Cognitive Psychology 62 (2010) 1–40

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

# Learning language from the input: Why innate constraints can't explain noun compounding

Michael Ramscar *, Melody Dye

*Department of Psychology, Stanford University, CA 94305, United States*

A R T I C L E  I N F O

*Article history:*
Accepted 8 October 2010
Available online xxxx

*Keywords:*
Language
Learning
Distributional information
Conventions
Semantics
Compounding
Inflection
Morphology
Phonology
Informativity

A B S T R A C T

Do the production and interpretation of patterns of plural forms in noun-noun compounds reveal the workings of innate constraints that govern morphological processing? The results of previous studies on compounding have been taken to support a number of important theoretical claims: first, that there are fundamental differences in the way that children and adults learn and process regular and irregular plurals, second, that these differences reflect formal constraints that govern the way the way regular and irregular plurals are processed in language, and third, that these constraints are unlikely to be the product of learning. In a series of seven experiments, we critically assess the evidence that is cited in support of these arguments. The results of our experiments provide little support for the idea that substantively different factors govern the patterns of acquisition, production and interpretation patterns of regular and irregular plural forms in compounds. Once frequency differences between regular and irregular plurals are accounted for, we find no evidence of any qualitative difference in the patterns of interpretation and production of regular and irregular plural nouns in compounds, in either adults or children. Accordingly, we suggest that the pattern of acquisition of both regular and irregular plurals in compounds is consistent with a simple account, in which children learn the conventions that govern plural compounding using evidence that is readily available in the distribution patterns of adult speech.



* Corresponding author.
*E-mail address:* ramscar@gmail.com (M. Ramscar).

0010-0285/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.cogpsych.2010.10.001

## 1. Introduction

> "For several days I carried in my pocket a small white card on which was typed understander. On suitable occasions I would hand it to someone. 'How do you pronounce this?' I asked. He pronounced it. 'Is it an English word?' He hesitated. 'I haven't seen it used very much. I'm not sure.' 'Do you know what it means?' 'I suppose it means one who understands.' I thanked him and changed the subject...
>
> Probably no one but a psycholinguist captured by the ingenuous behavioristic theory that words are vocal responses conditioned to occur in the presence of appropriate stimuli would find anything exceptional in this... [but] if one believes there is some essential difference between behavior governed by conditioned habits and behavior governed by rules, [a conditioning theory of language] could never be more than a vast intellectual pun." (Miller, 1967, pp. 80–82)

George Miller's (1967) essay, *The Psycholinguists*, laid the terrain for what has become a cornerstone debate in the psychology of language: Do the rules described in linguistic theories correspond to mental structures that implement the same kind of formal mechanisms that are used in descriptive grammars, such that "the mental parser [makes] basically the same distinctions as the grammar" (Clahsen, 1999, p. 995; see also Jackendoff, 1997)? Or are linguistic conventions the product of processes that are neither governed by, nor even necessarily correspond to, the rules put forward by grammarians? Historically, much of the focus of this debate has been on inflectional morphology, and the productive way in which markers such as *–er*, and *–ed* can be added to verbs, or *–s* to plurals (for reviews see Pinker & Ullman, 2002; McClelland & Petterson, 2002). In response to Miller's challenge, an array of probabilistic models that do *not* make use of explicitly encoded rules have been put forward to capture the kind of flexible production and comprehension of inflections he described (e.g. Hahn & Nakisa, 2000; Haskell, MacDonald, & Seidenberg, 2003; MacWhinney & Leinbach, 1991; Plunkett & Marchman, 1993; Rumelhart & McClelland, 1986; see also Ramscar & Dye, 2009; Ramscar & Yarlett, 2007; see also Baayen, 2010; Bybee, 1988; Bybee, 1995; Bybee & Slobin, 1982; Köpke, 1993; Milin, Kuperman, Kostic, & Baayen, 2009; Plaut & Booth, 2000 for other distributional approaches to language). In these models, "rules" arise out of regularities in the phonological (and, sometimes, semantic) distributions of languages, and can be seen as epiphenomenal.

However, the idea underlying these models—that inflection can be explained by analogy to previously learned forms—has been fiercely criticized for failing to capture the essential nature of *regular* inflection, which, it is argued, is governed by rules in just the way that Miller (1967) suggests. In particular, it has been argued that systematic constraints apply to regular but not irregular inflection, and that this is evidence that the "language processor" formally, and perhaps innately, distinguishes rule-based regular forms from irregulars (Alegre & Gordon, 1996; Berent & Pinker, 2007; Clahsen, 1999; Gordon, 1985; Marcus, Brinkmann, Clahsen, Wiese, & Pinker, 1995; Pinker, 1991, 1999, 2001; Pinker & Prince, 1988; Prasada & Pinker, 1993).

On this view, biases that explicitly distinguish between regular and irregular forms, and rules that apply categorically to grammatical classes (such as noun and verb), are necessary to account for the way people process inflectional morphology (see Pinker, 1991, 1999). For example, with regards noun inflection—the focus of this paper—it has been claimed that the patterns of children's behavior as they learn to compound nouns offers clear evidence for the existence of rule-based, categorical mechanisms (Alegre & Gordon, 1996; Gordon, 1985; Pinker, 1999; see also Teichmann et al., 2005; Ullman, 2004). Gordon (1985) argues that children's tendency to produce *mice-eater* but not *rats-eater* in elicitation studies "strongly supports the notion that level ordering [a grammatical constraint specific to regular noun plurals] constrains the child's word-formation rules, *independent of the input received* [our emphasis]" (Gordon, 1985, p. 73). Level ordering, it is suggested, reveals the existence of an innate constraint on compounding which *prevents* regular plurals from entering compounds, while allowing irregular plurals to do so.

### *1.1. Level ordering in lexical development*

According to the level ordering hypothesis, a series of constraints on word formation affect the production (and thus bias the acceptability) of plural forms in compounds (Kiparsky, 1982, 1983). Level ordering proposes that lexical processes in English are assigned to one of three levels. Level One is essentially a repository of stored forms – such as idiosyncratic forms, the irregular forms of verbs and nouns, pluralia tantum nouns such as *scissors*, etc. – and includes some basic derivational processes – such as those that tend to affect the phonology of hosts to which they apply and or forms that are semantically unpredictable – , as well as affixes that change the stress patterns or vowel structures of stems. At Level Two, processes such as "neutral" derivational processes – those that do not affect phonology and are semantically predictable – and noun compounding are found. Finally, inflection takes place in Level Three.

Level ordering maintains that word formation proceeds in an orderly unidirectional manner. Compounding can only occur at Level Two; inflection can only occur at Level Three. Importantly, once a compound is formed at Level Two, its constituents cannot be inflected at Level Three; only the compound itself can be (holistically) inflected. These sequencing constraints predict important differences between regular and irregular plurals in compounding.

For example, since singular forms and any irregular forms are all stored at Level One, compounding for these words takes place at Level Two, *before* regular inflection at Level Three. Thus, the word formation system can compound *mice*, *mouse* or *rat* to *eater* (and pluralize these compounds at Level Three to produce *mice*-, *mouse*-, or *rat-eaters*). However, the same cannot hold true for regular plural nouns. Regular plurals, such as *rats*, are only formed (i.e., inflected) at Level Three, *after* compounding takes place at Level Two. Since the constituents of compounds cannot be inflected at Level Three (e.g., *rat-eater* cannot be inflected to *rats-eater*) and since regular plurals can only be formed *after* compounding has taken place, regular plurals cannot enter into compounds (see Fig. 1).

This theory naturally lends itself to explaining a number of behavioral phenomena that are thought to be associated with noun compounding. For example, in an elicitation task, Gordon (1985) found that 3–5 year-old children produced compounds containing irregular plurals in the non-head (i.e. left) position of noun-noun compounds, such as *mice eater*, but *avoided* compounds containing regular plurals in the same position, such as *rats eater* (this has often been replicated; see e.g., Nicoladis, 2003). Gordon (1985) argues that the patterns of inflection obtained in his study provide evidence that children are sensitive to the difference between irregular and regular plurals in compounds, and to different constraints applying to them (even if the specific details of level ordering theory do not provide the best account of the organization of lexical processing; see Alegre & Gordon, 1996; Pinker, 1991).

Further support for this view comes from Clahsen, Rothweiler, Woest, and Marcus (1992), who replicated Gordon's study in German. Although the question of "regularity" with regards the German plural is controversial (see Behrens, 2001; Clahsen, 1999; Clahsen et al., 1992; Hahn & Nakisa, 2000; Köpcke, 1998; Marcus et al., 1995), Clahsen and colleagues showed that the productive plural forms –*s* and –*n* were similarly avoided in compounds (even when, as in the case of –*n*, this led to children producing incorrect forms such as *blume-vase* as opposed to the correct *blumen-vase*).

Finally, Alegre and Gordon (1996) found evidence consistent with what is perhaps the most intriguing and suggestive prediction of level ordering: that when regular plurals *are* encountered in compounds,

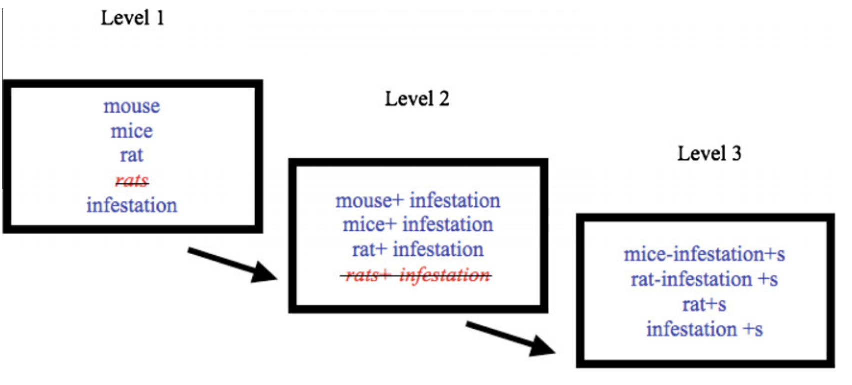


**Fig. 1.** Unlike "mice," "rats" is not formed until Level 3, and so it is not available for compounding at Level 2 (Gordon, 1985).

they are evidence of 'recursion' inside the word formation system—i.e., evidence that words that have completed lexical processing have been later fed back into the system—and thus will be subject to *recursive interpretation.* Alegre and Gordon (1996) argue that in certain circumstances, the recursive feeding back of output forms from the lexical processing system is warranted, allowing the production of forms such as *seat-of-the-pants-decision*, in which the phrase *seat-of-the-pants* is first produced by lexical processing and syntax at Level Three, and then fed back into Level Two to be compounded onto *decision*.

This process leads to a specific, testable prediction with regards to compounds comprising an adjective followed by two nouns, such as *red rat eater*. When this phrase contains the non-head noun *rat* in its singular form, it is structurally ambiguous: it could either have been formed by compounding the adjective *red* onto the noun-noun compound *rat-eater*, or else it could have been formed by first forming the noun-phrase *red-rat*, and then feeding this back into Level Two and compounding it to *eater* (see Fig. 2). These different compounds have different semantic properties: the former would describe a red monster that eats rats, whereas the latter would describe a monster that eats red rats.

Theoretically, the same kind of structurally ambiguous compound could arise from any singular noun or any irregular plural noun. However, once the regular plural *rats* is inserted into the compound in the non-head position (i.e., *red rats eater*), then according to level ordering, only one interpretation is possible. Because the noun-noun compound *rats eater* violates level ordering constraints, *red rats eater* can only be a noun-phrase (*red rats*) compounded onto a noun (*eater*), thereby ruling out the adjective—noun-noun interpretation. In other words, in the regular plural instance, the only viable semantic interpretation is a monster that eats red rats.

Alegre and Gordon (1996) examined this prediction by comparing how 3, 4 and 5 year olds interpreted *red rat eater* and *red rats eater*. Consistent with the predictions of level ordering, children produced significantly more recursive interpretations in response to *red rats eater* than to *red rat eater*. Alegre and Gordon argue that this is evidence that children are instinctively aware of the fact that *rats eater* is "disallowed in the grammar" and that there is "some constituent comprising *red* and *rats* which, by their categorical assignments, would constitute an NP (noun phrase) when combined" (Alegre & Gordon, 1996, p. 76).

### 1.2. Level ordering and innate grammatical constraints

The evidence described so far has been put forward to argue that children are sensitive to formal grammatical distinctions between regular and irregular forms, and that this sensitivity cannot be reduced to simple experiential factors such as frequency or the distribution of forms (Pinker, 1999; see also Alegre & Gordon, 1996; Clahsen, 1999; Gordon, 1985; Berent & Pinker, 2007, 2008). If the constraints described by level ordering theory reflect strong, possibly even innate constraints, on how people process language, then this is of enormous significance—not only for theories of language

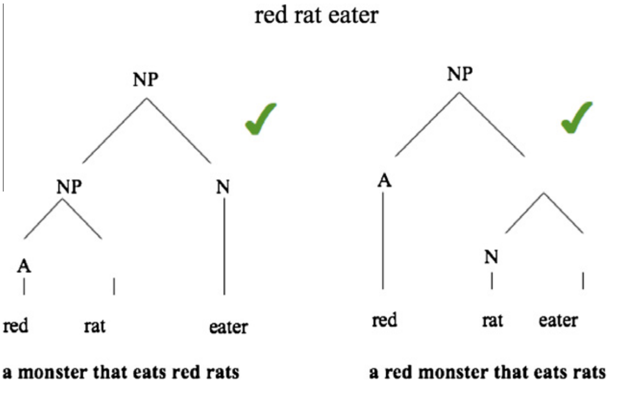


**Fig. 2.** The two possible ways of parsing *red rat eater* – "a monster that eats red rats" or a "red monster that eats rats" (Alegre & Gordon, 1996).

acquisition and linguistic development—but also for the theory and practice of identifying and treating language-learning deficits (see e.g. Clahsen, 1999; Gopnik, 1990; Rice & Wexler, 1996; van der Lely & Christian, 2000; van der Lely & Ullman, 1996) and even understanding cognitive decline (Teichmann et al., 2005). For these reasons, it is important to note that many difficulties in proposing a coherent theory of level ordering have been enumerated, particularly in accounting for the slew of apparent exceptions to level ordering in compounding (see e.g., Haskell et al., 2003). Haskell et al. write:

> "[Proponents do] not make clear how the data that are problematic for level-ordering can be explained within a theory that retains its 'logic' but not its substantive proposals. The validity of the level-ordering account has to be considered with respect to all the phenomena it was intended to address, not merely the restricted subset involving regular plural modifiers in compounds." (124)

Haskell, MacDonald & Seidenberg point out that level ordering cannot be considered a coherent scientific theory absent specific predictions about how inflection should work across a full range of cases, as opposed to being considered only in relation to findings with which it is compatible. In what follows, we take a different, albeit complementary approach to this, by examining whether the phenomena that are usually taken to best exemplify the idea of level ordering in noun compounding are themselves actually explained by—or even compatible with—the theory (or some yet to be articulated variant of it).

We will assume that many of the arguments made on behalf of level ordering are logically valid—or at least plausible and coherent—and in the following seven experiments, examine whether the premises on which these arguments are founded are *sound*—i.e., whether the assumptions they make are supported by the evidence. We will also consider whether a distributional learning approach might better explain the overall patterns of data relating to inflection in noun compounding. We begin by outlining and analyzing the arguments presented by Gordon (1985) and Alegre and Gordon (1996).

The basic structure of Gordon's (1985) "learnability" argument can be broken down as follows:

(1a) In English spoken by fluent adults, there are virtually no compounds containing either regular or irregular plurals as the non-head noun (p. 76).
(2a) However, children will still use irregular plurals as non-head nouns in compounds, whereas they will never use regular plurals in this way (pp. 81–83).
(3a) This discrepancy is not reflected in the input (child-directed speech), since adults prefer to use singular non-head nouns in compounding for both regulars and irregulars (p. 86).
(4a) Therefore, this discrepancy is not something children could have learned from experience (Gordon, 1985). Rather, these differences must reflect an "innate constraint on lexical structure such that the ordering effects follow from the system" (Alegre & Gordon, 1996, p. 67).

Alegre and Gordon's (1996) work further fleshed out the details of level ordering, particularly as it related to apparent exceptions to the theory (namely, regular plurals in compounds). The logic of the argument is as follows:

(1b) If level ordering is correct, regular plurals should *only* occur in compounds if they are recursively compounded as part of a noun phrase (e.g., *equal rights amendment* or *American cars exposition*, p. 69).[1]
(2b) For this reason, any adjective-regular plural noun-noun compound can *only* be interpreted as a noun-phrase noun compound, and not as an adjective—noun-noun compound (e.g., red rats eater must be a "red rats" eater and not a red "rats eater"). In other words, the adjective must be perceived as modifying the regular plural specifically, instead of the compound as a whole (p. 70).

[1] While Alegre and Gordon (1996) acknowledge the existence of examples that do not fall under (1b)—e.g., "parks commissioner" or "buildings inspector," they do not explain how their existence is to be accounted for within the level ordering framework.

(3b) Children interpret regular plurals in adjective-noun-noun compounds in line with these predictions (p. 75).
(4b) This counts as further evidence of level ordering in general, and recursion, in particular (p. 75–6).
(5b) *Corollary*: These predictions do not apply to either singular nouns or irregular plurals. Since irregular plurals can be compounded freely, adjective-irregular plural noun-noun phrases (e.g., red *mice* eater) should be ambiguous in their interpretation (p. 66).

Our examination begins with this last argument: although Alegre and Gordon (1996) tested how children interpreted *regular* plurals in adjective-noun-noun compounds (3b), they did not test whether these same children interpreted *irregular* plurals any differently. To make the case that regular plurals are uniquely subject to this compounding constraint (2b), it is necessary to show that it doesn't apply to irregular plurals (5b). With this in mind, we asked: will interpretations of *red rats/rat eater* differ qualitatively from interpretations of *red mice/mouse eater*? Contrary to the predictions of level ordering, the results of Experiments 1 (adults) and 3 (children) suggest that the way English speakers process and interpret irregulars embedded in compounds does not differ qualitatively from the way that they process and interpret similarly embedded regulars. Following up on this idea, Experiments 2 (adults) and 4 (children) examined whether the bias towards 'recursive' interpretations was fixed, or whether it could be over-ridden by semantic considerations. The results in both cases strongly support the idea that it can indeed be over-ridden by semantics.

Next we ask: if level ordering 'optionally allows' irregular plurals to take the non-head place in compounding (e.g., "feet lover"), why is this so rarely seen in the input (see 1a)? Indeed, why do adults seem to produce regular *and* irregular plurals in compounds as if they were governed by the same constraints? Experiment 5 tests the claim that regular and irregular plurals are treated in qualitatively different ways when adult speakers form noun compounds, and finds—contrary to this notion—that adults strongly prefer to include the singular forms of irregular nouns in compounds (just as they do with regulars), and that adults do this even when the semantics of the situation indicate that the noun refers to more than one item. Although these findings are not inconsistent with observations made by proponents of level ordering, we suggest that the consistency of these results, and the necessity of explaining where the conventions adults apply when processing irregular plurals in compounds *come from*, favor an account based on learning rather than on strong constraints that apply in qualitatively different ways to regular and irregular plurals.

Next, does children's inclusion of plural forms in compounds stem from their 'natural' (perhaps innate) inclination to include such forms (2a) or from artifacts of testing? The results of Experiment 6 indicate that the claim that children tend to include irregular plurals in compounds is heavily dependent on a priming bias in the Gordon (1985) elicitation task. (It is unsurprising that irregulars are more susceptible to priming than regulars; the type frequency of irregulars is vastly smaller than that of regulars in English, and thus children get less exposure to—and practice with—them.) Experiment 6 reveals that in the absence of priming, children tend to compound irregular plurals in much the same way that they do regulars: using their singular forms. Taken together, the results of Experiments 5 and 6 suggest that there is no qualitative difference to be found in the way children and adults produce noun-noun compounds, and further, that there is no qualitative difference between the way they produce compounds of regular and irregular plurals. These findings appear to fundamentally undermine conclusion (4a) above.

Finally, in Experiment 7 we ask whether semantic/phonetic conventions might be a better indicator of perceived "acceptability" in compounding than supposed grammatical constraints (see also Haskell et al., 2003). In particular, we examine whether the presence of medial sibilance (the sound of –s in the suffix) in non-head nouns is a better predictor of acceptability than the regular/irregular distinction, and find support for this view.

## 2. Experiment 1

Although Alegre and Gordon's (1996) study of children's interpretations of *red rats eater* presented evidence that is consistent with a level ordering, it did not examine the conditions that would have actually tested its main prediction: namely, that because people perceive that *rats eater* violates the constraints governing normal word formation, they *therefore*—as a result of recursion, and not a learned semantic convention—judge *red rats eater* to be an NP-noun compound. In order to examine this question, and rule out a more general explanation for Alegre and Gordon's results, it is necessary to compare interpretations of adjective-noun-noun compounds that are subject to the constraints of level ordering with interpretations of those that are *not* (see Corollary 5b).

Fortunately, since level ordering—and Alegre and Gordon—assume that regular and irregular inflection are processed in different ways, and since irregular plural non-head nouns in compounds do not violate level ordering constraints, testing the level ordering hypothesis in controlled circumstances is a straightforward affair. Level ordering predicts that *red rats eater* is interpreted "recursively" because *rats eater* violates level ordering constraints (2b). However, level ordering makes no such prediction in the case of *mice eater*, which does not violate level ordering constraints (5b). According to level ordering theory, both *mice* and *eater* are free to enter into the compounding process at Level Two, and thus grammatically, *mice eater* is as equally acceptable a form as *mouse eater*. Accordingly, while level ordering predicts that people's sense that *rats eater* is 'disallowed in the grammar' (Alegre & Gordon, 1996, p. 76) means that *red rats eater* should be interpreted "recursively" as an NP-noun compound (i.e. as a monster that eats red rats), *red mice eater* is ambiguous grammatically, and therefore ought to be ambiguous semantically: a *red mice eater* can be *either* a monster that eats red mice, or a red monster that eats mice, with seemingly equal plausibility (see Fig. 3). Thus level ordering predicts that people should show a preference for interpreting *red rats eater* "recursively." However, if this pattern is to be taken as evidence for "recursion," they should show a different pattern of preference when interpreting *red mice eater*.

Since our goal here is to test whether the patterns of interpretation predicted by level ordering do or do not obtain in English, we first tested competent adult speakers in Experiment 1. As Pinker (1984, 1989) notes, in order to hypothesize about language, we need to have an idea of *what* develops; it is not enough simply to examine *when* certain competencies are present or not:

> "A good model of grammar is necessary [in order to answer] even the most elementary and tentative answers to questions about what it is that is being acquired, how to sort children's utterances into categories, and what mechanisms children are born with that allows them but not kittens to learn language." (Pinker, 1989, p. 461).

There is a dearth of evidence in the literature regarding the compounding constraints that fully competent *adult* speakers manifest. Testing adults enabled us to get a picture of what develops, prior to considering how and when it develops.

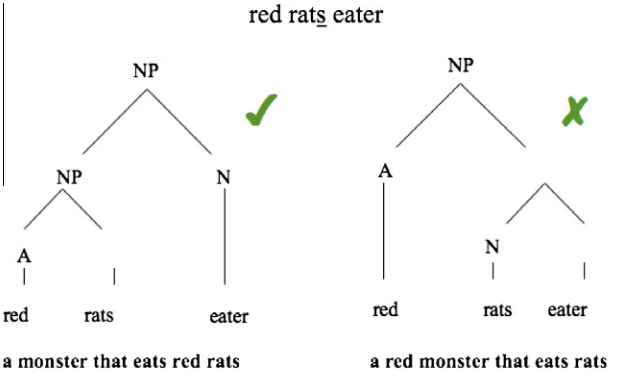


**Fig. 3.** Because the noun-noun compound *rats eater* violates level ordering, the interpretation on the right is not supposed to be valid for *red rats eater* (Alegre & Gordon, 1996). However, both interpretations would be allowed for *red mice eater*.

### 2.1. Participants

Eighty native-English speaking students at Stanford University participated for class credit. Two native-English speaking students at the University of Edinburgh volunteered as raters in the manipulation check.

### 2.2. Materials

This experiment examined eight adjective-noun-noun compound pairs. Each pair comprised two items, which shared the lead adjective and the head noun (e.g., *red* and *eater*) and a regular or irregular non-head noun. The non-head nouns were semantically matched to one another (e.g., *rat* and *mouse*), and the items were designed so as to be plausible in both 'recursive' and 'non-recursive' interpretations (see Figs. 2 and 3). The items used are shown in Table 1.

### 2.3. Procedure

To replicate the forced choice task in Alegre and Gordon's original study, participants were presented with booklets containing eight compound nouns, followed by their 'recursive' and 'non-recursive' interpretations. For example, a *red rats eater* might be a monster that eats *red rats* ('recursive' interpretation) or a *red monster* that eats rats ('non-recursive' interpretation). Participants were asked, "What is the obvious meaning that strikes you? Circle the meaning you would *most likely* assign to each."

To prevent participants from transferring patterns of interpretation based on the characteristics of the non-head noun between the different morphological categories—i.e. to avoid the possibility that interpreting compounds containing regular nouns might influence the subsequent interpretation of compounds containing irregular nouns—participants were tested on either regular or irregular non-head nouns *only*. Thus each booklet contained either regular or irregular non-head nouns in compounds in both their plural and singular forms. The order of items was randomized, and participants saw only one item from any given pair.

The questionnaire was embedded in a larger questionnaire packet that contained many other pages unrelated to this study. Participants completed the questionnaires in their own time.

### 2.4. Manipulation check

Two independent raters were asked to assess whether the adjectives could plausibly be applied to the nouns tested in this experiment. The raters indicated on a 3-point scale (1 = plausible; 2 = implausible; 3 = don't know) whether, e.g. *heavy ox*, *heavy cow* or *heavy register* were plausible adjective noun combinations in English. The raters judged all of the adjective noun combinations studied here to be plausible, and inter-rater agreement was 100%.

**Table 1**
Items examined in Experiment 1. The percentage of NP-noun compound interpretations (singular then plural) is shown for each compound.

| Irregular singular/plural | NP-N Rate S/P | Regular singular/plural | NP-N Rate S/P |
|---|---|---|---|
| New doormouse/mice expert | 0%/31% | Big toe(s) crusher | 57%/58% |
| Big foot/feet crusher | 43%/53% | New gerbil(s) expert | 5%/26% |
| Heavy ox/oxen register | 53%/71% | Heavy cow(s) register | 63%/80% |
| Fat goose/geese chaser | 42%/62% | Fat duck(s) chaser | 63%/76% |
| Old fireman/firemen list | 52%/58% | Old farmer(s) list | 71%/68% |
| Small octopus/octopi scale | 38%/79% | Small flamingo(s) scale | 53%/48% |
| Awful child/children scarer | 21%/14% | Awful girl(s) scarer | 10%/42% |
| Red mouse/mice eater | 63%/67% | Red rat(s) eater | 58%/62% |

### 2.5. Results

Consistent with Algere and Gordon's (1996) findings, in this experiment participants interpreted the items containing a regular plural non-head noun (*fat ducks chaser*) "recursively" (as NP-noun compounds) more often ($M$ = 58%, SEM = 4.9) when the non-head noun was presented as a singular (*fat duck chaser*, $M$ = 47%, SEM = 4.6). However, this trend was repeated for the irregular items: participants interpreted items containing an irregular plural non-head noun (*fat geese chaser*) as NP-noun compounds more often ($M$ = 54%, SEM = 3.8) than when the non-head noun was presented as a singular (*fat goose chaser*, $M$ = 38%, SEM = 4.2). Mean interpretation rates for the items are shown in Table 1. A 2 form (singular vs. plural) $\times$ 2 type (regular vs. irregular) repeated measures ANOVA of our participants' interpretations confirmed that these differences resulted from a tendency to adopt NP-noun compound interpretation based on form, whenever a non-head noun was marked plural rather than singular ($F(1, 78) = 11.463$, MSe = .72, $p < .001$). No interaction between form and type ($F(1, 78) = .304$, MSe = .02, $p > .5$) was observed.

Post-hoc paired $t$-tests revealed that subjects preference for NP-noun compound interpretations when the non-head noun was plural was significant both for irregulars ($t(39) = 3.32$, $p < .001$) and regulars ($t(39) = 1.8$, $p = .05$).

### 2.6. Discussion

Level ordering does not explain the findings obtained here, nor can it. Although these data appear to confirm one aspect of Alegre and Gordon's (1996) findings—that plural non-head nouns in compounds such as *red mice eater* tends to bias an NP-noun interpretation (2b)—they roundly contradict previous interpretations of what this implies: it does not appear from these results that the bias towards an NP-noun interpretation of *red rats eater* results from the grammatical considerations put forward by Kiparsky (1982, 1983) and Gordon (1985).

In these results participants demonstrated a clear bias to interpret *irregular* plurals in compounds as NP-noun expressions, even though irregular non-head noun plurals in compounds *do not* violate the proposed constraints of level ordering. Indeed, in all of the measures taken here, the plural bias towards an NP-noun interpretation was actually clearer in the irregular items than it was in the regular ones. Since violations of level ordering were supposed to explain the bias towards an NP-noun interpretation when regular non-head nouns were plural, and since in the case of irregular plurals in compounds level ordering quite clearly *cannot* explain this, some other explanation for this effect is needed.

One potential explanation lies in the variable pattern of participants' responses to different items in Experiment 1. This variability seems more compatible with the idea that the NP-noun interpretation of plurals in compounds may be driven *contextually*, by local semantic (and perhaps phonological and prosodic) considerations and conventions. This would suggest that people's interpretations were driven by the properties of specific items rather than by a categorical rule. Accordingly, Experiment 2 was designed to contrast the rule-based account of the NP-noun interpretation of plurals in compounds with a semantic explanation.

## 3. Experiment 2

According to level ordering theory, the presence of a regular plural non-head noun in an adjective-noun-noun compound should prompt an NP-noun interpretation (2b). However, while an NP-noun interpretation fits cases such as *chemical weapons inspector* and *identical twins project* (where one would expect the twins to be identical, and the weapons to be chemical), it appears to run into problems when changes are made to the adjectives heading these compounds. In both *portly weapons inspector* and *controversial twins project*, the NP-noun interpretation (which suggests that the weapons are portly, and the twins are controversial) though *possible*, is not the most plausible one semantically. The adjective-noun-compound interpretations (in which the inspector is portly, and the project is controversial) seem to make more sense.

Are NP-noun interpretations of non-head noun plurals in compounds driven by semantic or grammatical factors? Experiment 2 was designed to contrast the predictions of level ordering with predictions arising from simple semantic considerations. Experiment 2 examined the way that people interpreted adjective-noun-noun compounds when the adjective was more strongly associated with one noun rather than another, but where interpreting the adjective as modifying either the head or non-head noun would not directly contravene their world knowledge. Could the bias towards NP-noun interpretations when plural non-head nouns are present in compounds (confirmed in Experiment 1) be *over-ridden* by manipulating the semantic relationships between the adjective and the head and non-head nouns?

### 3.1. Participants

Sixty-six native-English speaking students at Stanford University participated for class credit. Four native-English speaking students at the University of Edinburgh volunteered as raters in the manipulation check. (This was a different group of participants to that in Experiment 1; separate sample groups were used in all of the studies reported here.)

### 3.2. Materials

This experiment examined eight adjective-noun-noun compound sets such as *relentless rats/mice eater* in which the non-head noun was always pluralized. Each set comprised four items, in two pairs: in the first pair, two alternate non-head noun plurals—a regular and irregular noun-plural that were semantically matched (e.g., *rats* or *mice*) were paired with an adjective that was more commonly associated semantically with the head noun (e.g., *relentless* and *eater*)—while in the second pair, the adjective was one that was more commonly associated semantically with the non-head noun (e.g., *helpless* and *rats/mice*—see Table 2). Notably, these materials were devised specifically to test for the possibility that compounding is responsive to semantics, and were not part of Gordon's (1985) original suite of experiments.

### 3.3. Procedure

Participants were presented with booklets containing eight compound nouns, followed by their 'recursive' and 'non-recursive' interpretations. Participants were asked, "What is the obvious meaning that strikes you? Circle the meaning you would *most likely* assign to each."

**Table 2**
The items examined in Experiment 2. The percentage of NP-noun compound interpretations for each alternate adjective is shown for each compound in the final column.

| | | NP-N Rate |
|---|---|---|
| *Regular* | | |
| Stinky vs. metal | Hands crusher | 83%/57% |
| Brave vs. long | soldiers list | 100%/14% |
| Prize-winning vs. printed | Bulls register | 64%/45% |
| Cute vs. slimy | Girls scarer | 100%/57% |
| Helpless vs. relentless | Rats eater | 43%/25% |
| Pink vs. precise | Flamingos scale | 67%/14% |
| Fat vs. modernized | Ducks farm | 100%/21% |
| Scuttling vs. knowledgable | Gerbils expert | 71%/8% |
| *Irregular* | | |
| Stinky vs. metal | Feet crusher | 83%/14% |
| Brave vs. long | Firemen list | 50%/83% |
| Prize-winning vs. printed | Oxen register | 71%/33% |
| Cute vs. slimy | Children scarer | 75%/14% |
| Helpless vs. relentless | Mouse eater | 71%/17% |
| Pink vs. precise | Octopi scale | 57%/8% |
| Fat vs. modernized | Geese farm | 75%/7% |
| Scuttling vs. knowledgable | Doormouse expert | 93%/8% |

As in Experiment 1, to prevent participants from transferring patterns of interpretation derived from items containing regular non-head nouns to those containing irregular non-head nouns, participants were tested on either regular or irregular non-head nouns only.

The presentation order of items was randomized, and participants saw only one item from each pair. The questionnaires were embedded in a larger questionnaire packet containing many other pages unrelated to this study. Participants completed the questionnaires in their own time.

### *3.4. Manipulation check*

Two independent raters were asked to assess whether the adjectives could plausibly be applied to the nouns tested in this experiment. The raters indicated on a 3-point scale (1 = plausible; 2 = implausible; 3 = don't know) whether, say, *heavy ox*, *heavy cow* or *heavy register* were plausible adjective noun combinations in English. The combined plausibility rating of the two judges was 94.9%, and at least one rater found each one of the adjective noun combinations plausible. Inter-rater agreement was 89.7%.

Another two independent raters assessed which of the target nouns were most commonly associated with the adjectives paired with them in this experiment. The raters were asked to indicate their preferences in pairs (e.g. say which of *feet* or a *crushing machine* they would most associate the adjective *stinky*) or else indicate that they didn't know. The raters selected the predicted associate (i.e. said they associated *stinky* more with *feet*) 94.6% of the time, and inter-rater agreement was 89.2%. At least one rater selected the predicted associate in each pair, and there were no instances of raters selecting the predicted non- (or lesser-) associate (i.e. neither said they associated *stinky* more with a *crushing machine*). The 5.4% of occasions where one of the judges did not select the predicted associate were "don't knows."

### *3.5. Results*

Consistent with level ordering (2b), when adjectives were semantically associated with a regular plural non-head noun in an adjective-noun-noun compound, 78% (SEM = 5) of participants preferred the 'recursive' interpretation (e.g., *helpless rats eater* was interpreted as a monster that eats helpless rats). However, contrary to the idea that this is evidence that supports level ordering theory (5b)—and consistent with the results of Experiment 1—the *same* pattern was revealed when the adjective was semantically associated with an irregular plural non-head noun in an adjective-noun-noun compound: 76% (SEM = 4) of participants preferred the NP-noun interpretation (e.g., *helpless mice eater* was interpreted as a monster that eats helpless mice).

Further, consistent with the idea that NP-noun interpretations of plurals in compounds result from *semantic* factors, participants did not favor an NP-noun interpretation when the adjective was more commonly semantically associated with the head noun, despite the plurality of the non-head noun in the adjective-noun-noun compound. In these cases, for both regular-plural (79%, SEM = 5) *and* irregular plural (86%, SEM = 4) non-head nouns, participants preferred the interpretation that interpreted the adjective as modifying a noun-noun compound to the NP-noun interpretation preferred in Experiment 1.

Confirming this overall pattern of results, a 2 semantic bias (to head noun vs. to non-head noun) $\times$ 2 type (regular or irregular) repeated measures ANOVA revealed a strong effect of semantics on our participants' interpretations of the adjective-noun-noun compounds ($F(1, 50) = 107.16$, MSe = 6.68, $p < .001$), but no interaction between type and semantics ($F(1, 50) = .57$, MSe = .04, $p > .45$). Post-hoc paired sample *t*-tests confirmed that the different interpretation patterns produced by the semantic manipulation were significant within individual subjects in both the regular ($t(26) = 6.51$, $p < .001$) and irregular ($t(26) = 8.21$, $p < .001$) groups.

### *3.6. Discussion*

The pattern of results obtained in this experiment is consistent with participants' interpretations of noun compounds being driven by semantic conventions. They do not, however, offer support for the claim that people's interpretations are governed by their sensitivity to level ordering (4b): there were no qualitative differences in the way that participants in Experiment 2 interpreted regular and irregular noun plural items, and—despite their being no particular semantic objection to it—the vast majority of participants rejected a recursive interpretation of *relentless rats eater* (as a monster that eats relentless rats) (5b). Although level ordering predicts that participants should prefer not to interpret *relentless rats eater* as a relentless monster that eats rats because *rats eater* violates level ordering (2b), 74.8% of participants did prefer this interpretation. To these participants, a *relentless rats eater* was simply a relentless monster that eats rats.

### 3.7. Interpreting noun compounds: what rules?

Level ordering maintains that qualitatively different processes are involved in the comprehension patterns noun-compounds depending on whether they contain regular or irregular plurals. If these comprehension differences do indeed stem from the innate structural factors that are assumed to affect the production and subsequent inflection of noun compounds (Kiparsky, 1982, 1983), it seems reasonable to expect that it should be possible to discern qualitative differences in the way that people actually interpret noun-compounds. Yet the results from Experiment 1 suggest that there actually *are no* qualitative differences to be found in the way adults comprehend noun-compounds that contain regular as opposed to irregular plural non-head nouns. Moreover, the results of Experiment 2 suggest that the supposedly innate, structural interpretation proposed by level ordering can be *over-ridden* simply by familiarity (local semantic associations), even when the 'recursive' interpretation is plausible.

Thus, when it comes to interpretation, the results of Experiments 1 and 2 do not support any of the interesting predictive claims of level ordering (Alegre & Gordon, 1996; Pinker, 1999). There appears to be no qualitative difference in the way that regular and irregular plurals in compounds affect adult English speakers' interpretations of these compounds.

### 3.8. Red rats, recursion, and the rules of hypothesis testing

As noted in Experiment 1, an obvious deficiency in Alegre and Gordon's study of children's comprehension of 'recursive' and 'non-recursive' noun compounds was that they only tested children on *regular* plurals. No observations or measurements were taken regarding children's interpretation of irregular noun compounds. Although level ordering maintains that irregular plurals are not subject to recursion, Alegre and Gordon did not test this prediction. Rather, their study simply sought confirmatory evidence for recursion in regular compounds.

Accordingly, Experiments 3 and 4 were designed to examine whether the interpretations of regular noun compounds made by children in Alegre and Gordon's study ought to be attributed to children's sensitivity to recursion in level ordering *or* whether these interpretations might be more readily explicable in terms of learning. If children are sensitive to recursion in level ordering, they ought to be biased toward recursively interpreting appropriate regular plural compounds in a way that is qualitatively different to the way they interpret irregular plural compounds. If, on the other hand, they learn general conventions about plurals in compounds, we would expect them to treat regular and irregular forms in a similar fashion.

## 4. Experiment 3

Will children prefer a 'recursive' NP-noun compound interpretation with both regular *and* irregular pluralized non-head nouns (as adults did in Experiments 1 and 2) or—as level ordering suggests—will they only prefer 'recursive' interpretations with *regular* nouns (2b, 5b)? To explore this question, Experiment 3 sought to replicate Alegre and Gordon's basic findings by testing children's interpretation of adjective-noun-noun compounds with singular and plural regular non-head nouns, while also examining whether children would adopt the same pattern of interpretation when the regular non-head nouns were replaced with irregulars.

One problem in studying children in the 3–5 age range is that they are notoriously perseverative in their behavior, and often repeat an initial response in the face of further questioning, even when the demands of a task are subtly changed (see e.g., Diamond & Goldman-Rakic, 1989; Kirkham & Diamond, 2003; O'Sullivan, Mitchell & Daehler, 2001; Ramscar & Gitcho, 2007). Thus it is likely that a child's response to given trial may be influenced as much by his or her response in a preceding trial as it is by a given presented item, and it is unclear to what degree studies of children in this age group should rely for their power on repeated measures designs. To address these concerns, the sample sizes in all of the cells of Experiment 3 were increased twofold over those in Alegre and Gordon (1996), and a nonparametric analysis of children's noun-compounding comprehension was conducted.

### 4.1. Participants

Seventy-eight native-English speaking children participated voluntarily in this study. As in Alegre and Gordon (1996), the children comprised three groups, 3 year olds (28 children, average age 3;6), 4 year olds (25 children, average age 4;6) and 5 year olds (25 children, average age 5;3). Two Stanford undergraduates participated in a control evaluation for class credit.

### 4.2. Materials

Following Alegre and Gordon (1996), materials comprised slides depicting monsters blocked in various colors (dark blue, pale blue, pink, gray, green and red) eating several smaller items blocked in a contrasting color (extensive pre-testing revealed that children of this age could easily identify and contrast these colors; the two blues were never contrasted in test pairs). Slides were arranged such that two contrasting scenes were depicted—a monster of color A eating items of color B, and a monster of color B eating items of color A—such that in one scene the monster (eater) was the color used in the descriptive sentence (e.g., red rats(s) eater) and in the other it was the eatees (rats) that were so colored.

Four regular items (rats, crabs, flowers and spiders) and four irregular items (mice, geese, feet and spacemen) were used to depict the dietary preferences of the various monsters.

### 4.3. Procedure

In the main test, each child was presented with eight pairs of pictures depicting a monster of some sort eating one of the various edible items. In each pair, one of the pictures depicted the monster in the target color, and one depicted the things being eaten in the target color.

Following Alegre & Gordon, in a between-subjects design, half of the subjects were randomly assigned to a plural condition, and half to a singular condition. Within those conditions, regular and irregular items were presented in blocks, with half of the children randomly assigned to a regular block followed by an irregular block sequence, and the other half assigned to an irregular block followed by a regular block sequence.

For testing purposes, items were presented on the screen of a laptop computer that allowed the presentation of items within blocks to be randomized. The children were presented with each of the pairs of pictures, and asked to identify the things depicted in them (monsters, mice, rats, etc.) and their colors (for a *red rats(s) eater* pair, the children were prompted to say that both pictures depicted a monster and some rats, and that in one picture the monster was red and the rats were blue, and in the other the monster was blue and the rats were red).

Following this, the children in the *plural condition* were prompted to "point to the picture that shows a *red rats eater*." Children in the *singular condition* were prompted to "point to the picture that shows a *red rat eater*." The same pattern of prompting and questioning was repeated for all eight of the test items.

An experimenter blind to the hypotheses conducted the testing. The experimenter was given extensive pre-training in pronouncing the test descriptions evenly so as not to adopt a stress pattern that might bias the children's interpretation of the descriptions (see Alegre & Gordon, 1996; a control analysis of the experimenter's spoken presentation is described in the results section). All sessions were videotaped and then transcribed.

### 4.4. Results

As anticipated, the distribution of data from our child participants was bimodal (the majority of children responded to all of the items with either one response pattern or the other). Accordingly, for the purposes of analysis, participants were binned according to whether they made either a majority of recursive or non-recursive interpretations on each test (cases where participants did not show a bias for one interpretation or another—12 out of 156 data—points were excluded). Fig. 4 shows the percentage of participants who provided recursive responses to regular and irregular stimuli by compound type and age group. As can be seen, contrary to the claim that level ordering reflects a unique set of constraints that apply only to regular nouns (2b, 5b), the overall pattern of data for both the regular and irregular items in this experiment is remarkably consistent: there appears to be no effect of age on the pattern of responses, whereas there *does* appear to be a qualitative effect of condition for both regular and irregular items.

Analyzing the *regular* items reveals a pattern of data that is consistent with the predictions of level ordering, and broadly replicates the findings of Alegre and Gordon (1996): In the *singular condition*, children did not show a preference for recursive interpretations (46% made a majority of recursive interpretations, whereas 54% made a majority of non-recursive interpretations). However, in the *plural condition* children did show a strong preference for recursive interpretations (78% made a majority of recursive interpretations, as opposed to just a 22% preference for non-recursive interpretations). A chi-square analysis revealed this difference to be significant ($\chi^2(1, N = 74) = 8.27$, $p < .0005$).

However, analysis of the *irregular* items showed exactly the *same* pattern of interpretative preference. Again, in the *singular condition*, children did not show a preference for recursive interpretations (43% made a majority of recursive interpretations, whereas 57% made a majority of non-recursive interpretations). And again, in the *plural condition*, children showed a strong preference for 'recursive' interpretations (76% made a majority of recursive interpretations, as opposed to just a 24% preference for non-recursive interpretations). Once again, a chi-square analysis revealed this difference to be significant ($\chi^2(1, N = 72) = 8.05$, $p < .0005$. (A further analysis which considered data only from the first block of items – either regular or irregular – that each participant completed, suggested that these differences resulted independently and reliably from both sets of items: regular items, $\chi^2(1, N = 37) = 3.26$, $p < .08$; irregular items, $\chi^2(1, N = 36) = 4.50$, $p < .05$).

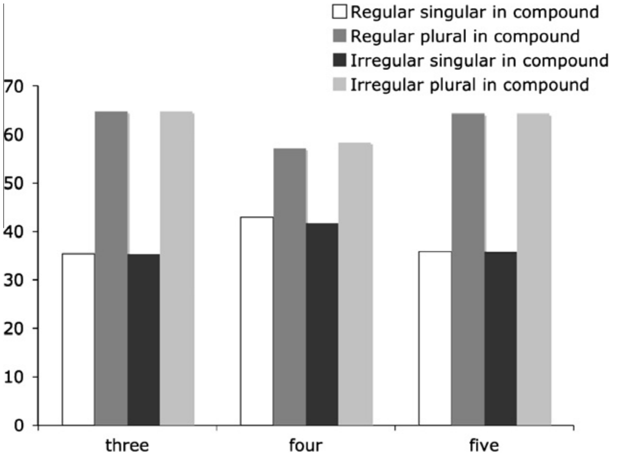


**Fig. 4.** Rates at which participants provided predominantly recursive interpretations in Experiment 3 by age and stimulus type.

Although our experimenter was blind to the hypotheses tested, it is still possible that her stress patterns in providing the descriptions to the children may have contributed to these results. To control for this possibility, two raters were provided with 60 excerpts from the trials randomly selected from the videos of the test sessions. Each excerpt began with the experimenter asking the child to point to the described picture (e.g., "the red rat(s) eater") and concluded prior to the child's indication of their preference. Thirty of the trials were from the plural condition and 30 were from the singular condition, and of each of these 30 trials, half were of instances where a child gave a recursive interpretation, and half were of instances where a child gave a non-recursive interpretation. The ambiguity of the descriptions was explained to the raters, as well as the way in which stress could be used to disambiguate these descriptions (e.g., "RED rat eater" vs. "red RAT eater"; see Alegre & Gordon, 1996). For each trial, the raters were then asked to judge which picture the child selected based on the stress pattern of the experimenter. Raters were given as much time as they needed to complete the task, and were encouraged to view the recording as often as they wished in order to reach the right judgment.

In spite of their being aware of the way in which stress could be used to disambiguate the descriptions, the raters proved to be poor predictors of the children's subsequent choices given a particular utterance from the experimenter. Rater 1 managed only a 48% success rate, while Rater 2 managed only a 57% success rate. Nor was there much agreement between the raters (inter-rater agreement was 58%), which suggests that there was little information in the experimenter's stress patterns to cue the children (or raters) to a particular response.

### 4.5. Discussion

Consistent with our adult findings in Experiments 1 and 2, the children in this study tended to interpret compounds containing regular or irregular plural nouns in the non-head position in much the same way. In the absence of any qualitative difference in the manner in which children and adults interpret irregular and regular plurals in compounds, it seems unlikely that their sensitivity to "recursion"—as described by level ordering—is responsible for shaping those interpretations.

## 5. Experiment 4

What then of the findings in Experiment 2, that adults' interpretations of compounds are less structural (recursive) than semantic, and that 'recursive' interpretations will be violated if semantic factors within a compound support a non-recursive interpretation? Will children's interpretations of plurals in compounds also show the same malleability when semantics are changed, and if so, what kind of semantic factors are children sensitive to? The original Alegre and Gordon study provides an ideal basis for investigating these matters. Alegre and Gordon's study comprised two cells (replicated in Experiment 3) of what is essentially a four cell design. Child participants saw a monster eating multiple rats, and either heard a plural (red *rats* eater) or singular (red *rat* eater) description. Filling in the remaining cells of Alegre and Gordon's design—by testing pictures of a monster eating a *single* rat (Fig. 5)—could enable further light to be shed on these questions.

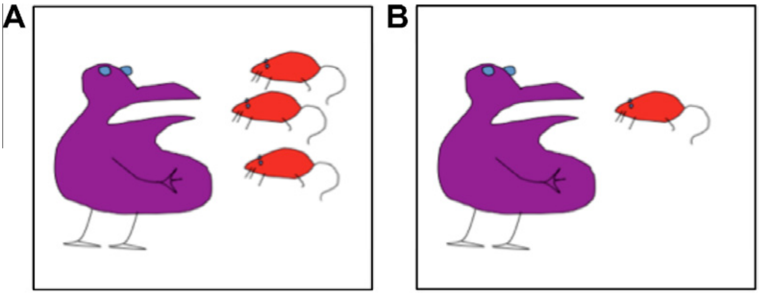


**Fig. 5.** Black and white rendering of the stimuli tested in Experiment 3 (A) and Experiment 4 (B). In the experiment, the monster was purple and the rats were red.

If level ordering theory is correct, Alegre and Gordon's findings should be replicated and supported even when the semantics of the pictures presented to the children are changed. According to level ordering theory, 'recursive' interpretations are determined by structural factors (1b, 2b), *not* local semantic considerations. Thus, changing the semantics of the pictures so that a child is looking at a picture of a red monster eating a green rat (singular) and a green monster eating a red rat (singular), should not affect the pattern of data obtained. In either case, level ordering predicts that *red rat eater* should still be ambiguous—because structurally the description is still ambiguous (*red rat eater* can be either an NP noun or an adjective modifying a noun-noun compound; 5b)—while *red rats eater* ought to still be *un*ambiguous (in terms of level ordering theory it can only be an NP noun; 2b).

A semantic processing account would make a very different set of predictions. Recently, a great deal of light has been shed on the way that semantic interpretations are formed incrementally as sequential inferences, as listeners attend to sentences (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003; Tanenhaus & Brown-Schmidt, 2008; Altmann & Mirković, 2009; Gennari & MacDonald, 2009; Ramscar, Yarlett, Dye, Denny, & Thorpe, 2010), and there is considerable evidence that children begin to use incremental, predictive information in comprehension in the earliest stages of language processing (e.g., Fernald, Swingley, & Pinto, 2001; Swingley, Pinto, & Fernald, 1999; Swingley, 2009). These findings suggest an alternative explanation for children's recursive interpretations of *red rats eater* vs. *red rat eater*.

Consider the demands on a child in Alegre and Gordon's *red rats eater* task. The child is looking at pictures of a red monster eating green rats, and a green monster eating red rats, and has to select a picture to match a description. Data from incremental processing studies suggests that on hearing *red*, the child has two possible places to look or inferences to make: the monster or the rats. On hearing *rats*, the child gets evidence that disambiguates the task: only one picture contains *red rats*, so the child can fixate on the red rats and infer that this is the object of the description (see also Tanenhaus & Brown-Schmidt, 2008; Altmann & Mirković, 2009; Ramscar et al., 2010). On the other hand, hearing *red rat* less clearly disambiguates the task, because *rats* not 'rat' (singular) are present in both pictures. In this instance, things remain somewhat ambiguous, which may be reflected in children's responses, which favor neither picture.

Now consider the demands on a child in the new condition of the task, in which one picture shows only a single rat being eaten. If the child is indeed determining her response to the task on-line, as the description is given to her, then *red rat eater* will be unambiguous (one picture contains a picture of a *red rat*) and so the children's responses to this description ought to *change* with the new semantics of the pictures in this task. With regards *red rats eater*, it is harder to determine exactly what an incremental semantic processing account would predict. This is because in English, the bare plural *rats* tends to be used as both a plural and a generic (as in "look a rat – I hate rats"), whereas bare singulars such as *rat* do not tend to be used as generics (people tend not to say "look a rat – I hate rat"; see Chierchia, 1998, for a full discussion of this distinction). If children treat *rats* as a plural, then one might expect them not to prefer recursive interpretations (one might expect a pattern of data analogous to that for *red rat eater* when multiple rats are depicted); on the other hand, if children treat *rats* as a generic, then a single rat should be sufficient for them to complete their inference about the object of the picture description, and they ought to stick with a 'recursive' interpretation.

### *5.1. Participants*

Seventy-seven native-English speaking children participated voluntarily in this study. For comparison with the results of Experiment 3, the children were divided into three groups: 3 year olds (24 children, average age 3;7), 4 year olds (29 children, average age 4;7) and 5 year olds (24 children, average age 5;4). One Stanford undergraduate participated in a control evaluation for class credit.

### *5.2. Materials*

The materials were the same as those used in Experiment 3 *except* that the slides depicted monsters (dark blue, pale blue, pink, gray, green and red) eating *single* smaller items *instead* of multiple items. Once again, the pictures of the monsters and their prey were presented in pairs, with the monster depicted in the target color in one picture and the prey depicted in the target color in the other. The colors and items used were the same as those in Experiment 3.

### 5.3. Methods

The methods were the same as for Experiment 3. Children in the *plural condition* were prompted to "point to the picture that shows a *red rats eater*," and children in the *singular condition* were prompted to "point to the picture that shows a *red rat eater*." The experimenter was blind to the experimental hypotheses.

### 5.4. Results

Consistent with the idea that *semantic* rather than structural factors were determining children's interpretation of the noun compound descriptions of the stimuli, when a single rat was depicted being eaten, children generally provided 'recursive' (NP-noun) interpretations in the singular condition (*red rat eater*), in contrast to Experiment 3 where the same description resulted in no preference. Seventy percent of the children's interpretations in the regular singular condition in Experiment 4 were 'recursive' as compared to 46% in Experiment 3. This change was significant ($\chi^2(1, N = 74) = 5.50$, $p < .05$.) Once again, the same pattern of data was obtained for irregular items: 69% of the children's interpretations in the irregular singular condition in Experiment 4 were 'recursive' as compared to 42% in Experiment 3 ($\chi^2(1, N = 71) = 5.10$, $p < .05$.)

In the plural condition, results were largely comparable to those of Experiment 3. Seventy-four percent of the children's interpretations in the regular plural condition in Experiment 4 were 'recursive,' similar to the 78% in Experiment 3, and 71% of the children's interpretations in the irregular plural condition in Experiment 4 were 'recursive,' in keeping with the 76% in Experiment 3.

As in Experiment 3, a rater was presented with 60 excerpts of the experimenter giving the noun compound descriptions to the children, and as before, the rater proved a poor judge of their choices, achieving just a 55% success rate (this was in part due to the rater guessing that the children would choose the recursive interpretation 65% of the time, despite being informed that half of the choices were of each interpretation). It is thus highly unlikely that the stress patterns adopted by the experimenter in Experiments 3 and 4 had any bearing on our findings.

### 5.5. Discussion

The findings of Experiment 4 offer support for the idea that children are sensitive to contextual and semantic factors when they interpret noun compounds. This is in keeping with the findings of Experiments 1 and 2. These data do not, in themselves, lend support to any particular model of the *way* in which contextual and semantic factors are incorporated into compound comprehension (for more detailed proposals, see Tanenhaus & Brown-Schmidt, 2008; Altmann & Mirković, 2009; Ramscar et al., 2010).

For our present purposes, however, one thing seems clear: In Experiment 4, the pattern of data obtained for both regular and irregular items were again qualitatively the *same*, in keeping with the adult data in Experiments 1 and 2. These results do not support the idea that different constraints apply to regular and irregular nouns in compounding. Given that level ordering theory was motivated by the assumption that these differences *do* exist in compounding, it is not clear what the theory adds to our understanding here, or whether the theory is even capable of making testable claims about this kind of behavior.

### 5.6. Whither production?

The data from four experiments examining comprehension of noun plural compounds, in both children and adults, has thus far lent little support to the level ordering account. Further, these findings suggest very different predictions regarding adult performance when it comes to the *production* of noun compounds than does level ordering.

Recall the questions we asked at the outset: If level ordering optionally allows irregular plurals to take the non-head place in compounding, why is this so rarely attested to in the input? Indeed, why do adults seem to produce regular and irregular plurals in compounds as if they were governed by the same constraints?

Alegre and Gordon argue that the structural constraint on regular plurals proposed in level ordering "seems to be a part of adult speakers' competence" (Alegre & Gordon, 1996, p. 67). Level ordering thus maintains that adults produce compounds involving irregular plurals in a manner that is qualitatively different from compounds involving regular plurals (Alegre & Gordon, 1996; Gordon, 1985). The obvious problem with this claim is Gordon's own observation that adults widely eschew both regular and irregular plural non-head nouns in compounding (1a, 3a).

This raises the question of exactly what (and how) level ordering is supposed to add to our understanding of the psychological processes involved in language production and comprehension. The findings presented above suggest that there simply *are no* qualitative differences in both child and adult speakers' interpretations of regular and irregular plurals in compounds. This would seem to indicate that (ordinarily) neither children nor adults differentiate between regular and irregular plural forms in production either. (Unless production is conventionalized, it is hard to see how comprehension can be conventionalized too; see Ramscar et al., 2010). If this is the case—i.e., if children and adults show roughly the same patterns of production—this would clearly call into question the scientific bona fides of the level ordering hypothesis. Of course, as Gordon rightly notes, "one cannot argue that a theory is disproved because it failed to predict a particular outcome" (P. Gordon, personal communication, 2010). However, if the evidence were to show that children and adults do not make a qualitative distinction between regulars and irregulars when compounding, then given that level ordering theory exists to *explain* such differences, it would seem that the theory has failed the usual measures of likelihood and plausibility that govern the everyday practice of science (Kuhn, 1962).

To explore this possibility, we now turn our attention to production. In keeping with the motivation underlying Experiments 1, 2, 3 and 4—which sought to illuminate how language develops by isolating *what* develops (Pinker, 1989)—Experiment 5 examines the constraints applying to noun-compound production by first exploring the performance of adult speakers in tasks that replicate and extend the paradigm used by Gordon (1985).

## 6. Experiment 5

Gordon (1985) argues that the patterns of inflection obtained in his study provide evidence that children are sensitive to the difference between irregular and regular plurals in compounds, and to the different constraints that apply to them (2a). Further, Alegre and Gordon (1996) argue that adult competence embodies the same constraints that are present in children. However, the findings presented here suggest that adults are sensitive to *both* regular and irregular non-head noun plurals in compounds. If, as our results suggest, adults conventionally omit regular plural forms from noun-noun compounds in line with their interpretative conventions, then they also ought to omit *irregular* plural forms from noun-noun compounds.

In order to test this, a replication of Gordon's (1985) elicitation task was conducted on fully competent adult Native English speakers. Noun-agentive compounds (e.g., *rat*-eater) were elicited from English speaking adults. As in Gordon's study, in the main replication, "the context was biased to predispose [participants] to use plural forms inside the compound. This was done both by having a plural referent for the non-head (left) noun, and by having participants produce the plural form (*rats*) prior to the compound form (*rat(s)-eater*)," Gordon, 1985, p. 78) Additionally, since our interest here is in what kinds of noun-noun compounds competent adult Native English speakers ordinarily produce (as opposed to the kind of speech-errors that can be induced from them), two other variations on Gordon's original task were examined. Since participants performing Gordon's task voice a plural immediately prior to compound production, two novel variations were designed to control for the possibility that this might result in *priming* that is not necessarily reflective of ordinary, spontaneous noun-compound production.

### 6.1. Participants

Forty-five native-English speaking students at the University of Edinburgh participated voluntarily in this experiment.

### 6.2. Materials

Following Gordon (1985), participants were trained in the compounding task using non-pluralizable mass nouns. The stimuli were pictures of mud, sand, bread, fruit and corn. The main test items were pluralizable count nouns. Gordon (1985) elicited the singular form for items before eliciting the plural in order to check that the children had lexical knowledge of the items (for some items the majority of children had to be supplied with the singular form). Since we were confident that this was not a problem in the adult condition, in the replication (and its variant) participants were supplied the name of the item with a picture of it in singular form, and then the plural was elicited using an unlabeled picture of multiple examples of the item. The items tested were: regular – *ear, hammer, shirt, knife,*[2] *hand, ring, cow, doll, duck, rat;* irregular – *goose, snowman, mouse, tooth, child, foot*; pluralia tantum nouns – *trousers, pliers, scissors, glasses*. (Note that, as in Gordon's original study, each irregular or pluralia tantum noun had a regular semantic complement.)

### 6.3. Procedure

Trials were videotaped and later transcribed. Participants were tested individually, and told that they were assisting in piloting a task that was subsequently to be conducted on children. They were told that the task might seem trivial, but that their answers were important, and that they should give the answers that seemed most natural to them.

As in Gordon's original study, the 14 participants in the *main replication condition* introduced to a "Biscuit[3] Monster" and told:

> "Do you know who this is? ... It's the Biscuit Monster. Do you know what he likes to eat? (Answer: Biscuits.) Yes. But this is a particularly greedy Monster – he eats all sorts of things ..."

Participants were then shown a picture of one of the training items, and asked, "What is this? (Answer: X) – What would you call someone who eats X?" (Answer: An X-eater). Replicating Gordon (1985), this procedure was used to elicit compounds of the form *corn-eater, fruit-eater*, etc. (see also Graves & Koziol, 1971).

For the main items, the singular, plural and compound forms were elicited. To elicit the singular form, participants were shown a labeled picture of a single item and asked, "What is this?" For the plural, an unlabeled picture of a group of the items was presented, and participants were asked, "And what are these?" The Monster was then made to eye the picture greedily, and the participant was asked, "And what would you call a Monster that eats X?" where X was the form of the noun produced by the participants in response to the plural elicitation.

For the 11 participants in the *first priming control condition*, the procedure was identical to the main condition, except that in the compound task, participants were asked "And what would you call a Monster that eats them?" This was to see whether the experimenter's voicing of the plural immediately prior to the compound production—in addition to the participants' having just done so—had any effect in priming the production of plurals in compounds.

The *second priming control condition* was designed to control for the likelihood that simply voicing the plural form of a noun immediately prior to compounding might prime plurals in compounds. In this condition, 16 participants first took part in a naming task, where singular forms of the items were elicited, after which they were shown two training items that were labeled (*sand* and *mud*). Following this, each participant was shown pictures of two other training items (*bread* and *fruit*), and asked to

[2] It is questionable whether *knife* is a regular noun. This point is discussed further below.

[3] To our British-English speaking participants "biscuit" was the equivalent of the American "cookie."

describe the kind of monster they thought might eat such items (answer: bread-monster, etc.). Participants were then shown pictures of large green Monsters that were depicted with their preferred dietary items (In all of the main test items, the Monster's dietary preferences were presented as multiple items, e.g., participants saw a monster and a bunch of rats; see Fig. 5A).

At this point, participants were asked to name the monster. They were then presented with a picture showing only its diet and were asked to state what it was that the monster ate. This condition thus served to test a prediction made about learning that was also tested in Gordon's original experiment. Gordon (1985) suggests that:

> [if someone] "were presented with a context in which there were a number of rats being referred to, one might expect him or her to denote an eater of such animals as a **rats-eater* rather than a *rat-eater*. However, if we assume that level ordering constrains pluralization in this case, then such errors should never be found... if [one] has to learn that reduction of regular plurals is required in compounds, there is little reason to assume that this rule would apply only to regular plurals. This is especially true considering the fact that... the input data tends not to include compounds containing irregular plurals... a natural induction from such evidence would be that irregular plurals are also subject to reduction inside compounds." (Gordon, 1985, pp. 77–78; our emphasis).

Participants produced a compound to describe a monster that was depicted looking hungrily at several examples of each test item *prior* to naming the plural form so as to control for any interference that voicing the plural might have on the compounding task.

### 6.4. Manipulation Check

In the second priming control condition, the pictures of monsters that eat multiple items were intended to semantically suggest a monster that eats items in the plural (a monster that is eating *mice/rats*). To test this, 3 participants were shown single items and then asked to go through the 'monster items' (Fig. 5A) and state what it was that the monster was eating. One hundred percent of these answers were in the plural form (e.g., "That is a monster that is eating *teeth*").

### 6.5. Results

All participants completed the test items without much difficulty. As Gordon (1985) notes, participants occasionally changed their responses as if to correct themselves. Gordon's (1985) practice of scoring the second 'corrected' response was adopted. All of the participants successfully produced the singular form of each noun, and 100% of participants successfully produced the correct irregular plural. One participant produced the overregularized *knifes* as the plural for *knife*, but otherwise all other regular plurals were successfully pluralized.

#### 6.5.1. Irregular and regular plurals

The response patterns for irregular nouns and their regular controls (and the corresponding data from Gordon, 1985) are given in Table 3. For the regularly pluralized nouns, participants overwhelmingly showed the predicted pattern of reducing plural to singular forms in compounds (e.g., *rat-eater*), with 86/90 such patterns (Gordon's figure with children was 161/164). As per Gordon (1985), participants were categorized as supporting the pattern predicted by level ordering if all regular plurals were reduced to singulars inside compounds, and chi-square analysis showed that the tendency to reduce regular plurals was significant ($\chi^2(1, N = 15) = 5.40$ $p < .025$).

Importantly however, the adult participants in this study also tended to reduce *irregular* plurals to singulars in compounds as well (if to a slightly lesser extent). 73/90 responses in this category were of the form *mouse-eater*, as compared to only 4/40 in Gordon's study. It is worth noting that the adults tested in Experiment 5 also succeeded in producing irregular plurals in the elicitation phase 100% of the time, whereas the children tested by Gordon produced correct irregular plural forms on fewer than 25% of trials (as compared to regular forms, which were produced correctly on over 99% of trials).

To examine whether the 'optional' inclusion of irregular plurals in compounds was at chance, as might be expected from a level ordering account (Gordon, 1985), chi-square values were calculated

**Table 3**
Mean response patterns for the irregular and regular plural compounds in Experiment 5 (replication condition) compared to Gordon (1985). The average number of responses in each category is given, with the absolute number of responses in parenthesis. The bottom row gives the rate at which singular forms of irregular plurals that were overregularized in the plural elicitation phase were included in compounds.

| Compound form | Regular nouns | | Irregular nouns | |
|---|---|---|---|---|
| | Experiment 5 | Gordon (1985) | Experiment 5 | Gordon (1985) |
| Singular | 95%<br>5.7/6 (86) | 98%<br>4.9/5 (161) | 82%<br>4.9/6 (73) | 2.5%<br>0.12/5 (4) |
| Plural | 5%<br>0.3/6 (4) | 1.8%<br>0.09/5 (3) | 18%<br>1.1/6 (17) | 22%<br>1.09/5 (36) |
| Singular form of overregularized irregular plural | Not applicable | Not applicable | 0%<br>0/6 (0) | 52%<br>2.6/5 (86) |

for participants who included irregular plurals inside compounds less than 50% of the time, and again the results were significant ($\chi^2(1, N = 15) = 5.40$ $p < .025$). (Gordon's children produced the exact opposite effect, showing a significant tendency to include > 50% of irregular plurals in compounds.) Further, the effect of participants *not* putting any irregular plurals in compounds was not significant ($\chi^2(1, N = 15) = .67$, $p > .75$).

There were no differences in the rates between the main replication condition and the first priming control in which the experimenter did not repeat the plural: the rate of regular reduction in compounds in the first priming control was 94% (62/66) as compared to 96% (86/90) in the replication condition. The figures for irregular reduction were 82% (54/66) in the first priming control and 81% (73/90) in the replication. Thus it did not appear that the experimenter's repeating the plural form uttered by the participants had any significant effect on whether they included an irregular plural in a compound or not.

However, the results of the second priming control condition, in which participants saw a picture of a Monster eating multiple items (thereby creating semantic conditions suggesting that a plural compound such as *teeth-eater* might be appropriate), underlined the extent to which the behavior of the participants in this experiment diverged from the predictions made by level ordering. Although they saw pictures of monsters eating *mice, geese, teeth*, etc., participants produced singular forms of the irregular non-head nouns in compounds (e.g., *tooth-eater*) at a rate of 99% (95/96), while singular forms of the regular non-head nouns in the compounds were produced 98% (94/96) of the time.

It is interesting to compare the findings in this condition to what we might have expected to find, given that level ordering claims that a *unique* set of constraints apply to regular plurals. First, as level ordering predicts, the effect of participants producing singular forms for regular plurals in compounds 100% of the time was significant ($\chi^2(1, N = 16) = 9.0$ $p < .005$). However, the second prediction, that the inclusion of irregular plurals is not governed by the same constraints – such that, given the semantics of the task, participants should prefer irregular plurals in compounds – (Gordon, 1985, pp. 77–78), was not confirmed. The effects of participants producing plural forms of the irregular plural nouns in compounds less than 50% of the time was highly significant ($\chi^2(1, N = 16) = 16$ $p < .0001$). Moreover, a highly significant majority of these participants produced *no* irregular plurals in compounds ($\chi^2(1, N = 16) = 12.2$, $p < .0001$).

### 6.5.2. Pluralia tantum nouns

Gordon (1985) argues that level ordering predicts that pluralia tantum nouns—like irregulars—should be optionally allowed inside compounds, while plural forms of their regular counterparts should *not* be (this prediction will be considered further below; in the replication condition, 100% of the regular counterparts were reduced to singulars).

No clear pattern emerged for three out of four of the pluralia tantum nouns: *trousers* was reduced in 53% of responses; *pliers* in 40% of responses and *scissors* in 60% of responses. Only *glasses*, which was reduced in only 7% (1/15) responses revealed any significant pattern ($\chi^2(1, N = 15) = 11.3$, $p < .001$). (The children studied by Gordon (1985), reduced *glasses* in 70% of cases.) The results of the first priming control condition were similar to those in the replication.

In the second priming control condition, however, where participants had to produce a compound to describe a monster eating multiple examples of *pliers, glasses, trousers* and *scissors*, clear patterns of reduction or plural inclusion did emerge: 88% (14/16) of these participants reduced *scissors* to *scissor* in compounds (i.e., *scissor-eater*; $\chi^2(1, N = 16) = 9.0$, $p < .005$), 81% (13/16) reduced *trousers* to *trouser* ($\chi^2(1, N = 16) = 6.25$, $p = .05$), and 75% of participants reduced *pliers* in compounds (preferring *plier-eater*; $\chi^2(1, N = 16) = 4.0$, $p < .05$). On the other hand, consistent with the replication condition (but not with the particular finding of Gordon, 1985, with regards children), 100% of the adults in this condition (barring one who produced the form *spectacles*) preferred *glasses* in its non-reduced form in compounds (i.e., *glasses-eater*; ($\chi^2(1, N = 15) = 15$, $p < .0001$).

It is not clear how level ordering—which predicts that pluralia tantum nouns, like irregular inflections, occur at Level 1—can account for their reduction in any of the conditions described. We return to this point below.

### 6.6. Discussion

The results of this experiment are surprisingly clear-cut in the case of both regular and irregular nouns. Gordon (1985) argued that if level ordering was a learned constraint, one would expect it to apply to both regulars and irregulars (“if [one] has to learn that reduction of regular plurals is required in compounds… a natural induction from [the] evidence would be that irregular plurals are also subject to reduction inside compounds,” pp. 77–78). Given Gordon’s very reasonable hypothesis, and the evidence from competent adult speakers of English, it would seem reasonable to conclude that compounding constraints are learned. The adults in Experiment 5 overwhelmingly produced compounds containing singular (or ‘reduced’) forms of *both* regular and irregular nouns, even in conditions that were specifically designed to elicit ‘erroneous’ plural forms in compounds (Gordon, 1985). Consistent with the findings of Experiments 1 and 2, adult speakers of English formed compounds in ways that appear to respect a systematic convention which they applied irrespective of whether a noun was regular or irregular.

Further, the results of Experiment 5 suggest that other factors have been confounded in previous studies of compounding behavior, and in expositions of level ordering. It is important to disentangle these factors: First, when comparing the conditions that replicated Gordon’s method of having participants voice the plural form prior to compounding to the condition that did not, there were *significant* differences in the number of participants who produced plural irregular compounds (this was notable even though the majority of irregular compounds produced in all of the conditions were singular). In the conditions where a plural was voiced first, 11/26 (42.3%) participants produced at least one irregular plural compound (8/15 in the straight replication) as compared to only 1/16 (6.3%) in the second control condition ($\chi^2(1, N = 42) = 4.667$, $p < .05$; comparing the straight replication only, $\chi^2(1, N = 31) = 4.663$, $p < .05$). The overall rate at which irregular plural compounds were produced was 19% in the replication and 18% in the first control condition, but only *1%* in the second control condition. On the other hand, the overall rate at which *regular* plurals were produced in compounds remained relatively constant (at 2–4%) across all three conditions for regular plurals.

While these data reveal that participants are more likely to say *mice-eater* than *rats-eater* (though note that participants *did* occasionally produce *rats-eater*), the overall pattern of data does not suggest that people apply different constraints when compounding regular and irregular plurals. Rather, it suggests that in elicitation tasks, irregular nouns (which are infrequent as types *and* tokens in English) are more susceptible to induced errors in compounding than more frequent regular forms.

It also appears from these results that Gordon’s (1985) elicitation paradigm *does* induce errors. This is hardly surprising. In this method, participants are first presented with a single item (say a mouse) and asked to name it:

- Participant: MOUSE.

  Participants are then presented with multiple items and asked to name them:
- Participant: MICE.
- Experimenter: “And what do you call a monster that eats MICE?”
- Participant: X-eater.

It is worth recalling that the point of this procedure was not to test whether regulars and irregulars are more or less subject to priming. Rather it was intended to test a hypothesis about learning: that "[if someone] was presented with a context in which there were a number of rats being referred to, one might expect him or her to denote an eater of such animals as a **rats-eater* rather than a *rat-eater*. However, if we assume that level ordering constrains pluralization in this case, then such errors should never be found..." (Gordon, 1985, pp. 77–78).

It follows from this hypothesis that people should produce *mice-eater* in this context. Since they do not, this shows that the constraints on compounding that apply to regular nouns must at the very least be learned for irregulars, since irregulars are not subject to the constraints of level ordering. Indeed, the results of Experiment 5 showed that when people are faced with situations where the semantics indicate that multiple items need to be referred to in a noun-compound (e.g., describing a monster eating mice), they apply the *same* convention to both irregular and regular compounds and, all other things being equal, they produce a singular non-head noun, as in *mouse-eater*. If we apply Gordon's (1985) logic, then there is nothing in adult's compounding behavior to suggest that the constraints that apply to both regular and irregular noun plurals are anything other than learned.

## 7. Experiment 6

If adults seem to learn compounding constraints, what does that suggest about children? An analysis of child directed and child speech (CHILDES) Murphy (2002) could find no examples of irregular plurals included in compounds. This is consistent with the results of Experiment 5, and with the observation that adult native speakers rarely include irregular plurals within compounds (Gordon, 1985). Why do young children include irregular plurals in compounds at such a high rate in Gordon's compound elicitation task (and in subsequent replications; see e.g., Nicoladis, 2003)? To explore the possibility suggested by the results of Experiment 5—that these results may be artifacts of a particular elicitation procedure, rather than a manifestation of a specific constraint on language production—we first decomposed Gordon's original method into its logical constituents. These comprised four basic elements:

(1) checking that children knew the singular form of nouns,
(2) checking to see that they knew the plural forms,
(3) training them in noun-compound production, and
(4) presenting them with a situation that had a pictorial representation of a single eater consuming multiple (plural) examples of items and asking children to produce a compound to describe the eater.

Since the singular naming was a manipulation check to see whether children could identify the items independently—and as Gordon reports, children required a lot of prompting in this task—in the following experiments, this procedure was *separated* from the plural naming and production part of the procedure. In both of the experimental conditions described below, children performed the singular naming condition *prior* to receiving training in compound production (in contrast to the procedure used by Gordon 1985, described above).

This meant that the main body of the experimental procedure comprised two elements: presenting a set of items to elicit a plural, and presenting a monster eating multiple items to elicit a compound in circumstances where the semantics suggested that the eater was an "items-eater" rather than an "item-eater" (as described in relation to Experiment 3, above).

This allowed three versions of Gordon's procedure to be tested. In the first, the plural form was elicited prior to the compound. This was to see whether Gordon's results could be replicated without the experimenter priming the children's responses to the compound task (e.g. to ensure that the result did not rely on the question "And what do you call a monster that eats MICE?").

The second version of the procedure *inverted* the order of plural elicitation and compound production. Children were asked to produce a compound to describe the eater eating multiple items, and then were presented with just the multiple items alone in order to elicit a plural. This second version was intended to ensure that the results of the compounding task were the result of factors inherent in Gordon's original hypothesis, and not from priming resulting from children having just uttered a plural form. If the hypothesis that "[if someone] were presented with a context in which there were a number of rats being referred to, one might expect him or her to denote an eater of such animals as a **rats-eater* rather than a *rat-eater*," did correctly predict children's behavior, then the order of presentation of the plural elicitation and compounding tasks ought to be irrelevant. On the other hand, if differences do result from changes to the *presentation order* of the tasks, these differences would be better explained by task demands, and the influence of those demands on the child's performance (because the semantic/linguistic composition of these tasks is unaffected by our presentation order changes).

Finally, to make sure that any differences in the inclusion of regular and irregular plurals in compounds resulted from level ordering, and not from different voicing constraints affecting regular and irregular compounds, the third version of the procedure employed a different head noun (*box*, as opposed to *eater*), but was otherwise identical to the first procedure. Although Gordon's original experiment utilized four nouns as non-head compounds, only one head noun was used. Thus it is possible that the different rates of plural inclusion might stem from differences in the phonotactic characteristics of regular and irregular nouns in their plural and singular forms. Whereas the final voicing patterns of all of the regular items used in the Gordon experiment were inconsistent between their single and plural forms (e.g. rat ends in a non-sibilant /t/, whereas *rats* ends in a sibilant), the majority of the irregulars were consistent (both foot and feet end in a non-sibilant /t/).

### 7.1. Participants

Participants were 71 children aged between 3 and 8 (average age 5:11) recruited from a school and a kindergarten in Edinburgh, Scotland.

### 7.2. Materials

Materials were the same as those used in Experiment 4, except in condition 3, where *box* was used as a head noun instead of *eater*.

### 7.3. Procedure

Trials were both transcribed as they occurred, and videotaped to aid with resolving any ambiguities. Participants were tested individually, and told that they were helping to teach a particularly slow monster to learn the names of his friends. Each participant was then shown a picture of one of the training items, and asked, "What is this? (Answer: X). This was repeated until the child had named all the items. As in Gordon (1985), feedback was given when children failed to spontaneously name items.

Participants were then introduced to a monster that had a terrible memory. They were told they were going to be introduced to a number of his friends, and they were asked to assist the monster by telling him the names of his friends.

After training on mass nouns as per Gordon (1985), participants were randomly assigned to one of three versions of the main experimental procedure.

#### 7.3.1. Condition 1

In this version of the procedure, plural forms were elicited prior to the formation of compounds. Participants were first shown a picture of multiple items to elicit the plural, and asked, "what are these?" Having named the plural, children were then shown a monster eating the plural items and simply asked, "So what is this monster called?" These superficial changes to the procedure were made to enable the task to be run without the experimenter having to ask, "What would you call someone who eats X," and thereby priming the child with the plural form of X.

*7.3.2. Condition 2*

This trial procedure was as per Condition 1, except that the order of compounding and plural production was reversed.

*7.3.3. Condition 3*

This trial procedure was as per Condition 1, except that a different head noun – *box* – was employed instead of *eater*. After they had named the plural picture, the children were shown a box into which the plural items were tidied after play and asked, "So what is this box called?" in order to elicit the compound form "X-box."

*7.4. Results*

*7.4.1. Condition 1*

*7.4.1.1. Regular nouns.* In Gordon (1985), participants consistently reduced regular plurals to single forms inside compounds (161/164 of responses were of this form). This was the case in the present study as well, with 135/142 of regular plurals reduced to singles when compounded in Condition 1.

*7.4.1.2. Correct irregular nouns.* In Condition 1, 58/86 of the responses giving the correct irregular plural carried this on into the compound word to form, producing *mice-eater*. Gordon (1985) found that 36/40 of correct irregular responses were of this form. As in Gordon's study, a chi-square calculated for subjects showing greater than 50% inclusion of the irregular plural within compounds showed this trend to be significant ($\chi^2(1, N = 24) = 8.2$, $p < .005$).

*7.4.2. Condition 2*

In this task, reversing the order of compound production and plural elicitation produced two significant alterations in the pattern of data observed in Condition 1. First, it was found that when participants produced the compound first, they tended to produce the singular form of irregular plurals in their compounds (i.e., saying *mouse-eater*, not *mice-eater*; see Fig. 6). A chi square test calculated for subjects—which showed greater than 50% inclusion of the *singular form* of the irregular noun within compounds—showed this trend to be significant ($\chi^2(1, N = 26) = 3.85$, $p < .05$).

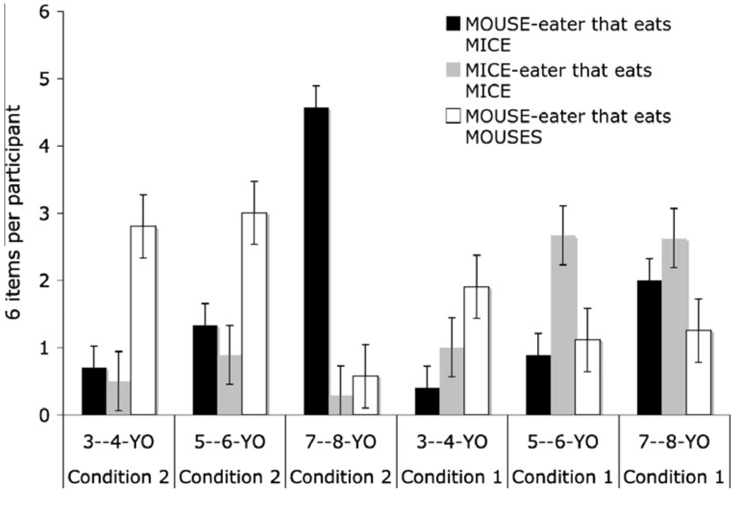


**Fig. 6.** Irregular plural compounding results from Conditions 1 and 2 in Experiment 6. In Condition 2 children formed a compound to describe a monster eating multiple items (e.g. *mice*) *before* identifying the items (e.g. stating the plural form), whereas in Condition 1 items were named prior to compound elicitation. Error bars are SEM.

Second, an increase in overregularization of irregular nouns was noted in Condition 2 as compared to Condition 1. The mean plural overregularization rate in Condition 1 was 24%; this figure rose to 48% in Condition 2, while the percentage of children producing 2 or more overregularizations rose from 30% to 62% significant ($\chi^2(1, N = 53) = 5.44$, $p < .05$).

#### *7.4.3. Condition 3*

Finally, we examined the number of plural compounds produced by participants in Conditions 1 and 3 to see whether the change in the head noun (from *eater* to *box*) had affected the number of regular and irregular plurals being included in compounds in the non-head noun position. This comparison showed that in Condition 1, 5% of the regular plurals were included in compounds, whereas in Condition 3, this figure rose to 21%. While only 15% of participants in Condition 1 included a plural regular noun in their compounds, this figure rose to 42% for the participants in Condition 3 ($\chi^2(1, N = 45) = 4.01$, $p < .05$). Although the number of regular plurals in compounds increased for *eater* as compared to *box*, the figures for irregular plural inclusion in compounds was less affected by this change. Inclusion rates for Conditions 1 and 3 were 34% and 32% respectively; 63% of participants in Condition 1 and 73% of participants in condition 3 included irregular plurals in compounds ($\chi^2(1, N = 45) = 4.01$, $p > .28$.

### *7.5. Discussion*

The main effects and trends reported in Gordon (1985) were replicated in Condition 1 of Experiment 6. However, the key finding in Gordon's original study—children's inclusion of irregular plurals in compounds (2a)—disappeared when children were asked to form compounds *before* plural elicitation. Moreover, this change occurred even though the semantics depicted in the stimuli (which were the driving force in Gordon's hypothesis) were unchanged.

Gordon's (1985) results were not supposed to depend simply on the order in which the experimental manipulations were conducted. Rather, singular, plural and compound forms were all elicited in order to verify that children knew both the singular and plural forms (plural elicitation was included as a manipulation check, not a manipulation in itself; Gordon, 1985). It was assumed that children included irregular plurals in compounds because the stimuli they were presented with were plural, and because there was no constraint against irregular plural inclusion in compounds. Consistent with the results of Experiment 5, it appears that children include irregular plurals in compounds (2a) because priming induces production errors, and that forms such as *mice-eater* do not reflect children's underlying understanding about the conventions of English when it comes to noun compounding; when not primed, children behave much like adults, and describe a monster that eats *mice* as a *mouse eater*.

This priming explanation gains further support from the increase in *over-regularization* observed in Condition 2 as compared to Condition 1. This finding suggests that the priming effect on children's irregular plural production is bi-directional: priming "mice" in Condition 1 caused children to say "mice-eater," despite the fact that all of the evidence collected here (and available to children) points to a rather general prohibition on plurals in compounds. On the other hand, saying "mouse-eater" immediately before producing a plural form caused children to say "mouses," another form that is not warranted by either the input or later production patterns.[4]

[4] Gordon (1985) interprets the effect of the bias built into his design in a very different way:

> "[that] the irregular plural appeared to be preferential inside the compound in the context of the present task, [was] presumably due to the biases set up in the design. That is, children would probably not be so biased to produce a form such as mice-eater in an ordinary everyday context. This would be especially unlikely if, like adults, the reduced singular form is preferred. **But the point is that such biasing was only effective in the allowable cases (i.e., irregulars and pluralia tantum)**... where the child is required to learn the appropriate restrictions from input, errors occur; but where the restrictions follow deductively from the structural constraints, then one finds no errors" (p. 86, our emphasis).

The results of our experiments do not license these conclusions. For one, children can be biased to produce *regular* plurals in compounds (Experiment 6, Condition 3). Gordon's claim also ignores the huge disparity in children's exposure to irregular plurals (in Gordon, 1985, children failed to produce an irregular plural *at all* in more than 60% of trials). Learning offers a far more plausible (and economical) account of why irregulars may be more susceptible to bias.

## 7.6. The great spectacle-case mystery—what happened to the /s/?

That the strong predictions of level ordering are not borne out by the data is apparent. In light of this, it is worth considering some other shortcomings of level ordering theory. First, with regards pluralia tantum nouns, it is not clear at all that level ordering *can* account for the pattern of data found in Experiment 5 (and in Gordon, 1985), in which pluralia tantum nouns were reduced in compounding.[5] Recall that according to level ordering theory, *scissors* and *trousers* are stored at Level 1, and enter in compounds at Level 2. Thus level ordering can give a good account of how *glasses-eater* and *pliers-eater* are formed (and *pants-eater* in Gordon, 1985). But what of *trouser-eater* and *scissor-eater*? The defining characteristic of pluralia tantum nouns like *scissors* and *trousers* is that they only have a plural form; no singular form is stored in the lexicon. That being the case, level ordering ought to predict that pluralia tantum nouns should *only* occur inside compounds in their plural form: The mechanism proposed by level ordering is not one in which regular nouns can be *reduced* (singularized); rather, level ordering proposes that in regular compound production, regular non-head nouns can be *extended* (pluralized). Thus, nothing in the level ordering process suggests that pluralia tantum nouns can be reduced, let alone proposes a mechanism by which this might be achieved. Further, if level ordering *were* to advance such a mechanism, it would introduce the problem of explaining whether the putative reduction mechanism *or* level ordering was responsible for regular noun reduction in compounding, and *why*.

Second, level ordering is somewhat selective in its treatment of irregular plural nouns: in particular, it fails to offer a convincing story regarding voicing change plurals such as *leaf-leaves*, *thief-thieves*, *loaf-loaves, life-lives*; *hoof-hooves*, etc. These nouns are irregular in that they are not marked for plurality by simply adding a sibilant. Rather, inflection modifies the singular 'stem' and then adds a sibilant. Further, like many irregular verbs, while they tend to cluster in families (Bybee & Slobin, 1982; Pinker, 1999), they all have regular soundalikes. The plural of *chief* is not *chieves* but the regular *chiefs*, likewise *brief-briefs*, *reef-reefs*, *fief-fiefs, motif-motifs*, etc.; the plural of *proof* is the regular *proofs* and not *prooves*, likewise*, spoof-spoofs, goof-goofs, oaf-oafs, fife-fifes*, etc. Voicing change irregulars can be over-regularized by children (Graves & Koziol, 1971) and regularized for particular semantic and pragmatic purposes (Pinker, 1999; e.g., in baseball it is the *Toronto Maple Leafs*). Thus it appears that voicing change irregulars also form part of Level 1 (Kiparsky, 1982, 1983), and hence in theory ought to 'optionally' enter into compounds, even though in practice it appears that they behave much like regulars in compounding (Gordon, 1985).

In short, it appears that not only does level ordering fail to predict the behavior we observed in compound production and comprehension with regards stem-change irregulars (e.g., *mouse-mice*), it further appears to lack a consistent explanation of the way in which English speakers compound pluralia tantum nouns and voicing change plurals (see also Haskell et al., 2003).

## 7.7. Of rats-eaters, mice-eaters and loaves-eaters

One final phenomenon associated with plurals in compounds that has been claimed in support of level ordering is the finding that regular plurals 'sound worse' than irregular plurals in compounds (e.g. Gordon, 1985; Pinker, 1991, 1999, 2001; Marcus, 1999). (Gordon 1985; see also Pinker, 1999) argues that intuitive differences in acceptability of *mice-infested* vs. *rats-infested* are explained by level ordering, since the former includes a Level 1 plural, whereas the latter includes a Level 3 plural. Alegre and Gordon describe a study by Senghas, Kim, Pinker and Collins (1991, cited in Alegre & Gordon, 1996) which showed that when adults had to judge the naturalness of novel compounds, those including irregular plurals were rated far more acceptable than those including regular ones. Indeed, Alegre and Gordon take these judgments as evidence that competent speakers of English possess qualitatively different constraints for compounding irregular and regular plurals. However, the evidence from

[5] A simple Google search reveals that certain pluralia tantum nouns are regularly reduced in compounds: e.g., *trousers* in "trouser(s) press," *scissors* in "scissor(s) lift" and *pliers* in "plier(s) stapler" are all reduced more than 90% of the time. On the other hand, *scissors* is only reduced 25% of the time in "scissor(s) kick" and *pliers* only about 30% of the time in "plier(s) clamp." Whether or not reduction occurs appears to hinge on the conventions of the particular expression, which suggests that these conventions are learned.

the experiments described so far suggests that competent speakers of English *do not* possess qualitatively different constraints for compounding irregular and regular plurals, and that even children will treat regular and irregular plurals in exactly the same way unless they are subject to priming. Thus, given that speakers of English appear to apply similar constraints when compounding both irregular and regular nouns, and given that adult performance is not at variance with childhood performance, it seems reasonable to consider whether the factors governing acceptability show evidence of having been *learned*.

## 8. Experiment 7

Experiment 7 examined adult participants' naturalness ratings for novel English compounds to determine whether a distinction between irregular and regular plural non-head nouns in compounds was the best way to explain the variance in these ratings. The hypothesis that participants might distinguish between regular and irregular nouns in their naturalness ratings was tested against the alternative hypothesis that participants might simply be more sensitive to sibilance (the sound of the *–s* suffix) in plural noun-head noun compounds. This alternative was suggested by the fact that the overwhelming majority of noun compounds that people encounter are regular (Gordon, 1985) and the overwhelming majority of these will have a singular non-head noun; the idea being that the question "*how natural does this sound?*" is equivalent to "*how well does this fit with what you have heard before?*"

The idea that sensitivity to siblance might contribute to 'naturalness' ratings was also explored in Haskell et al. (2003, Study 4). Haskell et al. (2003) propose that compounding is governed by separate—but interacting—semantic and phonological constraints, and that the acceptability of a given compound is a function of these constraints, which are probabilistic in nature:

> "One [factor] is whether a potential modifier is plural in meaning. This will be termed the semantic constraint. *Rats* and *mice* pattern together on this dimension and differ from *rat* and *mouse*. The second factor is whether a potential modifier has the phonological structure typical of a regular plural, hereafter the phonological constraint. On this dimension, *rat, mouse*, and *mice* pattern together and differ from *rats*. Only the latter contains the /s/ that is a proper realization of the plural inflection" (p. 131).

While the approach of Haskell et al. (2003) is broadly comparable to our own, we would suggest that a main function of phonology is to discriminate between semantic alternatives (Ramscar & Dye, 2009; Ramscar & Yarlett., 2007; Ramscar et al., 2010). For this reason, we expect that phonological forms that are more discriminable in any given context will be more useful (i.e., informative) in this regard. Given that the presence or absence of a final sibilant is used to discriminate between the plural and singular forms of most nouns, and that there are conventions that apply to plurals in compounds, we would expect that when the phonological forms of singular nouns are easily distinguished from forms with final sibilants, this will be more informative—both about plurality, and about the conventional status of a given compound—than when they aren't (i.e., *rats* may be a more informative plural form than both *mice* and *houses*).

Experiment 7 thus examined whether participants' ratings were reflective of the statistical properties of the linguistic input (i.e., consistent with learning), and in keeping with this, whether participants were sensitive to medial sibilance, and its impact on informativity.

### *8.1. Participants*

Thirty-four native-English speaking students at Stanford University participated in this study for credit. Two native-English speaking students at the University of Edinburgh volunteered to act as raters in the manipulation check.

### *8.2. Materials*

This experiment examined 30 noun-agentive compounds: 10 irregular noun compounds (spaceman-observer, foot-lover, fireman-list, tooth-saver, woman-chaser, foot-inspector, goose-keeper, louse-checker, tooth-remover and mouse-catcher); 10 regular noun compounds (moon-observer, bun-lover, award-lister, pet-saver, criminal-chaser, soul-inspector, bull-keeper, page-checker, shrub-remover and shrimp catcher); and 10 irregularly suffixed/voicing change plurals (spectrum-observer, loaf-lover, curriculum-list, life-saver, thief-chaser, self-inspector, ox-keeper, index-checker, cactus-remover and octopus-catcher). Five of the latter group were nouns that mark pluralization both with a voicing change and the –*s* suffix, while the other five marked pluralization with a voicing change and/or another suffix. Two versions of each compound were produced, one with a singular form of the non-head noun, and one with a plural form of the non-head noun.

Each compound head noun was matched by item in each of the three groups, i.e., the head noun *lover* was matched with the regular *bun*, the stem-change irregular *foot*, and the voicing change irregular *loaf*. The compounds were interspersed with 30 fillers and judged for their plausibility by two independent raters. All of the compounds were rated plausible 100% of the time, and inter-rater agreement was 100%.

The compounds were presented in questionnaires, with the order of presentation of individual items randomized.

### 8.3. Procedure

Participants were instructed to ‘‘study the following expressions. For each one, we'd like you to tell us whether it sounds to you like a natural English phrase. Indicate your judgment by marking the scale provided.'' Participants gave their responses on a seven point scale where 1 = ‘‘very natural,'' and 7 = ‘‘not natural at all.'' The questionnaires were completed by participants at home in their own time.

### 8.4. Results

The results of this experiment did not support a distinction between the acceptability of regular and irregular plural nouns in compounds (see Table 4); rather, they supported a distinction in acceptability based on whether the non-head noun plural ended in a sibilant in compounds. The irregular plurals, which level ordering predicts should be acceptable in compounds, showed mixed results, depending upon their type. As an overall category, participants' ratings of stem-change irregular plurals (*foot-feet*) showed no apparent sensitivity to non-head noun plurality (plural compounds, $M = 3.47$, SEM = .13; singular compounds $M = 3.44$; SEM = .1; $t(33) = .14$, $p > .8$). On the other hand, participants ratings of regular plurals (*pet-pets*) did show sensitivity to non-head noun plurality (participants preferred singular compounds, $M = 3.15$, SEM = .11 to plural compounds $M = 4.57$, SEM = .14; $t(33) = 8.27$, $p < .0001$). However, participants ratings of the irregularly suffixed/voicing change plurals (*curriculum-curricula; knife-knives*) *also* showed sensitivity to non-head noun plurality (participants preferred singular compounds, $M = 3.64$, SEM = .1 to plural compounds $M = 4.46$, SEM = .09; $t(33) = 4.307$, $p < .0001$).

To see whether grammatical type (regular vs. irregular) or plural sibilance best explained our participants' ratings of the acceptability of plural forms in compounds, we conducted a series of regression analyses on their ratings data. In keeping with our suggestion that the question ‘‘*how natural does this sound?*'' may be largely equivalent to ‘‘*how well does this fit with what you have heard before?*'' we considered four variables in these regressions:

1. *The summed frequency of the plural and regular frequencies of each noun* (e.g., *thief* + *thieves*) which provides an estimate of how often participants have encountered each noun. (Frequencies were taken from COCA; Davies, 2009).
2. *The proportion of plural to single tokens of each noun*, which provides an estimate of how often participants have encountered the singular form of each noun relative to its singular form.
3. *Regularity*: i.e., whether nouns were regular or irregular plurals (see below).
4. *Sibilance*: i.e., the presence (or absence) of a medial sibilant when the plural form is in a compound

Because, as we noted above, there is some debate over whether irregularly suffixed and voicing change plurals such as *life-lives* and *ox-oxen* should be treated as regular or irregular, we tested two kinds of ‘‘regularity'' in these models: in the first, all irregularly suffixed and voicing change plurals

**Table 4**
Subjects ratings and frequency counts for the items in Experiment 7. (Frequencies are the number of tokens of each plural (e.g. *thieves*) and single type (*thief*) in the Corpus of Contemporary English (Davies, 2009).

| Singular | Compound | Mean rating | COCA frequency | Plural | Compound | Mean rating | COCA frequency |
|---|---|---|---|---|---|---|---|
| Award | List | 2.38 | 15632.00 | Awards | List | 2.65 | 10579.00 |
| Moon | Observer | 3.03 | 23199.00 | Moons | Observer | 4.94 | 1625.00 |
| Page | Checker | 3.18 | 40484.00 | Pages | Checker | 5.44 | 19951.00 |
| Bull | Keeper | 2.68 | 7660.00 | Bulls | Keeper | 4.45 | 5073.00 |
| Pet | Saver | 4.06 | 7069.00 | Pets | Saver | 4.82 | 3537.00 |
| Shrimp | Catcher | 2.32 | 6861.00 | Shrimps | Catcher | 5.62 | 155.00 |
| Shrub | Remover | 2.29 | 841.00 | Shrubs | Remover | 4.00 | 2025.00 |
| Soul | Inspector | 3.94 | 19162.00 | Souls | Inspector | 5.15 | 4755.00 |
| Bun | Lover | 3.59 | 1198.00 | Buns | Lover | 3.74 | 802.00 |
| Criminal | Chaser | 3.97 | 23192.00 | Criminals | Chaser | 4.85 | 5688.00 |
| Goose | Keeper | 3.09 | 2716.00 | Geese | Keeper | 3.35 | 1789.00 |
| Louse | Checker | 2.44 | 133.00 | Lice | Checker | 3.44 | 606.00 |
| Mouse | Catcher | 2.00 | 6818.00 | Mice | Catcher | 3.18 | 3717.00 |
| Spaceman | Observer | 5.24 | 121.00 | Spacemen | Observer | 3.82 | 58.00 |
| Tooth | Saver | 4.53 | 3126.00 | Teeth | Saver | 3.68 | 18297.00 |
| Tooth | Remover | 3.53 | 3126.00 | Teeth | Remover | 5.09 | 18297.00 |
| Woman | Chaser | 1.74 | 150887.00 | Women | Chaser | 3.00 | 220351.00 |
| Fireman | List | 5.09 | 941.00 | Firemen | List | 3.35 | 940.00 |
| Foot | Inspector | 3.35 | 29569.00 | Feet | Inspector | 2.41 | 82969.00 |
| Foot | Lover | 3.44 | 29569.00 | Feet | Lover | 3.38 | 82969.00 |
| Index | Checker | 3.76 | 11592.00 | Indices | Checker | 5.35 | 1276.00 |
| Life | Saver | 1.06 | 293108.00 | Lives | Saver | 5.38 | 73144.00 |
| Loaf | Lover | 4.94 | 2272.00 | Loaves | Lover | 4.50 | 863.00 |
| Octopus | Catcher | 2.97 | 635.00 | Octopi | Catcher | 3.59 | 36.00 |
| Ox | Keeper | 3.56 | 776.00 | Oxen | Keeper | 2.29 | 463.00 |
| Self | Inspector | 3.06 | 16895.00 | Selves | Inspector | 5.79 | 1530.00 |
| Spectrum | Observer | 4.47 | 6644.00 | Spectra | Observer | 4.56 | 961.00 |
| Thief | Chaser | 3.76 | 2720.00 | Thieves | Chaser | 4.85 | 2407.00 |
| Cactus | Remover | 4.44 | 1444.00 | Cacti | Remover | 3.44 | 387.00 |
| Curriculum | List | 4.44 | 17085.00 | Curricula | List | 4.82 | 2652.00 |

were treated as irregular (Table 5, models 1 and 3), whereas in the second, only irregularly suffixed plurals such as *ox-oxen* were treated as irregular, while voicing change plurals such as *life-lives* were classed as regular (Table 5, models 2 and 4).

As Table 5 shows, even among this second set of models – which essentially give the level ordering hypothesis free reign to pick the items regularity applies to – grammatical status failed to account for the variance in participant ratings. Instead, variance was best explained by frequency and the presence (or absence) of a medial sibilance in noun compounds (irrespective of the "grammatical origins" of that medial sibilance).

### 8.5. Discussion

The results of Experiment 7 suggest that people are not sensitive to the kind of simple distinction between regular and irregular non-head nouns in compounds proposed by level ordering (2a). Instead, in judging naturalness, participants' appeared to be sensitive to whether plural forms of the non-head noun ended in a sibilant or not.

What might the cause of this be? If English speakers' and listeners are sensitive to the distribution of information in English (Altmann & Mirković, 2009), then one might expect them to acquire some sensitivity to the way nouns occur in compounds. To examine whether the distributional properties of English might help account for our participant data, we examined two apparent properties of English: (1) that the overwhelming majority of noun-compounds contain nouns with a singular non-head noun,[6] and (2) that the overwhelming majority of nouns mark plurality by adding an /s/ suffix.

[6] In English, a convention to avoid putting regular plurals in compounds makes good sense for the purposes of informativity, since it helps disambiguate compound nouns from possessives (Murphy & Hayes, 2010).

**Table 5**
The variance in the difference between plural and singular ratings as accounted for by: summed frequency of the plural and regular frequencies of each noun; the proportion of plural-to-single tokens of each noun; regularity; and the presence of a final sibilant when a plural form is in a compound.

| Model | $R^2$ | Predictors | Standard error | Standardized beta | *t* | *p*< |
|---|---|---|---|---|---|---|
| 1 | 0.56 | Summed Frequency | 0.00 | −0.44 | −3.36 | .00 |
| | | Plural Sibilance | 0.45 | 0.62 | 3.91 | .00 |
| | | Regular or Irregular | 0.47 | 0.06 | 0.37 | .71 |
| 2 | 0.57 | Summed Frequency | 0.00 | −0.43 | −3.26 | .00 |
| | | Plural Sibilance | 0.64 | 0.50 | 2.22 | .04 |
| | | "Modified regularity" | 0.62 | 0.19 | 0.85 | .41 |
| 3 | 0.56 | Summed Frequency | 0.00 | −0.45 | −3.30 | .00 |
| | | Frequency Proportion | 0.12 | −0.04 | −0.29 | .78 |
| | | Plural Sibilance | 1.80 | 0.63 | 3.81 | .00 |
| | | Regular or Irregular | 0.19 | 0.06 | 0.39 | .70 |
| 4 | 0.57 | Summed Frequency | 0.00 | −0.42 | −3.20 | .00 |
| | | Frequency Proportion | 0.12 | −0.07 | −0.50 | .62 |
| | | Plural Sibilance | 0.47 | 0.50 | 2.17 | .04 |
| | | "Modified regularity" | 0.49 | 0.22 | 0.94 | .36 |

Gordon (1985) found evidence for the first of these distributional properties in an examination of the stem-change irregulars *mouse, man, tooth, foot* and *goose* in Kucera and Francis (1967): stem-change irregulars occurred in noun compounds 153 times in the non-head (left) position in their singular forms but only 3 times in their plural form, suggesting a plural to singular ratio inside compounds of ~2:100. In a similar vein, Marcus (1995) has claimed that for nouns that mark plurality with a +/s/ suffix, there should be on the order of 40–50,000 noun compounds per million words where the non-head is singular, compared to only around 500 compounds per million where the non-head is plural (a ratio of ~1:100). While these estimates appear to have had considerable influence in shaping people's understanding of the way nouns and noun compounds are distributed, the usefulness of them is undermined by two important factors: first, both are made on the basis of very small language samples, which limits the confidence one can place in their accuracy (Tomasello & Stahl, 2004), and second, both appear to assume that events in language are normally distributed, which they most certainly are not (Zipf, 1949, 1935).

## 9. Comparing the distribution of regular and irregular nouns in compounds

To get a more accurate idea of the rates of inclusion of regular and irregular plurals in compounds, we compared the distribution of five irregular nouns (*mouse, man, tooth, foot* and *goose*) to that of five regular nouns matched for frequency and approximate semantics (*thing, eye, rabbit, nail*, and *swan*) in the 400+ million word COCA corpus (Davies, 2009). Table 6 shows the frequency of occurrence of these nouns, along with the percentage of bigrams in which the word following each target noun is a word that occurs as a noun in parts of speech (POS). This percentage estimate provides a measure of how often a particular noun is likely to be used as a non-head noun in noun-noun compounds. As can be seen, on average, the singular forms of our target nouns are followed by noun POS around 20% of the time, whereas the plural forms of nouns are only followed by noun POS around 1% of the time.[7] (This is consistent both with level ordering and with our suggestion that people learn a general constraint against including plurals in compounds).

In addition, an examination of these bigram distributions reveals a strong negative correlation between the frequency of singular forms and the likelihood of them being followed by a noun POS for both regulars ($r = -.73$) and irregulars ($r = -.88$). This suggests that the likelihood that a given noun

[7] Because many words, e.g. *dress*, can occur in multiple POS (*dress* is commonly used as both a noun and a verb) we should emphasize that this is a not a measure of the actual rate at which nouns occur in compounds; rather, it is a method which allows us to estimate the probability that a noun will occur in a compound given its frequency of occurrence. This provides estimates of the rates at which nouns occur in compounds, which can then be compared.

**Table 6**
The COCA frequencies for the singular and plural forms of five irregular nouns (*mouse, man, tooth, foot* and *goose*), and five regular matches (*thing, eye, rabbit, nail*, and *swan*), along with an estimate of the percentage of tokens of that noun-type that are followed by another noun POS. The negative correlation between the frequency of singular forms and the proportion of noun POS is easily observed.

| Singular | COCA frequency | % tokens followed by noun POS | Plural | COCA frequency | % tokens followed by noun POS |
|---|---|---|---|---|---|
| Man | 281,746 | 2.53 | Men | 163,663 | 0.78 |
| Foot | 29,569 | 13.98 | Feet | 82,969 | 1.42 |
| Mouse | 6818 | 22.31 | Mice | 3717 | 1.86 |
| Tooth | 3126 | 23.34 | Teeth | 18,297 | 1.02 |
| Goose | 2716 | 44.15 | Geese | 1789 | 1.23 |
| Thing | 191,681 | 2.07 | Things | 231,275 | 1.37 |
| Eye | 44,224 | 16.27 | Eyes | 133,291 | 0.79 |
| Rabbit | 4387 | 23.80 | Rabbits | 1958 | 0.61 |
| Nail | 4694 | 34.68 | Nails | 5126 | 1.66 |
| Swan | 1701 | 40.04 | Swans | 524 | 1.15 |

will be encountered in a compound *increases* as the overall frequency of the noun decreases. Why is this relevant? In short: because irregular nouns are so highly frequent compared to regulars. This might not be immediately obvious, since the combined token frequency of all of the types of +/s/ nouns (and in particular, regular +/s/ nouns) dwarfs that of the stem-change irregular types. However, *individually*, the token frequency of each individual irregular noun type is high, suggesting that on average, irregular nouns will occur in proportionally fewer compounds than regular nouns. This means that regular nouns – which mark plurality with a final sibilant – will not only be far more in evidence in the input overall, but they will also be more likely to occur in compounds. As a result, the practice English speakers get on the regular compound pattern will be far more frequent, and far more helpful than the practice they get on the irregulars (Ramscar et al., 2010). The likely consequence of all this practice is that speakers will become particularly sensitive to the presence (or absence) of final sibilants in the voicing of nouns.

## 10. Comparing the distribution of two types of regular nouns in compounds

To further explore this idea, we examined the distribution of the singular and plural forms of 18 regular nouns that ended with sibilants in their singular forms (sibiliant 'S' nouns), and two sets of 18 regular nouns that did *not* end with sibilants (non-sibilant 'NS' nouns), which were matched to the original 18 for semantics and frequency. As can be seen from Tables 7 and 8, the distribution of 'S' nouns is different to that of 'NS' nouns. Compared to 'NS' nouns, 'S' nouns are *less* likely to be followed by nouns in their singular forms, but *more* likely to be followed by nouns in their plural forms. While the plural and singular forms of the 'NS' nouns differ markedly – reflecting the conventional embargo on plurals in compounds – paired *t*-tests between the matched items revealed that that difference is significantly less clear cut for 'S' nouns: $M = 10.03\%$, SEM = 8.6; set 1 'NS' nouns: $M = 14.89\%$, SEM = 12.2, $t(17) = 2.2$, $p < .05$; set 2 'NS' nouns: $M = 15.95\%$, SEM = 13.06, $t(17) = 12.35$, $p < .05$. These findings suggest that 'S' regulars do not behave in the same way as 'NS' regulars in compounding. Moreover, given that all our 'S' and 'NS' nouns are regular, their different distributions cannot be explained in terms of the distinctions that level ordering makes on the basis of grammatical status. However, the finding that the compounding conventions apply less clearly to 'S' nouns *is* consistent with our suggestion that these conventions result in sensitivity to the presence of sibilants in compounds.

The rates at which 'S' and 'NS' nouns pluralize further underlines the very different distributions of these two different kinds of "regular" nouns. The rate at which the 'S' nouns occurred in their plural (as opposed to singular) forms was just 19.7% (SEM = 4.1), whereas for the matched 'NS' nouns, this rate was set 1, $M = 65.3$ % (SEM = 10.7) ($t(17) = 4.852$, $p < .0001$), set 2, $M = 61\%$ (SEM = 13) ($t(17) = 3.723$, $p < .001$). Since *treatises, tracts* and *monographs* are all regular plurals, there is no grammatical reason why the first should pluralize less than the other two; nor does any obvious semantic

**Table 7**
COCA frequencies and distributional data for the 18 regular nouns ending in a sibilant. From left to right: (a) the singular frequency, (b) the percentage of singular tokens followed by a noun POS, (c) the plural frequency, (d) the percentage of plural tokens followed directly by a noun POS, (e) the percentage difference between b and d, (f) the percentage difference between a and c. Note that the nouns selected occur predominantly as noun POS (unlike, e.g., *dress* which has a high occurrence rate in other POS, as in "He likes to *dress* well."

| Regular plurals ending in sibilants | Singular frequency | % of noun POS bigrams | Plural Frequency | % of noun POS bigrams | % difference in bigram distribution | % plural relative to singular |
|---|---|---|---|---|---|---|
| Appendix | 1685.00 | 10.15 | 124.00 | 0.81 | 9.34 | 7.36 |
| Ass | 6652.00 | 4.81 | 522.00 | 1.34 | 3.47 | 7.85 |
| Blouse | 2845.00 | 1.55 | 584.00 | 0.68 | 0.86 | 20.53 |
| Boss | 13265.00 | 8.80 | 2511.00 | 1.23 | 7.56 | 18.93 |
| Box | 42614.00 | 13.60 | 10794.00 | 0.46 | 13.14 | 25.33 |
| Buttress | 343.00 | 13.70 | 147.00 | 2.04 | 11.66 | 42.86 |
| Compass | 2244.00 | 17.07 | 141.00 | 0.71 | 16.36 | 6.28 |
| Cutlass | 263.00 | 19.77 | 37.00 | 2.70 | 17.07 | 14.07 |
| Fox | 19026.00 | 22.69 | 945.00 | 0.63 | 22.06 | 4.97 |
| Index | 11592.00 | 24.93 | 1276.00 | 0.71 | 24.23 | 11.01 |
| Noose | 659.00 | 2.73 | 107.00 | 0.09 | 2.64 | 16.24 |
| Spouse | 4741.00 | 6.03 | 3052.00 | 0.95 | 5.08 | 64.37 |
| Thesis | 3584.00 | 8.31 | 428.00 | 6.78 | 1.54 | 11.94 |
| Walrus | 257.00 | 29.18 | 106.00 | 0.94 | 28.24 | 41.25 |
| Purse | 5851.00 | 8.68 | 76.00 | 0.13 | 8.55 | 1.30 |
| Thesaurus | 280.00 | 7.50 | 21.00 | 4.76 | 2.74 | 7.50 |
| Briefcase | 2462.00 | 2.48 | 241.00 | 0.04 | 2.44 | 9.79 |
| Treatise | 906.00 | 4.08 | 383.00 | 0.52 | 3.56 | 42.27 |
| Averages | 6626.06 | 11.45 | 1194.17 | 1.42 | 10.03 | 19.66 |

consideration explain this difference (likewise for *walruses, sea lions* and *otters*, or *foxes, wolves* and *coyotes*). A more likely explanation is that because the singular forms of 'S' nouns are more confusable with plurals, this has a knock-on effect on pluralization (see also Ramscar et al., 2010).

To underline this point, compare the 'S' nouns *rhinoceros* and *hippopotamus* to their semantically indistinguishable 'NS' forms *rhino* and *hippo*. The pluralization rate of *rhinoceros* and *hippopotamus* is 20%, but that rate almost triples to 55% for *rhino* and *hippo*. If we then take a measure of how likely the plural and singular forms of these words are to be followed by a noun POS, the difference in this likelihood between the plural and singular forms of *rhino* and *hippo* is 23.3%. However, this difference drops to 16.8% for *rhinoceros* and *hippopotamus*. In other words, the conventions that govern the use of the various forms of *rhino* and *hippo* are applied less stringently to *rhinoceros* and *hippopotamus*. Given that *rhinoceros, hippopotamus, rhino* and *hippo* are all 'regular' nouns, this difference clearly cannot be explained in terms of a 'grammatical constraint' applying to all regular nouns.

The results of this corpus analysis suggest that the nature of the input makes conventions relating to sibilant-marked plurals *considerably* more easy to learn. The less noisy input for regular plurals makes them less prone to induced errors (Experiment 5), and more amenable to introspective consideration (Experiment 7), as comparted to conventions relating to irregular plurals that are not marked with a final sibilant *and* to regular plurals that end in a sibilant in their singular form. In light of this, it is hardly surprising that our participants' intuitions regarding stem-change irregulars (*foot-feet*) differed from their intuitions about those nouns that mark plurality with an +/s/ suffix, even though our data suggest that the same overall convention applies to all nouns.

## 11. Irregular plurals in compounds: few and far between

With this last point in mind, we return to a claim described at the start of this paper, that:

> "An apartment infested with mice is *mice-infested*, but an apartment infested with rats is not *rats-infested*; it is *rat-infested*, even though by definition, a single rat is not an infestation." (Pinker, 1999, p. 178; see also Gordon, 1985)

**Table 8**
COCA frequencies and corresponding distributional data for the two sets of semantic/frequency match nouns.

| | Singular frequency | % of noun POS bigrams | Plural frequency | % of noun POS bigrams | % difference in bigram distribution | % plural relative to singular |
|---|---|---|---|---|---|---|
| *Semantic/frequency matches – set 1* | | | | | | |
| Supplement | 3871.00 | 17.67 | 2978.00 | 2.32 | 15.35 | 76.93 |
| Butt | 5335.00 | 9.60 | 1333.00 | 3.45 | 6.15 | 24.99 |
| Shirt | 17780.00 | 7.77 | 4846.00 | 0.60 | 7.17 | 27.26 |
| Employer | 6523.00 | 11.77 | 9168.00 | 0.77 | 11.00 | 140.55 |
| Container | 5272.00 | 15.52 | 4178.00 | 0.55 | 14.97 | 79.25 |
| Reinforcer | 162.00 | 13.58 | 186.00 | 2.15 | 11.43 | 114.81 |
| Gyroscope | 104.00 | 16.35 | 122.00 | 0.08 | 16.26 | 117.31 |
| Scimitar | 122.00 | 21.31 | 64.00 | 0.16 | 21.16 | 52.46 |
| Wolf | 12149.00 | 26.12 | 4053.00 | 1.26 | 24.86 | 33.36 |
| Scale | 29617.00 | 5.68 | 7930.00 | 1.27 | 4.41 | 26.78 |
| Wreath | 1081.00 | 9.44 | 585.00 | 0.51 | 8.92 | 54.12 |
| Partner | 23281.00 | 8.88 | 16542.00 | 2.21 | 6.68 | 71.05 |
| Dissertation | 1635.00 | 20.43 | 341.00 | 0.29 | 20.13 | 20.86 |
| Sea lion | 284.00 | 54.23 | 475.00 | 0.02 | 54.20 | 167.25 |
| Handbag | 772.00 | 10.10 | 441.00 | 0.91 | 9.20 | 57.12 |
| Dictionary | 2775.00 | 8.29 | 374.00 | 0.80 | 7.49 | 13.48 |
| Portfolio | 6241.00 | 20.65 | 1788.00 | 1.45 | 19.20 | 28.65 |
| Monograph | 497.00 | 9.66 | 344.00 | 0.29 | 9.37 | 69.22 |
| Averages | 6527.83 | 15.95 | 3097.11 | 1.06 | 14.89 | 65.30 |
| *Semantic/frequency matches – set 2* | | | | | | |
| Footnote | 2648.00 | 9.33 | 1203.00 | 0.25 | 9.08 | 45.43 |
| Jerk | 2432.00 | 10.36 | 2432.00 | 0.66 | 9.70 | 100.00 |
| Jacket | 13245.00 | 7.40 | 3292.00 | 0.73 | 6.67 | 24.85 |
| Chief | 33417.00 | 21.19 | 14879.00 | 1.24 | 19.96 | 44.53 |
| Packet | 2060.00 | 8.50 | 1479.00 | 0.81 | 7.68 | 71.80 |
| Fortification | 155.00 | 16.13 | 366.00 | 1.09 | 15.04 | 236.13 |
| Oscilloscope | 54.00 | 24.07 | 17.00 | 0.59 | 23.49 | 31.48 |
| Claymore | 62.00 | 53.23 | 26.00 | 0.38 | 52.84 | 41.94 |
| Coyote | 2596.00 | 24.35 | 1565.00 | 1.53 | 22.81 | 60.29 |
| Indicator | 3431.00 | 10.78 | 4657.00 | 1.18 | 9.60 | 135.73 |
| Cuff | 1947.00 | 21.52 | 1280.00 | 1.17 | 20.35 | 65.74 |
| Lover | 6844.00 | 2.54 | 5146.00 | 1.55 | 0.99 | 75.19 |
| Essay | 9773.00 | 7.91 | 4750.00 | 1.18 | 6.73 | 48.60 |
| Otter | 671.00 | 34.58 | 539.00 | 1.30 | 33.28 | 80.33 |
| Pocketbook | 738.00 | 13.55 | 270.00 | 0.37 | 13.18 | 36.59 |
| Encyclopedia | 1606.00 | 9.78 | 299.00 | 1.00 | 8.77 | 18.62 |
| Backpack | 2714.00 | 6.56 | 758.00 | 0.92 | 5.64 | 27.93 |
| Tract | 2595.00 | 21.62 | 1280.00 | 0.31 | 21.31 | 49.33 |
| Averages | 4832.67 | 16.85 | 2457.67 | 0.90 | 15.95 | 66.36 |

This prediction was examined by searching the Google database of over 2 billion web pages (on the 2nd of January, 2002) for instances of *rat/rats-infested* and *mouse/mice-infested*. One hundred of the pages that matched each search pattern were sampled at random to derive an estimate for the number of appropriate matches (all pages explicitly referring to arguments about level ordering were excluded). This survey indicated that *rat-infested* appeared 7000 times to 10 instances of *rats-infested* ($\chi^2(1, N = 7100) = 6970$, $p < .0001$), and *mouse-infested* appeared 350 times to 120 instances of *mice-infested* ($\chi^2(1, N = 470) = 6110$, $p < .0001$). It appears from this that although the majority of English speakers agree with Pinker that an apartment infested with rats is *rat-infested* (indeed, *rat-infested* appears to be a relatively frequent and lexicalized compound), they do not agree that an apartment infested with *mice* ought to be *mice-infested*. Although *mouse-infested* is not nearly as frequent, or conventionalized as *rat-infested* (and thus strictly, it is clear that the two are not directly comparable), in a significant majority (74.5%) of the instances sampled here, the compound with a singular non-head noun—*mouse-infested*—was preferred.

Further support for this contention comes from an analysis of these words in COCA, which produced 44 instances of *rat-infested*, 5 instances of *mouse-infested* – revealing just how much more conventionalized *rat-infested* is as compared to *mouse-infested* – and no instances of either *rats-infested* or *mice-infested*. Once again, no marked qualitative difference was apparent in the way that irregular and regular plurals are compounded in English.

Returning to the massive imbalance between +/s/ suffixed plurals and stem-change plurals in the input, is it unsurprising that the intuitions measured in Experiment 7 are a poor guide to participants' linguistic behavior (measured in Experiments 1–4). That these data so comprehensively fail to accord with the intuition articulated by Pinker, above, suggests a fundamental flaw in the method that led to the claim that, ''An apartment infested with mice is *mice-infested*, but an apartment infested with rats is not *rats-infested*; it is *rat-infested*.'' Thought-experiments, by their very nature, run into serious problems when it comes to making hypothesis blind observations, and because of this, it seems reasonable to suggest that their results should be afforded less credence in considering the phenomena themselves.

## 12. General discussion

The experiments presented here have tested the specific predictions level ordering makes about adult and child comprehension, production and ratings of acceptability. First, with regards comprehension, level ordering predicted that the presence of regular plural non-head nouns in compounds violated a constraint that applied specifically—and only—to regular plural nouns (leading to the 'recursive' interpretation of *red rats eater* as an eater of red rats). Contrary to the predictions of level ordering, Experiments 1 and 2 showed that not only do adult speakers of English interpret the presence of *irregular* plural non-head nouns in compounds ''recursively'' – an effect that cannot, in principle, be accounted for by level ordering – but also that the supposedly in-built grammatical constraints that drive 'recursive' interpretations could be over-ridden by simple semantic factors. Further, as Experiments 3 and 4 demonstrate, children as young as age three appear to apply the same interpretative standards to these compounds, irrespective of whether they contain regular or irregular plurals.

Secondly, with regards production, level ordering predicted that in contexts in which there were a number of items being referred to, a speaker should be able to denote an eater of such items as an *items-eater* rather than an *item-eater*, unless level ordering constraints applied. Since level ordering constraints apply only to regular noun plurals, this means that adults should exclude regular noun plurals from the non-head position of noun compounds while allowing irregular plurals into this position. However, Experiment 5 showed that ordinarily, adult speakers of English exclude *all* plural forms from the non-head positions of compounds, and that adult speakers only included more irregular than regular plurals in compounds when they were induced to do so as a result of particular manipulations of the compounding task, and even then, only very rarely. This finding was then tested further in Experiment 6, which showed similarly, that children excluded plural forms of both regular and irregular nouns from compounds unless primed to do so. Indeed, unless primed to do otherwise, children will describe a monster shown eating mice as a *mouse-eater*.

Finally, with regards to the acceptability of plurals in compounds, level ordering predicted that it is the regularity of nouns (i.e. whether their plural form is produced at Level Three or not) that determines the acceptability of their plural forms in compounds. Contrary to this prediction, Experiment 7 showed that the regularity of plural forms did not predict their acceptability, whereas whether or not the form used to mark plurality ended in a sibilant *did*. This suggests that phonetic conventions in compounding seems to be a far better indicator of acceptability than supposed grammatical structure.

### 12.1. Level ordering vs. learning?

Do the phenomena associated with noun-compounding in English necessitate a nativist account? Since level ordering has been put forward in tandem with the claim that noun compounding *cannot* be learned from the input (Gordon, 1985; Marcus, 1999; Pinker, 1999, 2001), it is worth considering whether a nativist or learning account can best explain the phenomena detailed in the results of the experiments presented here.

First, level ordering fails to address the critical question of just *how* the child's input gets to be so out of kilter with the "grammar" in the first place: if the grammar optionally *allows* irregular plurals in compounds, but disbars regulars, then why is the input data for regulars and irregulars virtually the same? What has happened to all of the irregular plurals in cases where they ought to be allowed? Arguments for level ordering seem to be incompatible with any explanation of the phenomena they seek to address. If level ordering is right, then this apparent paradox ought never to arise; the input ought to contain proportionally far more irregular than regular plurals in compounds. Yet our corpus analyses (and those of others) did not suggest that there was a qualitative difference in the distribution of regular and irregular plurals in compounds; instead, to the degree qualitative differences in the distribution of nouns were seen, they arose out of differences in frequency and phonology.

Setting aside the claims of level ordering, for a moment, how might one account for why the distribution looks like this? The results of Experiment 4 suggest that the distribution looks this way simply because English speakers learn conventions that ordinarily exclude both regular and irregular plurals from compounds. To a competent, adult speaker of English, a monster that eats mice is a *mouse-eater*, *not* a *mice-eater*. (Given the child's input, it's hard to see how this might be otherwise.) Further, as Experiment 1 showed, adult speakers of English are keenly sensitive to the presence of irregular plurals in compounds, using this information to discern different conventionalized interpretations when confronted with either *red-mice-eater* or *red-mouse-eater*, a finding which supports the notion that the convention of excluding plurals in noun-noun compounds aids informativity. Moreover, these conventionalized interpretations—which were also evident in Experiment 2 and 3A and 3B—were neither predicted by level ordering, nor can they be explained in terms of it: level ordering has nothing at all to say about how people make sense of irregular plurals in compounds.

### *12.2. Much ado about priming?*

If, in spite of the predictions of level ordering, English speakers seem to learn and apply the same conventions with regular and irregular plural-in-compounds, then this raises two questions: first, how do they acquire "constraints" in compounding, and second, why has it seemed to some researchers that these constraints appear to apply differently to irregular plurals at various stages of acquisition?

The simplest answer to the first question is, as we have suggested throughout, that adults learn these conventions from the input. Since adults do not ordinarily *hear* plurals in compounds, they do not *put* them in compounds. This learning account easily explains the different strengths of adult intuitions about the acceptability of regular, voicing change plurals vs. stem-change irregular plurals in compounds. Simply stated, these intuitions reflect the input, in which data relating to nouns which mark plurality with an +/*s*/ suffix is richer, producing strong intuitions, whereas the sparser, noisier data relating to nouns which mark plurality by changing their stems produces weaker intuitions.

These considerations, in conjunction with the results of Experiments 5 and 6, provide one answer to the second question: how to explain why regular and irregular plurals appear—at least at various stages of acquisition—to embody different conventions (or constraints). The short answer is that, once one takes into account the different amount of exposure children get to different plural forms, and the difference in noise that results from the higher average frequency of irregular noun types, they *don't*. As the experiments presented above clearly indicate, young children learn conventions that apply to both regular and irregular plurals from the language that they hear. This is why they apply the same conventions when they are adults, and in turn, this is why these conventions are in what children hear in the first place (see also Ramscar et al., 2010).

Thus although we replicated Gordon's finding that children would describe a monster shown eating mice as a *mice-eater* (Condition 1 of Experiment 6), when plural-priming was removed (Condition 2 of Experiment 6)—so that the semantic prediction made by Gordon (1985) was tested in controlled circumstances—children ceased calling the monster a *mice-eater*, and instead called it a *mouse-eater*. Thus it appears that Gordon's original study did not test *whether* children apply the same conventions when forming and interpreting regular and irregular plurals; rather, it tested how *resistant* these conventions are to priming effects. Given that the token frequency of stem-change irregular nouns appear in compounds is massively less than regular nouns (our COCA analysis found just 17,445 bigrams in which *man/men*, *foot/feet, mouse/mice, tooth/teeth* and *goose/geese* preceded a noun POS, a token rate of just 42.5/million), and given that by 4 years-old a child will have only heard between 25–45 million words in total (Hart & Risley, 1995)—i.e. *at best* around 2000 such tokens—it is hardly surprising that the task demands of Gordon's elicitation method affected 3–5 year old children more than adults (for a much fuller treatment of this idea, see Haskell et al., 2003).

### *12.3. Learning the conventions of plural compounding*

According to Alegre and Gordon (1996, p. 65), "advances in our understanding of phenomena often arise from paradoxes." In keeping with this, it has long been claimed that explaining people's "paradoxical" behavior in forming and interpreting regular and irregular noun plurals in compounds, should somehow advance our understanding of the innate mechanisms that govern language (Alegre & Gordon, 1996; Clahsen, 1999; Gopnik, 1990; Gordon, 1985; Pinker, 1991, 1999, 2001; Rice & Wexler, 1996; Van Der Lely & Christian, 2000; Van Der Lely & Ullman, 1996).

The results of the experiments described here suggest that plurals in compounds present neither a paradox, nor evidence for native constraints. Putting these results together with those of Gordon (1985) yields a straightforward, non-paradoxical picture of learning: first, we have the input, which is rich in the case of regular nouns, and sparser and less informative in the case of stem-change irregular nouns. Unsurprisingly, the initial quality of the *out*put mirrors the quality of the input: as Gordon (1985) showed, children's ability to even *form* irregular plurals is far weaker than it is for regulars (in Gordon's 1985 study, children's correct rate of irregular plural production was only 24.3%, as compared to 99.4% for regular plurals). Such findings, coupled with what we know about the distribution of the input, suggests that the irregular constraints children extract from the environment will be learned less strongly than the regular constraints (Ellis & Schmidt, 1999; Ramscar & Yarlett., 2007), and more susceptible to priming.

However, once the priming in the Gordon task is controlled for (Experiment 5), it is quite apparent that over time, children readily learn compounding constraints for both regular and irregular nouns that reflect the patterns in the input. Indeed, given the sparseness of the input for irregular plural nouns, what is most striking is the very tight coupling between children ceasing to overregularize (describing a monster who eats *mouses* as a *mouse-eater*) and beginning to more consistently produce singular forms of irregular nouns in compounds (describing a monster who eats *mice* as a *mouse-eater*)—the correlation between these factors in Experiment 5 Condition 2 was $r = -.704$.

Finally, adult output *has* to match the input presented to the child. It should hardly be surprising that the production of adult speakers in the third (un-primed) control condition in Experiment 4 matched the proportions of regular and irregular nouns in plural and singular form in the linguistic environment (Gordon, 1985). Given how strikingly different these results were from what we found in the primed conditions, it follows that production influenced by priming is *not* representative of ordinary competence.

This observation helps us untangle one of the knottiest "paradox" of level ordering: how can a balanced input of regular and irregular plural compounds in the distribution result from very different constraints on linguistic production? The evidence suggests that (perhaps inevitably) the patterns governing the output ultimately align with those in the input. Thus, *there is no paradox*: people succeed in learning conventions for all compounds, and use these in producing the patterns of output that later serve as input to succeeding generations.

As with the past tense (Gordon & Miozzo, 2008; Ramscar, 2002), a detailed examination of the evidence leads us to conclude that innate constraints that distinguish between regular and irregular forms are not necessary to explain the behavior associated with inflection and morphology in compounding. We do not—and would not—suggest, that it follows that there are no innate constraints. Rather, our results show that there is no need to resort to language-specific innate constraints to account for plural compounding. Further, children's patterns of behavior as they learn plural compounding do not constitute evidence in support of these kinds of constraints. Whatever the claims regarding the "poverty of the stimulus" for other aspects of language acquisition, in both these cases we can give a simple, compelling and economical explanation of children's behavior in terms of the patterns that emerge naturally as they learn, through experience, the particular conventions embodied in their languages. Even when children's behavior apparently diverges from that of adults, this seems to reflect nothing more than differences in the frequencies of conventions, and the natural effect these frequencies have on the rates at which such conventions are learned.

Given that the participants in our experiments showed a remarkable sensitivity to the statistical properties of the English language, it would seem that the input – far from being an impoverished source of information about noun-compounding and other linguistic conventions – is by far the best predictor *of* such conventions.

## Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant Number 0547775 to Michael Ramscar. We are grateful to Lera Boroditsky, Daniel Yarlett, Michael Tomasello, Elena Leiven, Heike Behrens, Kirsten Smith, Annabel Harrison, Shamuna Ali and Ivan Pearson for insightful comments and discussions during the development of these ideas. Heartfelt thanks are offered to Annie Wilkinson for heroic feats of data collection. Correspondence concerning this article should be addressed to Michael Ramscar at the Department of Psychology, Jordan Hall, Stanford, CA 94305. Electronic mail may be sent to ramscar@gmail.com.

## References

Alegre, M. A., & Gordon, P. (1996). Red rats eater exposes recursion in children's word formation. *Cognition, 60*, 65–82.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 247–264.

Altmann, G. T. M., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science, 33*, 1–27.

Baayen, R. H. (2010). Assessing the processing consequences of segment reduction in Dutch with naive discriminative learning. *Lingue & Linguaggio*.

Behrens, H. (2001). Learning multiple regularities: Evidence from overgeneralization errors in the german plural. In A. H.-J. Do, L. Domínguez, & A. Johansen (Ed.), *Proceedings of the 26th Annual Boston University conference on language development*. Somerville, MA: Cascadilla Press.

Berent, I., & Pinker, S. (2007). The dislike of regular plurals in compounds: Phonological familiarity or morphological constraint? *The Mental Lexicon, 2*(2), 129–181.

Berent, I., & Pinker, S. (2008). Compound formation is constrained by morphology: A reply to Seidenberg, MacDonald & Haskell. *The Mental Lexicon, 2*(3), 176–187.

Bybee, J. L. (1988). Morphology as lexical organization. In M. Hammond & M. Noonan (Eds.), *Theoretical morphology* (pp. 119–141). San Diego: Academic Press.

Bybee, J. L. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes, 10*, 425–455.

Bybee, J. L., & Slobin, D. I. (1982). Rules and schemes in the development and use of the English past tense. *Language, 58*, 265–289.

Chierchia, G. (1998). Reference to kinds across languages. *Natural Language Semantics, 6*, 339–405.

Clahsen, H. (1999). Lexical entries and the rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences, 22*, 991–1060.

Clahsen, H., Rothweiler, M., Woest, A., & Marcus, G. (1992). Regular and irregular inflection in the acquisition of German noun plurals. *Cognition, 45*, 225–255.

Davies, M. (2009). The 385+ million word corpus of contemporary American English (1990–2008): Design, architecture and linguistic insights. *International Journal of Corpus Linguistics, 14*, 159–190.

Diamond, A., & Goldman-Rakic, P. S. (1989). Comparison of human infants and rhesus monkeys on Piaget's AB task: Evidence for dependence on dorsolateral prefrontal cortex. *Experimental Brain Research, 74*, 24–40.

Ellis, N. C., & Schmidt, R. (1999). Rules or associations in the acquisition of morphology? The frequency by regularity interaction in human and PDP learning of morphosyntax. *Language & Cognitive Processes, 13*, 307–336.

Fernald, A., Swingley, D., & Pinto, J. P. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child Development, 72*, 1003–1015.

Gennari, S. P., & MacDonald, M. C. (2009). Linking production and comprehension processes: The case of relative clauses. *Cognition, 111*, 1–23.

Gopnik, M. (1990). Feature-blind grammar and dysphasia. *Nature, 344*, 715.

Gordon, P. (1985). Level ordering in lexical development. *Cognition, 21*, 73–93.

Gordon, P., & Miozzo, M. (2008). Can word formation be understood or understanded by semantics alone? *Cognitive Psychology, 56*, 30–72.

Graves, M. F., & Koziol, S. (1971). Noun plural development in primary grade children. *Child Development, 42*, 1165–1173.

Hahn, U., & Nakisa, R. C. (2000). German inflection: Single route or dual route? *Cognitive Psychology, 41*(4), 313–360.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experiences of young American children*. Baltimore: Paul H. Brookes Publishing Company.

Haskell, T. R., MacDonald, M. C., & Seidenberg, M. S. (2003). Language learning and innateness: Some implications of compounds research. *Cognitive Psychology, 47*, 119–163.

Jackendoff, R. (1997). *The architecture of the language faculty*. MIT Press.
Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). Prediction and thematic information in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*, 133–156.
Kiparsky, P. (1982). From cyclic phonology to lexical phonology. In H. van der Hulst, & N. Smith (Eds.), *The structure of phonological representations – Part 1*. Foris.
Kiparsky, P. (1983). Word-formation and the lexicon. In F. Ingemann, (Ed.), *Proceedings of the 1982 mid-American linguistics conference*. Lawrence, KS: University of Kansas.
Kirkham, N. Z., & Diamond, A. (2003). Sorting between theories of perseveration: Performance in conflict tasks requires, attention, and inhibition. *Developmental Science, 6*, 474–476.
Köpcke, K-M. (1998). The acquisition of plural marking in English and German revisited: Schemata versus rules. *Journal of Child Language, 25*, 293–319.
Köpke, K-M. (1993). *Schemata bei der Pluralbildung im Deutschen*. Tübingen: Gunter Narr.
Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence: Brown University Press.
Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualizations: Revising the verb learning model. *Cognition, 40*, 121–157.
Marcus, G. F. (1995). Children's overregularization of English plurals: a quantitative analysis. *Journal of Child Language, 22*, 447–459.
Marcus, G. F. (1999). Poverty of the stimulus arguments. In R. Wilson & F. C. Keil (Eds.), *Encyclopedia of cognitive science*. Cambridge MA: MIT Press.
Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R., & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology, 29*, 189–256.
McClelland, J. L., & Patterson, K. (2002). Rules or connections in past-tense inflections: what does the evidence rule out? *Trends in Cognitive Sciences, 6*, 465–472.
Milin, P., Kuperman, V., Kostic, A., & Baayen, R. H. (2009). Paradigms bit by bit: An information theoretic approach to the processing of paradigmatic structure in inflection and derivation. In J. P. Blevins & J. Blevins (Eds.), *Analogy in grammar: Form and acquisition* (pp. 214–252). Oxford: Oxford University Press.
Miller, G. A. (1967). The psycholinguists. In G. A. Miller (Ed.), *The psychology of communication*. London: Penguin Books.
Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.
Murphy, V. A., & Hayes, J. (2010). Processing english compounds in the first and second language: The influence of the middle morpheme. *Language Learning, 60*(1), 194–220.
Nicoladis, E. (2003). Compounding is not contingent on level ordering in acquisition. *Cognitive Development, 18*, 319–328.
O'Sullivan, L. P., Mitchell, L. L., & Daehler, M. W. (2001). Representation and perseveration: Influences on young children's representational insight. *Journal of Cognition & Development, 2*, 339–366.
Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
Pinker, S. (1989). Review of Dan Slobin, "Crosslinguistic evidence for the language-making capacity" and Melissa Bowerman's "What shapes children's grammars". *Journal of Child Language, 16*, 456–463.
Pinker, S. (1991). Rules of language. *Science, 253*, 530–535.
Pinker, S. (1999). *Words and rules*. New York: Basic Books.
Pinker, S. (2001). Four decades of rules and associations, or whatever happened to the past tense debate? In E. Dupoux (Ed.), *Language, the brain, and cognitive development: Papers in honor of Jacques Mehler*. Cambridge, MA: MIT Press.
Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28*, 73–193.
Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences, 6*, 456–463.
Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review, 107*, 786–823.
Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition, 48*, 21–69.
Prasada, S., & Pinker, S. (1993). Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes, 8*, 1–56.
Ramscar, M. J. A. (2002). The role of meaning in inflection: Why the past tense doesn't require a rule. *Cognitive Psychology, 45*(2), 45–94.
Ramscar, M. J. A., & Dye, M. (2009). Expectation and negative evidence in language learning: the curious absence of mouses in adult speech. In *Proceedings of the 31st annual conference of the cognitive science society*, Amsterdam, Netherlands.
Ramscar, M., & Gitcho, N. (2007). Developmental change and the nature of learning in childhood. *Trends in Cognitive Science, 11*(7), 274–279.
Ramscar, M. J. A., & Yarlett, D. G. (2007). Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. *Cognitive Science, 31*, 927–960.
Ramscar, M., Yarlett, D. G., Dye, M. W., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science, 34*(6), 909–957.
Rice, M. L., & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of Speech and Hearing Research, 39*, 1239–1257.
Rumelhart, D., & McClelland, J. (1986). On learning the past tenses of English verbs. In J. McClelland, & D. Rumelhart (Eds.), *Parallel distributed processing* (Vol. 2). The PDP Research Group, MIT Press.
Senghas, A., Kim, J. J., Pinker, S., & Collins, C. (1991). Plurals-inside compounds: Morphological constraints and their implications for acquisition. In *Proceedings of the sixteenth annual Boston university conference on language development*, Boston, MA.
Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B, 364*, 3617–3622.
Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition, 71*, 73–108.

Tanenhaus, M. K., & Brown-Schmidt, S. (2008). Language processing in the natural worldB. C. M. Moore, L. K. Tyler, & W. D. Marslen-Wilson (Eds.). The perception of speech: From sound to meaning. *Transactions of the royal society B: biological sciences, 363*, 1105–1122.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. E. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.

Teichmann, M., Dupoux, E., Kouider, S., Brugieres, P., Boisse, M. f., Baudic, S., et al (2005). The role of the striatum in rule application: The model of huntington's disease at early stage. *Brain, 128*, 1155–1167.

Tomasello, M., & Stahl, D. (2004). Sampling children's spontaneous speech: How much is enough? *Journal of Child Language, 31*, 101–121.

Ullman, M. T. (2004). Contributions of neural memory circuits to language: The declarative/procedural model. *Cognition, 92*(1-2), 231–270.

Van Der Lely, H. J. K, & Ullman, M. T. (1996). The computation and representation of past-tense morphology in specifically language impaired and normally developing children. In *Proceedings of the 20th annual Boston University conference on language development*. Somerville, MA: Cascadilla Press.

Van Der Lely, H. J. K., & Christian, V. (2000). Lexical word formation in children with grammatical SLI: A grammar-specific versus an input-processing deficit? *Cognition, 75*, 33–63.

Zipf, G. K. (1935). *The psychobiology of language*. New York: Houghton-Mifflin.

Zipf, G. K. (1949). *Human behavior and the principle of least-effort*. Cambridge, MA: Addison-Wesley.