Cognitive Science 32 (2008) 222–255

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1080/03640210701703766

Is Structure Dependence an Innate Constraint? New Experimental Evidence From Children's Complex-Question Production

Ben Ambridge, Caroline F. Rowland, Julian M. Pine

Department of Psychology, University of Liverpool, UK

Received 30 March 2006; received in revised form 26 January 2007; accepted 26 January 2007

Abstract

According to Crain and Nakayama (1987), when forming complex yes/no questions, children do not make errors such as *Is the boy who smoking is crazy?* because they have innate knowledge of *structure dependence* and so will not move the auxiliary from the relative clause. However, simple recurrent networks are also able to avoid such errors, on the basis of surface distributional properties of the input (Lewis & Elman, 2001; Reali & Christiansen, 2005). Two new elicited production studies revealed that (a) children occasionally produce structure-dependence errors and (b) the pattern of children's auxiliary-doubling errors (*Is the boy who is smoking is crazy?*) suggests a sensitivity to surface co-occurrence patterns in the input. This article concludes that current data do not provide any support for the claim that structure dependence is an innate constraint, and that it is possible that children form a structure-dependent grammar on the basis of exposure to input that exhibits this property.

Keywords: Structure dependence; Yes/no questions; Complex syntax; Poverty of stimulus argument; Language acquisition; Relative clause

A central question in the cognitive sciences is how children come to acquire their native language. Dating back to Chomsky (1959), a popular view has been that children "attain knowledge of the structure of their language for which no evidence is available in the data to which they are exposed as children" (Hornstein & Lightfoot, 1981, p. 9) or, as in the title of a recent article by Lidz, Waxman, and Freedman (2003), that "infants know [facts] about syntax that they couldn't have learned" (p. B65), and hence that certain syntactic principles or constraints must be innately specified in the genome.

Perhaps the "parade case" (Crain, 1991) of such a constraint is that of *structure dependence*; the principle that grammatical rules apply not to items that occupy a certain position in the

Correspondence should be addressed to Ben Ambridge, Department of Psychology, University of Liverpool, Bedford Street South, Liverpool, L69 7ZA, UK. E-mail: ben.ambridge@liverpool.ac.uk

sentence with respect to linear order, but to items that play a specific syntactic role with respect to the hierarchical constituent-structure of the sentence. The example most often used to illustrate this principle is that of yes/no questions (e.g., see Piatelli-Palmerini, 1980, p. 40), which, by hypothesis, are formed by "moving" elements of an untransformed version of the utterance (approximating to a declarative statement). On the basis of simple questions (1), the child could hypothetically formulate a rule such as *move the first auxiliary*[1] *to sentence initial-position*:

1a. The boy is crazy → 1b. Is the boy crazy?

This rule, however, would output an ungrammatical utterance for questions with more than one auxiliary ("complex questions"):

2a. The boy who is smoking is crazy → 2b. *Is the boy who smoking is crazy?

In order to form such questions correctly, the child must formulate a structure-dependent rule (e.g., move the auxiliary *in the main clause* to sentence-initial position), rather than a rule based on the linear position of items in the sentence. This rule is structure dependent because it makes reference to the notion of "main clause"—a constituent of the hierarchical sentence structure—and not simply to items that occupy a certain position based on linear order.[2]

According to Chomsky (1980; see also Crain & Nakayama, 1987; and many others), the input does not contain sufficient evidence that the structure-dependent generalization is the correct one because "a person might go through much or all of his life without ever having been exposed to relevant evidence" [presumably questions with two auxiliaries] (pp. 114–115). Chomsky (1980) argued that because the child will "nevertheless unerringly employ the structure-dependent generalization on the first relevant occasion" (p. 145), she must be innately constrained to consider only structure-dependent rules when formulating her grammar.

The claim that complex yes/no questions are extremely rare in child-directed speech is well supported. In a search of approximately 3 million caregiver utterances in the CHILDES database, MacWhinney (2004) found only one complex yes/no question. Thus, unlike some authors (e.g., Pullum & Scholz, 2002), our goal in this article is not to dispute the first part of Chomsky's (1980) claim; that children virtually never hear complex yes/no questions. Rather, our goal is to reassess the second part of this claim—that children never make structure-dependence errors—in the light of some recent findings from the field of computer science. In doing so, we will also address the third part of Chomsky's (1980) claim; that the (putative) absence of such errors necessarily implies that structure dependence is an innately specified constraint.

There are two possible sources of evidence to support the second part of Chomsky's (1980) claim—that children never make structure-dependence errors: naturalistic spontaneous production data and experimental data. Although no structure-dependence errors are recorded in the entire CHILDES database, neither is there a single instance of a child producing a well-formed complex yes/no question (B. MacWhinney, 2004). Because we have no spontaneous production data in which children are actually attempting to produce complex yes/no questions, naturalistic data provide no evidence either for or against the claim that children "unerringly employ the structure-dependent generalization on the first relevant occasion" (Chomsky, 1980, p. 145).

Given the absence of appropriate contexts (i.e., attempts at complex yes/no questions) in the available naturalistic data, the claim that children never make structure-dependence errors rests critically on a single experimental study. Crain and Nakayama (1987) used an elicited-production paradigm to elicit a range of complex yes/no questions with two instances of copula/auxiliary BE (Experiment 1), or with one instance of copula *is* and one modal auxiliary (Experiment 2), from 30 children aged 3;2 through 4;7. Although overall rates of correct performance were only 60% and 21% for Experiments 1 and 2, respectively, no child made a structure-dependence error (e.g., see Question 2b). However, errors of auxiliary-doubling (e.g., *Is the boy who is smoking is crazy?*) and resumption (e.g., *Is the boy who is smoking, is he crazy?*) were common. Crain and Nakayama (1987, p. 142) therefore concur with Chomsky (1980) that "only structure-dependent rules are formulated in language acquisition."

There are, however, three features of the design of Crain and Nakayama's (1987) study that would seem to work against the possibility of observing structure-dependence errors. First, the study contained relatively few test trials. Second, the majority of the target questions used the auxiliary with which children are probably most competent at forming questions (*is*), perhaps reducing overall error rates. Third, the target questions were such that, in order to make structure dependence errors, children would have had to strongly violate the co-occurrence statistics of English.

With regard to the number of test trials, it is important to note that Crain and Nakayama's (1987) claim regarding the innateness of structure dependence depends crucially on the finding of no structure-dependence errors. In order to make a theoretical claim on the basis of a null effect, it is necessary to demonstrate that the phenomenon in question was not observed, despite a large number of trials having been conducted. Crain and Nakayama elicited six sentences from each of 30 participants (180 trials, Experiment 1) and four sentences from each of 10 participants (40 trials, Experiment 2). This means that if structure-dependence errors do occur, but only relatively rarely, say 5%, one would expect only about 10 errors across both studies even if there were no other factors in the studies militating against the child producing such errors (see below).

The majority of test sentences elicited by Crain and Nakayama (1987) included two instances of copula/auxiliary *is*. This might be expected to reduce overall error rates (and hence the likelihood of observing structure dependence errors), as young children seem to be relatively competent at forming questions with this auxiliary. The test sentences elicited in Crain and Nakayama's Experiment 1 are reproduced in 3 through 8 (from p. 528):

3. The dog that is sleeping is on the blue bench → Is the dog that is sleeping on the blue bench?
4. The ball that the girl is sitting on is big → Is the ball that the girl is sitting on big?
5. The boy who is watching Mickey Mouse is happy → Is the boy who is watching Mickey Mouse happy?
6. The boy who is unhappy is watching Mickey Mouse → Is the boy who is unhappy watching Mickey Mouse?

7. The boy who is being kissed by his mother is happy → Is the boy who is being kissed by his mother happy?
8. The boy who was holding the plate is crying → Is the boy who was holding the plate crying?

With the exception of (8), all these target sentences included two instances of copula/auxiliary *is*. This is potentially problematic because both naturalistic (e.g., Labov & Labov, 1978; Maratsos & Kuczaj, 1978; Rowland, 2007; Rowland, Pine, Lieven, & Theakston, 2005; Stromswold, 1990) and experimental (Ambridge, Rowland, Theakston, & Tomasello, 2006; Santelmann, Berk, Austin, Somashekar, & Lust, 2002; Valian & Casey, 2003) investigations of children's simple question production have found that *BE* (and in particular auxiliary *is*) is generally associated with lower error rates than auxiliaries such as *DO, CAN,* or *SHOULD* (although Santelmann et al. reported high error rates for copula *BE*). Ambridge et al. also found a trend toward a higher error rate for the plural (*are*) than the singular (*is*) form of auxiliary BE. This raises the possibility that structure-dependence errors might have been observed in the study of Crain and Nakayama (1987) had these investigators elicited questions with auxiliaries such as *are*, *can, should,* and *do* with which children are less proficient at forming even simple questions.

The third potential shortcoming of the study of Crain and Nakayama (1987) is that the target sentences were such that, in order to produce a structure-dependence error, children would have had to have been willing to produce strings that strongly violate their knowledge of the co-occurrence statistics of English; that is, knowledge of which words (or syntactic categories) may follow one another (e.g., *relativizer + auxiliary*) and which (on the whole) may not (e.g., *relativizer + participle*). This point is best illustrated by the computer simulations of Lewis and Elman (2001) and Reali and Christiansen (2005).

Lewis and Elman (2001) trained a simple recurrent network on a word prediction task: The task of the network at each step was to predict the word that would be input at the next step. More important, the input to the network consisted merely of strings of words. No additional information about either the words themselves (e.g., grammatical category, semantics) or the grammatical structure of the strings was given. In a training phase, the network was presented with strings corresponding to questions containing a single auxiliary (e.g., *Is Mummy beautiful?*) and declarative statements (e.g., *Mummy is beautiful*). Around 2% of declarative statements contained a relative clause (e.g., *The boy who is smoking is crazy*). Crucially, however, none of the questions presented to the network contained a relative clause. That is, the network did not see any questions such as *Is the boy who is smoking crazy?* that contain more than one auxiliary.

At test, the model was presented with successive words of complex questions such as *Is the boy who is smoking crazy?* with the task of predicting, at each pass, the next word in the string. Fig. 1 (reproduced from Lewis & Elman's, 2001, Fig. 3) summarizes the model's predictions for successive words of the question *Is the boy who is smoking crazy?* The target words are shown from left to right. The column corresponding to each target word displays the average activation value for each of the syntactic categories shown. Following the initial *is*, for example, the model predicts either a name/pronoun (e.g., *Mummy*) or a determiner (e.g., *the*). It is important to note, however, that the input to the model was not tagged with syntactic

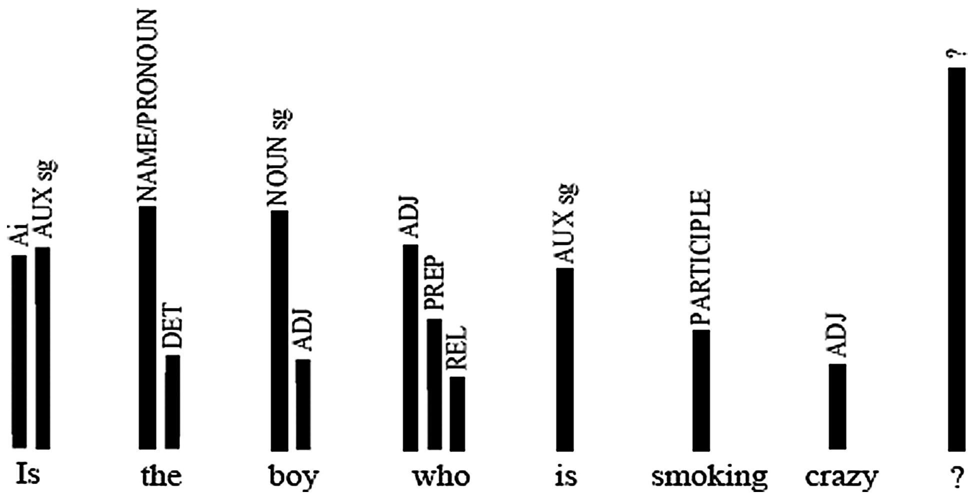

Fig. 1. Lewis and Elman's (2001) simple recurrent network's predictions for successive words of the question *Is the boy who is smoking crazy?* (reproduced from Lewis & Elman, 2001).

category information. The activation value shown for "DET," for example, does not represent the activation of a category of determiners—the model was given no such information—but the average activation across all items that would be classified as such in the adult grammar (the other category label whose referent may be unclear is REL for *relativizer*, e.g., *who*).

As Fig. 1 shows, the simple recurrent network (SRN) did not make predictions corresponding to structure-dependence errors: After a string such as *Is the boy who . . .* the model consistently predicted an auxiliary (e.g., *is*) and not a participle (e.g., *smoking*). Thus although the model had never encountered any complex yes/no questions, it was able to avoid structure dependence errors and, indeed, to make predictions corresponding to well-formed complex questions.

The input to Lewis and Elman's (2001) model was not real speech but a toy-grammar designed to "crudely approximat[e] the structure of child directed speech" (p. 4). Reali and Christiansen (2005; Experiment 3) replicated the SRN study of Lewis and Elman, using a small and noisy corpus of child-directed speech (Bernstein-Ratner, 1984), tagged for lexical category.[3] More important, this corpus contained no examples of complex questions. Nevertheless Reali and Christiansen found that the network predicted the correct form of complex questions over the corresponding structure-dependence error for 27 out of 30 test pairs.

In another study, Reali and Christiansen (2005; Experiment 1) used the transitional probability statistics of individual pairs and triplets (*bi-/tri-grams*) of untagged words from this corpus to calculate a measure of probability of occurrence (*cross-entropy*) for correctly formed and erroneous complex questions. On the basis of this corpus data (which contains no examples of this structure) correctly formed complex questions were shown to be statistically more probable than corresponding structure-dependence errors for 96 of 100 test sentence pairs and (Reali & Christiansen's Experiment 2) for each of the six sentences in Crain and Nakayama's (1987) Experiment 1 (using either bi- or tri-gram statistics). Together the studies of Lewis and Elman (2001) and Reali and Christiansen demonstrate that even when complex yes/no questions are not present in the input, both local and long-distance categorical co-occurrence statistics make structure-dependence errors unlikely to occur (although it should be acknowledged that these error-rates, 10% and 4%, are higher than one might expect, given the results of Crain & Nakayama, 1987; see also Kam, Stoyneshka, Tornyova, Fodor, & Sakas, 2005).

The studies of both Lewis and Elman (2001) and Reali and Christiansen (2005) highlight a feature of the test sentences used by Crain and Nakayama (1987) that might be expected to render the commission of structure-dependence errors particularly unlikely. An important reason why the SRN simulations discussed do not predict structure-dependence errors is that these errors contain bi-grams such as *who smoking* that virtually never occur in the input. The same is true of the target sentences in Crain and Nakayama's Experiment 1. In order to produce an error, children would have had to produce strings such as *that sleeping* or *who unhappy,* which occur with almost zero frequency in the input (at either the lexical level or the level of syntactic categories: *relativizer* + *particple/adjective*). It is possible, then, that structure-dependence errors would have been observed in the study of Crain and Nakayama had these authors elicited questions for which such errors contain strings that are less inconsistent with the bi-gram statistics of English.

Structure-dependence errors with modal auxiliaries such as CAN (9) or SHOULD are less inconsistent with the bi-gram statistics of English than formally equivalent errors with auxiliary BE (10):

9. *The boy who can run fast can jump high* → **Can the boy who run fast can jump high?*
10. The*boy who is running fast is jumping high* →**Is the boy who running fast is jumping high?*

This is because whilst structure-dependence errors with BE contain an extremely low frequency bi-gram (e.g., *who running*), formally equivalent errors with modal auxiliaries (e.g., CAN) contain a bi-gram (e.g., *who run*), which is likely to have occurred in the input with far greater frequency, as part of sentences with plural subjects (e.g., *Boys who run fast are popular*).[4] Thus Crain and Nakayama's (1987, Experiment 1) use of copula/auxiliary BE as opposed to modal auxiliaries would appear to render the commission of structure-dependence errors particularly unlikely.

In a second experiment, Crain and Nakayama (1987) did elicit questions containing the modal auxiliaries *can* and *should*. However, the nature of these questions (in particular the fact that they also contained one instance of copula *is*) meant that the relevant bi-grams (*who happy, who unhappy, who working*) were—except for the case of *who see*—of extremely low frequency. The fact that these questions contained two different auxiliaries (one instance of copula *is* and one modal) is also likely to have worked against the possibility of structure dependence errors being committed. To see why this is the case, consider the theoretically-possible structure-dependence errors shown below (derived from the sentences in Crain & Nakayama, 1987, p. 535):

11. The boy who is happy can see Mickey Mouse →*Is the boy who happy can see Mickey Mouse?

12. The boy who is unhappy should fix his TV set →*Is the boy who unhappy should fix his TV set?
13. The boy who can see Mickey Mouse is happy →*Can the boy who see Mickey Mouse is happy?
14. The boy who should be working is asleep →*Should the boy who be working is asleep?

In Crain and Nakayama's (1987) study, children were given a declarative statement (e.g., *the boy who is watching Mickey Mouse is happy* or *the boy who can see Mickey Mouse is happy*) in an indirect question frame (*I wonder if* . . .), with the task of producing utterances that question the truth value of these statements (e.g., *Is the boy who is watching Mickey Mouse happy?* or *Is the boy who can see Mickey Mouse happy?*). On the assumption that children understand the task, structure-dependence errors with two instances of the same auxiliary (e.g., *Is the boy who watching Mickey Mouse is happy?*) may be more likely to be produced than structure-dependence errors with different auxiliaries (e.g., *Can the boy who see Mickey Mouse is happy?*) as the former but not the latter can reasonably be construed as questioning the truth value of the declarative statement presented in the experimenter's indirect question frame. That is, only errors of the former type can reasonably be interpreted as utterances that are consistent with the semantics of the target question. Thus Crain and Nakayama's decision to use two different auxiliaries in Experiment 2 may have reduced the likelihood of structure-dependence errors being observed.

In summary, despite the widely held view to the contrary, evidence that children never produce structure-dependence errors, producing a correctly formed complex yes/no question "on the first relevant occasion" (Chomsky, 1980, p. 145) is extremely thin. Currently available naturalistic data are irrelevant to the argument, as even the largest corpora do not include any attempts at the relevant structure. The only evidence in support of this claim comes from a single experimental study (Crain & Nakayama, 1987) the design of which rendered the observation of structure-dependence errors particularly unlikely. Only six (Experiment 1) and four questions (Experiment 2) were elicited from 30 and 10 children, respectively. Experiment 1 used exclusively copula/auxiliary BE (in the form *is*, in all but one case). Compared to auxiliaries such as CAN, DO, and SHOULD copula/auxiliary BE is unlikely to attract structure-dependence errors for two reasons: (a) Even young children are relatively proficient at forming questions with this auxiliary (particularly *is*) and (b) such errors strongly violate the transitional probability statistics of English. Experiment 2 used modal auxiliaries, thus avoiding the first shortcoming. However, all but one of the four questions elicited suffered from the second shortcoming, and all suffered the potentially more serious shortcoming that any possible structure-dependence error would be semantically inconsistent with the target question.

The goal of this article is to reassess Crain and Nakayama's (1987) claim that children never make structure-dependence errors, in light of the evidence discussed above that certain specific design-features of this study may have contributed to this null finding. In particular, we investigate the possibility raised by the studies of Lewis and Elman (2001) and Reali and Christiansen (2005) that surface distributional properties of English might influence the likelihood of such errors being committed. Experiment 1 investigated whether structure-dependence errors would be observed for questions with the modal auxiliary CAN for which such errors violate the bi-gram statistics of English less strongly than for Crain and Nakayama's stimuli. Experiment 2 investigated whether the patterning of another error type was influenced by surface co-occurrence statistics, which would suggest some sensitivity to these statistics on the part of children.

1. Experiment 1

We argued earlier that structure-dependence errors are particularly unlikely to be observed for complex questions with two instances of copula/auxiliary *is* (Crain & Nakayama's, 1987, Experiment 1) because (a) children are relatively proficient at forming questions with this auxiliary and (b) such errors contain bi-grams (e.g., *who running*) that strongly violate the transitional probability statistics of English. We further argued that structure-dependence errors are unlikely to be observed for questions with one instance of auxiliary *is* and one modal auxiliary (Crain & Nakayama's, 1987, Experiment 2) because they result in a question that is semantically inconsistent with the target question. Experiment 1 of the present study investigated whether structure-dependence errors would be observed for complex questions with two instances of the modal auxiliary CAN. Such errors do not so strongly violate the transitional probability statistics of English because they contain bi-grams that appear as part of sentences with plural subjects (e.g., *who run*) and are readily interpretable (e.g., **Can the boy who run fast can jump high?*).

1.1. Method

1.1.1. Participants

Participants were 22 normally developing monolingual English-speaking children (11 girls and 11 boys) aged 6;3 to 7;9 ($M = 6;9$), recruited from schools in the North West of England. Compared to the children studied by Crain and Nakayama (1987; mean age of 4;3), these children were relatively old. Although older children might be expected to exhibit lower overall error rates, it was felt that they would be likely to produce more valid attempts the full target questions than younger children (who may give up, ask a simple question instead, or use a "restarting" strategy).

Table 1
Example target questions for Experiment 1 (CAN questions)

Transitivity	No. of Subject	Target Question
Intransitive	Singular	Can the boy who can run fast jump high?
Intransitive	Plural	Can the boys who can run fast jump high?
Transitive	Singular	Can the boy who can play the piano play the guitar?
Transitive	Plural	Can the boys who can play the piano play the guitar?

1.1.2. Design

Singular- and plural-subject versions of eight different *can . . . can questions*—four transitive and four intransitive—were constructed for a total of 16 trials per participant (see Appendix A). Four example questions are shown in Table 1. All children completed all 16 trials.

1.1.3. Materials

A toy dog with an internal loudspeaker connected to a minidisc player was used to provide responses to children's questions (see Ambridge et al., 2006), whilst a second minidisc player connected to a Shure SM58 Microphone was used to record children's questions. Sixteen two-sided cards were created; one for each target question. One side (the *child side*) showed the scene about which the question was to be asked (e.g., one boy running fast and another walking). The other (the *dog side*) depicted the answer (e.g., the boy who had been running now performing a high jump, the other standing still).

1.1.4. Procedure

The experimental procedure was based on that of Crain and Nakayama (1987) and of Ambridge et al. (2006). For each trial, the child selected a card (thus randomizing the order of the trials), which the experimenter then placed in a stand such that one side faced the child and the experimenter, whilst the other faced the dog. The experimenter then elicited a question from the child by using an embedded yes/no question of the form *I wonder if X*, and, subsequently, *Ask the dog if X*, where X was a statement containing a relative clause (e.g., *the boy who can run fast can jump high*). After the child had attempted to produce the relevant complex yes/no question, the experimenter operated the minidisc player to have the dog provide an appropriate response.

Trials were presented in two sittings of eight trials each, given on consecutive days. Before the first session, children completed three warm-up trials, which were designed to familiarize children with the experimental procedure, but to avoid training them on the actual question types used in the test trials. The target questions for the warm-up trials were *Is the boy happy?* (simple question) *Is the girl who has brown hair happy?* and *Is the boy who has a football happy?* (complex questions with different auxiliaries).

1.1.5. Transcription and coding

Responses were recorded onto minidisc, then transferred to computer for subsequent transcription. The second author served as both experimenter and primary transcriber. Forty of the 352 individual trials (11%, randomly selected) were additionally transcribed by the first author for reliability analysis. No mismatches were observed. Each response was then assigned to one of the following mutually exclusive categories:

- *Correct question*: A correctly formed question such as, *Can the boy who can run fast jump high?* Substitutions of *that* or *which* for *who* (the only such substitutions that occurred) were permitted. No other substitutions (e.g., of either auxiliary, of the main verb or of noun phrases [NPs]) were permitted for questions scored as *correct*. Although they slightly alter the semantics of the question, questions with a short-form relative clause (e.g., *Can the monsters who eat* [for *can eat*]*people eat elephants?*) were scored as a subset of correct questions. Short-form questions for singular subjects where children changed the form of the main verb to create agreement with the subject (e.g., *Can the man who drives* [for can drive] *cars drive tractors?*) were scored as another subset of correct questions (*short-form + agreement* questions). However, trials for which participants substituted *is + present participle* for *can + bare stem* (e.g., *Can the girl who is making* [for *can make*] *cakes make biscuits?*) were scored as *excluded*.

- *Auxiliary-doubling error*[5]: Questions such as *Can the boy who can run fast can jump high?* (Crain & Nakayama's, 1987, Type 1 errors) were coded as *auxiliary-doubling errors*. To be counted as an auxiliary-doubling error, the auxiliaries, both NPs and the lexical main verb had to be exactly of the form modelled (although, again, substitutions of *which* or *that* for *who* were permitted).
- *Structure-dependence error:* An error such as *Can the boy who run fast can jump high?* or *Can the boys who run fast can jump high?* For a number of would-be structure-dependence errors with singular sentence subjects, children replaced the bare-stem main-verb form with a simple verb form marked for "agreement" with the subject (e.g., **Can the girl who make-s cakes can make biscuits*). Note that the equivalent form for plural sentence subjects (e.g., **Can the girls who make cakes can make biscuits*) would be coded as a structure-dependence error (although see the discussion of this point in the results section). For this reason, errors of the former type were coded separately as *structure dependence + agreement* errors.[6]
- *Unclassified error:* Errors that did not fit into any of the aforementioned categories but that, nevertheless, preserved the semantics of the target question were coded as *unclassified*. Such responses include those with determiner omission or duplication, or omission of the relativizer (except for correct short form questions or structure-dependence errors).
- *Excluded:* Questions that did not preserve the semantics of the target question (e.g., where the child substituted nouns or main verbs, or used an NP with the incorrect number) or that did not have the syntax of a complex question (i.e., that did not have all the appropriate syntactic constituents in the correct positions) were excluded. Restarting errors (Crain & Nakayama's, 1987, Type 2 errors; e.g., *Can the boy who can run fast, can he jump high?* were scored as *excluded* for this latter reason (although a count was kept, to allow for comparison with Crain & Nakayama's, 1987, data). This category also includes trials for which children produced a statement rather than a question, began to ask a question and then gave up, and so on.

All transcribed responses were coded by both the first two authors. The two coders agreed on the classification of all but 13 of the 352 individual trials (96%). All of these disagreements potentially affected coding, but the transcribers were able to reach agreement for all trials, with disagreements representing either confusion regarding the grounds for a response to be scored as excluded, or simple misreadings of the transcribed responses.

1.2. Results and discussion

Table 2 summarizes the results of Experiment 1. Note that the mean percentages shown here (and referred to in the text) represent the means of percentages calculated for individual children, not of the group as a whole. For example, the figure shown for correct questions (full

Table 2
Classification of children's question productions for Experiment 1 (CAN questions)

			% of all Responses		% of Valid Attempts	
Coding Category	Example	Number	M	SD	M	SD
Correct questions (full form)	*Can the boys who can run fast jump high?*	151	42.90	37.12	58.78	35.07
Short-form "correct" questions	*Can the boys who run fast jump high?*	9	2.56	4.98	3.53	6.06
Short-form + agreement "correct" questions (singular subject only)	*Can the boy who runs fast jump high?*	7	1.99	4.48	2.68	5.62
Auxiliary-doubling errors	*Can the boys who can run fast can jump high?*	49	13.92	16.47	23.97	25.88
Structure dependence errors	*Can the boys who run fast can jump high?* *Can the boy who run fast can jump high?*	16	4.55	9.08	7.92	12.82
Structure dependence + agreement errors (singular subject only)	*Can the boy who runs fast can jump high?*	4	1.14	4.15	1.78	5.74
Unclassified errors	*Can boy who can run fast jump high?*	3	0.85	2.92	1.34	4.13
Restarting Type 2 errors (excluded)	*Can the boys who can run fast, can they jump high?*	38	10.80	23.56	na	na
Other excluded responses		75	21.31	27.25	na	na

form) as a percentage of valid attempts (58.78%) does NOT represent the percentage of the total number of valid trials for which any child gave a correct response; that is, 100*(151/(352 – 38 – 75)) = 63.2%. Rather this figure was arrived at by calculating a percentage-correct (full form) score for each child, and then taking the mean. These two different figures are the same for the column headed "% of all responses" because all children contributed the same number of TOTAL responses (i.e., 16), even if some were subsequently excluded, but not for that headed "% of valid responses" as many children contributed less than 16 VALID (i.e., non-excluded) responses.

The first point to note is that a mean of 79% of all questions were categorized as preserving the semantics of the target question (a mean of 32% were excluded, but 11% of these constituted Type 2 errors, which preserve the semantics of the target question). Questions that were either excluded or categorised as *unclassified errors* are shown in Appendix B (which can be found online at http://www.cogsci.rpi.edu/CSJarchive/Supplemental/index.html). Inspection of these errors reveals that even many of the excluded trials appear to constitute attempts at the appropriate question, even though they may contain several different errors. Thus we can be reasonably confident that children understood the semantics of the target questions. The mean rates of correct question production (59% of valid attempts), auxiliary-doubling (Type 1) error (25%) and restarting (Type 2) error (11% of all responses) were very similar to the rates observed in Crain and Nakayama's (1987) Experiment 1 (60%, 23%, & 9%, respectively).

Table 3
Distribution of structure dependence errors among participants for Experiment 1 (CAN questions)

Child	SD Errors (of Which Plural Questions)	Valid Attempts (Maximum = 16)	% StDep Errors
A	1 (0)	6	16.67
B	3 (2)	10	30.00
C	1 (1)	8	12.50
D	6 (3)	14	42.86
E	1 (1)	8	12.50
F	2 (2)	7	28.57

Overall, 16 structure-dependence errors were observed (not counting structure dependence + agreement errors); a mean rate across children of 5% of all responses and 8% of valid attempts at the appropriate question (i.e., as a percentage of all questions except those classified as *excluded*). Two children were excluded from this latter calculation, as they produced no valid attempts at the appropriate question. The 16 structure-dependence errors can be found in Appendix C.

Table 3 shows the distribution of structure-dependence errors amongst participants. It is notable that over a quarter of the participants made at least one such error. Thus whilst structure-dependence errors are by no means frequent, they would seem to be made by a reasonably high proportion of children, at least for questions with the modal auxiliary CAN. It should be noted that, for CAN questions, some structure-dependence errors with a plural subject (e.g., *can the boys who run fast can jump high?*) could alternatively be analysed as auxiliary-doubling errors (*can NP can jump high?*) with a *short-form* relative clause (*the boys who run fast* as opposed to *the boys who CAN run fast*).[7] However, if some of children's apparent structure-dependence errors are actually auxiliary-doubling + short-form errors, then one would expect to find that (errors that have the surface form of) structure-dependence errors are more frequent for plural than singular CAN questions. This is because an auxiliary-doubling + short-form error will yield (a string with the surface form of) a structure-dependence error for a plural CAN question (e.g., *Can the boys who run fast can jump high?*) but not a singular CAN question (e.g., *Can the boy who runs fast can jump high*). In fact, structure-dependence errors were almost as common for singular (7 errors) as plural (9 errors) CAN questions, corresponding to mean rates of 7% and 9% of all scorable responses, respectively. Thus if structure-dependence errors with plural CAN questions are excluded as ambiguous (and trials with plural CAN questions therefore removed from the denominator), the mean rate of structure-dependence errors (as a proportion of scorable responses) barely changes, dropping from 8% to 7% (the mean rate as a proportion of ALL responses drops from 6% to 4%).

Although the SRN studies discussed did not investigate questions with the modal auxiliary CAN, they would seem to predict more structure-dependence errors with plural than singular subjects. This is because, whilst both contain legal bi-grams in the crucial position (e.g., *who crawl*), structure-dependence errors with plural subjects (15), but not singular subjects (16) contain legal tri-grams (underlined) in this position:

15. The babies who can crawl can walk → Can the babies who crawl can walk?
16. The baby who can crawl can walk → Can the baby who crawl can walk?

As we have seen, structure-dependence errors were observed at a similar rate for singular and plural questions (seven and nine errors, respectively). Although this difference was not significant, evaluation of the above prediction is probably not possible with so few errors for each sentence type.

However, the study of Reali and Christiansen (2005) found that, when using the transitional probabilities of word bi- and tri-grams to calculate a measure of the probability of occurrence of particular sentences (*cross entropy*), correctly formed complex questions were assigned a higher probability value than corresponding structure-dependence errors, regardless of whether bi- or tri-gram statistics were used. In fact, although correctly-formed complex questions were significantly more probable than corresponding structure-dependence errors using either the bi- or tri-gram statistics (both at $p = .0001$), the magnitude of this difference in probability was slightly greater using the bi-gram ($t[99] = 15.03$) than tri-gram statistics ($t[99] = 11.74$).[8]

In summary, the findings of Experiment 1 suggest that Crain and Nakayama (1987) are premature in their conclusion that children *never* make structure-dependence errors in complex question formation (although such errors are rare). A further implication is that the absence of such errors in the study of Crain and Nakayama (1987) may partly reflect the fact that these authors elicited questions with copula/auxiliary *is,* for which structure-dependence errors strongly violate the transitional probability statistics of English. Implicit in this latter claim are the assumptions that young children are sensitive to the transitional probability statistics of the input data, and that their sentence productions are affected by this implicit knowledge. The goal of Experiment 2 was to investigate whether the patterning of a relatively frequent error type—auxiliary-doubling errors—would provide support for these assumptions.

2. Experiment 2

The aim of Experiment 2 was to investigate whether children's auxiliary-doubling errors would pattern as predicted by the computer simulations of Lewis and Elman (2001) and Reali and Christiansen (2005) and, in doing so, to test the plausibility of the claim that children—like these simulations—may be protected from structure-dependence errors by surface co-occurrence patterns in the input data.

As Table 2 shows, around a quarter of all valid responses made by children in Experiment 1 were auxiliary-doubling errors (e.g., **Can the boys who can run fast can jump high?*). The SRN models discussed also, in effect, make auxiliary-doubling errors: If given a string such as *Is the boy who is smoking ...* (or *Is the boy ...*) early in training, the models predict the next item to be an auxiliary (e.g., *is*) with relatively high probability. For the models, these "auxiliary-doubling errors" are a consequence of the fact that the learning mechanism is initially misled by the high frequency of strings such as [NP] [AUX] [NP] in the input (although the simulation of Lewis & Elman, 2001, does not "see" categories such as NP or AUX, but creates internal representations that may approximate to these categories). Although such errors are quite persistent, they eventually cease because questions in the input provide evidence that an auxiliary is not an appropriate prediction for strings beginning with an auxiliary followed by a noun phrase.

Table 4
Auxiliary-doubling errors in Crain and Nakayama (1987)

Auxiliary-Doubling Error	Lexical-Category String Created by Insertion of Extra Auxiliary/Copula	Auxiliary-Doubling Errors	Frequency (Manchester Corpus Mothers)
Is the dog that is sleeping is on the blue bench?	PART BE (COP) PREP	2	1
Is the ball that the girl is sitting on is big?	PREP BE (COP) ADJ	6	1
Is the boy who is watching Mickey + Mouse is Happy?	NOUN[a] BE (COP) ADJ	8	252
Is the boy who is unhappy is watching Mickey + Mouse?	ADJ BE (AUX) PART	4	2
Is the boy who is being kissed by his mother is happy?	NOUN[a] BE (COP) ADJ	8	252
Is the boy who was holding the plate is crying?	NOUN[a] BE (AUX) PART	11	757

[a]To give a conservative estimate of the frequency of these strings, only common nouns were included in the corpus search.

Since the SRN models make auxiliary-doubling errors because of the presence of certain category combinations in the input, it follows that, if learning proceeds in a similar manner for children, the rate of such errors will be high for questions where this error results in a high-frequency category combination (e.g., [NOUN] [BE] [ADJECTIVE]; *Is the boy who is being kissed by his mother is happy?*) and low for questions where this error results in a less frequent combination (e.g., [PARTICIPLE] [BE] [PREPOSITION]; *Is the dog that is sleeping is on the blue bench?*). This prediction is not shared by the innate-constraint view (Crain & Nakayama, 1987), under which children form yes/no questions using movement rules that are explicitly NOT formed on the basis of the input data.

In fact, Crain and Nakayama's (1987) own results provide some support for the prediction that rates of auxiliary-doubling error will be highest where the "inserted" auxiliary forms part of a relatively high frequency category combination. Table 4 shows the total number of auxiliary-doubling errors made for each of the test sentences elicited in Crain and Nakayama's Experiment 1. For each auxiliary-doubling error, the word tri-gram containing the "inserted" auxiliary is underlined; the corresponding lexical categories are shown in the second column. Table 4 reveals that auxiliary-doubling errors that resulted in the category combination of [NOUN] [BE] [ADJECTIVE] or [NOUN] [BE] [PARTICIPLE] were over twice as common as auxiliary-doubling errors that resulted in the combinations [PARTICIPLE] [BE] [PREPOSITION], [PREPOSITION] [BE] [ADJECTIVE] or [ADJECTIVE] [BE] [PARTICIPLE] (27 vs. 12 errors). A search of the maternal data of the Manchester Corpus (Theakston, Lieven, Pine, & Rowland, 2001) yielded 757 sentences of the form [PROPER NOUN] [BE3sg] [PARTICIPLE], and 252 of the form [PROPER NOUN] [BE3sg] [ADJECTIVE]. In contrast, this corpus contained only two instances of the combination [ADJECTIVE] [BE3sg] [PARTICIPLE] (*Somebody special is coming; Something strange is going on*) and only one each of the combi-

Table 5
Example target question types for Experiment 2 (BE questions)

Match–Mismatch of Number Features of NPs	Number Feature of NP in Main Clause	Target Question and (in Parentheses) Auxiliary-Doubling Error	Category Combination Generated by Auxiliary-Doubling Error	Relative Frequency of Category Combination
Match	Singular	*Is the boy that is washing the elephant (is) tired?*	$NP_{singular}$ is ADJ	High
	Plural	*Are the boys that are washing the elephants (are) tired?*	NP_{plural} are ADJ	High
Mismatch	Singular	*Is the boy that is washing the elephants (is) tired?*	NP_{plural} is ADJ	Low
	Plural	*Are the boys that are washing the elephant (are) tired?*	$NP_{singular}$ are ADJ	Low

nations [PARTICIPLE] [BE3sg] [PREPOSITION] and [PREPOSITION] [BE3sg] [ADJECTIVE] (*The whole point of going is for you to see if you like it; That bit underneath is yellow*).

It is impossible to be sure, however, that this pattern of results is driven by the frequency of particular category combinations, because these sentences differ along other dimensions (e.g., simple length, and whether the relative clause modifies the subject or the object). The aim of Experiment 2, then, was to elicit minimal pairs of sentences where the frequency of the relevant category combination varied, but sentence length and structure were held constant. In order to achieve this, we varied the frequency of the category combinations by manipulating the number features of the two NPs and their agreeing auxiliary verbs as illustrated by the example target sentences in Table 5.

Although the relevant category combinations are at a more fine-grained level than those previously discussed (e.g., $NP_{singular}$ is ADJECTIVE as opposed to NP AUXILIARY ADJECTIVE) this does not affect the prediction of the modelling accounts. The SRN model of Lewis and Elman (2001) is not presented with category information, but is able to generate its own fuzzy probabilistic categories at all levels from individual lexical items (e.g., *is*) to those approximating to higher order categories (e.g., BE → AUX → NP → VP and so on). It is clear that combinations such as $NP_{singular}$ is ADJECTIVE (e.g., *elephant is tired*) are more frequent than those such as NP_{plural} is ADJECTIVE (e.g., *elephant are tired*) as the former are common in simple sentences whereas the latter are illegal except in complex sentences or complex NPs.

The prediction from SRN models, then, is that auxiliary-doubling errors will occur at a higher rate for questions where such an error results in a higher frequency category combination ($NP_{singular}$ is ADJECTIVE or NP_{plural} are ADJECTIVE) than a lower frequency combination (NP_{plural} is ADJECTIVE or $NP_{singular}$ are ADJECTIVE). The primary goal of Experiment 2 was to test this prediction. A second aim was to investigate whether Crain and Nakayama's (1987) finding of no structure-dependence errors for BE questions would be replicated. As we have argued above, there is reason to expect that structure-dependence errors will be less frequent for BE than CAN questions, as such errors violate the transitional probability statistics of English more strongly for the former than the latter. A third aim was to investigate whether, should such errors in fact be observed, they would be more frequent for sentences with the plural (*are*) than the singular (*is*) form of copula/auxiliary BE. Although there is some evidence to suggest that children are more proficient at forming questions with auxiliary *is* than *are* (e.g., Ambridge et al., 2006), Crain and Nakayama used only the singular form; a fact that may have contributed to their finding of no structure-dependence errors.

2.1. Method

2.1.1. Participants

Participants were 33 normally developing monolingual English-speaking children (19 girls and 14 boys) aged 4;7 to 5:7 ($M = 5; 1$), recruited from schools in the North West of England. It was decided to recruit younger participants for Experiment 2 than Experiment 1 in order to increase the likelihood of observing high overall rates of auxiliary-doubling error (the elicited production study of Ambridge et al, 2006, found relatively low rates of doubling-error with auxiliary BE for simple questions). The participants for Experiment 2 were also closer in age to those studied by Crain and Nakayama ($M = 4;3$). A further three children did not understand the experimental procedure of putting questions to the dog and so were excluded, but all of the remaining 33 children produced at least one attempt at a complex question during the warm-up, and so were included.

2.1.2. Design

The experiment used a 2 × 2 within-subjects design. The first factor (match–mismatch) corresponded to whether the number features of the NP in the main clause and the NP in the relative clause were the same (both singular or both plural) or different (one singular and the other plural). The second factor was simply whether the NP in the main clause was singular or plural. Crossing these factors resulted in four target-question types, examples of which are shown in Table 5. Four questions of each type were generated, for a total of 16 items (see Appendix D).

2.1.3. Procedure

The procedure for Experiment 2 was identical to that for Experiment 1.

2.1.4. Transcription and coding

Reponses were again recorded onto minidisc, then transferred to computer for subsequent transcription. For this study, both the first two authors served as experimenters. Data collected by the second author were transcribed by the first author, while data collected by the first author were transcribed by a research assistant. One hundred and twelve of the 528 individual trials (21%, randomly selected) were additionally transcribed (by the first or second author, or the research assistant) for reliability analysis. The two transcribers reached agreement for 103 of these 112 trials (95%). Four of the nine mismatches affected coding as they involved a relevant auxiliary, NP or relativizer whilst five did not (e.g., superfluous determiners). In each case, the two transcribers were able to reach agreement on what the child had actually said.

All transcribed responses were then scored by both the first two authors. Each response was again assigned to one of a number of mutually exclusive categories. The basic coding scheme was the same as that used for Experiment 1, although a number of modifications were made in order to take account of a number of errors that were specific to one or other of the experiments:

- *Short-form (correct) questions* were not subdivided into short-form and short-form + agreement questions, as no child produced an utterance with agreement marked on the main verb (e.g., *Is the boy who wash-es elephants tired?*). Similarly, the category of structure dependence + agreement error was not used, as no child produced an utterance such as *Is the boy who washes the elephant is tired?*
- To be counted as an auxiliary-doubling error, both NPs and the lexical main verb had to be exactly of the form modelled (although, again, substitutions of *which* or *that* for *who* were permitted). However, for some auxiliary-doubling errors the child "changed" the form of the extra auxiliary to create a spurious local agreement (e.g., *Is the boy who is washing the elephants are tired?*). Such utterances were coded as auxiliary-doubling errors but were flagged to allow them to be removed from some analyses. For some trials, children created a spurious local agreement by changing not the auxiliary, but the final NP (e.g., *Is the boy who is washing the elephant is tired?* [where three elephants were pictured]). Because errors where the child changes the number of either NP do not preserve the semantics of the target question, all such errors (regardless of whether they would otherwise have been scored as auxiliary-doubling errors) were excluded. Auxiliary-doubling errors where the child changed the number of the first or second auxiliary (e.g., *Is the boys who are washing the elephants are tired?*) were counted simply as auxiliary-doubling errors (as such errors do not affect the strings under investigation, and do not alter the semantics of the target question).
- Errors with incorrect number marking on any auxiliary were scored as *unclassified* unless already classified as auxiliary-doubling or structure-dependence errors.

Again, all children's questions were coded by two coders (the first two authors). Although the authors initially disagreed on the coding of 46 of the 528 individual responses (an agreement rate of 91%) they were able to reach agreement for all of these responses, as all disagreements were simple cases of one coder failing to notice a particular error.

2.2. Results and discussion

Table 6 summarizes the percentage of children's questions that were classified into each of the different coding categories. One child was excluded from the mean percentage of valid attempts calculation, as he produced no valid attempts at the appropriate question.

For this study, a mean of 74% of trials constituted attempts at a semantically appropriate question (i.e., a mean of 28% of trials were excluded, with 2% of these Type 2 errors), although the rate of unclassified errors (17%) was much higher than for Experiment 1. This is probably due to the fact that the participants of Experiment 2 were around 20 months younger. Unclassified errors and excluded responses are shown in Appendix E (which can

Table 6
Classification of children's question productions for Experiment 2 (BE questions)

Coding Category	Example	Number	% of all Responses M	% of all Responses SD	% of Valid Attempts M	% of Valid Attempts SD
Correct questions (full form)	*Is the boy who is washing the elephant tired?*	141	26.70	28.61	33.83	34.17
Correct questions (short form)	*Is the boy washing the elephant tired?*	52	9.85	21.88	13.66	27.87
Auxiliary-doubling errors	*Is the boy who is washing the elephant is tired?*	74	14.02	22.64	18.20	28.87
Structure dependence errors	*Is the boy who washing the elephant is tired?*	24	4.55	8.88	6.64	12.65
Unclassified errors	*Is the boy who is washing elephant is tired?*	90	17.05	20.51	27.67	31.83
Restarting Type 2 errors (excluded)	*Is the boy who is washing the elephant, is he tired?*	11	2.08	5.10	na	na
Other excluded responses		136	25.76	23.06	na	na

be found online at http://www.cogsci.rpi.edu/CSJarchive/Supplemental/index.html). Again, there appear to be few questions that suggest that children did not comprehend the procedure or the target question. For this study, the rate of correct-question production (27%) was considerably lower than that observed in Crain and Nakayama's (1987) Experiment 1 (60%), with rates of auxiliary-doubling (14%) and restarting errors (2%) also somewhat lower than those observed by Crain and Nakayama (23% & 9%).

Surprisingly, given the results of the SRN simulations of Lewis and Elman (2001) and Reali and Christiansen (2005), structure-dependence errors (see Appendix F) were observed at a rate similar to that for CAN questions (5% of all questions, a mean rate across children of 7% of all valid attempts). Structure-dependence errors were predicted to be less frequent for BE questions, as such errors include very low frequency bi-grams (e.g., *who running*), whereas corresponding errors with CAN questions contain more frequent combinations (e.g., *who run*). Table 7 displays the distribution of these errors amongst participants. It is notable that again over a quarter of the children studied produced one or more structure-dependence error; a proportion similar to that observed for CAN questions. We return to this finding at the end of this section.

The fact that structure-dependence errors were observed allows us to investigate a second possibility; that such errors are more frequent with the plural (*are*) than the singular form (*is*) of copula/auxiliary BE. Data from studies of children's simple-question production (e.g., Ambridge et al., 2006) suggest that children are more proficient at forming questions with *is* than *are*, raising the possibility that the rate of structure-dependence errors in Crain and Nakayama's (1987) study may have been lowered by the decision of these authors to elicit only singular BE questions. Of the 24 structure-dependence errors observed in our second experiment, 15 had plural subjects with copula *are*, and six singular subjects with copula *is* (the remaining three included one instance of *is* and one of *are*). Although this difference was not statistically significant, the present findings are at least consistent with the

Table 7
Distribution of structure dependence errors among participants in Experiment 2 (BE questions)

Child	SD Errors (of the Form *are... are* vs. *is... is*)	Valid Attempts (Maximum = 16)	% StDep Errors
A	3 (3 vs. 0)	9	33.33
B	2 (2 vs. 0)	13	15.38
C	1 (1 vs. 0)	15	6.67
D	4 (0 vs. 2*)	13	30.77
E	1 (0 vs. 1)	5	20.00
F	5 (5 vs. 0)	11	45.45
G	4 (4 vs. 0)	12	33.33
H	1 (0 vs. 0*)	13	7.69
I	3 (0 vs. 3)	15	20.00

Note. Child D (2 errors) and Child H (1 error) made structure dependence errors with one singular (*is*) and one plural (*are*) form; see Appendix F.

suggestion that Crain and Nakayama's finding of no structure dependence errors may have been partly caused by their decision to use singular (*is*) BE questions, with which children are perhaps more proficient than plural (*are*) questions. Planned comparisons between *is... is* questions and questions with one or more auxiliary *are* were conducted to investigate this issue. In support of this suggestion, significantly more correctly-formed questions were produced for *is... is* questions ($M = 1.48$ out of a maximum of 4, $SD = 1.60$) than *are... are* ($M = 0.81$, $SD = 1.21$, $t = 2.37$, $p = .02$), *is... are* ($M = 1.06$ $SD = 1.34$, $t = 2.10$, $p = .05$) or *are... is* ($M = 0.90, SD = 1.35, t = 2.20, p = .04$) questions, which did not differ significantly.

The SRN models of Lewis and Elman (2001) and Reali and Christiansen (2005) predict that auxiliary-doubling errors will be more common when such errors result in the formation of higher frequency category strings (e.g., *elephant is tired, elephants are tired*) than lower frequency strings that are found only in sentences with complex syntax or complex NPs (e.g., *elephants is tired, elephant are tired*). Table 8 shows the mean number of auxiliary-doubling errors made by children for each question type elicited in Experiment 2 (out of a maximum of 4)

A 2 × 2 analysis of variance (ANOVA) revealed a significant main effect of match vs. mismatch, $F(1, 32) = 6.77$, $p = .01$, $\eta_p^2 = 0.18$, demonstrating that children made significantly more auxiliary-doubling errors for sentences where such errors resulted in the formation of a higher frequency category combination ($M = 0.67$, $SD = 1.13$) than a lower frequency combination ($M = 0.46$, $SD = 0.90$). This ANOVA also revealed a significant main effect of main-clause-NP-number, $F(1, 32) = 5.93$, $p = .02$, $\eta_p^2 = 0.16$, such that (perhaps surprisingly) children made significantly more auxiliary-doubling errors with singular ($M = 0.67$, $SD = 1.14$) than plural ($M = 0.45$, $SD = 0.88$) main clause subjects. These two factors were not found to interact significantly ($F = 0.66$, $p = .42$). The finding of more auxiliary-doubling errors for singular than plural subjects is potentially consistent with the Lewis and Elman (2001) and Reali and Christiansen (2005) models. Although in the input data, the transitional probability between a singular NP and a singular auxiliary is probably approximately equal to that between a plural NP and a plural auxiliary, it would seem likely that

Table 8
Mean number of auxiliary-doubling errors by question type for Experiment 2 (BE questions)

Match–Mismatch of Number Features of NPs	Number Feature of NP in Main Clause	Target Question and (in Parentheses) Auxiliary-Doubling Error	Category Combination Generated by Auxiliary-Doubling Error	Relative Frequency of Category Combination	Mean Number of Auxiliary-Doubling Errors (SD), Maximum = 4
Match	Singular	*Is the boy that is washing the elephant (is) tired?*	$NP_{singular}$ is ADJ	High	0.82 (1.31)
	Plural	*Are the boys that are washing the elephants (are) tired?*	NP_{plural} are ADJ	High	0.52 (0.91)
Mismatch	Singular	*Is the boy that is washing the elephants (is) tired?*	NP_{plural} is ADJ	Low	0.52 (0.94)
	Plural	*Are the boys that are washing the elephant (are) tired?*	$NP_{singular}$ are ADJ	Low	0.39 (0.86)

particular singular $NP_{singular}+ AUX_{singular}$ strings have a higher chunk frequency than similar $NP_{plural}+ AUX_{plural}$ strings (simply because, intuitively, singular subjects are more common that plural subjects).

If this turns out to be the case, then this offers a ready explanation for the finding of more auxiliary-doubling errors for *is* than *are*.

However, this pattern of results is somewhat distorted by the fact that children made more valid attempts at questions with a match (a mean of 3.00, $SD = 1.10$, out of a maximum of 4) than a mismatch ($M = 2.77$, $SD = 1.35$) between the numbers of the NPs in the main and relative clauses, and at questions with a singular ($M = 2.92$, $SD = 1.30$) than plural ($M = 2.85$, $SD = 1.15$) NP in the main clause (these figures include all children). Table 9 shows, for each of the four question types elicited, the mean number of auxiliary-doubling errors, the mean number of valid attempts at the appropriate question, and the mean proportion of auxiliary-doubling errors. Five children were excluded from these calculations, as they failed to produce at least one valid attempt for one or more of the four question types (including the one child who made no valid attempts at the appropriate question throughout the entire experiment).

A 2×2 ANOVA calculated for the proportion of valid attempts at questions that constituted auxiliary-doubling errors revealed a significant main effect of match vs. mismatch, $F(1, 27) = 8.37$, $p = .007$, $\eta_p^2 = 0.24$, demonstrating that children made a significantly greater proportion of auxiliary-doubling errors for target questions where such errors resulted in the formation of a higher frequency category combination ($M = 0.24$, $SD = 0.36$) than a lower frequency combination ($M = 0.17$, $SD = 0.30$). No other main effects or interactions were observed. This suggests that the previous finding of significantly more auxiliary-doubling errors for singular than plural questions might have been partly caused by the trend toward more valid attempts at singular than plural (with respect to the main-clause NP) questions.

While, as discussed in the method section, questions in which the number feature of either NP differed from the target were excluded (because this changes the semantics of the question), the figures reported so far include auxiliary-doubling errors where the child "changed" the form of the extra auxiliary to create a spurious local agreement (e.g., **Is the boy who is washing the elephants are tired*). Such errors in fact represent around 15% of all the auxiliary-doubling errors that occurred in Experiment 2. Such errors are consistent with the prediction that children will produce more auxiliary-doubling errors where this results in a higher frequency category combination, as "changing" the auxiliary to create a spurious local agreement results in a higher frequency category combination. However, the inclusion of such responses as auxiliary-doubling errors actually reduces the likelihood of the predicted effect being observed: When children "change" the form of the extra auxiliary, this more often has the effect of creating a spurious local agreement for a mismatch sentence (e.g., *the elephants is tired* $\rightarrow$ *the elephants are tired*), rather than removing such an agreement that would otherwise have been present for an auxiliary-doubling error with a match sentence (e.g., *the elephant is tired* $\rightarrow$ *the elephant are tired*). Thus the inclusion of such responses as auxiliary-doubling errors artificially inflates the rate of errors for mismatch sentences (because children are, in effect, converting them into match sentences). For a third analysis, then, we removed every auxiliary-doubling error where the inserted auxiliary was not of identical form to the auxiliary in the main clause of the target question. This led to the further exclusion of two children

Table 9
Auxiliary-doubling errors in Experiment 2 (BE questions)

Match–Mismatch of Number Features of NPs	Number Feature of NPs in Main Clause	Example Target Question and (in Parentheses) Auxiliary-Doubling Error	Mean Number of Auxiliary-Doubling Errors (SD), Maximum = 4	Mean Number of Valid Attempts (SD), Maximum = 4	Mean Proportion of Auxiliary-Doubling Errors (SD)
Match	Singular	*Is the boy that is washing the elephant (is) tired?*	0.96 (1.37)	3.58 (0.69)	0.27 (0.37)
	Plural	*Are the boys that are washing the elephants (are) tired?*	0.57 (0.96)	3.04 (0.74)	0.21 (0.36)
Mismatch	Singular	*Is the boy that is washing the elephants (is) tired?*	0.46 (0.92)	3.21 (0.92)	0.18 (0.29)
	Plural	*Are the boys that are washing the elephant (are) tired?*	0.61 (0.99)	3.04 (1.10)	0.15 (0.32)

(for a total of seven exclusions) who "changed" the inserted auxiliary for every auxiliary-doubling error for one particular question type (*mismatch, singular main clause NP,* in both cases). In practice, the exclusion of such errors made little difference to the pattern of results. Again, a 2×2 ANOVA calculated for the proportion of valid attempts at questions that constituted auxiliary-doubling errors revealed a significant main effect of match ($M = 0.21$, $SD = 0.35$) vs. mismatch ($M = 0.13, SD = 0.27$; $F[1, 25] = 7.64$, $p = .01$, $\eta_p^2 = 0.23$) but no other significant main effects or interactions (the effect was reduced in size despite the removal of auxiliary-change errors because 2 fewer subjects contributed scores to the analysis).

In summary, the results of Experiment 2 support the prediction of Lewis and Elman's (2001) and Reali and Christiansen's (2005) SRN simulations that auxiliary-doubling errors will be more frequent when such errors create higher frequency category combinations. A non-significant trend for a higher rate of structure-dependence errors with *are* than *is* questions was also observed.

We now return to the finding that, counter to the prediction derived from the models of Lewis and Elman (2001) and Reali and Christiansen (2005), structure-dependence errors appeared to be no more frequent for CAN than BE questions. It is important to note that our two studies are not directly comparable, as they differed with respect to the age of the participants, and to the particular questions elicited. Nevertheless, due to the importance of this prediction, we decided to compare the rates of structure-dependence errors observed with CAN (Experiment 1) and BE (Experiment 2) questions statistically. The raw rates of structure-dependence error were identical for both experiments (4.55%) and so were not compared statistically. The mean rates of structure-dependence error as a proportion of all valid attempts at the correct question were 7.92% ($SD = 12.82$) for CAN questions (Experiment 1) and 6.64% ($SD = 12.65$) for BE questions. An independent samples t-test found no significant difference between these error rates, $t(1, 50) = 0.35$, $p = .73$.

As noted in the results section of Experiment 1, comparing rates of structure-dependence error for plural questions is problematic, because such errors have an alternative analysis for CAN questions (as auxiliary-doubling errors with a short-form relative clause) but not BE questions. Excluding all questions with a plural subject or auxiliary form leaves seven structure-dependence errors for CAN questions, and six for BE questions. Although these numbers are too small to compare meaningfully, the finding of similar rates of structure-dependence error for CAN and BE questions therefore remains unchanged.

3. General discussion

The goal of the present studies was to reassess Chomsky's (1980) and Crain and Nakayama's (1987) claims that (a) children never make structure-dependence errors when forming complex questions and that (b) the absence of such errors necessarily implies that structure dependence is an innately specified constraint, in light of the results of the SRN studies of Lewis and Elman (2001) and Reali and Christiansen (2005).

Given that, in even the largest currently available child corpora, we have no recorded attempts at complex yes/no questions, the only previous evidence in support of the claim that children do not make structure-dependence errors is the elicited production study of Crain and Nakayama (1987). Leaving aside the questions with CAN (for which structure-dependence errors are inconsistent with the semantics of the target question), this study used only auxiliary *is*, with which even young children are relatively proficient at forming questions, and consisted of only 220 trials (as opposed to 880 in the present study). In fact, when a large number of trials were elicited with a wider range of auxiliaries (*is, are*, and *can*) in the present studies, 40 structure dependence errors were observed (only 6 of which contained 2 instances of auxiliary *is*, as in the questions elicited in Crain & Nakayama's, 1987, Experiment 1). Although rates of structure-dependence error were never high, over a quarter of all children studied produced at least one such error.

It would be possible in principle to explain away the observed structure-dependence errors as production errors. One could argue, for example, that such errors constitute auxiliary-doubling errors with the second of the three auxiliaries omitted or not phonologically realized (e.g., *Is the boy who 0is [*] washing the elephant is tired?*; see Footnote 7) or, in the case of plural CAN questions (e.g., *Can the boys who run fast can jump high?*), simple auxiliary-doubling errors (analogous to, for example, *Can the boys can jump high?*) that are indistinguishable on the surface from structure-dependence errors. Once one assumes, however, that children's elicited production data reflect performance factors rather than grammatical competence, empirical data such as those reported by Crain and Nakayama (1987) become irrelevant to the question of whether or not structure dependence is an innately specified constraint. If any structure-dependence errors observed can be dismissed as production errors, then the finding of an error rate of zero (as reported by Crain & Nakayama, 1987) has no different theoretical consequences to the finding of an error rate of 5%, or of 25%. We do not wish to claim, however, that the structure-dependence errors observed are a consequence of children having formed an incorrect rule such as *move the first auxiliary to sentence-initial position* (Chomsky, 1980); in fact we do not accept the claim that questions are formed by a movement rule (see Ambridge et al., 2006). Thus, whilst we accept that the observed errors are, in some sense, "performance errors," this does not entail acceptance of the claim that children have been born with (or even acquired) a structure-dependent movement rule.

This brings us to Chomsky's (1980) and Crain and Nakayama's (1987) second claim that the absence (or very low rate) of structure-dependence errors constitutes evidence that structure dependence is an innately specified constraint. The SRN studies of Lewis and Elman (2001) and Reali and Christiansen (2005) demonstrate that, in principle at least, even a learner that does not have structure dependence as an innately specified constraint can avoid making structure-dependence errors in complex question formation, on the basis of surface-distributional properties of the input data. The present Experiment 2 provides evidence that young children are sensitive to the surface distributional properties of their input, as such properties can explain the patterning of children's auxiliary-doubling errors. Thus it seems that children are indeed in possession of a learning and/or production mechanism that is influenced by transitional probability statistics of the input data; statistics that are, in principle, sufficient to protect the learner against the production of structure-dependence errors. It simply does not follow, then, that even the finding of an error rate of zero would necessarily imply the existence of an innately specified principle of structure dependence.

Leaving aside the empirical data, however, many authors have made the theoretical argument that structure dependence must be an innately specified constraint because the input does not contain sufficient information *in principle* for the learner to form the structure-dependent representations necessary for correct complex question formation (Chomsky, 1980). The studies of Lewis and Elman (2001) and Reali and Christiansen (2005) demonstrate that this is not the case. These models do not have structure dependence as an innate constraint, yet they do not make structure-dependence errors, and in fact learn to make predictions corresponding to well-formed questions. Whether or not these models are psychologically plausible, they nevertheless demonstrate that input that does not contain any examples of complex questions can in principle contain sufficient information for children to learn to ask correctly formed complex questions.

What, then, is the relationship between innate constraints and data-driven learning, with respect to structure dependence? We can see three possibilities. The first is that the principle of structure dependence is indeed, as claimed by Chomsky (1972), "an innate schematism applied by the mind to the data of experience" (p. 30). Neither the present study nor those of Lewis and Elman (2001) or Reali and Christiansen (2005) demonstrate that structure dependence is not an innate constraint (although it is difficult to see how any study could do so, even if it were not). What these studies do demonstrate, however, is that both the empirical and theoretical arguments that have previously been put forward to support the innate-constraint account (at least with regard to complex questions) are flawed.

The second possibility is that the input data interacts with some kind of innate structure to achieve the end-state grammar. Yang (2002) proposes a model of "language acquisition as a population of competing grammars, whose distribution changes in response to the linguistic evidence presented to the learner" (p. 12). Under this proposal, children's innate endowment consists of a set of parameters (e.g., whether a language has wh-movement or allows subjects to be dropped) that are set on the basis of the input data. This approach differs from other parameter-setting accounts (e.g., Fodor, 1998; Hyams, 1986) in that parameters are not "triggered" by a single input datum. Rather, at any one time the learner has a number of different grammars—each specifying a particular setting for each parameter—which compete stochastically: Grammars (or more properly their associated parameter settings) that successfully parse input sentences are rewarded, thus increasing the likelihood of their being selected to parse subsequent sentences, while those that do not are punished. Eventually the learner arrives at a stable grammar (or grammars) corresponding to a set of parameter settings that can correctly generate any required utterance.

Yang (2002, pp. 109–114) specifically addresses the question of structure dependence in yes/no questions, asking whether the input data contain sufficient evidence to allow the learner to make a "binary choice" (p. 110) between a rule such as "front the first auxiliary verb in the sentence" (p. 109) and the correct, structure dependent rule. Yang claims, on the basis of data presented in Valian (1991), that English-speaking children learn that English is not a subject-drop language by age 3;0, and that sentences that provide evidence for this ("signatures," Yang, 2002, p. 104), specifically those with pure expletive subjects (e.g., *There is a man in the room*), constitute 1.2% of all adult sentences in CHILDES. Yang goes on to argue that, in order for children aged 3;0 to have learned that the structure-dependent version of the question formation rule is the correct one, sentences providing evidence for this fact (e.g., *Is the boy who is in the corner smiling?* or *How could anyone that was awake not hear that?*, p. 100) would have to constitute around 1.2% of sentences encountered. Because the rate of such sentences is apparently only "approximately 0.03%" (Yang, 2002, p. 112), yet children aged 3;0 do not make structure-dependence errors (Crain & Nakayama, 1987), Yang (p. 114) concluded (quoting Chomsky, 1975, p. 33) that "the principle of structure dependence is not learned, but forms part of the conditions for language learning."

Yang's (2002) argument, however, is logically flawed. Even if one accepts the claim that an expletive-subject rate of 1.2% is *sufficient* for children to learn that English requires subjects, this does not entail that a similar rate of complex questions is *necessary* to acquire the principle of structure dependence. Indeed, the models of Lewis and Elman (2001) and Reali and Christiansen (2005) make predictions corresponding to correctly-formed complex questions, avoiding structure-dependence errors, in the absence of *any* complex questions in the input. In any case, the empirical findings of the present study render Yang's argument irrelevant. If, as we have shown, young children do make structure-dependence errors then there is no reason to posit an innate constraint to explain why they do not.

The third possibility is that structure-dependence is not an innate constraint, but that the child acquires a structure-dependent grammar because she forms her grammar on the basis of the input data, which exhibits this property (e.g., Stemmer, 1981). For example, there is no reason in principle why the child could not learn that simple NPs (e.g., *the boy*) and complex NPs (e.g., *the boy who is smoking*) can be substituted for one another on the basis of a distributional analysis of the input. For example, in declarative sentences both simple and complex NPs will appear before many of the same auxiliaries (e.g., *is, can, does*), before verbs marked for tense and agreement (e.g., *likes, played*), after prepositions (e.g., *I looked at/spoke to*), and so on. Then once the child has learned to ask simple yes/no questions (perhaps by acquiring a yes/no question construction such as *[AUXILIARY/COPULA] [NP] [VERB/ADJECTIVE]?*) she can form a complex question by substituting a complex NP for a simple NP. Many theorists reject out of hand the possibility that children could conduct such a distributional analysis (e.g., Pinker, 1987), but this is to ignore a considerable body of evidence that demonstrates that even pre-linguistic infants are sensitive to surface distributional properties of the input data (e.g., Gomez & Gerken, 1999; Maye & Weiss, 2003; Saffran, Aslin, & Newport, 1996). Experiment 2 of the present study demonstrates that young children are sensitive to such properties in the domain of complex question formation.

Furthermore, the SRN simulations of Lewis and Elman (2001) and Reali and Christiansen (2005) predict the correct structure for complex questions in a very similar way.[9] In effect, they learn that strings such as *the boy* and *the boy who is smoking* can be substituted for one another, on the basis that they share similarities in their surface distributions (i.e., both predict, and are predicted by, the same items). It is premature to reject out of hand the possibility that children could do something similar, particularly as they also have another source of evidence that simple and complex phrases can be substituted for one another that is not available to the SRN models; namely functional or semantic information. For many complex phrases (e.g., *the boy who is smoking*) there exists a simple phrase (e.g., *the boy*) that denotes the same referent and/or serves the same function. Of course, claiming that children can substitute complex NPs for simple NPs in questions is a restatement of the principle of structure dependence. The point is that it is a restatement that conceptualizes structure dependence as a learnable property of the input data, and not as an innate constraint.

Indeed, moving away from the specific example of complex yes/no questions, *any* evidence that simple and complex phrases can be substituted for one another constitutes evidence (at least in principle) that language has the abstract property of structure dependence. Even an utterance as simple as *I want the ice cream. I want it!* constitutes evidence that language has constituent structure, whereby strings of arbitrary length that share certain properties can be substituted for one another (and could be seen as a "signature" of structure dependence under accounts such as that of Yang, 2002).

One advantage of this kind of learning account is that the same mechanism that explains how children learn to correctly produce complex yes/no questions (on the whole, avoiding structure-dependence errors) also accounts for the precise pattern of auxiliary-doubling errors observed for such sentences. Although some versions of the innate-constraint account can explain why auxiliary-doubling errors are common (e.g., because they reflect a process such as movement without deletion; Crain & Nakayama, 1987) such accounts currently have no explanation for the particular patterning of these errors.

One potential problem for this learning account is that it cannot explain the present finding of approximately equal rates of structure-dependence error for CAN and BE questions, predicting instead a lower rate for the latter. One possibility is that this account is simply incorrect. An alternative possibility is that more structure-dependence errors might have been observed for BE than CAN questions had the participants of the former study not been older (by around 20 months) than those of the latter. A third possibility is that functional-semantic similarities between simple and complex phrases play a more important role than surface-distributional similarities in demonstrating to children that simple and complex phrases can be substituted for one another (i.e., that language is structure dependent).

The most significant implication of this learning account is that arguments about whether or not the input might contain some small percentage of complex yes/no questions (e.g., Pullum & Scholz, 2002; Yang, 2002) are not directly relevant to the question of whether the child could learn to form such questions without an innately specified principle of structure dependence. Under this account, the child does not need to hear *any* complex questions in order to be able to produce them correctly. The studies of Lewis and Elman (2001) and Reali and Christiansen (2005) demonstrate that even input that contains *no* complex questions can contain, in principle at least, sufficient evidence to allow the learner to infer the correct form of such questions.

It is also important to note that under a version of this learning account where children acquire questions as an independent construction (e.g., *[AUXILIARY/COPULA] [NP] [VERB/ADJECTIVE]?;* Ambridge et al., 2006; Rowland & Pine, 2000), movement rules that derive questions from declarative utterances (or similar) play no role. Under this view, the question of whether children might incorrectly formulate a structure-independent rule (e.g., *move the first auxiliary to the front of the sentence*) does not arise, because children do not form questions by "moving" *any* auxiliary. Thus the claim that because children do not (often) make structure-dependence errors, this principle must be an innately-specified constraint simply does not follow given a different set of assumptions about the adult endpoint.

Of course, no single study can address all the issues that arise from this controversial topic. We, therefore, conclude our discussion with some suggestions for future work in this area. One major finding of the present study was that, counter to the prediction derived from the SRN models of Lewis and Elman (2001) and Reali and Christiansen (2005), structure-dependence errors were no more frequent with BE than CAN questions. However, the present study did not compare rates of structure-dependence error for these auxiliaries with a single group of children. Future experimental work should address this shortcoming. Further computer modelling work should also be conducted in order to quantify the predicted difference in rates of structure-dependence error (and auxiliary-doubling error) for different auxiliaries.

Another unexpected finding was that rates of auxiliary-doubling error were relatively high, constituting around a quarter of all valid questions produced. Similar studies with simple questions (Ambridge et al., 2006; Guasti, Thornton, & Wexler, 1995) have found that auxiliary-doubling errors are virtually non-existent for positive questions with the auxiliaries BE and CAN. One possible reason for this discrepancy is that the questions in the present study included much longer noun phrases between the fronted auxiliary and the gap from which it notionally moved. This raises the possibility that the redundant auxiliary might represent some kind of resumptive form (similar to *Type 2* errors such as *Can the boy who can run fast, can he jump high?*). This NP-length explanation is consistent with an account whereby auxiliary-doubling errors represent memory failure, with the child "defaulting" to an item that has high transitional probability (e.g., an auxiliary following a NP). Because the strength of representation of a particular item in a SRN decays over time, one might expect models of the type investigated by Lewis and Elman (2001) and Reali and Christiansen (2005), which make auxiliary-doubling errors, to also show this NP-length effect. Future computer modeling work could be used to make predictions regarding the extent to which rates of auxiliary-doubling error increase as a function of NP length (and perhaps lexical auxiliary), which could then be tested on children.

As we discussed, the SRN models would appear to predict more structure-dependence errors with plural than singular subjects for questions with auxiliary CAN because, whilst both contain legal bi-grams in the crucial position (e.g., *who crawl*), structure-dependence errors with plural subjects contain legal tri-grams (e.g., *babies who crawl*). The present study found only a non-significant trend in this direction (9 vs. 7 errors). However, the study of Reali and Christiansen (2005) found that, for some question types at least, tri-gram statistics are no more useful than bi-gram statistics in predicting the correct structure. Future computer modelling work, focussing on auxiliary CAN, should be conducted to determine whether, in fact, a higher rate of structure-dependence errors is to be expected for plural than for singular CAN questions, and future experimental work with children should investigate whether this is the case. Any study comparing rates of structure-dependence error would have to include either a large number of trials or a large number of participants because such errors are so infrequent.

In conclusion, the present study has demonstrated that the claim that children simply do not make structure-dependence errors in complex question formation (Crain & Nakayama, 1987) is false. Our findings, together with those of Lewis and Elman (2001) and Reali and Christiansen (2005) also call into question the conclusion that low or zero rates of structure-dependence error necessarily imply that the principle of structure dependence is part of the child's innate endowment. The findings of Experiment 2 demonstrate that children's question production is influenced by surface distributional properties of their input; properties that, in principle at least, could be used to infer the correct structure of complex yes/no questions.

Notes

1. In this article, as is conventional in the relevant literature, the term *auxiliary* is used to refer to the copula and the modal auxiliaries as well as true auxiliaries. In fact, in the structure dependence literature, copula and auxiliary BE are generally treated as interchangeable. For example, Crain and Nakayama (1987) stated that "all test sentences contained two occurrences of the auxiliary verb *is*" (p. 528), although one of the two occurrences of *is* was in fact a copula in all but one of the test sentences.
2. Another possible movement rule based on linear order—move the final auxiliary to sentence initial position—outputs an ungrammatical question from statements such as, *The girl is kissing the boy who is smoking*. As Yang (2002, p. 109) noted, there are, in principle, an infinite number of possible rules (e.g., "Front the auxiliary verb whose position in the sentence is a prime number"), but this concern is not relevant here.
3. The categories presented to the model were *noun, verb* (including auxiliaries)*, adjective, numeral, adverb, determiner, pronoun, preposition, conjunction, interjection, complex contraction, abbreviation, infinitive marker,* and *proper name*.
4. To check that strings such as *who run* are indeed more frequent in the input data than strings such as *who running*, we searched the British National Corpus for the example strings *who run/who smoke* and *who running/who smoking*. The respective frequencies were 165, 53, zero, and zero. Thus, although strings like *who smoking* are not, strictly speaking, "illegal" (e.g., *Is the boy who smoking offends still here?;* Gualmini & Crain, 2004, p. 4), it is probably fair to say that they have a frequency of approximately zero. Note that these figures represent chunk frequencies (i.e., the frequency of a particular string) as opposed to transitional probabilities (i.e., the probability of occurrence of a certain word, given the previous word). The two measures are highly correlated, and the transitional probability between two items that, together, have a chunk frequency of zero (e.g., *who running*) will also be zero. Reali and Christiansen's (2005) bi-/tri-gram analysis used transitional probability statistics.
5. These errors actually have three auxiliaries. We term them *auxiliary-doubling errors* because they have the *appearance* of errors in which the child has duplicated the auxiliary from either the main or relative clause (although we explicitly do NOT wish to claim that this extra auxiliary has been duplicated via some process such as movement without deletion), and because they are analogous to errors in simple questions (e.g., *Does she does[n't] like Coke?*) that are commonly termed *auxiliary-doubling errors* in the literature (see Ambridge et al., 2006; Guasti, Thornton, & Wexler, 1995).
6. This term is used purely descriptively. We do not wish to imply that children are actually making a structure-dependence error and erroneously inserting the agreement-marking *–s* morpheme, and we do not include such errors in our counts of structure-dependence errors. We also agree with an anonymous reviewer that such errors are more plausibly

analysed as auxiliary-doubling errors with a *short-form* relative clause. The point in using the *descriptive* term *auxiliary – doubling + agreement error* is to highlight the fact that the equivalent form for a sentence with a plural subject does constitute a structure-dependence error (although, see the discussion of this point in the results section).

7. It is interesting to note that a similar proposal is put forward by Santelmann, Berk, Austin, Somashekar, and Lust (2002, p. 819) who argued that apparent non-inversion errors in simple yes/no questions (e.g., *dat's the owner* for *is dat the owner?*) could instead "represent both auxiliary 'overmarking' [i.e., doubling, B.A.] and 'omission' of the initial auxiliary: *is dat's the owner?*" and thus that such "errors ... cannot be taken as conclusive evidence either for or against grammatical knowledge of inversion."
8. Although this finding may seem somewhat counterintuitive, it is not difficult to think of an example where the transitional probabilities associated with a particular bi-gram are higher, on average, than those associated with a tri-gram containing that bi-gram. For example, the bi-gram *What did* will have a higher average transitional probability than the tri-gram *What did John*. This is because *What* is usually followed by one of a small number of closed-class items (i.e., auxiliaries) whereas *(What) did* can be followed by a potentially infinite number of different subjects.
9. Although in saying this we explicitly do NOT mean to claim that children necessarily actually employ SRN-like mechanisms. The point is simply that the SRN studies discussed demonstrate that the input data can contain information that is sufficient in principle to allow children to infer the correct structure of (at least some types of) complex yes/no questions.

Acknowledgments

This research was funded by an Economic and Social Research Council (ESRC) Post-doctoral Fellowship awarded to Ben Ambridge (PTA–026–27–0705) and an ESRC Grant to Ben Ambridge, Julian Pine, and Caroline Rowland (RES–000–22–1540). We thank the children, parents, and teachers of St. Peter's Church of England Aided Primary School and Barnston Primary School who took part in the research reported here. Thanks also to Morten Christiansen, Jeff Elman, Florencia Reali, William O' Grady, Janet Fodor, and five anonymous reviewers for helpful discussions and comments on a previous version of this manuscript.

References

Ambridge, B., Rowland, C. F., Theakston, A., & Tomasello, M. (2006). Comparing different accounts of children's non-inversion errors in object wh-questions: What experimental data can tell us? *Journal of Child Language, 30*, 519–557.

Bernstein-Ratner, N. (1984). Patterns of vowel modification in motherese. *Journal of Child Language, 11*, 557–578.

Chomsky, N. (1959). A review of B. F. Skinner's *Verbal Behavior. Language, 35*, 26–58.

Chomsky, N. (1972). *Language and mind.* New York: Harcourt Brace Jovanovich.

Chomsky, N. (1975). *Reflections on language.* New York: Pantheon Books.

Crain, S. (1991). Language acquisition in the absence of experience. *Behavioral and Brain Sciences, 14*, 597–650.

Crain, S., & Nakayama, M. (1987). Structure dependence in grammar formation. *Language, 63*, 522–543.

Fodor, J. D. (1998). Unambiguous triggers. *Linguistic Enquiry, 29*, 1–36.
Gomez, R. L., & Gerken, L. A. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*, 109–135.
Gualmini, A., & Crain, S. (2004). Operator conditioning. In *Proceedings of the 28th annual Boston University Conference on Language Development* (pp. 1–12). Somerville, MA: Cascadilla.
Guasti, M., Thornton, R., & Wexler, K. (1995). Negation in children's questions: The case of English. In *Proceedings of the 19th annual Boston University Conference on Language Development* (pp. 228–239). Somerville, MA: Cascadilla.
Hornstein, N., & Lightfoot, D. (1981). Introduction. In N. Hornstein & D. Lightfoot (Eds.), *Explanation in linguistics: the logical problem of language acquisition* (pp. 9–31). London: Longman.
Hyams, N. (1986). The theory of parameters and syntactic development. In T. Roeper & E. Williams (Eds.), *Parameter detting* (pp. 1–22). Dordrecht, The Netherlands: Reidel.
Kam, X. N. C., Stoyneshka, I., Tornyova, L., Fodor, J. D., & Sakas, W. G. (2005). *Non-robustness of syntax acquisition from n-grams: A cross-linguistic perspective.* Paper presented at the 18th Annual CUNY Sentence Processing Conference, April, University of Arizona.
Labov, W., & Labov, T. (1978). Learning the syntax of questions. In R. Campbell & P. Smith (Eds.), *Recent advances in the psychology of language* (pp. 1–44). New York: Plenum.
Lewis, J. D., & Elman, J. L. (2001). Learnability and the statistical structure of language: Poverty of stimulus arguments revisited. In *Proceedings of the 26th annual Boston University Conference on Language Development* (pp. 359–370). Somerville, MA: Cascadilla.
Lidz, J., Waxman, S. R., & Freedman, J. (2003). What infants know about syntax but couldn't have learned: Experimental evidence for syntactic structure at 18 months. *Cognition, 89*, B65–B73.
MacWhinney, B. (2004). A multiple process solution to the logical problem of language acquisition. *Journal of Child Language, 31,* 883–914.
Maratsos, M., & Kuczaj, S. (1978). Against the transformationalist account: A simpler analysis of auxiliary overmarking. *Journal of Child Language, 5*, 337–345.
Maye, J. & Weiss, D. (2003). Statistical cues facilitate infants' discrimination of difficult phonetic contrasts. In *Proceedings of the 27th annual Boston University Conference on Language Development* (pp. 508–518) Somerville, MA: Cascadilla.
MacWhinney, B. (2004). A multiple process solution to the logical problem of language acquisition. *Journal of Child Language, 31,* 883–914.
Piattelli-Palmarini, M. (Ed.). (1980). *Language and learning: The debate between Jean Piaget and Noam Chomsky*. Cambridge, MA: Harvard University Press.
Pinker, S. (1987). The bootstrapping problem in language acquisition. In B. Macwhinney (Ed.), *Mechanisms of language acquisition* (pp. 339–441). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
Pullum, G. K., & Scholz, B. C. (2002). Empirical assessment of stimulus poverty arguments. *Linguistic Review, 19*, 9–50.
Reali, F., & Christiansen, M. H. (2005). Uncovering the richness of the stimulus: Structure dependence and indirect statistical evidence. *Cognitive Science, 29*, 1007–1028.
Rowland, C. F. (2007). Explaining errors in children's questions. *Cognition, 104,* 106–134.
Rowland, C. F., & Pine, J. M. (2000). Subject-auxiliary inversion errors and wh-question acquisition: What children do know? *Journal of Child Language, 27*, 157–181.
Rowland, C. F., Pine, J. M., Lieven, E. V. M., & Theakston, A. L. (2005). The incidence of error in young children's wh-questions. *Journal of Speech, Language and Hearing Research, 48*, 384–404.
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.
Santelmann, L., Berk, S., Austin, J., Somashekar, S., & Lust, B. (2002). Continuity and development in the acquisition of inversion in yes/no questions: Dissociating movement and inflection. *Journal of Child Language, 29*, 813–842.
Stemmer, N. (1981). A note on empiricism and structure-dependence. *Journal of Child Language, 8,* 649–663.
Stromswold, K. (1990). *Learnability and the acquisition of auxiliaries.* Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge.

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. R. (2001). The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of Child Language, 28*, 127–152.
Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition, 40*, 21–81.
Valian, V., & Casey, L. (2003). Young children's acquisition of wh- questions: The role of structured input. *Journal of Child Language, 30*, 117–143.
Yang, C. D. (2002). *Knowledge and learning in natural language*. Oxford, England: Oxford University Press.

Note: Appendices 2 and 5 are not included here, but are available at http://www.cogsci.rpi.edu/CSJarchive/Supplemental/index.html

Appendix A

Target questions for Experiment 1

Transitivity	No. of Subjects	Target Question
Intransitive	Singular	Can the boy who can run fast jump high?
		Can the girl who can dance sing?
		Can the baby who can crawl walk?
		Can the bird that can swim fly?
Intransitive	Plural	Can the boys who can run fast jump high?
		Can the girls who can dance sing?
		Can the babies who can crawl walk?
		Can the birds that can swim fly?
Transitive	Singular	Can the boy who can play the piano play the guitar?
		Can the monster that can eat people eat elephants?
		Can the man who can drive cars drive tractors?
		Can the girl who can make cakes make biscuits?
Transitive	Plural	Can the boys who can play the piano play the guitar?
		Can the monsters that can eat people elephants?
		Can the men who can drive cars drive tractors?
		Can the girls who can make cakes make biscuits?

Appendix C: Structure dependence errors observed in Experiment 1

Can babies who crawl can walk?
Can birds that swim can fly?
Can girls that make cakes can make biscuits?
Can girls who dance can sing?
Can monsters that eat people can eat elephants?
Can <monsters eat> [//] monsters that eat people can eat elephants?
Can the baby which crawl can walk?
Can the bird that swim can fly?
Can the boy that run fast can jump high?

Can the boy who jump [/] who run fast can jump high?
Can the boys that jump high can run fast?
Can the boys who run fast can jump high?
Can the girl who dance can sing?
Can the girl who make cakes can make biscuits?
Can the men that drive cars can drive tractors?
Can the monster who eat people can eat elephants?

Appendix D

Target questions for Experiment 2

Match–Mismatch of Number Features of NPs	Number Feature of NPs in Main Clause	Target Question
Match	Singular	Is the boy that is washing the elephant tired?
		Is the girl that is brushing the dog kind?
		Is the cat that is chasing the mouse naughty?
		Is the lion that is eating the man hungry?
	Plural	Are the boys that are washing the elephants tired?
		Are the girls that are brushing the dogs kind?
		Are the cats that are chasing the mice naughty?
		Are the lions that are eating the men hungry?
Mismatch	Singular	Is the boy that is washing the elephants tired?
		Is the girl that is brushing the dogs kind?
		Is the cat that is chasing the mice Is the lion that is eating the men hungry?
	Plural	Are the boys that are washing the elephant tired?
		Are the girls that are brushing the dog kind?
		Are the cats that are chasing the mouse naughty?
		Are the lions that are eating the man hungry?

Appendix F: Structure dependence errors observed in Experiment 2

Are cats who chasing the mice are naughty?
Are cats who chasing the mouse are naughty?
Are lions who scaring the man are hungry?
Are the boys that washing the elephants are tired? (3 children)
Are the boys who washing the elephant are tired?
Are the cats that chasing the mice are naughty? (2 children)
Are the cats who chasing the mice are naughty?
Are the cats who chasing the mouse are naughty?
Are the girls that brushing the dog are kind?
Are the girls that brushing the dogs are kind?
Are the girls that washing the dog is kind or not?
Are the girls who brushing the dog are kind?

Are the lions who eating the man are hungry?
Is cat what catching the mouse is naughty?
Is the cat that chasing the mice are # naughty?
Is the cat that chasing the mice is naughty?
Is the cat that chasing the mouse is naughty?
Is the girl that brushing the dogs is kind?
Is the lion that eating the man is hungry?
Is the lion that eating the men are hungry?
Is the lion that eating the men is hungry?