 2015. Inhibitory control predicts grammatical ability. *PLoS ONE* 10(12). e0145030. doi:10.1371/journal.pone.0145030

Jakubowicz, Celia & Nelleke Strik. 2008. Scope-marking strategies in the acquisition of long distance *wh*-questions in French and Dutch. *Language and Speech* 51(1–2). 101–132. doi:10.1177/00238309080510010701

Kampen, Jacqueline van. 1997. *First steps in* wh-*movement*. Utrecht, The Netherlands: Rijksuniversiteit dissertation.

Kipp, Katherine & Steffen Pope. 1997. The development of cognitive inhibition in streams-of-consciousness and directed speech. *Cognitive Development* 12(2). 239–260. doi:10.1016/S0885-2014(97)90015-0

Labelle, Marie. 1990. Predication, *wh*-movement, and the development of relative clauses. *Language Acquisition* 1(1). 95–119. doi:10.1207/s15327817la0101_4

Lezak, Muriel Deutsch. 1995. *Neuropsychological assessment*, 3rd edn. Oxford: Oxford University Press.

Luciana, Monica, Heather M. Conklin, Catalina J. Hooper & Rebecca S. Yarger. 2005. The development of nonverbal working memory and executive control processes in adolescents. *Child Development* 76(3). 697–712. doi:10.1111/j.1467-8624.2005.00872.x

Lutken, Carolyn Jane. 2021. *Cross-linguistic investigations of syntactic creativity errors in children's* wh-*questions*. Baltimore, MD: Johns Hopkins University dissertation.

Lutken, C. Jane & Géraldine Legendre. 2020. Immature syntax or processing? What causes "scope marking errors" in English-speaking 5-year-olds? Poster presented at the 45th Annual Boston University Conference on Language Development [BUCLD 45], November 5–9, Boston University.

Lutken, C. Jane, Géraldine Legendre & Akira Omaki. 2020. Syntactic creativity errors in children's *wh*-questions. *Cognitive Science* 44(7). doi:10.1111/cogs.12849

Mazuka, Reiko, Nobuyuki Jincho & Hiroaki Oishi. 2009. Development of executive control and language processing. *Language and Linguistics Compass* 3(1). 58–89. doi:10.1111/j.1749-818X.2008.00102.x

McDaniel, Dana. 1986. *Conditions on* wh-*chains*. New York: The City University of New York dissertation.

McDaniel, Dana. 1989. Partial and multiple *wh*-movement. *Natural Language & Linguistic Theory* 7(4). 565–604. doi:10.1007/BF00205158

McDaniel, Dana, Bonnie Chiu & Thomas L. Maxfield. 1995. Parameters for *wh*-movement types: Evidence from child English. *Natural Language & Linguistic Theory* 13(4). 709–753. doi:10.1007/BF00992856

McDaniel, Dana, Cecile McKee, Wayne Cowart & Merrill F. Garrett. 2015. The role of the language production system in shaping grammars. *Language* 91(2). 415–441. doi:10.1353/lan.2015.0021

McDaniel, Dana, Cecile McKee & Merrill F. Garrett. 2010. Children's sentence planning: Syntactic correlates of fluency variations. *Journal of Child Language* 37(1). 59–94. doi:10.1017/S0305000909009507

McKee, Cecile & Dana McDaniel. 2001. Resumptive pronouns in English relative clauses. *Language Acquisition* 9(2). 113–156. doi:10.1207/S15327817LA0902_01

Miller, Earl K. & Jonathan D. Cohen. 2001. An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience* 24. 167–202. doi:10.1146/annurev.neuro.24.1.167

Miyake, Akira, Naomi P. Friedman, Michael J. Emerson, Alexander H. Witzki, Amy Howerter & Tor D. Wager. 2000. The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology* 41(1). 49–100. doi:10.1006/cogp.1999.0734

Momma, Shota & Colin Phillips. 2018. The relationship between parsing and generation. *Annual Review of Linguistics* 4. 233–254. doi:10.1146/annurev-linguistics-011817-045719

Novick, Jared M., Erika Hussey, Susan Teubner-Rhodes, J. Isaiah Harbison & Michael F. Bunting. 2014. Clearing the garden-path: Improving sentence processing through cognitive control training. *Language, Cognition and Neuroscience* 29(2). 186–217. doi:10.1080/01690965.2012.758297

Nozari, Nazbanou & Jared Novick. 2017. Monitoring and control in language production. *Current Directions in Psychological Science* 26(5). 403–410. doi:10.1177/0963721417702419

Nozari, Nazbanou & Akira Omaki. 2018. Syntactic production is not independent of inhibitory control: Evidence from agreement attraction errors. In Timothy T. Rogers, Martina Rau, Jerry Zhu & Charles W. Kalis (eds.), *Proceedings of the 40th annual conference of the Cognitive Science Society*, 822–827. Austin, TX: Cognitive Science Society.

Oiry, Magda. 2006. Direct versus indirect *wh*-scope marking strategies in French child grammar. *University of Massachussetts Occasional Papers in Linguistics*.

Pierce, Amy E. 1992. *Language acquisition and syntactic theory: A comparative analysis of French and English child grammars* (Studies in Theoretical Psycholinguistics 14). Dordrecht, The Netherlands: Kluwer Academic Publishers. doi:10.1007/978 - 94 -011-2574-1

Peirce, Jonathan W. 2007. PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods* 162(1–2). 8–13. doi:10.1016/j.jneumeth.2006.11.017

Peirce, Jonathan W. 2009. Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics* 2(10). doi:10.3389/neuro.11.010.2008

Peirce, Jonathan, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman & Jonas Kristoffer Lindeløv. 2019. Psychopy2: Experiments in behavior made easy. *Behavior Research Methods* 51. 195–203. doi:10.3758/s13428-018-01193-y

Plessis, Hans du. 1977. *Wh* movement in Afrikaans. *Linguistic Inquiry* 8(4). 723–726.

Qi, Zhengan, Jessica Love, Cynthia Fisher & Sarah Brown-Schmidt. 2020. Referential context and executive functioning influence children's resolution of syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 46(10). 1922–1947. doi:10.1037/xlm0000886

R Core Team. 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/.

Rizzi, Luigi. 1993. Some notes on linguistic theory and language development: The case of root infinitives. *Language Acquisition* 3(4). 371–393. doi:10.1207/s15327817la0304_2

Roeper, Thomas & Edward Matthei. 1975. On the acquisition of "some" and "all." In *Papers and reports on child language development*, 63–74. Stanford, CA: Stanford University.

Roeper, Tom & Jill G. de Villiers. 2011. The acquisition path for *wh*-questions. In Jill G. de Villiers & Tom Roeper (eds.), *Handbook of generative approaches to language acquisition* (Studies in Theoretical Psycholinguistics 41), 189–246. Dordrecht, The Netherlands: Springer. doi:10.1007/978-94-007-1688-9_6

Scontras, Gregory, William Badecker & Evelina Fedorenko. 2017. Syntactic complexity effects in sentence production: A reply to MacDonald, Montag, and Gennari (2016). *Cognitive Science* 41(8). 2280–2287. doi:10.1111/cogs.12495

Scontras, Gregory, William Badecker, Lisa Shank, Eunice Lim & Evelina Fedorenko. 2015. Syntactic complexity effects in sentence production. *Cognitive Science* 39(3). 559–583. doi:10.1111/cogs.12168

Shao, Zeshu, Antje S. Meyer & Ardi Roelofs. 2013. Selective and nonselective inhibition of competitors in picture naming. *Memory & Cognition* 41. 1200–1211. doi:10.3758/s13421-013-0332-7

Shipley, Elizabeth F., Carlota S. Smith & Lila R. Gleitman. 1969. A study in the acquisition of language: Free responses to commands. *Language* 45(2). 322–342. doi:10.2307/411663

Simon, J. Richard. 1969. Reactions toward the source of stimulation. *Journal of Experimental Psychology* 81(1). 174–176. doi:10.1037/h0027448

Simon, J. Richard & Alan P. Rudell. 1967. Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology* 51(3). 300–304. doi:10.1037/h0020586

Stromswold, Karin J. 1990. *Learnability and the acquisition of auxiliaries*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Thornton, Rosalind. 1990. *Adventures in long-distance moving: The acquisition of complex* wh-*questions*. Storrs, CT: University of Connecticut dissertation.

Thornton, Rosalind. 1995. Referentiality and *wh*-movement in child English: Juvenile D*Link*uency. *Language Acquisition* 4(1–2). 139–175. doi:10.1080/10489223.1995.9671662

Thornton, Rosalind & Stephen Crain. 1994. Successful cyclic movement. In Teun Hoekstra & Bonnie D. Schwartz (eds.), *Language acquisition studies in generative grammar: Papers in honor of Kenneth Wexler from the 1991 GLOW workshops* (Language Acquisition and Language Disorders 8), 215–252. Amsterdam, The Netherlands: John Benjamins. doi:10.1075/lald.8.11tho

Thothathiri, Malathi, Christine T. Asaro, Nina S. Hsu & Jared M. Novick. 2018. *Who* did *what*? A causal role for cognitive control in thematic role assignment during sentence comprehension. *Cognition* 178. 162–177. doi:10.1016/j.cognition.2018.05.014

Tomasello, Michael. 2000. Do young children have adult syntactic competence? *Cognition* 74(3). 209–253. doi:10.1016/S0010-0277(99)00069-4

Torrego, Esther. 1983. More effects of successive cyclic movement. *Linguistic Inquiry* 14(3). 561–565.

Trude, Alison M. & Nazbanou Nozari. 2017. Inhibitory control supports referential context use in language production and comprehension. In Glenn Gunzelmann, Andrew Howes, Thora Tenbrink & Eddy Davelaar (eds.), *Proceedings of the 39th annual meeting of the Cognitive Science Society*, 1218–1223. Austin, TX: Cognitive Science Society.

Villiers, Jill G. de, Peter A. de Villiers & Thomas Roeper. 2011. *Wh*-questions: moving beyond the first phase. *Lingua* 121 (3). 352–366. doi:10.1016/j.lingua.2010.10.003

Villiers, Jill G. de, Jessica Kotfila & Madeline Klein. 2019. Parsing, pragmatics, and representation: Children's comprehension of two-clause questions. In Tania Ionin & Matthew Rispoli (eds.), *Three streams of generative language acquisition research: Selected papers from the 7th meeting of Generative Approaches to Language Acquisition—North America, University of Illinois at Urbana–Champaign* (Language Acquisition and Language Disorders 63), 85–105. Amsterdam, The Netherlands: John Benjamins. doi:10.1075/lald.63.05vil

Villiers, Jill G. de & Thomas Roeper. 1995. Relative clauses are barriers to *wh*-movement for young children. *Journal of Child Language* 22(2). 389–404. doi:10.1017/S0305000900009843

Villiers, Jill G. de, Thomas Roeper, Linda Bland-Stewart & Barbara Pearson. 2008. Answering hard questions: *Wh*-movement across dialects and disorder. *Applied Psycholinguistics* 29(1). 67–103. doi:10.1017/S0142716408080041

Villiers, Jill G. de, Thomas Roeper & Anne Vainikka. 1990. The acquisition of long-distance rules. In Lyn Frazier & Jill G. de Villiers (eds.), *Language processing and language acquisition* (Studies in Theoretical Psycholinguistics 10), 257–297. Dordrecht, The Netherlands: Kluwer Academic Publishers. doi:10.1007/978-94-011-3808-6_10

Wagers, Matthew W. & Colin Phillips. 2014. Going the distance: Memory and control processes in active dependency construction. *The Quarterly Journal of Experimental Psychology* 67(7). 1274–1304. doi:10.1080/17470218.2013.858363

Wahba, Wafaa Abdel-Faheem Batran. 1991. LF movement in Iraqi Arabic. In Cheng-the James Huang & Robert May (eds.), *Logical structure and linguistic structure: Cross-linguistic perspectives* (Studies in Linguistics and Philosophy 40), 253–276. Dordrecht, The Netherlands: Kluwer Academic Publishers. doi:10.1007/978-94-011-3472-9_9

Welsh, Marilyn C., Bruce F. Pennington & Dena B. Groisser. 1991. A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology* 7(2). 131–149. doi:10.1080/87565649109540483

Woodard, Kristina, Lila R. Gleitman & John C. Trueswell. 2016. Two- and three-year-olds track a single meaning during word learning: Evidence for propose-but-verify. *Language Learning and Development* 12(3). 252–261. doi:10.1080/15475441.2016.1140581

Yuile, Amanda Rose & Mark A. Sabbagh. 2020. Inhibitory control and preschoolers' use of irregular past tense verbs. *Journal of Child Language* 48(3). 480–498. doi:10.1017/S0305000920000355

Appendix: Elicited production materials

The materials for the elicited production task are reported in this appendix. There were six videos that the child and experimenter watched together from the Brazilian animated cartoon show, Turma da Mônica 'Monica's Gang'. The clips lacked dialogue. In what follows, we briefly describe the contents of each clip before giving the prompts for the questions we tried to elicit after each clip.

The first video involved a top hat that was moving around on the ground. A boy went to investigate what was causing the top hat to move. After picking it up, it is revealed that Mônica was under the top hat (and she was holding a bunny stuffed animal). She then kisses the boy on the cheek. The following prompts were used to elicit the following three intended questions.

(18) Prompt:We know that it was a girl under the top hat, but let's ask Snuggles what he thinks.
Intended:What do you think was under the top hat?
(19) Prompt:We know that it was a bunny that the girl was holding, but let's ask Snuggles what he thinks.
Intended:What do you think the girl was holding?
(20) Prompt:We know that it was the girl that kissed the boy, but let's ask Snuggles who he thinks.
Intended:Who do you think kissed the boy?

The second video involved Mônica and a friend eating watermelon together. A boy in yellow steals Mônica's bunny and runs away, with Mônica chasing him. He proceeds to kick the bunny into the air, where it is caught by a boy in green, floating by on balloons. Mônica's friend spits watermelon seeds at the balloons, popping them, the boy in green falls to the ground, and Mônica retrieves her bunny. The following prompts were used to elicit the following four intended questions.

(21) Prompt:We know that it was the watermelon that the girls were eating, but let's ask Snuggles what he thinks.
Intended:What do you think the girls were eating?
(22) Prompt:We know that it was the bunny that the yellow boy kicked, but let's ask Snuggles what he thinks.
Intended:What do you think the yellow boy kicked?
(23) Prompt:We know that it was with balloons that the green boy flew, but let's ask Snuggles how he thinks.
Intended:How do you think the boy flew?
(24) Prompt:We know that it was with watermelon seeds that the girl popped the balloon, but let's ask Snuggles how he thinks.
Intended:How do you think the girl popped the balloons?

The third video involved the boy in green and the boy in yellow fleeing from Mônica, who is chasing them across a tightrope because they have stolen her bunny. The boys reach the platform at the end of the tightrope before Mônica, where they proceed to use a pair of scissors to cut the tightrope, which would putatively cause Mônica to fall to the ground. They cut the rope, but it turns out that the platform falls to the ground, while the rope remains suspended in the air. As the boys fall the ground, Mônica grabs her bunny back from them. The following prompts were used to elicit the following four intended questions.

(25) Prompt:We know that it was on the rope that the kids were walking, but let's ask Snuggles where he thinks.
Intended:Where do you think the kids were walking?
(26) Prompt:We know that it was the girl that was chasing the boys, but let's ask Snuggles who he thinks.
Intended:Who do you think was chasing the boys?
(27) Prompt:We know that it was with scissors that the boys cut the rope, but let's ask Snuggles how he thinks.
Intended:How do you think the boys cut the rope?
(28) Prompt:We know that it was the boys that fell, but let's ask Snuggles which kids he thinks.
Intended:Which kids do you think fell?

The fourth video involved Mônica chasing the boy in green and the boy in yellow. She chases them to the edge of a cliff with water below. The boy in yellow grabs Mônica's bunny from her and whips it around like a propeller in order to fly away and escape. The boy in green jumps into the water below in order to hide from Mônica, but Mônica gets a fishing pole and fishes him out of the water. Mônica walks away with the boy in green hanging on the fishing line, and the boy in yellow is seen flying in the sky above with her bunny as the clip ends. The following four prompts were used to elicit the following four intended questions.

(29) Prompt:We know that it was the yellow boy that flew in the sky, but let's ask Snuggles who he thinks.
Intended:Who do you think flew in the sky?

(30) Prompt:We know that it was in the water that the green boy tried to hide, but let's ask Snuggles where he thinks.
Intended:Where do you think the green boy tried to hide?

(31) Prompt:We know that it was the green boy that jumped in the water, but let's ask Snuggles which kid he thinks.
Intended:Which kid do you think jumped in the water?

(32) Prompt:We know that it was the girl that got the green boy out of the water, but let's ask Snuggles which kid he thinks.
Intended:Which kid do you think got the green boy out of the water?

The fifth video involved the boy in green and the boy in yellow playing together with a soccer ball. Mônica tries to join them, but the two boys don't want to let her play. Mônica ends up stealing the ball from them and demonstrating that she is really good at soccer, much better than the two of them. The clip ends with Mônica kicking the ball and hitting the boy in green with the ball. The following four prompts were used to elicit the following four intended questions.

(33) Prompt:We know that it was a ball that the boys were playing with, but let's ask Snuggles which toy he thinks.
Intended:Which toy do you think the boys were playing with?

(34) Prompt:We know that it was soccer that the boys were playing, but let's ask Snuggles which game he thinks.
Intended:Which game do you think the boys were playing?

(35) Prompt:We know that it was the girl that was really good at soccer, but let's ask Snuggles who he thinks.
Intended:Who do you think was really good at soccer?

(36) Prompt:We know that it was the green boy that the girl hit, but let's ask Snuggles which boy he thinks.
Intended:Which boy do you think the girl hit?

The sixth video involved a boy looking through a hole in a fence. He sees something that he thinks is funny. It is revealed that Mônica was behind the fence, as she knocks over the fence on top of the boy and walks away. The following two prompts were used to elicit the following two intended questions.

(37) Prompt:We know that it was a girl that the boy saw, but let's ask Snuggles who he thinks.
Intended:Who do you think the boy saw?

(38) Prompt:We know that the it was behind the fence that the boy saw the girl, but let's ask Snuggles where he thinks
Intended:Where do you think the boy saw the girl?