The Fleeting Isomorphism Effect
Anastasia Conroy, Jeffrey Lidz and Julien Musolino

1 Introduction

To comprehend a sentence in real time, a listener must parse a string of words into a syntactic structure and obtain a meaning, all while taking into account the conversational goals of the speaker and the relevance of the utterance to the discourse context. It is no surprise, then, that children frequently exhibit non-adult behavior in sentence comprehension. As has often been observed, such non-adult behavior may derive from a variety of sources, including (but not limited to) a grammar that diverges from adults’, processing difficulties, limitations in the ability to relate a statement to a context, or limitations in the ability to determine the discourse function of an utterance. Thus, understanding language development requires an understanding of the contribution of each of these factors to non-adult behavior. Sources of evidence regarding such non-adult behavior will include information about the developmental timecourse for the particular phenomenon in addition to independent evidence about the grammar, the parser and the child’s pragmatic abilities.

Much research suggests that adults, in processing a sentence, entertain a number of alternative readings simultaneously, and use a combination of structural preferences, lexical frequency, and discourse information to select from them, although the exact nature of these interactions is debated (Frazier and Fodor, 1978; Boland and Blodgett, 2001; McRae, Spivey-knowlton, and Tanenhaus, 1998). In this paper, we will refer to the generation of alternatives and the mechanisms of selection as the ‘parser’, and view the discourse as one source of information that contributes to parsing decisions, but is separate from the parser. In children, there is evidence that discourse integration abilities are emerging, and are not fully adult-like (Trueswell, Sekerina, Hilland, and Logrip, 1999), preliminarily motivating the view that, in children, discourse
integration abilities should be viewed separately. With this view in mind, if we observe a behavioral difference between children and adults, we are interested in determining if this difference is due to inherent processing limitations of the children’s linguistic processing mechanism, or due to their abilities to determine the linguistic function of contextual information.

One class of constructions which has been used to investigate children’s language processing abilities is scopally ambiguous sentences containing a quantificational NP and negation. These sentences contain multiple possible interpretations, and are a domain in which children’s behavior is observed to differ from that of adults’. A long-standing finding is that children prefer to interpret sentences with a quantified NP and negation according to their surface scope interpretation, dispreferring interpretations that require covert movement (Musolino, Crain and Thornton, 2000). This observation is called the ‘isomorphism effect’, and is particularly striking because, in some cases, children’s preference for surface scope interpretations persists despite an overwhelming preference for inverse scope in adult usage, and therefore, the child’s input.

Given that children appear to differ from adults in this domain, we would like to know how the isomorphism effect can inform our understanding of the developing parser, the developing pragmatic system and the interactions between them. Although across many experiments, children fail to obtain inverse scope interpretations, this restriction is not absolute. In particularly rich discourse situations, children can overcome their preferences and achieve inverse scope interpretations ((Musolino and Lidz, 2006), (Gualmini, 2004)). In addition, in sentences lacking negation, children appear to behave like adults in their understanding of sentences involving quantification (Lidz, McMahon, Syrett, Viau, Anggoro, Peterson-Hicks, Sneed, Bunger, Flevaris, Graham, Grohne, Lee, and Strid, 2004). Both findings, together, provide strong evidence that children possess grammars that permit an adult-like understanding of sentences containing quantification. However, the exact cause of children’s non-adult behavior has not been pinpointed, as previous studies have focused on the relative contributions of the grammar and processing abilities, and have not investigated the different contributors to processing abilities.

It is possible that children’s non-adult behavior can be attributed to an immature parser (Lidz and Musolino, 2002). There are different ways in which a child’s parser may be immature. One, while we know that the adult parser is capable of handling multiple representations, children,
who are limited in working memory resources, may be unable to handle multiple representations simultaneously (Grodzinsky and Reinhart, 1993). Two, while adults maintain the ability to select (according to probabilistic mechanisms) an interpretation, children may be restricted in selecting only the interpretation that mimics the surface structure in scope. Three, if acquiring adult-like interpretations requires revision, it may be the case that children lack the resources to perform this operation. These possibilities, for this paper, will be considered ways in which children’s parsers may differ from adults.

A second point of view claims that children’s shortcomings derive from an inability to integrate contextual information in an adult-like way (Musolino and Lidz, 2006). In this domain, there are also a variety of explanations for why children may exhibit non-adult behavior. One, children may be unable to accommodate pragmatic infelicities, as it has been shown that children are extremely sensitive to pragmatic manipulations (Gualmini, 2004). Two, it may be the case that children do not have any access to discourse information in their environment, thus resulting in a failure to use such information in guiding their choice of linguistic representation. Three, even if children have access to discourse information, they may not know the linguistic function of this information. Four, if children do recognize the discourse function of an utterance, children may still fail to use this type of information as a trigger to perform revision of their initial linguistic analysis, thereby ultimately failing to integrate the discourse information into their language processing (Novick, Trueswell, and Thompson-Schill, 2005). Pragmatic and discourse information, in this paper, will be considered as external to the capability of the parser. We leave open the question of how these components precisely interact in adults, but simply note that, at the conclusion of sentence interpretation, all components are ultimately integrated.

We cannot distinguish between all of the hypotheses we have outlined above, but we can distinguish between parsing and discourse factors at the most general level. Differentiating between these two natural classes would lend insight to understanding the faculties that the child brings to the task of understanding language. This paper presents an experiment designed to weigh the impact of these contributing factors to the isomorphism effect.

Previous research has identified both the presence and absence of an isomorphism effect across a wide variety of constructions and developmental timepoints, suggesting that the iso-
morphism effect has a developmental trajectory that is poorly understood. An assumption that is implicit in most previous studies is that the presence of non-adult-like behavior represents the starting point of the developmental trajectory. This, however, is an assumption subject to empirical investigation. In this paper, we attempt to determine the developmental trajectory of the isomorphism effect. We test children across a wide age range on sentences containing a subject universal quantifier and negation, a construction which has demonstrated a powerful isomorphism effect (Musolino, Crain, and Thornton, 2000). We find that although there exists a time during which children obtain mainly isomorphic interpretations, this occurs after a period of adult-like behavior, suggesting that the isomorphism effect does not inform us about a child’s starting state. Any explanation for the isomorphism effect must account both for why it occurs and also for why it shows a late onset. Any theory treating the effect as due to an inherent limitation of immature parsers or conversational agents must surely be incorrect. We argue that children have the ability to calculate inverse scope at a very early age, and the isomorphism effect occurs as a by-product of children’s increasing ability to integrate discourse information in sentence comprehension.

2 Scope Ambiguity and the Isomorphism Effect

Sentences that contain a quantifier and negation yield interpretive ambiguity.

(1) Two horses didn’t go to the show

(2) Every horse didn’t jump over the fence

The sentences in (1) and (2) are scopally ambiguous. For example, (2) can mean either none of the horses jumped (the isomorphic, or surface scope, interpretation) or not every horse jumped (the non-isomorphic, or inverse scope, interpretation). However, not all sentences with a quantified expression and negation are ambiguous.

(3) The detective didn’t find someone

(3) has two potential interpretations: it’s not the case that someone was found or someone remained unfound. However, in adult English, only the inverse scope interpretation obtains.
While both children and adults prefer the surface scope interpretation for (1), children show non-adult behavior with the constructions in (2) and (3). For these constructions, children show a preference for the surface scope interpretation, while adults prefer the inverse scope interpretation. We will now review the studies that show these findings.

The adult’s preferred interpretation for sentences containing negation and a universal quantifier in subject position is the inverse scope interpretation, but Musolino, Crain and Thornton (2000), found that five year olds exclusively access the surface scope interpretation.

Musolino et al. conducted a study using a Truth Value Judgment Task (TVJT), in which a story is acted out with characters, and a puppet attempts to describe what happened in the story. The child’s task is to judge whether the puppet was right or wrong about what happened in the story. The task is designed so that the truth of the two interpretations of the target sentence differ, so that the child’s judgment can be taken as a reflection of their interpretation of the target sentence.

Musolino et al. tested sentences as in (4).

(4) Every horse didn’t jump over the fence

Children are presented with a story in which there are three horses, and ultimately, two jump over a fence, and one fails. In this scenario, the inverse scope reading, not all of the horses jumped over the fence, is true because one horse failed to make it over the fence. The surface scope reading, all of the horses failed to jump over the fence, is false, because two horses succeeded in jumping the fence.

Musolino et al. found that 0% of 5 year olds and 15% of 6 year olds interpreted the target sentence under the inverse scope interpretation. This contrasts with the adults, who accepted the inverse scope interpretation 100% of the time. Musolino and Lidz (2006) replicated this effect with children aged 5;0-5;11, who accepted the inverse scope interpretation only 15% of the time.

The observation of isomorphism has also been found in sentences like (5).

(5) Two horses didn’t go to the show
Musolino and Lidz (2005) showed that children prefer the surface scope interpretation for sentences like (5), but this is uninformative with respect to the observation of isomorphism, because adults also prefer the surface scope interpretation. Additionally, it is possible that numbers, as indefinites, behave differently than other quantifiers (Kraemer, 2000). For this reason, we primarily focus on constructions like (4).

The observation of isomorphism persists with sentences like (6).

(6) The detective didn’t find someone

Musolino et al. tested sentences like (6) in a scenario where a detective found some boys who were hiding, but failed to find others. The inverse scope interpretation of (6) is true, because there exist some boys the detective failed to find. The surface scope interpretation is false, because it is not true that the detective found no one. Musolino et al. tested 2 groups of children. The younger group, aged 3;10-5;2 (mean 4;7), accepted the inverse scope interpretation 35% of the time. The older group, aged 5;2-6;6 (mean 5;7), accepted the inverse scope interpretation 65% of the time. Adult controls accepted the target sentence 100% of the time. Musolino et al. conclude that there exists an isomorphism effect, as younger children more frequently reject the inverse scope interpretation than older children and adults. This result is perhaps even more surprising than the universal quantifier results, as the surface scope interpretation is unavailable in the adult grammar.

2.1 The Source of the Effect

The studies we have reviewed show that children refrain from interpreting scopally ambiguous sentences with inverse scope interpretations. This leaves open the question of the source of this deficiency. These results could derive from a grammatical deficiency, in which children have not acquired the grammatical competence to generate inverse scope interpretations (Musolino et al., 2000). Alternatively, children could differ from adults in other ways, such as with respect to their parsers or pragmatic abilities. Multiple experiments have been conducted to show that the children’s shortcomings are not grammatical.

Musolino and Lidz (2006) showed that the preference for isomorphism could be overcome by demonstrating that when the presence of negation was made more felicitous in the target
sentence, children more readily accepted the inverse scope interpretation. Musolino and Lidz performed a TVJT with stories similar to those in Musolino et al. (2000). However, in these modified stories, all three horses jumped over a log, but only two succeed in jumping over the fence. The target sentence is shown in (7).

(7) Every horse jumped over the log, but every horse didn’t jump over the fence

In this scenario, where the presence of negation is contrastive and therefore more felicitous, acceptance rates of the inverse scope interpretation rose to 60%, significantly higher than the children in the control condition, who received target sentences like (4) and accepted the inverse scope interpretation only 15% of the time. This suggests that children’s inability to obtain inverse scope cannot be grammatical. Relatedly, Lidz et al. (2004) show that sentences requiring the same grammatical operations, but which lack negation, are interpreted by children in an adult-like fashion, suggesting that the relevant grammatical operations are present in children’s grammars.

Gualmini (2004) finds a similar result; that children can obtain inverse scope under certain pragmatic situations. However, he suggests that children’s adult-like behavior in certain contexts reflects the true nature of children’s interpretive abilities. In other words, he claims that the isomorphism effect is an experimental artifact deriving from the appropriateness of the critical sentences to the context of presentation. Gualmini noticed that the goals of the characters in the stories presented in the Musolino et al. experiments were not consistent across trials. Trials differed in the degree to which the target sentence was appropriate for the discourse context in the experiment, and Gualmini suggested these infelicitous targets gave the illusion of an isomorphism effect. Gualmini performed a TVJT with the target sentences in (8) and (9), two sentences which differ in felicity with respect to the same story, to systematically test the effects of felicity on interpretation. These sentences followed a scenario in which multiple guys hide, and the firefighter finds some, but misses others.

(8) The firefighter didn’t find some guys

(9) The firefighter didn’t miss some guys

The inverse scope interpretations (there are some guys the firefighter didn’t find for (8) and there are some guys the firefighter didn’t miss for (9)) are both true. Likewise, the surface scope inter-
interpretations are both false. However, the two target sentences differ in their felicity conditions. In order for the use of negation to be felicitous, it must be the case that, in the story, the character was attempting to perform an action, but then failed to. In the case of (8), the felicity conditions are clear, as it is the case in the story that the firefighter attempted to find all of the guys, but could not succeed due to the difficulty of the task. However, this felicity condition is not met in (9). That is, it is not the case that the firefighter’s goal was to miss the guys, and that he failed. Therefore, the target sentence in (9) is less felicitous than (9) with respect to the story shown.

Gualmini tested 30 children between 4;0 and 5;0 and found that children accepted (9) only 50% of the time, consistent with the isomorphism effect. However, children accepted (8), the more felicitous target sentence, 90% of the time, meaning the children accessed inverse scope. Therefore, Gualmini suggests the isomorphism effect is not grammatical or a result of a different parser, but a result of children’s inability to fit a sentence to context that is not pragmatically aligned.

Gualmini has shown that for quantifiers headed by some, it is possible to reduce the isomorphism effect by manipulating the contextual features. However, the same manipulations reduced, but did not eliminate the isomorphism effect in sentences with a universal quantifier phrase in subject position (Musolino and Lidz, 2006). Therefore, the question remains as to whether the integration of discourse information can entirely account for the isomorphism effect. Interestingly, Gualmini tested children nearly a full year younger than Musolino & Lidz, leaving open the possibility that the high rates of acceptance of inverse scope are a result of age-related differences. Furthermore, Musolino and Lidz (2006) claim that the isomorphism effect stems from difficulties in integrating discourse information, but provide no concrete evidence that these difficulties do not arise due to limitations in the parser, as claimed in Lidz and Musolino (2002). That is, children’s abilities to obtain the inverse scope interpretation may have been a result of discourse information aiding their ability to revise, or due to overt discourse information that compensates for their difficulty with integrating contextual information. We are interested in determining whether the isomorphism effect, if it exists as reported in numerous studies, reflects the starting point of development or whether it occurs after a period of adult-like behavior. If it is the case that the isomorphism effect appears only after a period of adult-like behavior, then this observation would undermine any account in which isomorphism is a reflection
of immaturity in sentence processing. Gualmini’s account, in which isomorphism is an artifact that occurs only when children have difficulty with integrating the proper discourse information, would also fail, unless younger children are more able to accommodate infelicity than 5 year olds. The account by Lidz and Musolino in which children are restricted to isomorphism as a result of an inability to revise fails on similar grounds. If the observation holds that 4 year olds show adult-like behavior, but 5 year olds demonstrate the isomorphism effect, then we must appeal to an account which permits declining overall performance as a product of the maturational process. We suggest that such data could only be explained by an account in which children appear to be non-adult-like as a by-product of the process of their developing discourse integration abilities (Musolino and Lidz, 2006), resulting in a temporary drop in adult-like performance as these abilities come online.

3 Current Study

We want to conduct a pragmatically felicitous experiment across a range of ages to determine the extent of the isomorphism effect. In this experiment, we test a construction which contains the universal quantifier and negation, which has been reported to show a strong isomorphism effect (Musolino et al. (2000); Lidz and Musolino (2002)). However, this construction has primarily been tested in 5 year olds, meaning we need to test a wider age range to determine when this effect emerges and if the effects seen by Gualmini are due to age-related effects.

3.1 Hypotheses

One hypothesis is that children, either due to cognitive limitations or the structure of their parser, are unable to access the inverse scope interpretation except in cases with significant contextual assistance (Lidz and Musolino, 2002). One implementation of this idea is that children have great difficulty performing the revision required to access the inverse scope interpretation. According this hypothesis, there is a genuine isomorphism effect, deriving from the difficulty required in calculating inverse scope. This hypothesis predicts difficulty with inverse scope at the earliest possible age, persisting until the child is old enough to have developed adult-like parsing resources. According to this hypothesis, there need not be any shortcoming in the child’s discourse integration abilities.
A second hypothesis is that children’s parsers are completely adult-like, and the isomorphism effect derives completely from children’s inability to accommodate infelicitous experimental conditions (Gualmini, 2004). Accommodation is a specific instance of failure to integrate discourse information in an adult-like way. According to this hypothesis, given fair experimental conditions, children will obtain inverse scope at a very early age, and we expect to find no isomorphism effect across children.

A third hypothesis is that children, although they have adult-like parsing capabilities, have difficulty integrating the proper discourse information that would yield inverse scope (Musolino and Lídz, 2006), as has been observed in other domains (Trueswell et al., 1999). The predictions of this view depend on the capabilities of the child. If it is the case that young children completely ignore discourse information, and children must learn to incorporate context into sentence comprehension, then we may expect non-adult behavior early on, increasing to adult-like performance as discourse capabilities improve. Alternatively, if it is the case that children initially attempt to match the behavior of adults, but must do so by non-adult means (due to their discourse abilities being incomplete), we may expect U-shaped behavior as their discourse integration abilities come online. If children initially appear adult-like by using an alternative mechanism, we may expect a period of non-adult behavior as children reanalyze their linguistic analysis. That is, during the period when children are first integrating discourse information into sentence comprehension, children may experience difficulty using that information to decide between two representations, creating an isomorphism effect.

### 3.2 Design

We conducted a TVJT, modeled after the universal quantifier experiment in Musolino et al. (2000). Musolino et al. report results for 5 and 6 year olds, but Gualmini tested 4 year olds (although with a different construction). Because we are attempting to determine the developmental trajectory of children’s interpretations, we require consistent data across the 4 and 5 year old age ranges with a single construction.

We conducted a TVJT, with target sentences similar to the sentence in (10). The target sentence followed a story in which three cats attempt to hide behind the sofa, but only two succeed. The
final scene is shown in Figure (1)\textsuperscript{1}.

(10) Every cat didn’t hide behind the sofa

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure1.png}
\caption{TVJT, final scene}
\end{figure}

Each child received 2 warm up stories, 6 targets and two fillers. The fillers were dynamic, so that the experimenter could choose either a true or false target sentence, depending on the child’s previous responses. This ensured that the child answered true and false roughly an equal number of times across the experiment.

Children were tested one at a time in a quiet testing room away from other children. The child was introduced to the computer, and told how to play the game. The puppet, a dog, was also on the computer screen, and gave an introduction before beginning the experiment. The entire session lasted less than 20 minutes.

\subsection{3.3 Participants}

Children were recruited from a preschool at the University of Maryland. There were two groups of children, four year olds and five year olds, each group consisting of 15 children. One child was replaced in the design for incorrectly answering both fillers. The four year old group ranged from

\textsuperscript{1}The TVJT was conducted on a computer screen, with animated characters. To the best of our knowledge, these stories mimic tasks done with act-out toys, but have the advantage of requiring only one experimenter. We performed a within-subjects experiment with 24 children comparing the computer and live-action TVJTs, and found no difference in interpretations.
and the five year old group aged 5:3-5:7 (mean 5:4). The ages of these two groups are statistically significant (p < .01). These age groups were chosen to maximally match the previous studies conducted. Gualmini (2004) tested children younger than 5:0, while Musolino and Lidz (2006) tested children older than 5:0. Furthermore, Musolino et al. (2000) found an age difference in their study around the age of 5:2. Therefore, we selected 5:2 as a separation point, although we make no claims that critical changes occur at this exact age. Further research must be done to determine the causes underlying these observed age differences.

Additionally, this study was conducted with a control group of 12 adults from the University of Maryland. Participants received payment for their participation.

3.4 Results

Adults accepted the puppet’s statement 76% of the time. This acceptance rate is lower than is standardly reported for a TVJT, and we have no explanation for this fact. Performance with fillers across all groups is over 90%, suggesting overall good performance on the task.

We found that the four year olds accepted the inverse scope interpretation 81% of the time. The four year olds accepted the sentence significantly more often than the five year olds (t(14)=2.04, p < .05), but at the same rate as adults. All children were asked for justification for their responses. All children gave responses consistent with their true/false answers.

The five year olds accepted the inverse scope interpretation 44% of the time. This is rate is significantly different from the four year olds, and marginally significantly different from the adults (t(25)=2.05, p < .06). The marginal significance results from the lower adult rates of acceptance. The five year old group contains 7 children who never access the inverse scope interpretation. This suggests that these children are failing to access the inverse scope interpretation, and the low acceptance rate is not distributed evenly across children.

3.5 Discussion

Musolino et al. and Musolino and Lidz report low rates of acceptance of the inverse scope interpretation, with data from 5 year olds. We replicated this ‘isomorphism effect’ with 5 year
olds with the universal quantifier in subject position. However, we found that the isomorphism effect holds only in a limited window in language development, and is not an adequate picture of the starting point of acquisition.

Let’s recall the hypotheses and predicted results. If it is the case that children have non-adult parsers, we expect younger children to behave just like the five year olds, in being restricted to the isomorphic interpretation. If it is the case that the isomorphism effect is a result of infelicitous materials, then we too, expect similar rates of acceptance to persist into the four year olds. If it is the case that children’s development in integrating discourse impacts interpretation, we may expect to see preferences fluctuate over time.

The four year olds accessed inverse scope at the same rates as adults, leading us to conclude two things. First, because younger children can obtain the inverse scope interpretation, and presumably younger children do not have more parsing resources than older children, we conclude that the isomorphism effect in five year olds cannot be due to immaturity of the sentence parser, as claimed in Lidz and Musolino (2002). Although previous research has shown that there is certainly an effect of felicity on interpretation, in this experiment, we see interpretation differences across age, and not as a product of methodology. Therefore, we conclude that the isomorphism effect does not solely derive from a failure to experimentally meet felicity conditions. Instead, it appears that children adhere to a U-shaped development in the domain of scope ambiguity resolution. If it is the case that children are developing the ability to integrate discourse information into their processing of language, it is possible that emerging abilities in this area may temporarily limit their ability to access inverse scope. This research gives further support to the hypothesis put forth in Musolino and Lidz (2006).

This new evidence comports well with data from frozen-scope languages. Goro (2007) has found that although Japanese adults and five year olds do not permit inverse scope interpretations in sentences with multiple quantificational NPs, four year olds do, similar to English-speaking four year old children. This type of data is clearly incompatible with a hypothesis which states that children’s parsers have extreme difficulty computing inverse scope. Instead, it seems that children compute inverse scope, even preferring inverse scope interpretations, regardless of evidence of these interpretations, at very young ages.
4 Conclusions

Previous research on the acquisition of scope ambiguity presents a fractured view of children’s performance. Using a subject universal quantifier and negation, we replicate the long-standing ‘isomorphism effect’ with five year old children. However, we find that four year olds readily obtain the inverse scope interpretation at adult levels. This finding casts doubt on the hypothesis that the isomorphism effect is a result of immaturity of the parser, as it is standardly believed that one’s parsing capacity would not decrease with age. Furthermore, this finding casts doubt on the view that the isomorphism effect derives strictly from children’s difficulty in accommodation (Gualmini, 2004). It is no doubt that children’s interpretations are determined by a complex combination of pragmatic information and parsing costs: likely causing shifts in preferred interpretations over time.

We suggest that children’s parsers are fully adult-like, and that children can process inverse scope at the earliest testable age. We have suggested that children’s temporary non-adult behavior results from children’s emerging abilities to integrate discourse information. While younger children appear adult-like, they may produce this behavior with non-adult means. As the child develops adult-like resources for discourse integration, this process produces a temporary drop in adult-like performance. This hypothesis leaves open the question of why the integration of discourse information causes a change in interpretation, and also warrants an investigation of the differing discourse abilities between 4 and 5 year olds.

Acknowledgments

This research was funded by an NSF grant to the second author (#BCS-0418309). We would like to thank all of the subjects at the University of Maryland for participating in these studies. Also, thanks to all the kids and parents at the CYC who make this research possible! We especially thank Kristin Koch, who created the materials for, and tested all of the children for, the follow-up experiment comparing children’s interpretations with toys and computers.
References


