The effect of infelicity on children's sensitivity to Weak Crossover

Evidence from Dutch and English

Marie-Elise van der Ziel and William Philip
UiL-OTS, Utrecht University

1. Introduction

The only possible interpretation of the pronoun in sentences like (1a) is the deictic interpretation. This sentence can only mean that a certain mother tickled all of some set of boys. The pronoun cannot be interpreted as a variable bound by the quantifier phrase *each boy*; the reading whereby each boy is tickled by his own mother is unavailable. The same applies to sentences like (1b): the pronoun can only be interpreted deictically; a bound-variable interpretation of the pronoun is ungrammatical.

(1) a. *$**His**_i$ **mother** tickled each boy$_i$.
     b. *Who$_i$ did **his**$_i$ **mother** tickle t$_i$?

What the sentences in (1) have in common is that at the level of logical form (LF) the operator, the pronoun, and the trace of the operator occur in a configuration of the type schematically represented in (2). In both sentence types an operator moves (either overtly or covertly) across a pronoun to a position from which it c-commands this pronoun and is co-indexed with it.

(2) LF: $*_{CP}[QP_{i\ TP}[\ _{DP}[\ Pron_i\ NP]\ _{VP}[V\ t_i]]]$

The prohibition against interpreting the pronoun as coindexed with the operator in such a configuration, i.e. the prohibition against a bound-variable interpretation of it, is known as the Weak Crossover (WCO) Constraint.[1]

Many researchers have attempted to reduce the WCO-constraint to more general grammatical constraints on binding (e.g. Koopman & Sportiche 1982; Reinhart 1983; Safir 1984), to constraints on scope (Pica & Snyder 1995; Ruys 2000), or to constraints on prominence relations (Bresnan 1998). Regardless of the question which of these approaches is most appropriate, given the widely accepted assumption in competence theory that the WCO-constraint derives from

Universal Grammar (UG), and given the equally widely accepted assumption in UG-based first language acquisition (L1A) theory that children have full access to UG from the onset of L1A (Strong Continuity, see e.g. Crain 1991), children should acquire knowledge of the WCO-constraint once they have correctly identified the grammatical morphemes in the target language that function as quantifiers and pronouns. This much is predicted by the standard theory. Since children have completed this mapping task by the age of 3 (if not earlier), they should have knowledge of the WCO-constraint by this age, and certainly by the late age of 5. Any experimental evidence which contradicts this therefore presents a challenge to both UG theory and to UG-based L1A theory.

This paper offers a critical examination of experimental observations in the literature which would seem to be problematic for the prediction of early L1A. However, we will argue that this problem is only apparent. We suggest that children's violations of the WCO-constraint are in fact due to the infelicitous presentation of test sentences. Hence, we conclude that the Strong Continuity hypothesis can be maintained. In Section 2 we review previous experimental research. Next, in Section 3, we present a pair of TVJ experiments with Dutch and English children in which we test the prediction of our felicity hypothesis, i.e. that children will correctly reject WCO-sentences when these are presented felicitously.

2.   Review of Prior Experimental Research

2.1 Thornton (1990)

Using a Truth Value Judgment paradigm, Thornton (1990) tested five English-speaking children (age range 3;7–4;2) on their knowledge of the WCO-constraint in wh-questions such as (1b). The participants were presented with three trials of experimental conditions of the kind in (3). After establishing a discourse context by telling a story and acting out scenes using toy props, the experimenter delivered the test sentence (in boldface) as a question to be answered by a puppet. The child's task was to judge the correctness of the puppet's answer.

(3)   Discourse Referents: 4 Ninja Turtles and their respective trainers (8 in total)

      Discourse Context: 3 Ninja Turtles, Michelangelo, Raphael and Leonardo, are each given a weapon by their own trainer, and 1 Ninja Turtle, Donatello, is given a weapon by the trainer of another Ninja Turtle.

      Experimenter: **Who did his trainer give a weapon to?**

      Puppet: This one and this one (= Raphael and Leonardo)

Since children who had acquired the WCO-constraint should not be able to allow a bound-variable interpretation of the pronoun in the test sentence, they were expected to judge the puppet's answer to this question as incorrect. In contrast, children without knowledge of this constraint should be able to assign a bound interpretation, and hence consider the puppet's response to be correct, given that Raphael and Leonardo did each receive a weapon from their respective trainers. Thornton tested questions with *who* and with *which N*. For the *who*-questions the children correctly rejected the puppet's answer about 80% of the time, while for the *which N*-questions they correctly rejected the puppet's answer only 40% percent of the time. It should be noted that in adult performance, too, the acceptability of bound-variable readings in these two sentence types appears to be different: *which N*-questions are more readily assigned an ungrammatical bound-variable interpretation of the pronoun (e.g. Lee 1996). In sum, Thornton's findings suggest that by the age of 4, children have fully acquired the WCO-constraint; their performance is essentially adult-like, even with respect to a systematic adult error arising with the D-linked operator *which*.

Lewis (2000) argues that Thornton's findings are unclear, since the answer to the test question contains only a *subset* of the referents for which a bound-reading would be true. In (3) the puppet refers to Raphael and Leonardo only; he does not mention Michelangelo, who also received a weapon from his own trainer. The puppet's answer could therefore be perceived as incorrect even under a bound-variable interpretation of the pronoun in the test sentence. The children's correct rejection of the puppet's answer, then, may not have been due to knowledge of the WCO-constraint, but rather to an adult-like pragmatic principle ruling out incomplete responses to questions. In hypothesis testing this problem constitutes a type 1 error, since it biases the experiment towards adult-like falsity judgments, causing one to reject the null hypothesis too quickly.

2.2 Lewis (2000)

Lewis (2000) offers a slight modification of Thornton's design so as to eliminate the potential confounding effect of 'incomplete answers'. In Lewis' study the puppet's answer included all the referents in the domain presupposition under a bound-variable interpretation of the question. (The children were presented with a single trial of the experimental condition only.) This is shown in (4), with the test sentence represented in boldface.

(4)   Discourse Referents: Elmo, Grover, Oscar, Cookie Monster, and their four respective mothers (8 in total)

Discourse Context:	All four protagonists are kissed by their own mother, except for Oscar who refuses to be kissed by his own mother. Subsequently, Oscar is kissed by Grover's mother. At this point Sherlock, the Sesame Street detective, enters the story, inspects lipstick marks, and delivers the test sentence.
Sherlock:	**I think I know who his mother kissed: Elmo, Grover and Cookie**.

Lewis presented a total of 58 children (age range 2;11–7;1) with the experimental task.[2] Of those children no fewer than 49, i.e. 84.5% of the sample, accepted Sherlock's statement as being correct, suggesting that they did not apply the WCO-constraint in their interpretation of the test sentence. Moreover, all correct rejections of the answer were made by children older than 5. Lewis' findings thus cast doubt on Thornton's, and they suggest that children have not fully acquired the WCO-constraint before the age of 5.

2.3 Su (2001)

Unlike Thornton and Lewis, Su (2001) tests children's comprehension of Quantifier WCO-sentences (as in (1a)). Using a TVJ paradigm similar to that of the previous studies, Su tested 17 English children (age range 4;4–6;6, mean age 5;6) on sentences like *His dog licked each dwarf*. The discourse context involved a story about a boy and his dog who meet three dwarves, each of whom also has a dog. The boy's dog refuses to lick the dwarves, and in the end each dwarf is licked by his own dog. The children correctly judged the test sentence to be false 94% of the time. At first blush the children's adult-like performance would seem to indicate complete acquisition of the WCO-constraint, at least as far as quantifier sentences are concerned. However, a potential problem of the experimental condition is that the contextual support for a grammatical deictic interpretation of the pronoun is too strong: the story was "about a boy" and the things that happened to him. It is therefore questionable whether Su's experiment truly addresses the question of whether the children can assign a bound-variable interpretation. Rather, Su's findings may simply show that children whose grammar allows both deictic and bound-variable interpretations can be guided by discourse factors to select the former rather than the latter when disambiguating pronoun reference. Like Thornton's study, Su's experiment is strongly biased towards the expected adult-like response, and thus also constitutes a type 1 error.

Su's study highlights an intrinsic problem for L1A research on the WCO-constraint which uses sentences of the kind in (1a). On the one hand, if the context

provides a salient antecedent under a deictic reading, as in Su's study, then an adult-like response does not demonstrate that a child has knowledge of the WCO-constraint. On the other hand, if the context does not provide a clear antecedent under a deictic reading, then the test sentence will be infelicitous under an adult-grammatical reading. As has been noted on many occasions (e.g. Grimshaw & Rosen 1990; Crain et al. 1996), an infelicitous test sentence in a TVJ experiment may elicit responses which do not reflect children's grammatical knowledge. When confronted with an infelicitous pronoun, the listener — child or adult — must make inferences about the speaker's intended meaning in order to assign verifiable truth-conditions to the sentence in question. For adults this "accommodation" (Lewis 1979) is generally restricted to inferences about the speaker-intended antecedent of the pronoun. For children, however, inferences about the intended meaning of the speaker can go beyond the constraints imposed by the lexical and structural content of the linguistic input. As Donald & Lloyd (1974: 82–83), in their discussion of children's miscomprehension of sentences with quantifiers, note, what can happen is that "the child derives from the experimenter's words a notion of the kind of question he is to consider; but he derives the precise question that he does consider from his own encoding of the physical array." In other words, the risk of this kind of set-up is that children will make judgments about what the puppet *meant* to say, rather than about what the puppet actually said.

3.   Experiment 1: Dutch Children

This study attempted to finesse the felicity problem for WCO-sentences of the kind in (1a) by presenting the test sentence in a discourse context that neutrally supported both a deictic and a bound variable reading of the pronoun. This was achieved by having the puppet mention two discourse referents immediately prior to delivering the test sentence, as illustrated in (5). The causal relation between the two sentences in (5) was not overtly signalled by the presence of a word like *so*; rather, it was inferable from the discourse context preceding (5).

(5)   [Tom and Pete] were sad; (so) **his mother kissed each boy**.

In (5) pronoun disambiguation still requires some accommodation. However, since two possible antecedent choices are made salient by their close proximity to the pronoun and since the selection of either would falsify the test sentence under a grammatical deictic reading, it was hypothesized that this presentation of the test sentence would significantly reduce the number of wild inferences about the speaker-intended meaning. At the same time, this context does not introduce a bias in favor of the adult-grammatical deictic reading. If the child's grammar does

not rule out a bound-variable reading of the pronoun, the plural interpretation of *Tom and Pete* could also function as the presupposed set for *each boy*.

3.1 Participants

A group of 44 Dutch preschool children drawn from an elementary school in Utrecht were interviewed. The experiment included only those children who showed a 100% adult-like performance on control conditions (see Table 1 below); this amounted to a total of 21 "attentive" children (8 boys, 14 girls; age range 4;4–6;1; mean age 6;3). It excluded children who failed to master the experimental task (the main cause of the "yes-response" bias), leaving a reduced sample of 21 "attentive" children (8 boys, 14 girls; age range 4;4 to 6;1; mean age 6;3). The experiment also included an additional 8 native adult speakers of Dutch who functioned as a control group.

3.2 Design and Procedure

The experiment used a picture-story guessing game version of the TVJ task in which the child judges the guesses of a puppet about pictures which are visible to the child but not to the puppet. The material consisted of a set of short stories about the adventures of two boys, Tom and Pete. In each story Tom and Pete interact with a pair of secondary characters related to them by kinship. As the story unfolds, the puppet makes guesses about subevents at various points. These guesses function as a control for the children's attention, their mastery of the experimental task as well as their required linguistic knowledge (see below). The test condition, which occurred towards the end of the story, is given in (6). (The italicized text following each picture represents the storyteller's lines.)

(6)   Discourse Referents: Tom and Pete and their two mothers (4 in total)

Discourse Context:	Picture 1: *Two boys, Tom and Pete, went to the fair with their mothers. There the boys got some cotton candy.* Picture 2: *However, each boy thought the other had gotten more cotton candy, and they started fighting.* Picture 3: *Tom's mother got angry and scolded both boys.* Picture 4 (showing Tom's mother kissing Tom and Pete's mother kissing Pete): *The boys became sad, and so their mothers tried to comfort them by kissing them.*
Puppet: (actual input)	Tom and Pete were sad. **His mother kissed each boy**. Tom en Piet waren verdrietig. **Zijn moeder kuste elke jongen**

3.3 Results and Discussion

The results are shown in Table 1, where CT and CF represent two control conditions and WCO is the test condition. Note that the correct response to CF was a judgment of falsity, given that each boy was held by his own father. Children were presented with three trials of each of the test and control conditions.

Table 1. Results of Experiment 1

Condition	target input (= puppet's guess)	correct response	% children giving correct response	% adults giving correct response
CT	Each boy was given a peanut by his grandmother.	TRUE	100 %	100%
CF	Tom's father held onto each boy.	FALSE	100 %	100%
WCO	His mother kissed each boy.	FALSE	61.9%	100%

Although the children's performance under the WCO condition differed significantly from that of the adults ($p < 0.001$), they showed much greater sensitivity to the WCO-constraint than the children in Lewis' study. A direct comparison of the two studies is not possible due to differences in design and implementation. However, on the assumption that Dutch and English do not differ significantly with respect to the WCO-constraint, and that both studies involved comparable samples, the marked improvement in the children's performance suggests that the experimental design of the present study was on the whole quite successful. One remaining question is why the children in the present study did not show 100% adult-like performance under the WCO condition. However, before addressing this, we must first establish whether the results of experiment 1 are representative of the target child population. We therefore replicated this experiment study with a larger sample of children of the same age range. The results of this follow-up study are described in Section 4.

4. Experiment 2: Dutch and English Children

The aim of this experiment was to replicate the results of experiment 1 using different materials of the same type, a slightly less stringent inclusion criterion and a larger sample, which included both English and Dutch children of the targeted age range.

4.1 Participants

The children for this experiment were drawn from 2 Dutch elementary schools near Utrecht (urban environment) and a daycare centre in rural upstate New York. None of the Dutch children had participated in the previous study. As in Experiment 1, not all of the children who were interviewed were included in the study. However, the inclusion criterion was relaxed slightly: only those children who gave incorrect responses under more than one control condition (the same controls as used in experiment 1) were excluded.[3] The age and gender statistics of the samples are shown in Table 2 below.

Table 2. Included Participants

language	group	N	age range	mean age	#male	#female
Dutch	children	93	4;4–6;6	5;9	44	49
	adults	20	14–49	25	10	10
English	children	20	3;8–5;7	4;6	12	8

4.2 Design and Procedure

Experiment 2 had the same mixed design as experiment 1 and used the same experimental paradigm and similar materials, but included only one trial of the test and control conditions. The Dutch children were originally divided into two groups, with one being tested on a version of the materials with the quantifier *elk* 'each/every' and one on a version with the quantifier *ieder* 'each/every'. Since the two groups did not differ significantly in their performance under any
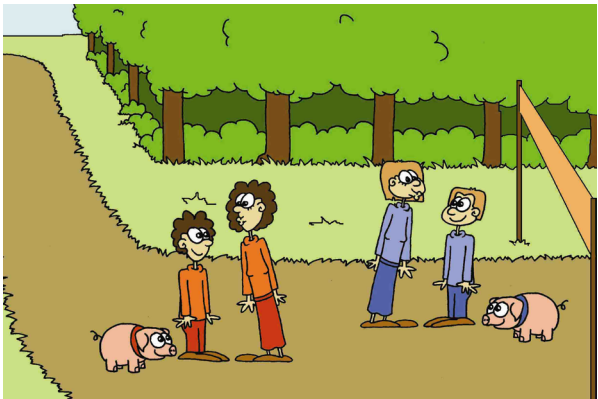

Figure 1. Picture for WCO Condition

Puppet: Tom en Piet waren blij. **Zijn moeder kuste elke/iedere jongen.**
　　　Tom and Pete were happy. **His mother kissed each boy**

experimental condition, they have been pooled together as single large sample. The English children received a translation of the materials which used *each* as the quantifier in the test and control conditions. (*Every* could not be used since the domain presupposition always had a cardinality of 2.) The picture accompanying the WCO-condition is shown in Figure 1, together with the puppet's guess (= test sentence).

4.3 Results

The results are summarized in Figure 2. While the Dutch adults all showed a perfectly grammatical performance under the WCO-condition, only 65% of the Dutch children and only 74% of the English children correctly judged the puppet's guess to be false. The performance of the Dutch and English children did not differ significantly ($p < .5$). The performance of the Dutch children was significantly different from chance ($p \geq .038$). The performance of the English children was not, which is most likely due to the small size of the sample. The combined performance of the Dutch and English participants also differed significantly from chance performance ($p \geq .010$).
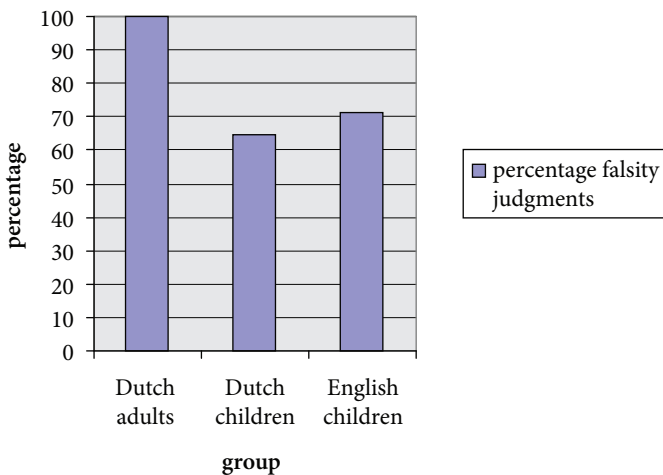

Figure 2. Results for Dutch and English

4.4 Discussion and Conclusion

The findings of experiment 2 replicate almost exactly those of experiment 1. Moreover, the English children that we tested on a translation of the Dutch materials showed the same performance as the Dutch children. We can therefore make the tentative empirical generalization that in our experimental set-up the performance

of children between the ages of 5 and 6 indicates knowledge of the WCO-constraint at least about 65% of the time. This is a greater-than-chance adult-grammatical performance, and it therefore partly bears out the prediction of our felicity hypothesis as well as the general claim of standard UG-based L1A theory (which includes the Strong Continuity Hypothesis). However, the question remains why the children's performance at this age is only 65% adult-like rather than 100%. This is not what UG-based L1-acquisition theory predicts.

We do not yet have an explanation for why the prediction of UG-based L1A theory is not fully borne out. However, we think that this should not detract from the importance of the results of our experiment. False predictions are more useful than true ones, as they stimulate further empirical and theoretical research. With this in mind, we conclude by offering two speculative hypotheses about why children between the ages of 5 and 6 do not always reject bound-variable readings of WCO-sentences in off-line experimental contexts.

One possible explanation for why the children in our study obeyed the WCO-constraint only 65% of the time is that the technique that we adopted to reduce infelicity did not eliminate this confound completely, such that for some children (roughly 35%) the residual infelicity of the pronoun in contexts such as (5) was sufficient to sabotage the TVJ task. If this is indeed what happened, it raises the question whether TVJ tasks, or any other off-line comprehension task for that matter, are suitable for L1A research on the WCO-constraint. This does not imply, of course, that the prediction of UG-based L1A theory is untestable, only that it must be tested by other means, e.g. with an on-line comprehension task.[4]

Another possible explanation for our results is that knowledge of the WCO-constraint is not fully acquired in all children until a relatively late age. This would have significant implications for UG-based L1A theory, since it entails that at least one aspect of UG "matures" (cf. Borer & Wexler 1987) or "emerges" (like the Cardinality Principle of counting, cf. Sarnecka & Carey 2008). In either case, the Strong Continuity Hypothesis would have to be relativized in such a way that only some aspects of UG are available from the onset of L1A. This would in turn lead to new empirical questions (e.g. to which aspects does the Strong Continuity Hypothesis apply?) and new theoretical questions (e.g. why does the Strong Continuity not apply to all aspects of UG?).

Notes

1. The constraint on interpretation is generally taken to be weaker than a related effect known as "strong crossover", which is observed in sentences containing a pronoun in argument position c-commanding and coindexed with an operator trace position, rather than one from within a

complex DP occupying an argument position. Strong Crossover is generally reduced to Principle C of canonical Binding Theory.

2. According to Lewis (2000:516), "there were 5 six-year olds, 18 five-year-olds, 26 four-year-olds, 7 three-year-olds [and] two children at the age extremes".

3. The inclusion criterion for this experiment was relaxed because a closer look at the data of experiment 1 revealed that the performance of the excluded participants, e.g. those participants who failed on more than one control item, was very similar to the performance of those subjects that showed "perfect performance" on control conditions.

4. Use of a production rather than a comprehension task would not solve the problem since an observation that children never produced WCO sentences with an ungrammatical bound variable reading in an experimental context would not show that they were in principle unable to do so.

References

Borer, H. & K. Wexler. 1987. "The maturation of syntax". *Parameter Setting* ed. by T. Roeper & E. Williams, 123–172. Dordrecht: Reidel.

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

Crain, S, R. Thornton, C. Boster, L. Conway, D. Lillo-Martin & E. Woodams. 1996. "Quantification without qualification". *Language Acquisition* 5.83–153.

Grimshaw, J. & S. T. Rosen. 1990. "Knowledge and obedience: the developmental status of the binding theory". *Linguistic Inquiry* 21.187–122.

Koopman, H. & D. Sportiche. 1982. "Variables and the bijection principle". *The Linguistic Review* 2.139–160.

Lee, G. 1996. "Referentiality and weak crossover". *CLS 32* ed. by L.M. Dobrin, K. Singer & L. McNair, 225–236. Chicago: University of Chicago Press.

Lewis, D. 1979. "Score keeping in a language game". *Journal of Philosophical Logic* 8.339–359.

Lewis, J. D. 2000. "Functional wh and language acquisition: weak crossover and wh/quantifier asymmetries in child language". *BUCLD 24* ed. by S. C. Howell, S. Fish & T. Keith-Lucas, 510–521. Somerville, MA: Cascadilla Press.

Reinhart, T. 1983. *Anaphora and semantic interpretation*. London: Croom Helm

Ruys, E. 2000. "Weak crossover as a scope phenomenon". *Linguistic Inquiry* 31.513–539.

Safir, K. 1984. "Multiple variable binding". *Linguistic Inquiry* 15. 603–539.

Sarnecka, B.W. & S. Carey. 2008. "How counting represents number: what children must learn and when they learn it". *Cognition* 108.662–674.

Su, Y.C. 2001. Scope and specificity in child language: a cross-linguistic study of English and Chinese. PhD diss., University of Connecticut.

Thornton, R. 1990. Adventures in long-distance moving: the acquisition of complex wh-questions. PhD diss., University of Connecticut.

Authors' addresses

Marie-Elise van der Ziel
UiL-OTS, Utrecht University
Janskerkhof 13a
3512 BL Utrecht
The Netherlands

m.e.vanderziel@uu.nl

William Philip
UiL-OTS, Utrecht University
Trans 10
3512 JK Utrecht
The Netherlands

W.C.H.Philip@uu.nl