J. Child Lang. **31** (2004), 431–457.
Printed in the United Kingdom

REVIEW ARTICLE

Editor's Acknowledgement
The Editor would like to acknowledge the help given by Dr Evan Kidd of the Max Planck Institute Child Study Centre in Manchester in preparing this section of the Journal for publication.

DOI: 10.1017/S0305000904006051

Learnability and linguistic performance*

KENNETH F. DROZD
University of Aarhus, Denmark

(*Received* 14 *August* 2001. *Revised* 7 *January* 2003)

CRAIN, S. & THORNTON, R. *Investigations in Universal Grammar*. Cambridge: MIT, 1998.

Stephen Crain (C) & Rosalind Thornton (T) have garnered a well-deserved reputation for their unwavering commitment to language learnability as a constraint not only on theories of child language and language development but also on experimental design and the interpretation of experimental findings. In his well-known defense of children's early knowledge of syntactic constraints, Crain (1991) argued for the widely-held position that the best solution to the learnability problem is to assume that grammatical knowledge which cannot be learned on the basis of experience is specified in advance as part of the human biological endowment for language in the form of a UNIVERSAL GRAMMAR (UG) (Chomsky, 1965). With respect to experimental design, C&T have strongly maintained that even young children know UG constraints but perform poorly in some experiments-due to the extralinguistic demands associated with experimental tasks, particularly those involved in presupposition accommodation and complex response planning. C&T specifically design their experiments to reduce the impact of extralinguistic demands on children's linguistic performance while at the same time providing felicitous environments for adultlike performance.

[*] Address for correspondence: Kenneth F. Drozd, Institute of Language, Literature and Culture, University of Aarhus, Nobelparken, Building 467, Jens Chr. Skousvej 7, DK-8000, Aarhus C, Denmark. e-mail: engkd@hum.au.dk

In *Investigations in Universal Grammar* (IUG), C&T extend Crain's (1991) programme by adopting learnability as a constraint on models of children's language performance systems, rather than on language acquisition theories alone. IUG is intended primarily as a guide to Crain & Wexler's MODULARITY MATCHING MODEL of the language processing system (henceforth, Modularity Matching) and to the appropriate way, given the model, to design experimental tasks and interpret experimental results. However, the book is also a comprehensive summary of the theoretical assumptions, experimental design features, and experimental techniques which C&T have been promoting and developing for decades. IUG provides C&T's most explicit, up-to-date discussions of the motivations and logic behind their experimental techniques as well as extremely useful step-by-step instructions for implementing them. It also provides the most comprehensive survey available of C&T's solutions for children's non-adultlike linguistic performance and their opinions of alternative models of children's performance and alternative experimental techniques.

A thorough evaluation of the many issues discussed in IUG is beyond the scope of this article. Instead, I evaluate the two central components of C&T's present research programme. The first is Modularity Matching. I argue that although the model is much more restrictive than alternative models, C&T have no coherent approach to demonstrating either the correctness of the model or its usefulness in explaining children's linguistic performance. The second component is C&T's experimental designs and techniques. I argue that C&T's experimental designs have little explanatory power because fundamental components of experimental design are entirely missing from C&T's experimental programme. I also argue that C&T's recent experimental design innovations, the Condition of Falsification, the Condition of Plausible Dissent, and the Condition of Plausible Assent, can and should be dismissed as felicity conditions on experiments.

I begin by summarizing the main assumptions of Modularity Matching, as presented in IUG, and discussing its predictions regarding children's language performance. I then briefly summarize C&T's experimental designs.[1]

[1] IUG consists of 40 short chapters, divided into 3 parts. Part 1 (16 chapters) is devoted to the Modularity Matching Model. This section discusses the two main components of the model, the theory of UG (Chapters 2 and 3) and the Referential Theory of language processing (Chapters 12 and 13), and includes a detailed comparison of the model with the Competing Factors Model (Chapters 4–11). This section also includes a discussion of principles and preferences (Chapter 14) and the nature of performance errors (Chapter 15), and ends with a chapter on methodological preliminaries (Chapter 16). Parts 2 and 3 (Chapters 17–24 and 25–39, respectively) discuss in depth the design features for the Elicited Production and Truth Value Judgement tasks, respectively. The conclusion (Chapter 40) is followed by notes, references, and an index.

The Modularity Matching model

Modularity Matching is a strong nativist model of 'the interrelations between linguistic representations and linguistic performance' (p. 6). The model consists primarily of the Principles and Parameters Theory of UG, a model of the language processing system, the well-known Innateness Hypothesis and Modularity Hypothesis (Fodor, 1983), and what I will call the Full Processing assumption (or simply Full Processing). I assume familiarity with UG and the Innateness and Modularity hypotheses.

The language processing system assumed under Modularity Matching combines Crain, Shankweiler, Macaruso & Bar-Shalom's (1990) model of the verbal working memory system and Crain & Steedman's (1985) Referential Theory of the language processing device. Briefly, Crain *et al.* (1990) proposed that the verbal working memory system consists of a storage buffer for phonological information and a control mechanism which recodes buffer contents as structural descriptions and transfers these descriptions to different levels of linguistic representation, allowing new phonological information to fill the buffer (p. 101). Crain & Steedman's (1985) Referential Theory provides the processing principles that regulate how the parser deals with ambiguous sentences. Under this theory (p. 107), the control mechanism of verbal working memory relays each alternative syntactic description of a structurally ambiguous sentence in parallel to the semantic processor, which attempts to select the description whose meaning best fits the perceiver's current mental model of the discourse. The selection process is mediated by three processing principles. The PRINCIPLE OF REFERENTIAL SUCCESS directs the parser to favour the syntactic description that succeeds in referring to entities already established in the perceiver's mental discourse model. The PRINCIPLE OF *A PRIORI* PLAUSIBILITY directs the parser to favour the more plausible reading with regard to either general world knowledge or knowledge of the discourse universe. The PRINCIPLE OF PARSIMONY directs the parser to favour the reading that carries fewer unsatisfied presuppositions or entailments.

What gives Modularity Matching its bite is its central assumption, Full Processing. Full Processing is the assumption that the language processing components of children and adults 'match' (p. 30). They share not only the same language performance apparatus and language processing principles, but also the same language processing capacities and memory limitations (pp. 30, 34, 51, 124, and elsewhere). According to C&T, Full Processing is necessary for explaining language learnability. Allowing the language processing systems of children and adults to differ, according to them, would open 'a Pandora's box of possible processing explanations' (p. 30) for children's linguistic performance. Full Processing 'skirts' (p. 51) the problem altogether by assuming the absence of both qualitative AND QUANTITATIVE differences in the language performance systems of children and adults.

C&T explicitly invite colleagues assuming the UG framework to adopt Modularity Matching as a working research model (p. 8). But jumping on board would involve a substantial shift away from widely-accepted assumptions in UG-based developmental psycholinguistics.

One firmly entrenched assumption is that grammatical competence is only one of a variety of factors that contributes to linguistic performance. Chomsky (1965), for example, specifically claimed that linguistic performance can never actually directly reflect linguistic competence and mentions memory limitations, distractions, shifts of attention and interest, and errors as intervening factors. Another widely held assumption is the CONTINUITY ASSUMPTION (Pinker, 1984). Pinker (1984: 6–12) proposed that developmental psycholinguistics should assume as the null hypothesis that children and adults share the same grammatical rule types and symbol classes, and that child and adult grammars are realized in on-line performance using the same parsing mechanisms, scheduling routines, and memory structures. These two assumptions, together with the Principles and Parameters Theory of UG, have given rise to the popular view that children should be credited with grammatical knowledge on the basis of adultlike performance whenever possible – e.g. if their adultlike performance rate is significantly above chance by statistical measures. Equally popular is the view that children's non-adultlike performance should be attributed to parametric or maturational development in UG principles, to processing limitations, or to performance factors. Pinker's Continuity Assumption is a particularly restrictive constraint on the second strategy. Countless patterns of children's non-adultlike production and comprehension performance have been attributed to memory or processing limitations or performance factors. These include the inconsistent use of functional categories in early (telegraphic) speech (e.g. Brown, 1973), subject drop (e.g. Valian, 1990), medial *Wh* constructions (O'Grady, 1997), root infinitives (Avrutin, 1999), overregularized past tense (Marcus, Pinker, Ullman, Hollander, Rosen & Xu, 1992), and pronoun interpretation (e.g. Grodzinsky & Reinhart, 1993). Processing limitations also figure prominently in many accounts of children's non-adultlike performance found in the sentence processing literature. Children's failure to accurately recall multi-clause sentences, to exploit prosodic patterns for identifying information structure, to revise incorrect referential commitments, and to recover multiple interpretations of syntactically and semantically ambiguous sentences, all have been attributed to limitations in children's processing capacities (e.g. Tyler & Marslen-Wilson, 1978; Lorsbach, Katz & Cupak, 1998; Trueswell, Sekerina, Hill & Logrip, 1999).

Modularity Matching adopts quite different background assumptions and a different research agenda. Whereas Chomsky (1965: 4) explicitly rejected the view that actual linguistic performance directly reflects competence,

C&T (pp. 89–90, 97–8) deny that factors other than linguistic competence play any significant role in linguistic performance. With respect to children, they explicitly reject the assumption that memory, processing, and performance factors significantly (and negatively) interact with linguistic competence in children's performance (Chapters 11–13), arguing instead that 'grammatical principles alone dictate children's responses' (p. 89) in experiments. Second, whereas Pinker (1984: 6) suggested that developmental changes can be attributed to quantitative differences in language performance systems if necessary, C&T propose that children share the same processing and memory capacities as adults and 'should only be subject to the same 'processing difficulties' that curtail adult performance' (p. 34). Moreover, performance factors play only a small role in adult linguistic performance and should also have little to do with children's actual linguistic performance, contributing 'only a small level of ''noise''' (p. 89), which is expected in any experimental context.

These assumptions set the stage for a research programme unlike those typically adopted by developmental psycholinguists. Unlike most researchers, C&T are not particularly interested in the varieties of often unexpected interpretations children actually assign to sentences or in explaining children's actual linguistic performance – what children actually say. Rather, their primary concern is with 'what children do not say and on what meanings they do not assign to sentences' (p. 10) and with uncovering the experimental conditions which elicit 'optimal' (p. 6) (adultlike) performance from children. Furthermore, the Full Processing assumption prohibits many widely-accepted techniques for eliciting and interpreting child language data. C&T find act-out tasks, imitation tasks, and reaction time tasks, as well as most alternative versions of their elicitation and truth value judgement tasks, unsuitable on the grounds that they are sensitive to irrelevant task demands and, therefore, do not directly tap children's grammatical knowledge (Chapters 8–10). C&T also reject the view that children should be credited with grammatical knowledge on the basis of merely statistically significant rates of adultlike performance. In their view, only near-perfect performance is expected from both adults and children with unambiguous sentences, rendering statistical analysis superfluous.

One solid virtue of Modularity Matching is that it imposes exceptionally strong restrictions on the interpretation of children's non-adultlike behaviour. Under this model, appeals to processing limitations other than those experienced by adults to explain non-adultlike performance by children are unacceptable. The only acceptable sources of children's non-adultlike performance are (1) experimental design flaws, (2) processing tasks which either exceed the boundaries of the adult language performance system or require access to extralinguistic knowledge, (3) performance

factors/noise, and (4) parametric variation or maturation of UG principles (pp. 121–5).

Experimental design flaws, such as the presentation of test sentences in infelicitous discourse contexts, force children either to proceed with the intended grammatical analysis in violation of a pragmatic principle or to choose an unintended linguistic analysis that is consistent with pragmatic principles but may violate syntactic principles. Experimental tasks that require children to build and execute complex cognitive plans or process repeated self embedding or garden paths may exceed children's cognitive abilities and lead to degraded or non-adult performance in children (pp. 23, 98). The processing of ambiguous sentences may also tax working memory and lead to errors if alternative analyses of an ambiguous sentence need to be maintained in verbal working memory for excessive amounts of time before the ambiguity is resolved, e.g. by querying extralinguistic knowledge. Performance factors that may affect linguistic performance include fatigue, distraction, lack of attention, lack of interest, and confusion about task demands. The effects of these factors are characterized as ' noise.' Errors are attributable to performance factors if they are (1) deletions which occur (2) relatively more often for younger children, (3) for only a subset of items and (4) only at the beginning or end of an experimental session (pp. 198–9). Lastly, children's nonadult language performance may be attributed either to parametric variation in UG (p. 31) or, as a last resort, to the maturation of UG principles (p. 121).

Importantly, C&T assume that any non-adultlike performance traceable to these sources does not threaten Modularity Matching. Infelicitous experimental designs do not affect children's language processing capacities or memory capacities but create no-win situations in which a child must choose whether to violate a syntactic or a pragmatic principle (p. 240). With respect to processing, children and adults should experience the same processing limitations and therefore exhibit degraded linguistic performance under the same conditions. Performance factors affect speakers' abilities to focus on or execute extralinguistic tasks and, by assumption, have no influence on the operations of the language performance system itself. Finally, the choice of a parametric option in UG, even one that is inconsistent with the target language, sufficiently narrows down a child's hypothesis space to include those options that can later be rejected by positive evidence (given the appropriate theory of parameter setting). This therefore does not introduce learnability problems.

Experimental design and techniques

One central theme of C&T's experimental programme, the roots of which date back to C's earliest research, is to identify and fulfil the felicity

conditions that lead to adultlike performance by children under experimental conditions. C and colleagues' well-known success at improving children's performance with relative clauses by fulfilling their presuppositions in context (e.g. Hamburger & Crain, 1982) provided early solid support for children's early knowledge of UG constraints and led to a new appreciation of how pragmatic conditions influence children's performance. In IUG, C&T claim that truth-value judgement tasks must satisfy two additional felicity conditions to be considered properly designed, the CONDITION OF FALSIFIABILITY and the CONDITION OF PLAUSIBLE DISSENT (see also Crain *et al*., 1996). To see what these conditions say and how they are satisfied, consider C&T's Principle C experiment.

In this experiment, a child is asked to judge whether Kermit the Frog correctly describes what happened in a story previously acted out by a second experimenter. In one such story, three characters each attempt to win a 'Best Jumper' contest. Of the three participants, only the Troll and Grover successfully perform the three jumps. Robocop, the judge, first tells the Troll that he performed well and could be the winner. Then, he declares Grover the winner of the contest on the basis of his performance. The Troll then objects on the grounds that he is the best jumper. Immediately after the story is acted out, Kermit says *I know one thing that happened* followed by the test sentence *He said that the Troll is the best jumper*.

Principle C, one of Chomsky's (1981) Binding Conditions, prohibits coreference between the referential (*r*-) expression *the Troll* and the pronoun *He* which c-commands it in *He said that the Troll is the best jumper*. The experimental hypothesis is that children know the constraint, correctly assign disjoint reference to *the Troll* and the pronoun *He* in the test sentence, and deny Kermit's description of the story. The null hypothesis is that a child does not know the constraint and allows *the Troll* and the pronoun *He* to be either coreferent or disjoint in reference. Under the coreference reading, a child should agree with Kermit since the Troll did indeed think that he was the best jumper.

Negative judgments are associated with the experimental hypothesis and positive responses with the null hypothesis in order to avoid Type 1 errors. Negative judgments are more informative than positive (*yes*) judgments because children have a normal bias to agree to sentences when they misunderstand an experimental task or miscomprehend a sentence. Selecting positive judgments to support an experimental hypothesis runs the risk of artificially inflating the level of support for that hypothesis, consequently increasing the likelihood of making a Type 1 error. Using negative judgments for this purpose helps to avoid this risk.

C&T's two felicity conditions are designed to make negative judgments felicitous. The Condition of Falsification says that if a child is being asked to deny that a sentence describes a story context, that context should make

the negation of the test sentence true (p. 223). The Principle C experiment satisfies this condition because Robocop chooses Grover as the winner, making the sentence *It is not the case that he$_2$ said that the Troll$_1$ is the best jumper* true. The Condition of Plausible Dissent states that 'it is felicitous to ask whether a sentence S is false on a reading only if the discourse context is such that S has been under consideration on that reading' (p. 237). Plausible Dissent is satisfied by first partitioning the test sentence into a background B and an assertion A, as in (1a) and (1b) (p. 226).

(1) a. Background (B): He (Robocop) said that *so-and-so* is the best jumper.
b. Assertion (A): the Troll.
c. Possible Outcome: He (Robocop) said that the Troll is the best jumper.
d. Actual Outcome: He (Robocop) said that Grover is the best jumper.

The test sentence is placed under consideration by making B(A), 'B applied to A', a possible outcome during a trial. An appropriate B(A) pair is one that 'enables the child to explain ''what really happened'' with least effort.' The Best Jumper Contest story satisfies Plausible Dissent because Robocop briefly considers the Troll as the winner of the contest (1c).

EVALUATION

Modularity Matching

Initially, Modularity Matching seems unlikely to be a realistic model of linguistic performance. First, the model adopts what most psycholinguists would view as unrealistic assumptions about linguistic performance, especially the linguistic performance of children. Contra Chomsky and others, C&T reject the view that 'linguistic theory is ''pure'' and that (children's) performance data are ''messy''' (p. 5). In their view, 'children's performance should parallel their linguistic competence' (p. 7). But the evidence against their view seems overwhelming. A central assumption underling Modularity Matching is that language processing impedes performance only when it results in unacceptability due to nesting, branching direction, self-embedding, or ambiguity (p. 99). But even the normal processing of acceptable sentences hinders performance. Eye movement during reading, for example, is one component of linguistic performance. As C&T correctly point out, eye-tracking experiments have shown that 'acceptable sentences may impose measurable processing difficulties at certain junctures … indicating the relative amount of computation that occurs at various points during sentence comprehension' (p. 99). Although C&T never make the connection, one might reasonably argue on this basis

that processing difficulties impede sentence comprehension in these experiments and that performance does not parallel comprehension. Eye movement latencies are significantly longer when processing difficulties are encountered, even during the processing of run-of-the-mill, acceptable sentences. If processing difficulties did not impede performance, the eyes would not pause at particular points in a sentence where additional processing is required. Eye movement latencies may not count as 'messy' performance data. But they certainly show that grammatical competence does not parallel linguistic performance and that processing tasks impede performance. There are countless similar arguments. But even if one accepts C&T's assumptions about linguistic performance, Modularity Matching models only the relations between grammar (narrowly construed as syntax and semantics) and the parsing principles and processing apparatus. The model has nothing to say about the interrelations between either phonological or prosodic principles and representations and the processors that implement them. Nor does it have anything to say about the various sorts of bookkeeping, monitoring, and reanalysis that take place during normal language processing. In actuality, Modularity Matching carries little weight as a predictor of linguistic performance.

Second, Modularity Matching is neither sufficient nor necessary for circumventing the learnability problem, as C&T claim. Modularity Matching 'skirts' the learnability problem by virtue of the assumption that children and adults share identical processing components and processing capacities, what I have called the Full Processing assumption. According to C&T, one benefit of the model is that questions like 'How does the child attain the adult processing system?' and 'How does the processing system of the child change so as to converge on the adult system?' which C&T take to be learnability questions (pp. 30, 51), never arise. But such questions are not learnability questions. The learnability problem is concerned specifically with how the acquisition device is able to construct the right grammar for a language in response to linguistic experience (e.g. O'Grady, 1997: 245) independently of how grammatical representations are implemented in language performance. If C&T were right, we should be able to ask questions like 'How does the language acquisition device converge on the adult processing system in response to linguistic experience?' But the language acquisition device doesn't choose among parsers or language processing systems in response to linguistic experience. It chooses among possible human grammars. And grammars are not parsers.

Modularity Matching is unnecessary for circumventing the learnability problem because, under the standard view, the existence of UG alone circumvents the problem (O'Grady, 1997: 265). For example, Chomsky (1981: 11) argued that the existence of UG, in making available only a finite number of core grammars, 'trivializes' investigations of learnability.

Chomsky & Lasnik (1993/1995) also suggest that 'if proposals within the P&P approach are close to the mark, then it will follow that languages are in fact learnable' (Chomsky, 1995: 18).

Third, C&T make no effort to compare Modularity Matching to alternative models of adult language processing or language performance (see e.g. Tanenhaus & Trueswell, 1995 for review) or children's language processing or language performance (e.g. Berwick & Weinberg, 1984; MacWhinney, 1987) or alternative approaches to learnability (e.g. Tesar & Smolensky, 2000). Instead, they opt for comparing Modularity Matching to the 'Competing Factors' model, a fictitious model which assumes 'no notion of grammaticality in any absolute sense' (pp. 33–4) and for which 'the grammatical module does not have any special status' (p. 36). C&T's list of Competing Factors advocates includes Paul Bloom, Jane Grimshaw, Helen Goodluck, Tom Roeper, Jill de Villiers, Barbara Lust, Dana McDaniel, and Cecile McKee, none of whom would endorse the model.

Nor do C&T discuss arguments against fundamental assumptions of Modularity Matching. There are by now well-articulated and compelling arguments for the view that the absence of environmental evidence, universality, and early emergence, C&T's three hallmarks of innate specification (p. 10, Crain, 1991), are insufficient as diagnostics of innate mechanisms (e.g. Elman, Bates, Johnson, Karmiloff-Smith, Parisi & Plunkett, 1996). There are also compelling arguments that Fodor's (1983) Modularity Hypothesis, which C&T adopt, is unlikely to provide an accurate description of the organization of the language performance system (e.g. Marslen-Wilson & Tyler, 1989; Elman *et al.*, 1996; Jackendoff, 2000). Moreover, it has become increasingly clear that the classical Principles and Parameters Theory of UG (Chomsky, 1981) has a rather poor track record as a theory of language acquisition (see Fodor, 1998*b* and Tomasello, 2000 for recent discussion). C&T never address these or other related issues, making it difficult to properly assess the costs and benefits of the model.

Fourth, Modularity Matching leads to unrealistic predictions about language learnability. One interesting learnability issue which C&T never address is what Fodor (1998*b*) labelled the 'catch-22 of learnability theory' (Fodor, 1998*b*: 344). Fodor (1998*a*) argued that the parser plays a necessary and pivotal role in language acquisition – 'learners must parse sentences in order to learn from them' (Fodor, 1998*a*: 286). But, by definition, grammars feed parsers. This raises the following paradox – 'The sentences from which the learner should learn are the ones that the current grammar does *not* yet license, but they're not available because they are the ones the learner cannot yet parse' (Fodor, 1998*b*: 344). To get around the paradox, Fodor proposed that syntactic parameters are 'treelets', innately supplied pieces of syntactic trees which are available to the parser for processing sentences beyond its current capacity. If patching a treelet into a parse tree

enables the parse to succeed, it is adopted into the learner's current grammar as the correct parameter value for the target language.

C&T would agree with Fodor that the parser is innate and that sentences must be parsed before they can be used by the learning mechanism. But then Modularity Matching seems untenable. If children and adults have equivalent parsers and parsing capacities, then, given Fodor's reasoning, children and adults must also share the same grammatical competence. This is because the parser is only capable of parsing sentences which are already licensed by the grammar (Fodor, 1998*b*: 344). But this incorrectly predicts 'instantaneous' language acquisition.

Fifth, C&T undermine Modularity Matching by repeatedly contradicting its central assumptions. C&T state that 'children assign the same syntactic representations as adults do' (p. 123) and that 'children's grammar matches that of adults' (p. 212). But they also claim that 'children are language learners and do not yet have the full repertoire of grammatical knowledge that adults do' (pp. 125–6, see also pp. 113, 115). C&T propose that children and adults have the same processing capacities, but concede that 'children's search space of grammatical representations is sometimes influenced by performance factors … because they are still in the throes of language learning' (pp. 126–7). If performance factors influence children's linguistic performance because they are in the throes of language learning, children and adults do not share the same linguistic performance capacities.

C&T also consistently misread passages they retrieve from published material. They observe that 'it is frequently lamented that there is an inherent tension between linguistic theory and the kind of performance data that are obtained in experimental research with children' (p. 5). But they then say that 'we take issue with the notion that there is an inherent tension between the competence grammar and the performance system in which it is embedded' (p. 6), misreading performance DATA as performance SYSTEM. C&T take issue with Lust, Eisele & Mazuka (1992), who claim that 'behavioural research must assume that data derived from experimental tasks are modulated to some degree (at least by chance) by performance factors' (Lust *et al.*, 1992: 340). But C&T state that 'according to the Modularity Matching Model, grammatical principles … are not ''modulated'' by performance factors in most instances' (p. 51), when Lust *et al.*, made no claim about grammatical principles.

Children's linguistic performance

Modularity Matching predicts that children's language performance with unambiguous sentences should match the language performance of adults at least 90% of the time (p. 51), given that they know the relevant grammatical

principles. Anything less must be traceable to one of the acceptable sources of non-adultlike performance mentioned above.

Several results impressively meet these expectations. For example, Thornton (1990: 68–70, 189) reported that children correctly produced *want to* (rather than *wanna*) in wh-questions like *Who do you want to* (**wanna*) *take a walk?* and correctly rejected two-clause crossover statements like **I know who$_1$ he$_1$ said has the best food* more than 90% of the time (but see below). But C&T have no compelling explanations for any pattern of children's non-adultlike performance.

Many error patterns receive no explanations whatsoever. These include children's incorrect acceptance of forwards anaphora (**Papa Bear$_i$ covered him$_i$*), backwards anaphora (**He$_j$ ate the hamburger when the Smurf$_j$ was inside the fence*), discourse binding (*Every mouse$_i$ came to Simba's party*. **He$_i$ wore a hat*), and distributive readings of sentences like *They are lifting four cans*, error patterns which occur from, 16% to 85% of the time (e.g. Crain & McKee, 1985; Miyamoto & Crain, 1991; Koster, 1994; Conway & Crain, 1995).

Other errors receive only a cursory discussion. Avrutin & Thornton (1994) reported that children incorrectly accepted sentences like *The Smurf and the Troll covered them* an unexpectedly high 27% of the time in 'distributive' contexts in which the Smurf and the Troll each covers himself (IUG: 316–17). C&T's only discussion of this error appears in a footnote, where they argue that the story contexts presented to the children were not always sufficient to cause children to generate the relevant distributive operator. They conclude on this basis that 'it is highly unlikely that children's acceptances … represent violations of Principle B' (p. 329). One problem here is that generating distributive operators has nothing to do with satisfying Principle B. Principle B rules out coreferential interpretations of *The Smurf and the Troll covered them*, whether the denotations of the two DPs are construed distributively or collectively. Second, C&T have no principled reason for suggesting that children do not generate distributive operators in Avrutin & Thornton's contexts. In discussing Miyamoto & Crain's (1991) results, C&T suggest that a similar if not identical context *was* a sufficient condition for children to generate a distributive operator for sentences like *They are lifting 4 cans* (p. 316).

Two other errors, auxiliary doubling errors and overt complementizer errors, receive more attention. But C&T seem unprepared for dealing with them.

Auxiliary doubling errors

One old, well-known finding reported by Nakayama (1987) and Crain & Nakayama (1987), is that children consistently make AUXILIARY DOUBLING

ERRORS about 70% of the time when asked to produce questions like *Is the boy that is watching Mickey Mouse happy?* These include COPULA DOUBLING ERRORS like **Is the boy that is watching Mickey Mouse is happy?* and *Is the boy who is happy can see Mickey Mouse?* (RELATIVE CLAUSE COPULA DOUBLING ERRORS), RESTART ERRORS like **Is the boy that is watching Mickey Mouse, is he happy?*, and MODAL DOUBLING ERRORS like **Can the boy who is happy can see Mickey Mouse?* Relative clause copula doubling, in particular, clearly threatens Modularity Matching because it suggests that children move or copy relative clause auxiliaries rather than main clause auxiliaries to CP, in violation of structure dependence constraints.

After more than 15 years, C&T have no coherent perspective on how to analyse auxiliary doubling. At one point, they suggest (pp. 172–3) that the important relative clause copula doubling involves prefixing an auxiliary to a well-formed declarative as in *Is+the boy who is happy can see Mickey Mouse*, an operation similar to *ka* prefixation in Japanese questions and *est-ce que* prefixation in French questions, and do not violate structure dependence constraints. But this can't be right. Japanese *ka* and French *est-ce que* are neither prefixes nor auxiliaries. Nor do they occur as doubled constituents in their respective languages. At another point, C&T suggest that relative clause copula doubling IS in fact evidence of a structure independent operation. Extending an argument made by Crain (1991), C&T point out that children know the structure dependence constraints because they never produced questions like *Can the boy who can see Mickey Mouse is happy?* which would have 'clearly exhibited a structure independent operation' (p. 174). But if *Can the boy who can see Mickey Mouse is happy?* clearly exhibits a structure independent operation, then relative clause auxiliary doubling like *Is the boy who is happy can see Mickey Mouse?* also does, since both feature sentence-initial auxiliaries which are identical only to their respective relative clause auxiliaries.

C&T have little of substance to say about double auxiliary errors more generally either. Nakayama (1987) reported that the number of double auxiliary errors positively correlated with relative clause length in the children's questions, suggesting that increases in relative clause length may induce processing difficulties for children, a view which directly challenges Modularity Matching. Surprisingly, C&T propose that 'the children tested in this experiment produced yes/no questions with doubling of the auxiliary only when processing demands were strenuous' (p. 174). They then go on to suggest 'that children respond to increases in processing load by reverting to an earlier grammar' (p. 174) for which auxiliary doubling is grammatical.

But this proposal makes little sense. First, C&T explicitly state elsewhere that 'adults have little, if any, difficulty interpreting two-clause sentences' and that such sentences 'should also fall within the processing capacities of

children' (p. 34). This suggests that children should have no trouble with sentences with relative clauses. Second, children produce auxiliary doubling even when processing demands are not 'strenuous'. Crain & Nakayama (1987: 538–40) also reported that children instructed to *Ask Jabba if that boy should have an umbrella* responded with double auxiliary errors like **Is that boy should have an umbrella?* and **Is that boy need an umbrella* an unacceptable 22% of the time, even though the target sentences did not include relative clauses and, presumably, did not strain children's processing resources. Third, C&T seem to have no principled basis for claiming that children revert to earlier grammars. Modularity Matching makes the apparently false prediction that adults, like children, respond to increases in processing load by reverting to earlier grammars. C&T suggest that adults do not revert to earlier grammars when they encounter the processing demands of relative clauses because they 'no longer have access to earlier grammatical representations' (p. 127). But when they discuss an adult subject who produced medial-*Wh* errors (see below), they suggest that 'a vestige of his early child grammar appeared, causing him to produce some medial-*Wh* questions' (p. 197), a suggestion which ALLOWS adults to access earlier grammars.

Overt complementizer errors

Thornton (1990) reported that children (2;10–5;5) asked to produce long distance *Wh* questions with subject extraction like *What do you think is in the box?* do so correctly only 63% of the time. Thornton reported that children made insertion errors like MEDIAL-*WH* ERRORS, as in *Who do you think who is in the box?*, 21% of the time, THAT-TRACE ERRORS like *What do you think that is in the box?* 14% of the time, and PARTIAL MOVEMENT ERRORS like *What do you think which boy ate the cookie?* 2% of the time. C&T refer to these errors as OVERT COMPLEMENTIZER ERRORS.

To explain these errors, C&T turn to Thornton's (1990) proposal that children adopt a grammar of *Wh*-questions consistent with the adult grammar of French. Children who adopt the French value of the relevant *Wh* parameter should use medial-*Wh* (1) only in subject-extraction *Wh*-questions (2) with either simple *Wh*-phrases like *who* or complex *Wh*-phrases like *Which boy* (3) either all the time or not at all. This is because the French solution for medial-*Wh* questions applies only to subject-extraction questions, whether the *Wh*-phrase is simple or complex. Furthermore, medial-*Wh* in French is rule-governed and should apply to all subject extraction long distance *Wh*-questions.

Predictably, the variation in children's performance with medial *Wh*-questions conflicts with the French medial-*Wh* account and with Modularity Matching more generally. Of the 21 children whose performance C&T

summarize in their Table 22.1, only 6 produced overt medial complementizers 100% of the time. Several children obligatorily use medial-*wh* in both subject and object extraction *Wh*-questions, a pattern which C&T link to Irish (p. 194), which also requires medial--*wh* in both question types. But C&T have no cogent explanations for the performance of the other children.

Many children apparently produced medial-*Wh* in both subject and object extraction *Wh*-questions before they enter a second stage where only medial-*Wh* in subject extraction *Wh*-questions is retained. C&T's explanation of this pattern only renames medial *Wh* object extraction questions as ''overkill'' and describes the developmental scenario from a UG perspective.

> 'In object-extraction questions, for example, the *wh*-trace is properly head-governed by the verb in the embedded clause, and overt expression of spec-head agreement in the intermediate Comp might be considered ''overkill''. After some time, during which they express spec-head agreement overtly in object-extraction questions as well as in subject-extraction questions, children ascertain that it is only necessary for the latter' (p. 195).

Furthermore, this explanation contradicts C&T's explanations of other errors. According to C&T, children do not always analyse medial-*Wh* with agreement features in object extraction questions. For example, when *that* appears in negative object-extraction *Wh*-questions like *What food that the Spaceman didn't like?*, C&T claim that *that* 'usually appears when no spec-head agreement has taken place in its projection' (p. 202). But in questions like *What do you think that is in the box?*, C&T claim that *that* is inserted to MARK spec-head agreement. Another puzzle is why children learning English should ascertain that overt spec-head agreement is necessary for subject extraction questions when English doesn't choose this option.

C&T report that the overt medial complementizers of two children, KK and SR, fell dramatically across testing sessions (for KK, from 86 to 42%) (Thornton, 1990: 328). C&T suggest that 'it may be that in the course of producing a variety of long distance questions, these children became aware of the adult question forms, which caused them to reformulate the rule for spec-head agreement' (p. 197). Like the previous explanation, this one does little more than summarize what needs to be explained. But even if a child did correctly reformulate her agreement rule, as C&T propose, her adultlike performance rate should have been 90% or above, according to Modularity Matching, rather than 58%, the average error rate for KK in the 2nd and 3rd sessions of Thornton's experiment (IUG: 198). C&T also seem unaware that rule reformulation creates a familiar learnability problem. Before they get busy reformulating their spec-head agreement rules, children

produce both adult and non-adult subject extraction *Wh*-questions. In order for children to converge on the adult grammar, they would have to unlearn the spec-head agreement rule(s) generating both overt and null medial complementizers. A child learning English would not learn anything from positive evidence, which only reinforces the use of null medial complementizers. The child would have to learn that *Wh*-questions with overt medial complementizers do not occur in (Standard) English, an achievement which requires negative evidence which they do not receive.

Interim discussion

I have shown that C&T have no compelling explanations for the kinds of children's non-adultlike performance they discuss in IUG. The primary reason for the poor performance seems to be that C&T never properly tested the model and were actually never in a position to defend it. Much of Part 1, *The Modularity Matching Model* (138 pages), is devoted to spelling out the model's predictions and its assumptions about the language performance architecture. But little of this theoretical firepower is ever exploited in experimental design. One reasonable approach to assessing the model would have been to check if the components of the language performance system operate in the way that the model says they should and that children's nonadult linguistic performance is traceable to the sources which the model specifies (see above). But C&T apparently never designed experiments to investigate whether children's performance could be traced to a bottleneck in verbal working memory or whether the verbal working memory systems of adults and children are subject to the same limitations. Nor did they consider investigating whether children's nonadult performance was due to immature storage capacities, contrary to Modularity Matching, rather than excessive demands on the control mechanism. None of C&T's experiments were designed to investigate whether children utilize the parsing principles of the Referential Theory or to explore the influence of performance factors on children's performance. But these are precisely the kinds of experiments one would expect to find in a guide to experiments assuming Modularity Matching.

Instead, C&T chose to support the model by reexamining previous findings, all of which were specifically designed to investigate children's knowledge of UG principles, not the interrelations between grammatical representations and language performance. Throughout IUG, C&T focus on illustrating that children's linguistic performance is UG-compatible. C&T remind the reader that 'the appropriate research strategy is to maintain the strongest view of language development that is consistent with the theory of Universal Grammar' (p. 306), and 'that processing demands do not cause children to adopt grammatical analyses that are not made available

by Universal Grammar' (p. 327), and that children's non-adultlike performance with medial-*Wh* was 'compatible with Universal Grammar' (pp. 185, 193).

C&T even endorse proposals which undermine Modularity Matching on the grounds that they are UG-compatible. For example, Grodzinsky & Reinhart (1993) proposed that children make coreference errors because, unlike adults, they do not have the memory capacity to successfully execute the pragmatic rule (Rule I) responsible for disallowing coreference. Although this proposal is clearly inconsistent with Modularity Matching, C&T endorse it on the grounds that it 'salvages children's grammatical knowledge by placing blame for the problem elsewhere, in the domain of linguistic performance' and thereby 'avoids the problem of learnability' (p. 270). But if all we have to do to avoid the learnability problem is salvage children's grammatical knowledge, then we certainly don't need Modularity Matching.

Experimental design and task design

C&T no doubt have an exceptional ability to design and present experimental techniques for children. But building an appropriate experimental design involves more than designing appropriate techniques. In designing tasks, researchers need to worry about establishing reliable and valid measures. In designing experiments, researchers need to worry about establishing causal relationships between (independent and dependent) variables. This is because such relationships provide the evidence in support of an experimental hypothesis. Designing experiments involves choosing appropriate independent and dependent variables, appropriate participant groups, appropriate numbers and types of experimental and control conditions, and an appropriate decision criterion for rejecting a null hypothesis and evaluating the possible effects of independent variables.

C&T's experiments generally lack many fundamental components of experimental design which give researchers the right to call a study experimental. Moreover, C&T's new felicity conditions, the Condition of Falsification, the Condition of Plausible Dissent, and the Condition of Plausible Assent, seem incoherent and redundant.

Experimental design

In Chapter 16, *Methodological preliminaries*, C&T discuss how to set up null and experimental hypotheses, the nature of Type 1 and Type 2 errors and how to avoid them, the importance of pretraining, and the purpose of fillers. They also make 'rule of thumb' suggestions, such as using a minimum of 4 items per condition. These are all welcome discussions in a guide to

experiments. But other fundamental and necessary features of experimental design are missing.

Missing decision criterion. One necessary and important feature of every experimental design is a criterion for deciding if a null hypothesis can be rejected in favour of an experimental hypothesis. The standard statistical decision criterion in scientific research is the ALPHA LEVEL, the probability that an observed difference between means (or any other sample statistics) is at least as large as the difference that could have occurred purely by sampling error. A difference between means which yields a probability (p) value which exceeds a chosen alpha level (typically 0·05) is usually interpreted as evidence of a significant contribution of the manipulation of independent variables to behaviour (e.g. Bordens & Abbott, 2002: 385–91).

C&T don't make use of a statistical decision criterion and have no other objective criteria for deciding for or against an experimental hypothesis. C&T explicitly reject (parametric) statistics as a means of evaluating children's performance with unambiguous sentences (see also Crain, 1991) on the grounds that Modularity Matching anticipates all-or-none (90% correct) linguistic behaviour by children who have the relevant grammatical knowledge and little or no variance in their performance. But there are at least three serious problems with this assumption. First, although C&T claim that they are simply following conventional wisdom (fn. 6, p. 322) in choosing the 90% benchmark (see Brown, 1973), most researchers would now agree with Stromswold (1996), who notes that 'there is nothing magical about 90% correct use in obligatory contexts' (Stromswold, 1996: 44). Second, the 90% correct response rate is a PREDICTION of Modularity Matching and cannot logically serve as an independent CRITERION for evaluating that prediction without circularity. Third, neither C&T nor Modularity Matching acknowledge sampling error as a general source of variation in linguistic performance. If children have the relevant grammatical knowledge and the experiments are properly designed, the model doesn't anticipate variation in performance with unambiguous sentences (p. 46). But without a decision criterion, C&T can only guess at how far below 90% an observed performance rate has to be before it should be taken seriously as an explicandum.

Unmet methodological criteria. A quick glance at only some of C's and T's previous research discussed in IUG shows that they apply their methodological criteria, at best, in an inconsistent fashion. For example, of the 3 experiments discussed by Crain & Nakayama (1987), none were conducted with fillers, and only two with control sentences, and Conway & Crain (1995: 190) used only two items per construction and provided no pretraining for their experimental participants. Of the 7 experiments discussed by Crain *et al*. (1996), 4 were conducted without fillers and with less than

4 trials per condition. None of the 7 provided pretraining. These facts suggest that C&T do not take their own methodological criteria seriously.

Inappropriate design. C&T are often reduced to armchair theorizing either because a completed experiment was not properly designed or because the proper experiment was never conducted. For example, in discussing their research in long distance questions, C&T state that they were unable to carry out a 'definitive' experiment which would have demonstrated children's knowledge of the EXTENDED CATEGORY PRINCIPLE (ECP), 'but have assembled an array of facts that lead to the same conclusion' (p. 198). This is a striking admission in a guide to experiments.

One general problem is that the experiments C&T discuss were not properly designed to address the PRINCIPLES VS. PREFERENCES PROBLEM, the problem of ruling out the possibility that a child's adultlike performance reflects a parsing preference rather than knowledge of a grammatical constraint. C&T promote themselves as actively addressing the problem, noting that 'the need to distinguish preferences and principles has not taken hold, even among researchers who work within the generative framework' (p. 307). But C&T show no commitment to resolving the problem. At one point, they admit that they used Thornton's (1990) *wanna* experimental design inappropriately to test children's knowledge of the ECP. The problem was that the design of the *wanna* contraction experiment 'cannot succeed for testing children's knowledge of the ECP' (p. 189) using *Wh* questions because there is no way to distinguish satisfaction of the ECP and satisfaction of a parsing preference for reduced forms. A proper procedure would have been to design and conduct a new experiment. Instead, C&T assume 'that the goal of the experiment is simply to elicit subject and object extraction long distance questions from children' (p. 196) rather than *wanna* contraction sentences. C&T then focus on explaining how children's use of overt medial complementizers reflects their knowledge of the ECP, ignoring the still live possibility that children's correct performance may have reflected a parsing preference.

No adult controls. Although one core assumption of Modularity Matching is that the language systems of children match those of adults, adult data are never reported and, apparently, were never obtained.

No manipulation of independent variables. Exposing subjects to different levels of an independent variable makes it possible to determine if the variable has an effect on participants' behaviour. These manipulations are often absent in C&T's experimental designs. For example, C&T claim that children's scores improve on Crain *et al*.'s (1996) experiments with universal quantification because felicity conditions like the Condition of Plausible Dissent are satisfied (see below). But they never compare their felicity conditions with alternative conditions which may have also resulted in children's improved scores. One of Modularity Matching's showcase

predictions is that linguistic performance of adults in children degrades due to in the same ways. Here, task demands and processing complexity count as natural independent variables. But none of C&T's designs treat them as such. There are numerous other examples.

Falsification and Plausible Dissent

The Condition of Falsification and the Condition of Plausible Dissent are C&T's most recent felicity conditions on experiments with children. The logic, plausibility, and effectiveness of Plausible Dissent have already been strongly challenged on the grounds that children not only perform well in experimental conditions where Plausible Dissent is not satisfied but perform poorly under experimental conditions where it is satisfied (e.g. Brinkmann, Drozd & Krämer, 1996; Drozd & van Loosbroek, 1999). However, other researchers have adopted Plausible Dissent as a felicity condition and have argued for its effectiveness (e.g. Guasti & Chierchia, 2001). I want to argue here is that there are very good reasons for dismissing these conditions altogether as felicity conditions on experiments.

Falsification, Plausible Dissent, and rational discourse

Both Falsification and Plausible Dissent are general conditions on rational discourse and need not be stated specifically as felicity conditions for experiments. C&T present Falsification as part of a research strategy designed to avoid Type 1 errors. But it is also an intuitive, necessary condition for rational discourse involving negative judgments. Speakers (and experimenters) can expect negative judgments from an addressee only if the addressee is in a position to infer that the proposition presented is false. C&T (p. 226) equate Plausible Dissent with RUSSELL'S MAXIM, the claim that 'perception only gives rise to a negative judgment when the correlative positive judgment has already been made or considered' (Russell, 1948: 138). But Russell surely intended his remarks as a general condition on negative judgments, not as an experimental felicity condition. C&T also state that Plausible Dissent is intended to make experimental contexts consistent with Grice's relevance maxim (p. 225). But we don't need Plausible Dissent to stipulate that Russell's Maxim or the Relevance Maxim applies to experimental contexts when they already do.

PLAUSIBLE DISSENT AND PLAUSIBLE ASSENT

In IUG, C&T describe how Plausible Dissent was satisfied and led to adultlike performance by children in Crain *et al.*'s (1996) universal quantification experiments, Thornton's (1990) Strong Crossover experiments, and

Conway & Crain's (1995) discourse binding experiments. But satisfying Plausible Dissent was unlikely to have had any effect on children's performance in any of these experiments. Due to space limitations, I restrict my discussion to Crain *et al.*'s (1996) and Thornton's (1990) experiments. Similar arguments apply to Conway's experiments.

Plausible Dissent, Plausible Assent and universal quantification

C&T first introduced Plausible Dissent as a felicity condition in Crain *et al.* (1996), though the general notion appears earlier as Plausible Denial in Crain (1991). Crain *et al.* (1996) were primarily interested in explaining why children in Philip's (1995) universal quantification experiments incorrectly denied that sentences like *Every boy is riding an elephant* correctly describe a perceptual array in which three boys are each riding an elephant if an additional unridden elephant also appears in the array, an error referred to as the SYMMETRICAL INTERPRETATION (SI) (Philip, 1995; see Drozd, 2001 for review).

Crain *et al.*, proposed that Philip's (1995) experiments elicited the SI because his designs did not satisfy Plausible Dissent, which they define as follows:

> 'Now let us suppose ... that we have chosen to make the test sentences TRUE on the adult interpretation ... for the question to be felicitous, then, the assertion must be IN DOUBT at some point during the trial. Therefore, some outcome other than the actual one should be conceivable at some point during the trial. Let us call this feature of the design the CONDITION OF PLAUSIBLE DISSENT' (Crain *et al.*, 1996: 116).

Crain *et al.*, reported that children's performance with similar universally quantified sentences considerably improved when children were first presented with the stories which satisfied Plausible Dissent, and concluded on these grounds not only that SI is an experimental artifact but also that children have full competence with universal quantification.

In IUG, C&T (p. 300) back up Crain *et al.*'s conclusions but propose that the relevant felicity condition responsible for the improvement in children's performance in Crain *et al.*'s experiments should be called the CONDITION OF PLAUSIBLE ASSENT rather than the Condition of Plausible Dissent. Closer inspection of Crain *et al.*'s design reveals why. Crain *et al.*'s Plausible Dissent is the felicity condition appropriate when asking children for a POSITIVE judgment (when the assertion is true in context). C&T's Plausible Dissent is the felicity condition appropriate when asking children for a NEGATIVE judgment (when the test sentence is false in context) (see (2) above and surrounding text).

One might be tempted to write off this confusion as harmless, but there is a serious error in reasoning here. Plausible Assent is not Russell's Maxim. Nor is it a 'corollary' of Plausible Dissent, as C&T claim, since it is neither inferable nor deducible from Plausible Dissent. Most importantly, the reasoning behind Plausible Dissent does not apply to Plausible Assent. Recall that presenting the application of background to assertion as a possible rather than an actual outcome is crucial to satisfying Plausible Dissent because it makes clear to a child why the test sentence is false. According to C&T (p. 300), Plausible Assent, like Plausible Dissent, is satisfied if a possible outcome other than the actual outcome is introduced during a trial. But when the sentence is true, applying the background to the assertion places the ACTUAL outcome, not a POSSIBLE outcome, under consideration. It is the actual outcome which makes it clear to a child why a test sentence is true. Introducing a possible outcome during a trial when the test sentence is true is superfluous.

If this reasoning holds, then Crain *et al.*'s (1996) experiments were misconceived since the reasoning behind their Condition of Plausible Dissent was seriously flawed. Remarkably, C&T ignore the flaws (and the confusion over terms) and promote Crain *et al.*'s experiments as appropriately designed. I would argue instead that both Crain *et al.*'s experiments and C&T's Plausible Assent should be dismissed.

Plausible Dissent and crossover

One seemingly intractable problem with C&T's experimental designs which is never addressed is that children sometimes find sentence interpretations felicitous even if Plausible Dissent has not been satisfied for these interpretations. In one of Thornton's (1990) strong crossover experiments discussed in IUG (Chapter 30), 12 children were presented with a story in which the Joker chooses the Ninja Turtle as the winner of a 'best food' contest after first considering Grover and Yogi Bear as having the best food. At the end of the story, Grover and Yogi Bear each claim that the Joker is wrong on the grounds that he, not the Ninja Turtle, has the best food. When the story is over, Kermit says either *I know who$_i$ he$_{*i/j}$ said t$_i$ has the best food. Grover and Yogi Bear* or *I know who$_i$ t$_i$ said he$_{i/j}$ has the best food. Grover and Yogi Bear*, which I have annotated for convenience. C&T reported that children correctly rejected the first test sentence 92% of the time on the grounds that the Joker said that the Ninja Turtle had the best food. When given the second sentence, the children either assigned a 'bound variable' reading to the sentence (for which *who* and *he* are coindexed) and correctly agreed with Kermit, or deictically linked *he* to Ninja Turtle and correctly rejected the sentence on the grounds that the Joker, rather than Grover and Yogi Bear said that Ninja Turtle has the best food (pp. 262–3).

C&T correctly point out that the story context satisfies Plausible Dissent for the intended interpretation of *I know who$_i$ he$_{*i/j}$ said t$_i$ has the best food. Grover and Yogi Bear*. The problem is that C&T's story context does not satisfy Plausible Dissent for either the bound variable or the 'deictic' interpretation of the other sentence. The bound variable reading, which was true in context, is only felicitous if some possible outcome consistent with an alternative bound variable reading of the sentence was presented during the story. But neither the Joker nor Ninja Turtle, the only other story characters, ever said that he has the best food. The deictic interpretation of the sentence, which was false in context, would have been felicitous if the event described by the sentence was presented as a possible outcome during the story. But nowhere in the story do Grover and Yogi Bear say that Ninja Turtle has the best food. According to C&T, children should not have found these readings felicitous. But they did.

Plausible Dissent and information structure

False propositions can be raised for consideration or rendered plausible in any number of ways. But C&T specify that Plausible Dissent is satisfied by first partitioning a test sentence into a background B and an assertion A. The partitioning is important because applying B to A is the means by which the possible outcome is registered in context and distinguished from the actual outcome, which involves an alternative assertion. The partitioning is so important, in fact, that C&T feature B and A as core components of experimental design.

But the satisfaction conditions for Plausible Dissent actually present a distorted view of how sentences are rendered felicitous in discourse contexts, and, more specifically, how sentence partitioning and the identification of possible outcomes each contribute to the felicitous use of sentences in such contexts. Once the distortion is clarified, Plausible Dissent is easily seen as misleading and superfluous.

What C&T refer to as 'assertion' in IUG is typically called the FOCUS, that part of a sentence which specifies what is new or contrastive for the hearer (e.g. Bolinger, 1961; Lambrecht, 1994). Focus, like presupposition and topic, is a category of INFORMATION STRUCTURE (also called FOCUS STRUCTURE), a core grammatical system in which propositional information is (pragmatically) structured to reflect a speaker's hypotheses about the mental states of other discourse participants (Lambrecht, 1994), or more generally, about how information is organized at the discourse level (Kadmon, 2001). Focus marking, either of the structural variety (clefting) or the prosodic variety (pitch accent), induces the partitioning of a sentence into a BACKGROUND (B) and focus (F). Informally, focal pitch accent on the DP *the Troll* in the sentence *Robocop thinks the Troll is the best jumper*

partitions the proposition conveyed by the sentence into a focus, the denotation of the DP *the Troll*, and the background, the property denoted by the open proposition *Robocop thinks that x is the best jumper*. The denotation of the focus identifies the particular individual from some presupposed set of focus-alternatives which the speaker wishes to highlight as new or contrastive. The sentence is interpreted as true if the backgrounded property applies to the focus, written B(F) (see e.g. Krifka, 1991).

One would expect any contemporary model of linguistic performance to have an informed approach to information structure (see e.g. Levelt, 1989). Focus, in particular, plays an integral role in organizing the operations speakers perform in interpreting sentences during sentence comprehension and in managing discourse information in verbal working memory (Frazier, 1999; Bosch & van der Sandt, 1999). Moreover, principles of information structure describe how propositional information is packaged for felicitous integration into discourse contexts and should play a major role in any description of felicity conditions for experiments.

Surprisingly, neither information structure nor intonation structure plays any role in C&T's research programme. This is perplexing for two reasons. First, C&T's sentence partitioning is identical to the sentence partitioning induced by focus marking. C&T's B and A are the information structure categories B and F, respectively, and C&T's possible outcomes are just applications of B to F (for satisfying Plausible Dissent) and B to a focus-alternative for F (for satisfying Plausible Assent). Although these identities are transparent, C&T never acknowledge them. Second, information structure provides everything that satisfying Plausible Dissent is intended to provide but in a principled and intuitive way. C&T require that sentences be partitioned to satisfy Plausible Dissent, but they never say why. Nor do they describe how the child is to recognize the experimenter's intended partition. But sentence partitioning has a clear role to play in information structure. It informs speakers about the discourse status (background, new/contrastive) of propositional information and, as a result, helps to identify the class of contexts in which a sentence can be felicitously uttered. This enables a child to explain 'what really happened' because it informs a child where the speaker wants her to find propositional information in her mental discourse model. The more salient the clues to the mental location of information are for a child, the easier it will be for that child to recall a story line and respond correctly when asked to judge a sentence about the story.

What should now be clear is that C&T have distorted the contributions of sentence partitioning and focus alternatives to sentence interpretation by miscasting them as satisfaction conditions for nonlinguistic felicity conditions. This is certainly a misleading, if not an inappropriate, method for accounting for children's linguistic performance. More importantly, once

the natural role of sentence partitioning in the pragmatic structuring of propositional information is acknowledged, Plausible Dissent is superfluous. Information structure already explains how and why sentence partitioning and the introduction of focus alternatives makes certain sentences felicitous in context. It already provides the information children need to explain not only why a sentence is false when they are asked to deny it but also why a sentence is true when they are asked to confirm it.

FINAL REMARKS

I have reviewed C&T's current research programme in language acquisition as described in IUG by evaluating its two most important components: (1) the Modularity Matching Model and (2) its accompanying experimental design components. I argued that Modularity Matching adopts unrealistic assumptions about linguistic performance, is too narrow in scope to make predictions about linguistic performance, and is neither necessary nor sufficient for circumventing the learnability problem. I showed that C&T present no coherent approach to demonstrating the correctness of the model. They offer no experimental programme for testing the model, seem unprepared to deal with children's non-adultlike language performance, and consistently misrepresent alternative approaches to UG- and non-UG-based developmental research.

I also argued that C&T's experiments overall are, at best, poorly designed. They lack the components of standard experiments which allow sophisticated treatments and comparisons of data and which constrain the interpretation of experimental results. C&T offer no replacements for the missing components and no discussion for why learnability considerations would require their exclusion. I argued that C&T's new felicity conditions on experiments are superfluous and should be dismissed. I argued that Plausible Dissent, in particular, was originally misconceived as a condition on positive, rather than negative, judgments, and that its satisfaction conditions distort the role of information structure in sentence interpretation and assessment in context.

REFERENCES

Avrutin, S. (1999). *Development of the syntax-discourse interface*. Dordrecht: Kluwer.

Avrutin, S. & Thornton, R. (1994). Distributivity and binding in child grammar. *Linguistic Inquiry* **25**, 265–71.

Berwick, R. C. & Weinberg, A. S. (1984). *The Grammatical basis of linguistic performance*. Cambridge: MIT.

Bolinger, D. (1961). Contrastive accent and contrastive stress. *Language* **37**, 83–96.

Bordens, K. S. & Abbott, B. B. (2002). *Research design and methods: a process approach* (5th edition). Boston: McGraw-Hill.

Bosch, P. & van der Sandt, R. (1999). *Focus: linguistic, cognitive, and computational perspectives*. Cambridge: CUP.

Brinkmann, U., Drozd, K. F. & Krämer, I. (1996). Physical individuation as a prerequisite for children's symmetrical interpretations. In A. Stringfellow, D. Calia-Amitay, E. Hughes & A. Zukowski (eds), *Proceedings of the 20th Annual Boston University Conference on Language Development*. Boston: Cascadilla Press.
Brown, R. (1973). *A first language*. Cambridge: Harvard University Press.
Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT.
Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
Chomsky, N. & Lasnik, H. (1993). The theory of principles and parameters. In J. Jacobs, A. von Stechow, W. Sternefeld & T. Vennemann (eds), *Syntax: an international handbook of contemporary research*. Berlin and New York: Walter de Gruyter. Reprinted in Chomsky, N. (1995) *The minimalist program*, Cambridge: MIT.
Conway, L. & Crain, S. (1995). Dynamic acquisition. In D. MacLaughlin & S. McEwen (eds), *Proceedings of the 19th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.
Crain, S. (1991). Language acquisition in the absence of experience. *Behavioral and Brain Sciences* **14**, 597–650.
Crain, S. & McKee, C. (1985). The acquisition of structural restrictions on anaphora. In S. Berman, J. Choe & J. McDonough (eds), *Proceedings of NELS 15*. Amherst: University of Massachusetts, GLSA.
Crain, S. & Nakayama, M. (1987). Structure dependence in grammar formation. *Language* **63**, 522–43.
Crain, S., Shankweiler, D., Macaruso, P. & Bar-Shalom, E. (1990). Working memory and comprehension of spoken sentences: investigations of children with reading disorder. In G. Vallar & T. Shallice (eds), *Neuropsychological disorders of short-term memory*. Cambridge: CUP.
Crain, S. & Steedman, M. (1985). On not being led up the garden path. The use of context by the psychological parser. In D. R. Dowty, L. Karttunnen & A. Zwicky (eds), *Natural language parsing: psychological, computational, and theoretical perspectives*. Cambridge: CUP.
Crain, S., Thornton, R., Boster, C., Conway, L., Lillo-Martin, D. & Woodams, E. (1996). Quantification without qualification. *Language Acquisition* **5**, 83–153.
Drozd, K. F. (2001). Children's weak interpretation of universally quantified sentences. In M. Bowerman & S. C. Levinson (eds), *Language acquisition and conceptual development*. Cambridge: CUP.
Drozd, K. F. & van Loosbroek, E. (1999). Weak quantification, Plausible Dissent, and the development of children's pragmatic competence. In A. Greenhill, H. Littlefield & C. Tano (eds), *Proceedings of the 23rd Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.
Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D. & Plunkett, K. (1996). *Rethinking Innateness*. Cambridge: MIT.
Fodor, J. A. (1983). *The modularity of mind*. Cambridge: MIT.
Fodor, J. D. (1998*a*). Learning to parse. *Journal of Psycholinguistic Research* **27**, 285–319.
Fodor, J. D. (1998*b*). Parsing to learn. *Journal of Psycholinguistic Research* **27**, 339–75.
Frazier, L. (1999). *On sentence interpretation*. Dordrecht: Kluwer.
Grodzinsky, Y. & Reinhart, T. (1993). The innateness of binding and the development of coreference. A reply to Grimshaw and Rosen. *Linguistic Inquiry* **24**, 69–103.
Guasti, M. T. & Chierchia, G. (2001). Backward versus forward anaphora: reconstruction in child grammar. *Language Acquisition* **8**, 129–70.
Hamburger, H. & Crain, S. (1982). Relative acquisition. In S. Kuczaj (ed.), *Language development: syntax and semantics*. Hillsdale, NJ: Erlbaum.
Jackendoff, R. (2000). *Foundations of language*. Oxford: Oxford University Press.
Kadmon, N. (2001). *Formal pragmatics*. Oxford: Blackwell.
Koster, C. (1994). *Errors in anaphora acquisition*. Utrecht: Research Institute for Language and Speech (OTS).

Krifka, M. (1991). A compositional semantics for multiple focus constructions. *Proceedings of SALT*.
Lambrecht, K. (1994). *Information structure and sentence form*. Cambridge: CUP.
Levelt, W. J. M. (1989). *Speaking: from intention to articulation*. Cambridge: MIT.
Lorsbach, T. C., Katz, G. A. & Cupak, A. J. (1998). Developmental differences in the ability to inhibit the initial misinterpretation of garden path sentences. *Journal of Experimental Child Psychology* **71**, 275–96.
Lust, B., Eisele, J. & Mazuka, R. (1992). The binding theory module: evidence from first language acquisition for Principle C. *Language* **68**, 333–58.
MacWhinney, B. (1987). The competition model. In B. MacWhinney (ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum.
Marcus, G., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J. & Xu, F. (1992). *Overregularization in language acquisition*. Monographs for the Society for Research in Child Development. Number 57.
Marslen-Wilson, W. & Tyler, L. K. (1989). Against modularity. In J. L. Garfield (ed.), *Modularity in knowledge representation and natural language understanding*. Cambridge: MIT.
Miyamoto, Y. & Crain, S. (1991). Children's interpretation of plural pronouns: collective vs. distributive interpretation. Paper presented at the 16th Annual Boston University Conference on Language Development.
Nakayama, M. (1987). Performance factors in subject-auxiliary inversion by children. *Journal of Child Language* **14**, 113–25.
O'Grady, W. (1997). *Syntactic development*. Chicago: Chicago University Press.
Philip, W. (1995). *Event quantification in the acquisition of universal quantification*. Doctoral dissertation, University of Massachusetts, Amherst.
Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
Russell, B. (1948). *Human knowledge: its scope and limits*. London: Allen and Unwin.
Stromswold, K. (1996). Analyzing children's spontaneous speech. In D. McDaniel, C. McKee & H. S. Cairns (eds), *Methods for assessing children's syntax*. Cambridge, MA: MIT Press.
Tanenhaus, M. K. & Trueswell, J. (1995). Sentence comprehension. In P. D. Eimas & J. L. Miller (eds), *Speech, language, and communication*. New York: Academic Press.
Tesar, B. & Smolensky, P. (2000). *Learnability in optimality theory*. Cambridge: MIT.
Thornton, R. (1990). *Adventures in long-distance moving. The acquisition of complex* wh-*questions*. Doctoral dissertation, University of Connecticut, Storrs.
Thornton, R. (1995). Referentiality and *wh*-movement in child English. Juvenile D-linkuency. *Language Acquisition* 4, 139–75.
Tomasello, M. (2000). *The cultural origins of human cognition*. Cambridge: Harvard University Press.
Trueswell, J. C., Sekerina, I., Hill, N. M. & Logrip, M. L. (1999). The kindergarten-path effect: studying on-line sentence processing in young children. *Cognition* **73**, 89–134.
Tyler, L. K. & Marslen-Wilson, W. (1978). Some developmental aspects of sentence processing and memory. *Journal of Child Language* **5**, 113–29.
Valian, V. (1990). Syntactic subjects in the early speech of American and Italian children. *Cognition* **40**, 21–81.