Title
Modeling scope ambiguity resolution as pragmatic inference:Formalizing differences in child and adult behavior

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 39(0)

Authors
Savinelli, K. J.
Scontras, Gregory
Pearl, Lisa

Publication Date
2017

Peer reviewed

Modeling scope ambiguity resolution as pragmatic inference: Formalizing differences in child and adult behavior

K.J. Savinelli, **Gregory Scontras**, and **Lisa Pearl**
{ksavinel, g.scontras, lpearl} @uci.edu
University of California, Irvine

Abstract

Investigations of scope ambiguity resolution suggest that child behavior differs from adult behavior, with children struggling to access inverse scope interpretations. For example, children often fail to accept *Every horse didn't succeed* to mean not all the horses succeeded. Current accounts of children's scope behavior involve both pragmatic and processing factors. Inspired by these accounts, we use the Rational Speech Act framework to articulate a formal model that yields a more precise, explanatory, and predictive description of the observed developmental behavior.

Keywords: Rational Speech Act model, pragmatics, processing, language acquisition, ambiguity resolution, scope

Introduction

If someone says "*Every horse didn't jump over the fence*," do you think any horses made it over the fence? If you think not, then you've interpreted this utterance as something like (1a). In contrast, if you think it's possible some horses made it, you've interpreted this utterance as something like (1b). These two different interpretations are possible because the utterance is scopally ambiguous. That is, it contains two scope operators: a quantifier (*every*=$\forall$) and a negation (*n't*=$\neg$). Either element can take scope over the other (indicated as >> in (1)), and so yield two different interpretations.

(1) *Every horse didn't jump over the fence.*
  a. $\forall >> \neg$ (surface scope):
     None of the horses jumped over the fence.
  b. $\neg >> \forall$ (inverse scope):
     Not all of the horses jumped over the fence.

While adults can access both interpretations given appropriate context, 5-year-old children typically struggle to obtain the inverse scope in (1b) (Musolino, 1998; Lidz & Musolino, 2002; Musolino & Lidz, 2006; Musolino, 2006; Viau et al., 2010; Tieu, 2015). For example, in a context where two out of three horses did in fact jump over the fence, only the inverse scope interpretation in (1b) is true. Adults charitably interpret the ambiguous utterance in a way that makes it a true statement (i.e., with the inverse scope in a two-out-of-three scenario), but 5-year-olds stick with the surface interpretation in (1a), which is false. Why does children's behavior differ from adults' in this context?

Previous accounts of children's scope interpretation behavior have recognized that both processing and pragmatic factors may contribute to non-adult-like behavior. Musolino (1998, 2006) observed that the surface scope interpretation in (1a) may be easier to process because the scope relationship in the semantics (i.e., $\forall$ scopes over $\neg$) aligns with the linear order of these elements in the utterance (i.e., *Every* precedes *n't*). In contrast, for the inverse scope interpretation in (1b), this isomorphism does not hold, with the scope relationship (i.e., $\neg$ scopes over $\forall$) opposite the linear order of the elements in the utterance. Musolino hypothesized that this lack of isomorphism would make the inverse scope interpretation more difficult to access. In line with this prediction, Conroy et al. (2008) found that when adults are time-restricted, they favor the surface scope interpretation. We thus see a potential role for processing factors in children's inability to access the inverse scope. Perhaps children, with their still-developing processing abilities, can't allocate sufficient processing resources to reliably access the inverse scope interpretation.

In addition to this processing factor, Gualmini et al. (2008) noted that discourse properties, such as what children consider the *question under discussion* (QUD), may impact their scope interpretation behavior. Formal theories of pragmatics suggest that all discourse transpires with respect to some QUD, whether implicit or explicit; utterances in the discourse need to (at least partially) answer the QUD to be pragmatically felicitous (Roberts, 2012). Gualmini and colleagues (Hulsey et al., 2004; Gualmini et al., 2008) suggest that children are very sensitive to this requirement. In particular, children may be able to access the inverse scope interpretation but nonetheless choose the surface scope interpretation because it better answers the perceived QUD in the contrived experimental setups. So, children's observed behavior would derive from a still-developing ability to manage the contextual information available and correctly infer the intended QUD.

Thus, children's developing processing and pragmatic abilities may both be a source of the observed non-adult-like behavior (Viau et al., 2010), though current experimental studies have struggled to clearly isolate the influence of each type of factor. To this end, we formally articulate the mechanism of scope ambiguity resolution using the Bayesian Rational Speech Act (RSA) computational modeling framework (Frank & Goodman, 2012; Goodman & Frank, 2016) in order to identify the separate contributions of processing and pragmatic factors.

We first summarize key experimental results from the literature on child scope ambiguity resolution, noting three core variables (one processing, two pragmatic) that affect children's scope disambiguation behavior. We also highlight the nature of the task children are being asked to engage in, which we then formally articulate using an RSA model that specifies the role of each of these three variables. Our results suggest that pragmatic factors play a larger role than processing factors in explaining children's non-adult-like scope ambiguity resolution behavior, and the computational modeling framework allows us to understand exactly why that's so. These results additionally suggest targeted future behavioral experiments to verify the impact of the specific pragmatic factors we identify. More generally, our model yields a more precise, explanatory, and predictive description of the observed developmental scope ambiguity resolution behavior.

Background: Experimental results

Children's ability to access the inverse scope interpretation has been shown to be sensitive to manipulations of experimental context. The methodology typically used to assess children's scope disambiguation is the Truth Value Judgment Task (TVJT; Crain & McKee, 1985). In the basic TVJT, children are presented with a background story about the actors—for example, horses engaging in some activities. After this background story, children watch as the horses attempt to complete an action, such as jump over a fence. The critical **not-all** result state meant to prompt the inverse scope interpretation is illustrated in Figure 1, where the white horse fails to jump over the fence.

Figure 1: Sample **not-all** scenario from Musolino and Lidz (2006): 2 of 3 horses succeed at jumping over the fence.

In this scenario, the surface scope interpretation of the scopally-ambiguous utterance *Every horse didn't jump over the fence* (i.e., **none** of the horses jumped over the fence) is false, and the inverse scope interpretation (i.e., **not all** of the horses jumped over the fence) is true. A puppet then says the scopally-ambiguous utterance, and the child is asked to state if the puppet is right. That is, the child is asked whether s/he would endorse the puppet's utterance as a true description of the scenario. Typically, children refuse to endorse the puppet's utterance, saying that the puppet is wrong. This behavior has been interpreted as children failing to access the inverse scope interpretation that would make the utterance true.

Interestingly, various alterations to the TVJT setup have yielded more adult-like behavior in children, namely greater rates of endorsing the puppet's ambiguous utterance in not-all scenarios. Musolino and Lidz (2006) observed that negation in an utterance might require certain felicity conditions to be met. In particular, negated utterances require a preceding *affirmative context* to contrast with (Wason, 1965). Musolino and Lidz augmented the basic TVJT to include an additional contrast condition in which the puppet precedes its negative scopally-ambiguous utterance with a contrasting affirmative clause. This additional clause describes a previous successful story action (i.e., *early-success*), such as *Every horse jumped over the log, but every horse didn't jump over the fence*. This early-success contrast manipulation increased children's willingness to accept the scopally-ambiguous utterance in the not-all scenario: Children in the baseline condition endorsed the puppet's statement just 15% of the time, while children in the early-success affirmative context condition endorsed the puppet's statement 60% of the time. Viau et al. (2010) later replicated this increase in utterance endorsement using only an early-success story context. That is, the utterance endorsement rate was maintained by an early-success story context alone, and children didn't need an explicit contrast clause in the test utterance.

Notably, the early-success affirmative context manipulation potentially changes several aspects of the experimental context. First, it can shift participants' *expectations* about successful outcomes in the experimental world. This shift then potentially increases the salience of a QUD targeting this success, such as *"Did all the horses succeed?"* (`all?`). Recognizing this QUD's potential significance, Gualmini (2004) attempted to manipulate the experimental context so it favored the `all?` QUD. With `all?` as the salient QUD, children's endorsement of a scopally-ambiguous utterance that perfectly answers `all?` in the critical not-all scenario increased to 90%. Even for a scopally-ambiguous utterance that does *not* answer the `all?` QUD, children's endorsement rate was at 50%—markedly higher than the 15% baseline from the original study by Musolino and Lidz (2006). This finding highlights that privileging the `all?` QUD increases children's utterance endorsement in these scenarios.

A third potential impact of the affirmative context manipulation involves scope access. By altering the experimental world expectations and/or QUD to increase access to the inverse scope, the inverse scope interpretation may also become more accessible for later use. Viau et al. (2010) term this *structural priming*. Children who are better able to access the inverse scope are then more likely to endorse the scopally-ambiguous utterance in subsequent not-all scenarios. Viau et al. investigated structural priming explicitly by attempting to directly alter the accessibility of the inverse scope interpretation. In one modified TVJT, they attempted to prime the access of the inverse scope interpretation, and in another modified TVJT, they attempted to directly prime the inverse scope's logical structure (e.g., $\neg >> \forall$).

The first structural priming manipulation was implemented via the now-familiar affirmative context (i.e., pragmatic) manipulation. For the first three trials, the prior experimental context indicated successful outcomes and the effect was that children endorsed the scopally-ambiguous utterance 50% of the time. Crucially, the subsequent three trials removed the supportive affirmative context manipulation—yet children continued to not only endorse the scopally-ambiguous utterance, but to endorse it more than they had before (80%). Viau et al. (2010) attribute this result to a priming effect of the inverse *interpretation* from the first three trials. Interestingly, the increase in utterance endorsement could be due to priming multiple factors that are products of the affirmative context manipulation: (i) the expectations about successful outcomes in the experimental world, (ii) the salience of the `all?` QUD, or (iii) the ease of access to the inverse scope interpretation.

The second structural priming manipulation removed the affirmative context story in the first three trials. In its place, children were asked whether they would endorse a scopally-*unambiguous* utterance (e.g., *"Not every horse jumped over the fence"*) whose interpretation had logical operators in the same order as the inverse scope interpretation of the scopally-ambiguous utterance (e.g., $\neg >> \forall$). Children endorsed this utterance 80% of the time. In the subsequent three trials, children were asked if they would endorse the scopally-ambiguous utterance in the same experimental scenario—and their endorsement rate remained at 80%. Viau et al. (2010) interpret this effect as priming of the relevant logical form: The inverse scope was easier to access in the scopally-ambiguous utterance because it was so recently accessed in the unambiguous utterances. The authors argue that this priming effect proceeded in the absence of manipulations to the pragmatic context, yet even here, there may still be pragmatic factors at work. The unambiguous utterance accomplishes three things: (i) it provides an instance of the $\neg >> \forall$ configuration, (ii) it provides information about successful outcomes, and (iii) it suggests the `all?` QUD, answering it with *no*. Thus, in this attempt to prime the inverse logical form, the authors may have also altered expectations about the pragmatic context of the experiment, as related to the successful outcomes and relevant QUDs.

These experimental studies highlight at least three core factors (two pragmatic, one processing) that underlie children's utterance endorsement behavior in the TVJT: (i) pragmatic: expectations about the experimental world (e.g., how likely successful outcomes are), (ii) pragmatic: expectations about the QUD (e.g., whether all outcomes were successful), and (iii) processing: the accessibility of the inverse scope (i.e., the ease by which the logical form is accessed). These experimental studies have also supported different theoretical proposals for the source of children's differences. The proposals split on whether they attribute the differences solely to an inability to manage contextual information (i.e., pragmatic factors; Gualmini, 2008) or whether processing deficits also significantly contribute (i.e., difficulty accessing inverse scope; Viau et al., 2010). Importantly, it is not obvious from any of the existing experimental manipulations how to separate the independent contributions of these components. To capture and independently manipulate the contributions of each of the pragmatic and processing factors, we formalize their role in the interpretation of scopally-ambiguous utterances, using tools from probabilistic modeling.

The model

We model ambiguity resolution within the Bayesian Rational Speech Act (RSA) framework (Goodman & Frank, 2016). This framework views language understanding as a social reasoning process. A *pragmatic listener* L_1 interprets an utterance by reasoning about a cooperative *speaker* S_1 who is trying to inform a *literal listener* L_0 about the world. Our model is a "lifted-variable" extension wherein the ambiguous utterance's literal semantics is parameterized by interpretation-fixing variables (e.g., the relative scope of the quantificational elements; Lassiter & Goodman, 2013). Hearing an ambiguous utterance, a pragmatic listener reasons jointly about the true state of the world (e.g., how many horses jumped over the fence), the scope interpretation that the speaker had in mind (i.e., surface vs. inverse), as well as the likely QUD that the utterance addresses (e.g., `all?`).

To connect our model's predictions with the available TVJT data, we follow Degen and Goodman (2014) and Tessler and Goodman (2016), modeling participants' TVJT behavior as the (relative) endorsement of a *pragmatic speaker* S_2 for an utterance about an observed situation. That is, we model whether a speaker would endorse the scopally-ambiguous utterance as a description of the observed state, or whether the speaker would prefer to say nothing at all. The pragmatic speaker S_2 makes this decision by reasoning about the probability that a pragmatic listener L_1 (who is reasoning about a speaker S_1 reasoning about a literal listener L_0) would arrive at the correct world state after hearing the utterance.

We take world states $w \in W$ to correspond to the number of successful outcomes, for example, the horses that successfully jumped over the fence ($W = \{0, 1, 2, 3\}$). We assume a simple truth-functional semantics where an utterance u denotes a mapping from world states to truth values ($Bool = \{\texttt{true}, \texttt{false}\}$). We parameterize this truth function so that it depends on the scope interpretation $i \in I = \{\texttt{inverse}, \texttt{surface}\}$, $[\![u]\!]^i$: $W \rightarrow$ Bool. We consider two alternative utterances $u \in U$: the `null` utterance (i.e., saying nothing at all, and so choosing *not* to endorse the utterance) and the scopally ambiguous utterance `amb` (e.g., *"Every horse didn't jump over the fence"*). So, $U = \{\texttt{null}, \texttt{amb}\}$. The utterance semantics appears in (2), where the parameterization only impacts the truth value for utterance `amb` (since that's when multiple interpretations are available). If `inverse` is active, this corresponds to the **not all** reading, and so is true as long as not all (i.e., $w \neq 3$) outcomes were successful. If `surface` is active, this corresponds to the **none** reading, and so is only true in world state 0.

(2) *Utterance semantics* $[\![u]\!]^i$:

a. $[\![\texttt{null}]\!]^i = \texttt{true}$

b. $[\![\texttt{amb}]\!]^i =$ if $i = \texttt{inverse}$ $[\![\texttt{inverse}]\!]$, else $[\![\texttt{surface}]\!]$

where:

$[\![\texttt{inverse}]\!] = \lambda w.\ w \neq 3$

$[\![\texttt{surface}]\!] = \lambda w.\ w = 0$

We consider three QUDs $q \in Q$: (i) "How many horses made it over?" (`how-many?`), (ii) "Did all the horses make it over?" (`all?`), and (iii) "Did none of the horses make it over?" (`none?`). The QUDs serve as projections from the inferred world state to the relevant dimension of meaning, $q : W \rightarrow X$ (Kao, Wu, et al., 2014; Kao, Bergen, & Goodman, 2014). In practice, the QUDs establish partitions on the possible world states, as shown in (3): `how-many?` is an identity function on world states, `all?` returns `true` only if all three outcomes were successful, and `none?` returns `true` only if none of the outcomes were successful.

(3) *QUD semantics* $[\![q]\!]$:
  a. $[\![$`how-many?`$]\!]$ = λw. w
  b. $[\![$`all?`$]\!]$ = λw. w = 3
  c. $[\![$`none?`$]\!]$ = λw. w = 0

The literal listener L_0 has prior uncertainty about the true state, $P(w)$, and updates beliefs about w conditioned on the the literal semantics. That is, L_0 restricts prior beliefs to those worlds that $[\![u]\!]^i$ maps to `true`. The function $\delta_{[\![u]\!]^i(w)}$ maps the Boolean truth value to a probability, 1 or 0.

$$P_{L_0}(w|u,i) \propto \delta_{[\![u]\!]^i(w)} \cdot P(w)$$

To capture the notion that communication proceeds relative to a specific QUD q, L_0 must infer not only the true world state w, but also the value of the QUD applied to that world state, $[\![q]\!](w) = x$.

$$P_{L_0}(x|u,i,q) \propto \sum_w \delta_{x=[\![q]\!](w)} \cdot P_{L_0}(w|u,i)$$

The speaker S_1 chooses an utterance u in proportion to its utility in communicating about the true state of the world w with respect to the QUD q, $[\![q]\!](w) = x$. Thus, the speaker maximizes the probability that L_0 arrives at the intended x from u. This selection is implemented via a softmax function (*exp*) and free parameter α, which controls how rational the speaker will be in utterance selection.

$$P_{S_1}(u|w,i,q) \propto exp(\alpha \cdot log(L_0(x|u,i,q)))$$

Utterance interpretation happens at the level of the pragmatic listener L_1, who interprets an utterance u to jointly infer the world state w, the interpretation i, and the QUD q. We therefore model *ambiguity resolution* as pragmatic inference over an under-specified utterance semantics (i.e., the interpretation variable i). To perform this inference, L_1 inverts the S_1 model by Bayes rule, and so the joint probability of w, i, and q is proportional to the likelihood of S_1 producing utterance u given world state w, interpretation i, and QUD q, as well as the priors on w, i, and q.

$$P_{L_1}(w,i,q|u) \propto P_{S_1}(u|w,i,q) \cdot P(w) \cdot P(i) \cdot P(q)$$

To model the utterance endorsement implicit in TVJT, we need one more level of inference. The pragmatic speaker S_2 observes the true world state w and selects u by inverting the L_1 model, thus maximizing the probability that a pragmatic listener would arrive at w from u by summing over possible interpretations i and QUDs q that accompany world w.

$$P_{S_2}(u|w) \propto exp(log \sum_{i,q} P_{L_1}(w,i,q|u))$$

To generate model predictions, we must fix various model parameters. The S_1 speaker rationality parameter $\alpha > 0$ is set to 2.5. The priors $P(w)$ and $P(q)$ correspond to expectations for the discourse context (i.e., likely world states or QUDs). In the default case, we set these priors to be uniform over their possible values: $P(w{=}0) = P(w{=}1) = P(w{=}2) = P(w{=}3) = \frac{1}{4}$; $P(\texttt{how-many?}) = P(\texttt{all?}) = P(\texttt{none?}) = \frac{1}{3}$. The interpretation prior $P(i)$ corresponds to how easy it is to access the inverse scope interpretation. Experimental literature on scope ambiguity resolution suggests that speakers more readily access the surface interpretation (Anderson, 2004; Conroy et al., 2008). We model this tendency by setting these default values: $P(\texttt{surface})$=0.7 and $P(\texttt{inverse})$=0.3. Importantly, to better understand children's utterance endorsement behavior with scopally-ambiguous utterances, we can independently manipulate the values of the priors on W, Q, and I, and observe their impact on utterance endorsement.

Results

To test how pragmatic and processing factors contribute to non-adult-like utterance endorsement in the TVJT, we systematically manipulate the relevant priors to favor specific parameter values, shown in Figure 2.

For the world state prior (Figure 2, *left*), we systematically favor specific world states by setting their prior probability to 0.9; if a world state is not favored, it receives a prior probability of $0.1/3 = 0.033$. Holding the QUD and scope priors at their default values, we see a marked increase in endorsement of the ambiguous utterance in the not-all scenario as beliefs about horse success increase. Utterance endorsement is at its lowest (0.25) when prior knowledge suggests that horses are particularly unlikely to succeed at jumping; utterance endorsement is at its highest (i.e., most adult-like: 0.90) when we believe horses are very likely to succeed.

Just as with the world state prior, we can systematically manipulate the QUD prior (Figure 2, *center*). Favored QUDs receive a prior probability of 0.9; other QUDs receive a prior probability of 0.05. Holding the other priors at their default values, we see an increase in utterance endorsement from the `none?` (*did no horses succeed?*; 0.28) to `how-many?` (*how many horses succeeded?*; 0.37) to `all?` (*did all horses succeed?*; 0.64) QUDs. The model predicts the most adult-like behavior when the QUD concerns whether all the horses succeeded.

Finally, for the binary scope prior (Figure 2, *right*), we systematically manipulate the prior probability of `inverse` from 0.1 to 0.9. Holding the other priors at their default values, we see a monotonic increase in utterance endorsement as the probability of `inverse` increases. At its most adult-like, the

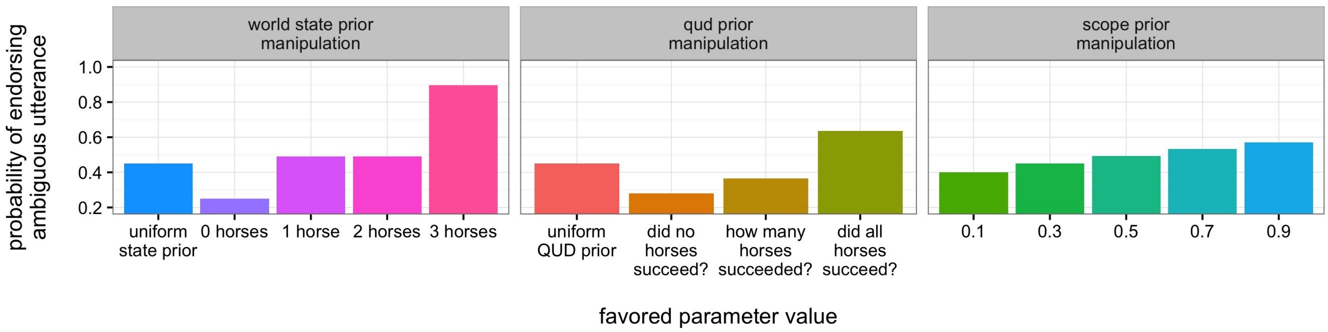

Figure 2: Model predictions for ambiguous utterance endorsement (e.g., *Every horse didn't jump over the fence*) in a not-all scenario (e.g., two-out-of-three horses jump over the fence). Lower endorsement probability corresponds to less adult-like (i.e., more child-like) behavior. For the pragmatic variables (world state, QUD), the favored parameter value receives most of the prior probability weight ($P(favored) = 0.9$). For the processing variable (scope), the prior corresponds to how strongly the inverse scope is favored.

model predicts an endorsement probability of 0.57 when the prior probability of `inverse` is at its highest (0.9)—at its *lowest* (0.1), endorsement only drops to 0.4.

To summarize, the world state and QUD priors have a more dramatic impact on utterance endorsement than the scope prior. There are two main reasons for this. First, for the world state prior, when expectations favor success (i.e., $w = 3$), the ambiguous utterance is maximally informative regardless of the scope interpretation it receives: `amb` communicates to a listener that prior expectations do not hold (i.e., ***None/Not all** of the horses succeeded* goes against the expectation that all three horses would succeed). So, `amb` is particularly useful for communicating about the *a priori* unlikely not-all world states that appear in the experimental scenarios. Second, for the QUD manipulation, when `all?` is favored, either interpretation of `amb` fully resolves the QUD: whenever `amb` is true (i.e., whether **None** or **Not all** the horses succeeded), it is not the case that all the horses succeeded. A pragmatic speaker recognizes the utility of `amb` as an answer to `all?` in a not-all world state, irrespective of the intended scope interpretation.

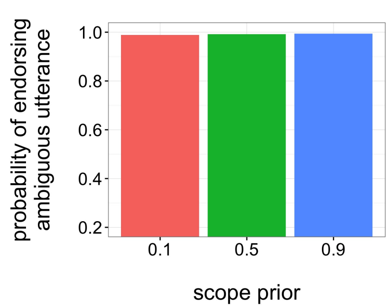

Figure 3: Model predictions for ambiguous utterance endorsement when optimal world state ($w = 3$) and optimal QUD (`all?`) are favored ($P(favored) = .9$).

So far, we have considered independent manipulations to the factors of interest. Figure 3 shows the interaction of all three factors for utterance endorsement when $w = 3$ and `all?` are favored. Here we see the additive effects of the world state and QUD priors; together, they lead to near-total endorsement of the ambiguous utterance. We also see more clearly the relatively small contribution of the scope prior, where changing the prior probability of `inverse` from 0.1 to 0.9 leads to just a 0.01 increase in endorsement probability. Thus, we see how the priors on the pragmatic factors overwhelm the processing scope prior. When the optimal (i.e., optimal for endorsement) QUD and world state are favored, even when `inverse` is highly inaccessible (i.e., $P(\texttt{inverse}) = 0.1$), we still predict massive utterance endorsement (0.99).

Discussion

Our model of ambiguity resolution qualitatively captures the changes in children's utterance endorsement from the experimental literature; our results suggest that when it comes to understanding non-adult-like behavior in the TVJT, there is a stronger role for the pragmatics of context management (as realized in priors on world state and QUD) than for grammatical processing (as realized in the prior on scope interpretations), although there is likely a role for both. So, the observed failure of children to endorse scopally-ambiguous utterances in not-all scenarios likely stems more from children's beliefs about the world of the experiment (e.g., whether horses are *a priori* likely to succeed) and about the topic of conversation (e.g., whether the conversational goal is to determine if all the horses succeeded), than their inability to grammatically derive the inverse scope interpretation in real time. Indeed, our model predicts the highest rates of utterance endorsement to occur when resolving the scope ambiguity is *irrelevant* for communicating successfully about the not-all world—that is, when expectations favor total success (i.e., $w = 3$), or when the QUD asks if `all?` of the horses succeeded. In either case, both scope interpretations serve to inform a listener, either that the *a priori* likely $w = 3$ isn't true or that the answer to the `all?` QUD is *no*.

These results also underscore the need for well-defined mapping hypotheses from observed experimental behavior to the psychological processes they inform, particularly for the sophisticated reasoning that occurs in tasks like the TVJT. In our brief review of the experimental literature, we were careful to point out alternative interpretations of the various experimental manipulations and their potential, unintended pragmatic consequences. At the very least, we hope to have demonstrated that utterance endorsement is not simple. A TVJT participant must reason recursively about the potential informativity of the utterance, attending to knowledge about the world of the experiment and the likely topic of conversation. That children stumble when attempting to perform these complex recursive inferences isn't so surprising. We suggest that a plausible source of differences in child and adult behavior on the TVJT is children's inability to successfully manage pragmatic information. We therefore propose to move the discussion away from the *fragility* of accessing inverse scope in children as a grammatical processing deficit and toward the *complexity* of behavior that scope interpretations require.

In addition to formalizing the pragmatics of ambiguity resolution in context, our results also motivate future experimental investigations that explicitly measure children's (and adults') expectations about the world and topic of conversation. We saw how past experiments did not completely deconfound the relevant factors. Perhaps the most straightforward way of testing these factors' effects is to *measure* the prior knowledge that participants bring to bear in the TVJT. These explicit measurements of pragmatic context can then form the basis of future modeling studies in this framework that could quantitatively match the behavioral results.

More generally, our results provide the foundation for more complete theories of the developmental process underlying scope ambiguity resolution. Children's relative lack of experience managing world and conversational knowledge likely contributes to their sensitivity to the experimental context. In short, five-year-olds may know the right interpretation, but they're still figuring out whether it's the best answer in the context of the experimental conversation.

References

Anderson, C. (2004). *The structure and real-time comprehension of quantifier scope ambiguity* (Unpublished doctoral dissertation). Northwestern University.

Conroy, A., Fults, S., Musolino, J., & Lidz, J. (2008). Surface scope as a default: The effect of time in resolving quantifier scope ambiguity. In *Poster presented at the 21st cuny conference on sentence processing, march* (Vol. 13).

Crain, S., & McKee, C. (1985). The acquisition of structural restrictions on anaphora. In *Proceedings of nels* (Vol. 15, pp. 94–110).

Degen, J., & Goodman, N. D. (2014). Lost your marbles? the puzzle of dependent measures in experimental pragmatics. In *Proceedings of the 36th annual conference of the cognitive science society* (pp. 397–402).

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*, 998-998.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*, 818-829.

Gualmini, A. (2004). Some knowledge children don't lack. *Linguistics*, 957–982.

Gualmini, A. (2008). The rise and fall of isomorphism. *Lingua*, *118*(8), 1158–1176.

Gualmini, A., Hulsey, S., Hacquard, V., & Fox, D. (2008). The question–answer requirement for scope assignment. *Natural language semantics*, *16*(3), 205–237.

Hulsey, S., Hacquard, V., Fox, D., & Gualmini, A. (2004). The question-answer requirement and scope assignment. *MIT working papers in Linguistics*, *48*, 71–90.

Kao, J. T., Bergen, L., & Goodman, N. D. (2014). Formalizing the pragmatics of metaphor understanding. In *Proceedings of the 36th annual meeting of the cognitive science society* (pp. 719–724).

Kao, J. T., Wu, J. Y., Bergen, L., & Goodman, N. D. (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences*, *111*(33), 12002–12007.

Lassiter, D., & Goodman, N. D. (2013). Context, scale structure, and statistics in the interpretation of positive-form adjectives. In *Semantics and Linguistic Theory (SALT) 23* (pp. 587–610).

Lidz, J., & Musolino, J. (2002). Children's command of quantification. *Cognition*, *84*(2), 113–154.

Musolino, J. (1998). *Universal grammar and the acquisition of semantic knowledge: An experimental investigation into the acquisition of quantifiernegation interaction in english* (Doctoral dissertation). University of Maryland, College Park.

Musolino, J. (2006). Structure and meaning in the acquisition of scope. In *Semantics in acquisition* (pp. 141–166). Springer.

Musolino, J., & Lidz, J. (2006). Why children aren't universally successful with quantification. *Linguistics*, *44*, 817-852.

Roberts, C. (2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, *5*(6), 1–69.

Tessler, M. H., & Goodman, N. D. (2016). *A pragmatic theory of generic language.* (http://arxiv.org/abs/1608.02926)

Tieu, L. (2015). Isomorphism for all (but not both): Floating as a means to investigate scope. *Language Acquisition*(ahead-of-print), 1–16.

Viau, J., Lidz, J., & Musolino, J. (2010). Priming of abstract logical representations in 4-year-olds. *Language Acquisition*, *17*(1-2), 26–50.

Wason, P. C. (1965). The contexts of plausible denial. *Journal of verbal learning and verbal behavior*, *4*(1), 7–11.