Some (Wh-) Questions Concerning Passive Interactions

Christopher Hirsch\textsuperscript{1} & Jeremy Hartman\textsuperscript{2}
\textsuperscript{1}Massachusetts Institute of Technology & \textsuperscript{2}Harvard University

1. Introduction

English-speaking children are known to be delayed in their comprehension of passive sentences (Slobin, 1966; Turner & Rommetveit, 1967; Bever, 1970; Maratsos & Ambramovitch, 1975; Maratsos, Fox, Becker, & Chalkley, 1985; Gordon & Chafetz, 1990; Fox & Grodzinsky, 1998; Hirsch & Wexler, 2004a). Additionally, it has been noted that children’s level of passive comprehension is in part determined by the type of verb used; children have significantly greater difficulty with passives involving “psychological” (subject-experiencer) verbs (1) than with passives of “actional” verbs (2), despite having no problems with either verb type in the active voice. This finding is confirmed in every study crossing voice and verb type (Maratsos, Kuczaj, Fox, & Chalkley, 1979; Maratsos, Fox, Becker, & Chalkley, 1985; Sudhalter and Braine, 1985; Gordon & Chafetz, 1990; Fox & Grodzinsky, 1998; Hirsch & Wexler, 2004a; Hirsch & Wexler, 2006).

(1) The boy was loved / seen / remembered by the girl.

(2) The boy was kissed / kicked / held by the girl.

It is this interaction of voice and verb type, which we refer to as the Maratsos effect after the seminal work by Michael Maratsos and colleagues, that primarily concerns us here. We examine three classes of theories offering potential explanations for this interaction. We concern ourselves with classes of theories, and not particular accounts, as we believe much can be gained from abstracting away from the details of specific theories, and focusing instead on various general approaches to the relevant phenomenon.

Three classes of theories are to be investigated: interaction/processing theories, frequency theories, and grammatical theories. The interaction/processing theories state that both non-canonical word orders and psychological verbs incur processing costs greater than canonical word orders and actional verbs, but that alone neither causes noticeable comprehension difficulties for children, while both factors together impose a great processing burden. Frequency theories generally claim that children only comprehend structures to which they have been sufficiently exposed. Given evidence that children hear many more actional passives than psychological passives, children’s comprehension asymmetry is taken to follow directly from the input asymmetry. Finally, grammatical theories claim that young children cannot
comprehend certain grammatical structures because they cannot syntactically represent them. With respect to passives, such theories predict a general difficulty for all passives, irrespective of verb type, and as such, these theories tend to offer evidence that children are able to make use of a compensatory linguistic heuristic that provides a reasonable analysis for actional passives, but fails to do so for psychological passives. The predictions of each class of theories will be considered below, and subjected to experimental and conceptual review.

2. Interaction/Processing Theories Considered

Many theorists have argued that children have difficulty comprehending sentences with non-canonical word orders (for English, non-SVO word order). This might account for children’s general difficulty with English passives (e.g., Bever, 1970; Sinclair & Bronckart, 1972\(^1\)), but such a simple canonicity theory leaves the interaction between voice and verb type unexplained. If it could be demonstrated, however, that psychological verbs incur a greater processing cost than actional verbs, then an interaction/processing story could perhaps be maintained.\(^2\) The hypothesis would be that while neither non-canonicity nor psychological verbs alone sufficiently tax children’s processing resources, the combination of both proves too much for children, where it is further assumed that children have more limited processing resources than adults.

While we know of no particular theorists who have explicitly posited such an account in the published literature, it is by no means an obvious strawman theory. In fact, this sort of explanation is eminently plausible if the following are true: (1) non-canonical sentences and psychological verbs incur greater processing costs than their canonical and actional counterparts, (2) these costs draw upon the same pool of resources, and (3) children’s pool of processing resources is great enough to handle either non-canonicity or psychological verbs, but not both simultaneously.

To examine whether such a class of theories can account for the Maratsos effect, we tested children’s comprehension of a more common non-canonical construction: object-extracted \(wh\)-questions. If non-canonical word order alone is responsible for children’s difficulties with passives, we should see a parallel deficit with object-extracted questions. Furthermore, if the Maratsos effect is due to a processing interaction of non-canonical word order and use of a psychological verb, then children should have significantly more difficulty with psychological object-extracted \(wh\)-questions compared to actional object-extracted \(wh\)-questions.

In order to test such predictions, we conducted a picture-question verbal-response task with four conditions, crossing question type (canonical subject-extracted vs. non-canonical object-extracted \(who\)-questions) and verb type (actional vs. psychological). For each experimental item, children were shown a set of three cards, depicting three characters interacting such that each character was both the
agent/experiencer and patient/stimulus for a particular actional or psychological verb. For example, for the actional verb *wash*, the three pictures might consist of Piggy washing Bunny, Bunny washing Kitty, and Kitty washing Piggy (Figure 1). In an item testing the psychological verb *see*, the three pictures might depict scenes where Piggy sees Bunny, Bunny sees Kitty, and Kitty sees Piggy (Figure 2). After the three pictures were displayed, the child was then asked either a subject-extracted or object-extracted *who*-question about one interaction among the characters (e.g., *Who does Piggy wash?* or *Who sees Kitty?*). The experiment used two actional verbs (*push* and *wash*) and two psychological verbs (*see* and *hear*). To help minimize task demands, only three characters (*Piggy*, *Bunny*, and *Kitty*) were used throughout the experiment. These characters were familiarized during the introduction to the experiment before any experimental items were presented. The entire experiment consisted of 24 total questions (six items per condition).

Each question was asked twice before children were allowed to respond, after which a pause was given for an answer. If no response was forthcoming, the question was repeated for the child indefinitely at 4- or 5-second intervals until an answer was received. Alternate names for the animals (e.g., *Rabbit* for *Bunny*, or *Cat* for *Kitty*) were accepted as correct answers. In the rare instance when a child responded that he was unsure, or took an unreasonable amount of time to provide an answer, he was asked whether he would like to come back to the item later; there was usually no problem the second time around. Rarely, a child would simply point to the correct character on the relevant card. In this case, the child was prompted to actually name the character, and the correct verbal response nearly
always followed. Condition type was pseudo-randomized, with items of the same condition never appearing back to back.

42 children were tested, ranging in age from 3;1 to 5;8 (mean age of 4;6). Children were divided by age into two equal groups of 21 subjects each. The younger group ranged in age from 3;1 to 4;4, with a mean age of 3;10. The older group ranged from 4;6 to 5;8, with a mean age of 5;1. Subjects were recruited from daycares in the Boston area. All children were native English speakers and came from families of varying socioeconomic status.

The experimental results are summarized below in Table 1. Children performed extremely well on all four conditions. Across all subjects, all four conditions were answered at better than 92% accuracy. Importantly, there was no interaction of question type and verb type ($F(1,1004) = 0.082$, $p = 0.775$), nor even a main effect of structure ($F(1,1004) = 0$, $p = 1$).³ When the children were split into the two equal groups by age, there was no hint of an interaction even in the younger (mean age 3;10) group ($F(1,500) = 0.514$, $p = 0.474$). In contrast to with non-canonical psychological passives, children perform extremely well on non-canonical psychological object-extracted who-questions.

<table>
<thead>
<tr>
<th>Group</th>
<th>Question Type</th>
<th>Actional</th>
<th>Psychological</th>
</tr>
</thead>
<tbody>
<tr>
<td>Younger Children</td>
<td>Subject</td>
<td>96.0%</td>
<td>87.3%</td>
</tr>
<tr>
<td></td>
<td>Object</td>
<td>97.6%</td>
<td>92.1%</td>
</tr>
<tr>
<td>Older Children</td>
<td>Subject</td>
<td>98.4%</td>
<td>97.6%</td>
</tr>
<tr>
<td></td>
<td>Object</td>
<td>96.0%</td>
<td>93.7%</td>
</tr>
<tr>
<td>Total</td>
<td>Subject</td>
<td>97.2%</td>
<td>92.5%</td>
</tr>
<tr>
<td></td>
<td>Object</td>
<td>96.8%</td>
<td>92.9%</td>
</tr>
</tbody>
</table>

Children’s excellent comprehension of object-extracted who-questions (both actional and psychological) demonstrates that non-canonicity itself is not responsible for poor performance. This finding fits neatly with previous corpus work demonstrating that children have no difficulty forming object-extracted questions, and that if anything, object-extracted questions appear slightly earlier in children’s natural productions than subject-extracted questions (Stromswold, 1995). The total lack of an interaction between question type and verb type clearly speaks against any general interaction/processing explanation for the Maratsos effect.

Proponents of an interaction/processing theory, however, might object that we have not correctly defined “canonical” word order. While an object-extracted wh-question is certainly non-canonical compared to active transitive sentences, perhaps what matters for interaction theories is the type of non-canonical order involved. Passives, for example, have a non-canonical argument order of patient-verb-agent.
Object-extracted *wh*-questions, on the other hand, have a non-canonical word order of patient-agent-verb. Since the Maratsos effect is noted in passives, but not in object-extracted *wh*-questions, one might try to “save” the interaction theory by refining it so that it says young children will only have problems with structures involving non-canonical word order of the passive type.

While this is certainly logically possible, the evidence available in the literature speaks strongly against such a hypothesis. The German word order for passives is patient-agent-verb, and it is known that children have trouble with passives in German (Grimm, Schöler, & Wientemantel, 1975; Bartke, 2004). Non-passive topicalized structures with the same word order, however, are known to be produced by very young German children (Poeppel & Wexler, 1993). Similarly for Japanese, passives are known to be delayed (Sugisaki, 1998; Sano, 2000), while sentences with the same surface word order derived by scrambling are comprehended without difficulty (Otsu, 1994). Thus, simple appeals to word order as accounts of children’s difficulties with passives prove untenable.

3. **Frequency Theories Considered**

Some theorists have argued that syntactic structures are only acquired once enough examples have been heard in the input. Proponents of such frequency theories often tout “verb-by-verb” accounts of structure learning as part of the package (see Gordon & Chafetz, 1990; Tomasello, 1992, 2003). It is not hard to see why verb-by-verb acquisition is attractive to frequency theorists. If one assumes that children do not formulate general rules for structures, but rather verb-specific ones, it becomes easier to view frequency as the determining factor in children’s comprehension. That is, if children represent rules on the level of the verb, it only makes sense that the verbs they hear should condition comprehension patterns.

As applied to passive acquisition, frequency theories claim that the passive will only be acquired once children hear enough token passive examples. The Maratsos effect is to be accounted for by evidence that the passives children hear are overwhelmingly actional, as first reported by Gordon and Chafetz (1990). These authors use corpus searches to demonstrate that while passives are rare in the input to children, the vast majority of the passives children do hear (92%) are actional passives. The rarity of psychological passives in the input is thus taken as the explanation for children’s problems comprehending them.

To examine whether a frequency-based theory could satisfactorily account for the Maratsos effect, we conducted a parallel corpus analysis to that of Gordon and Chafetz, searching the parental input for *wh*-questions. Just as Gordon and Chafetz compared the input frequencies of actional and psychological passives with children’s comprehension of these same structures, we examined the degree to which the input of actional and psychological *wh*-questions predicts children’s
comprehension of such structures. If a frequency theory of any sort is to remain a
tenable explanation for the Maratsos effect, it must account for any differences
between the input-comprehension relationship in passives and the input-
comprehension relationship in wh-questions.

We searched the parental input to Adam and Sarah from the Brown corpus
(Brown, 1973), the same data used by Gordon and Chafetz, for wh-questions
involving transitive verbs, classifying them by extraction position (subject vs.
object) and verb type (actional vs. psychological).

Adam’s data consisted of utterances he heard from 2;3 to 4;11 years-old, and comprised 26,178 total input
utterances. Sarah’s data consisted of utterances she heard from 2;3 to 5;1, and comprised 44,827 total input utterances. In both cases, the vast majority of the input was from the child’s parents. So that our input counts would be most relevant
to the hypotheses under consideration, and thus comparable to the comprehension
data gathered in the who-question comprehension experiment, we adhered to a
strict search procedure. Since the comprehension experiment only tested wh-
questions with transitive verbs, we only counted object-extracted and subject-
extracted who and what questions in the input containing transitive verbs. What
when used like which (e.g., What color did Bill see?) was not counted. Likewise,
NPs beginning with whose were not counted as legitimate instances of who (e.g.,
Whose car did Mary take?). Embedded questions were counted. Finally, immediate
repetitions of utterances were not counted.

While subject-extracted and object-extracted questions with both who and what
were analyzed, we were most interested in the input of actional and psychological
object-extracted who-questions. This particular sentence type makes for a nice
contrast with passives, as both are non-canonical structures involving displacement
of the logical object of a transitive verb to sentence-initial position. Furthermore,
who-questions are exactly the type of structure examined in our earlier
comprehension experiment, and thus straightforward contrasts can be made
between input frequency and comprehension levels. In order to make concrete
numerical comparisons for passives, Gordon and Chafetz’ passive input
frequencies were compared to the passive comprehension accuracies in Hirsch &
Wexler (2004a). These latter authors examined, among other things,
comprehension of actional and psychological passives in 60 children aged 3 to 5,
the same age range as the data for the Brown corpus input counts. The input
frequencies for the who-questions were compared with the comprehension of these
structures as determined by our who-question experiment. Again, the input searches
for passives conducted by Gordon and Chafetz, and the input searches for who-
object-extracted questions we carry out here are done for the very same children
(Adam and Sarah of the Brown corpus). Furthermore, the comprehension data for
passives taken from Hirsch & Wexler (2004a) and the comprehension data for who-
questions all come from children of the same age range (3-5 years-old).
Before comparing the input frequencies of passives and who-questions with their respective comprehension levels, we must first settle on the relevant level of comparison. As noted earlier, Gordon and Chafetz find that 92% of the passives Adam and Sarah hear are actional passives. They take this asymmetry in the input to account for the noted asymmetry in passive comprehension (the Maratsos effect). Our experimental results for object-extracted who-questions demonstrate no significant comprehension asymmetry (96.8% actional vs. 92.9% psychological). Interestingly, there is no asymmetry in the input ratio of actional object-extracted who-questions and psychological object-extracted who-questions (59% actional vs. 41% psychological). At first glance, then, this would seem to constitute support for frequency theories: an asymmetry in input frequency leads to a comprehension asymmetry (passives), while a lack of asymmetry in input frequency leads to a lack of a comprehension asymmetry (object-extracted who-questions). This conclusion, however, is predicated upon the assumption that what determines the relevant comparison is the ratio of one type of structure to another, as opposed to simply the absolute frequency of a particular structure in the input.

Frequency theories, however, must work over absolute frequencies, and not mere ratios. This is obvious when one considers that, if what determined the Maratsos effect was simply that children hear more actional passives compared to psychological passives, one would expect children never to acquire psychological passives, since the ratio of the two passives remains the same into adulthood. Thus, the notion of frequency that must be used by such theories is that of absolute frequency, which Gordon and Chafetz readily accept. The relevant purpose of the data under consideration, then, is to determine the absolute frequency of actional object-extracted who-questions (16 examples in the corpus) and psychological object-extracted who-questions (11 examples). The absolute frequencies of the who-questions share much more in common with the absolute frequency of psychological passives (7 examples in the corpus) than with actional passives (76 examples). Therefore, according to frequency theories, comprehension of both types of who-questions should pattern like psychological passives, which given the comprehension data, simply is not the case.

Once absolute frequencies are examined, it is clear that frequency accounts, at least as currently stipulated, cannot account for children’s varying comprehension across different sentence structures. Comparing input and comprehension for who-questions with input and comprehension for passives reveals numerous instances where similar input frequencies simply do not produce anything resembling similar comprehension levels. By the age of 3:8, for instance, Adam and Sarah (henceforth A&S) have heard 11 psychological object-extracted who-questions in the corpus, and our data indicate that children at this age comprehend these questions perfectly (98% correct). At this same age (3;8), however, A&S have heard 7 non-actional passives in the corpus, yet children score very poorly on this sentence type (54%
correct in Hirsch & Wexler, 2004a). To give another example: by age 5;1, A&S have heard 11 psychological object-extracted who-questions and 7 psychological passives (similar numbers), but while children score extremely well on psychological object-extracted who-questions at this age (94% for mean age of 5;1), they do very poorly on psychological passives (39% for a mean age of 5;3). Conversely, lower input frequencies can co-occur with higher comprehension levels: by age 5, A&S have heard 76 actional passives in the corpus, and children are 70% correct at comprehending these sentence types. By this age, A&S have heard many fewer actional object-extracted who-questions (only 30), but children are above 95% correct at comprehending these sentences. These data thus argue strongly against Gordon and Chafetz’ claim that the Maratsos effect is a cumulative reflex of the types of passives children hear.

Perhaps we have been too conservative in counting only who questions (i.e., ignoring what questions). This is a real possibility, especially given that the input frequencies rise dramatically (by a factor of nearly 40) if the latter are included, and perhaps this increase is what accounts for the comprehension differences. But this raises a number of questions about the theoretical and psychological foundations of frequency theories, and points to the need for greater specificity in explicating just what allows language learners to generalize across structures.

The assumption here is that children acquiring a language can collapse who and what object-extracted questions for the purposes of formulating a general rule, but cannot collapse object-extracted wh-questions and passives. The first question for frequency theorists is, why draw the line here? After all, both object-extracted questions and passives involve movement of an object to sentence initial position. There is nothing inherent to frequency theories of rule-learning in the literature that posits a barrier to generalization here. Indeed, there is nothing inherent to frequency-based theories of rule-learning that posits particular barriers to generalization in specific places. (Of course, without any barriers to generalization, a frequency theory would not have much explanatory power at all, but that is a more fundamental concern.) Furthermore, if children can collapse wh-question types, what accounts for children’s reluctance to collapse verbal and adjectival passives, which unlike who and what questions, are homophonous? As Gordon and Chafetz note, if adjectival passives are included in the input frequency counts, the strong actionality effect disappears. If nothing else, these issues highlight the need for frequency theorists to explain what counts as relevant input.

4. Grammatical Theories Considered

What are needed are theories able to distinguish the syntactic types of structures being tested. Once children are acknowledged to have deeper syntactic knowledge than is suggested by word order and frequency theories, motivated explanations
arise for which structures children find un/grammatical. Clearly passives and wh-questions are different syntactically, and the acquisition data strongly suggest that children are sensitive to these syntactic differences. A satisfactory account of the course of acquisition will thus by necessity involve explicit grammatical theories. Many grammatical theories have been posited over the years in an attempt to account for young children’s general difficulty comprehending passive sentences: the A-Chain Deficit Hypothesis (Borer & Wexler, 1987), the External Argument Requirement Hypothesis (Babylonyshev, Ganger, Pesetsky, & Wexler, 2001), the Universal Phase Requirement (Wexler, 2004), and Smuggling (Hyams & Snyder, 2005). All of these particular grammatical theories are maturational theories, positing that children’s early grammar lacks the syntactic means to represent passives due to biological immaturity. That is, while the syntactic mechanisms licensing passives in the adult grammar are assumed to be innately specified, it is hypothesized that these mechanisms are subject to biological development, appearing only in later childhood. Evidence that passives mature (as opposed to being “learned”) comes from behavioral genetics (Ganger, Dunn, & Gordon, 2004), the apparent cross-linguistic universality of passive delay (see Crawford, 2005), and the particular detailed pattern on passive acquisition (Hirsch & Wexler, 2006).

If children’s early grammar lacks the ability to represent/derive passives, then what accounts for the Maratsos effect, since the grammatical theories would appear to predict all passives to be uniformly delayed? The grammatical theories considered above all accept that English-speaking children make use of a strategy deriving a “passive-like” representation for actional passives, but not for psychological passives. The particular strategy involves children attempting to analyze verbal passives as adjectival passives. Since actional verbs, but not psychological verbs, form licit adjectival passives, children are able to arrive at a syntactic representation for actional passives, but not for psychological passives. There is much evidence that children’s early (actional) passives are adjectival. Horgan (1978) provides evidence that children’s early passives describe states, not events, as predicted on an adjectival analysis. Terzi & Wexler (2002) demonstrate that for a language in which the verbal passive and adjectival passive are not homophonous (Greek), the Maratsos effect does not obtain (Greek children perform very poorly on actional passives at ages where English-speaking children have nearly no problems). Finally, Hirsch & Hartman (2006) argue from experimental evidence that the class of passives acquired first is not that of paradigmatic actional verbs (e.g., hit), but that of object-experiencer verbs (e.g., scare). This is understandable in the context of an adjectival strategy, where object-experiencer verbs make even better adjectives than many actional verbs.

Further recommending (certain) grammatical theories is the fact that they straightforwardly account for (and predict) problems with other constructions involving similar syntactic dependencies, such as subject-to-subject raising (Froud,
Wexler, & Tsakali, in preparation; Hirsch & Wexler, 2004b; Hirsch & Wexler, to appear) and unaccusatives (Babyonyshev, Ganger, Wexler, & Pesetsky, 2001; Lee & Wexler, 2001; Ito & Wexler, 2002; Hirsch & Hartman, 2006). Acquisition theories that lack linguistic sophistication fail to account for these correlations, while such correlations follow naturally from many grammatical theories.

5. Conclusions

Three classes of theories were considered in an attempt to explain why children find psychological passives more difficult to comprehend than actional passives. Interaction/processing explanations, it was shown, simply make the wrong predictions about comprehension of object-extracted wh-questions and thus offer little hope. Frequency theories, at least as currently construed in the literature, suffer from a number of conceptual difficulties and also make incorrect comprehension predictions based on the comparative frequencies of various constructions in the child’s input. It is grammatical theories that best explain the available data and offer the most promising prospects of an empirically and conceptually satisfying account of the course of language acquisition.

Notes

* We gratefully acknowledge Ken Wexler for his detailed comments and discussion concerning this work. We would also like to thank Nadya Modyanova, Robyn Orfitelli, Alexandra Perovic, the entire Wexler ab/Normal Language Lab, the audiences at the Lisbon Workshop on Production versus Comprehension in the Acquisition of Syntax, and GALA 2005, as well as all the children who participated. The preparation of this article was supported in part by an NSF Graduate Fellowship awarded to the first author.

1 We mean the base ordering of logical arguments. Another way of referring to such orderings is by thematic ordering, where canonical order for English is agent-verb-theme.

2 While we know of no direct evidence that psychological verbs are harder (more costly) to process than actional verbs, this seems possible on the intuitively reasonable assumption that psychological verbs are less frequent than actional verbs, where token frequency relates to ease of retrieval (e.g., Howes & Soloman, 1951).

3 A main effect of verb type was obtained ($F(1,1004) = 9.87$, $p = 0.002$). This appears to be driven by some children’s isolated poor performance on questions involving the verb hear. Questions with the other verbs, including psychological see were all comprehended at above 95% correct, while accuracy for hear was at only 90.1%. This was likely due to some children having a difficult time interpreting the pictures associated with hear, which involved a relatively non-intuitive pictorial representations for hearing involving earmuffs.

4 Sentences with non-actional, non-psychological verbs like have and fit were not counted.

5 We omitted Eve, because her input data is only from 1;6 to 2;3. In comparing our numbers to those of Gordon and Chafetz, we have used only their combined Adam and Sarah totals.
References


Christopher Hirsch
Department of Brain and Cognitive Sciences, MIT
46-3027C
Cambridge, MA 02139 USA
ckh@mit.edu

Jeremy Hartman
Department of Linguistics, Harvard University
Boylston 3rd Floor
Cambridge, MA 02138 USA
jihartm@fas.harvard.edu