Chapter 6

Conclusion

> With the unknown, one is confronted with danger, discomfort, and care; the first instinct is to abolish these painful states. First principle: any explanation is better than none... The causal instinct is thus conditional upon, and excited by, the feeling of fear. The "why?" shall, if at all possible, not give the cause for its own sake so much as for a particular kind of cause – a cause that is comforting, liberating, and relieving.
>
> –Friedrich Nietzsche, *Twilight of the Idols*

6.1 Goals

The aim of this dissertation has been to analyze the logical and empirical foundations of Baker's Paradox. To review, Baker's Paradox comprises three premises: *productivity, arbitrariness,* and *no negative evidence.* Under the premise of productivity, an unbounded number of lexical items can instantiate a given syntactic pattern, in principle. Under the premise of arbitrariness, there are certain lexical items that cannot instantiate the pattern despite meeting the criteria governing the productivity of the pattern (when there is no preempting form). Under the "no negative evidence" premise, language learners do not have access to evidence that such forms are

ungrammatical.

On the empirical side, the emphasis of this dissertation has been on denying the premise of arbitrariness. The goal has been to demonstrate that this premise is not well-founded, by showing that putative examples of arbitrary exceptions to productive patterns in a range of domains fall under larger generalizations delimiting the productivity of the patterns in question. Although it would have been impossible to address the universe of examples of arbitrary exceptions, I hope that this work has been able to show that the premise of arbitrariness cannot be taken as given. Upon detailed linguistic investigation into the studied domains, the arbitrary exceptions have disappeared; the lexical items failing to instantiate the patterns in question are all subject to general criteria governing the productivity of the pattern.

On the logical side, the focus has been on the implications of the existence of arbitrary exceptions, or the lack thereof, for the learning of restrictions on productive patterns. The existence of arbitrary exceptions is often taken to have deep implications regarding the nature of learning, and it is therefore important to establish what follows logically from their existence or non-existence and what does not. The relationship between the "arbitrariness" and "no negative evidence" premises is one of the issues under this heading. The tripartite structure of Baker's Paradox might suggest that the three premises correspond to mutually exclusive solutions to the paradox, so denying arbitrariness requires asserting the absence of negative evidence. Baker's Paradox can be doubly solved by denying both, however; these are compatible solutions to Baker's Paradox.

Another issue under this heading is the relationship between arbitrariness and attentiveness by the learner to the use of individual words in individual constructions, which Culicover (1999) argues to be a consequence of idiosyncrasies in language. Indeed, the notion of attentiveness is more closely related to the issue of arbitrariness, and could in principle provide an explanation for why it should not hold. However,

in this chapter, I advocate a different view of learning, according to which learners prefer explanatory generalizations to arbitrary stipulations, namely, *the explanation-seeking learner*. This view is consonant with a large number of experimental findings in psychology, and explains why it is so difficult to find a good example of an arbitrary exception.

6.2 Empirical foundations of Baker's Paradox

6.2.1 Range of phenomena

Potential cases of arbitrary exceptions have been considered in a wide range of domains, one that is wider than the range of empirical phenomena that have heretofore been considered in relation to Baker's Paradox. Pinker (1989), and subsequent writers on this topic, have focussed on verbal diathesis alternations such as the causative, dative, and active/passive alternations. Recall Schütze's (1997:122) reasoning by which the premises of Baker's Paradox lead to a contradiction, which is stated in terms of subcategorization (emphasis added):

> If *subcategorization* is unpredictable [Arbitrariness], then it must be learned for each *verb* individually. With no negative evidence available [No Negative Evidence], the child would have to limit herself to repeating *subcategorization frames* perceived in parental speech. But this contradicts [Productivity] (i.e., the child will use *verbs* with unattested *subcategorization frames*).

This dissertation has demonstrated that the problem extends beyond verbal subcategorization. Of course, issues of subcategorization cannot be ignored in a substantive discussion of Baker's Paradox, and the two verb alternations that have figured most prominently in the literature on this subject are addressed in this dissertation: Chapters 2 and 3 address the causative and dative alternations, respectively. However,

the empirical realm of Baker's Paradox has been expanded to include other aspects of grammar: preposition pied-piping and stranding (Chapter 4) and the prenominal and predicative uses of adjectives (Chapter 5).

The range of phenomena to which the central claim of this dissertation applies extends beyond the case studies addressed here. Many interesting cases remain even within English syntax. For example, the principles underlying the selection of *as* complements do not seem to be well-understood; some 'verbs of considering' allow them and others do not (Pollard and Sag 1994):

(1) Mary regards/*believes Sally as an acceptable candidate.

Another instance of Baker's Paradox concerns the determiners that allow the so-called "Big Mess Construction" (Berman 1974):

(2) I got so/*this big a raise that I stayed there.

The variation in the domain of such determiners was claimed by Van Eynde (2007:11) to be "a matter of lexical stipulation," counter to the central claim of this dissertation, assuming that the Big Mess Construction is productive.

There could in principle be instances of Baker's Paradox in the realm of morphology as well. A productive derivational affix that apparently arbitrarily fails to attach to certain stems would constitute an instance of Baker's Paradox in the morphological domain. The general claim that this dissertation aims to defend, therefore, extends far beyond the cases that have been addressed within it.

Along with expanding the range of empirical instantiations of Baker's Paradox, this dissertation has also delimited them. Some idiosyncrasies do not constitute arbitrary exceptions, and are therefore consistent with the claim that I am defending here: Any idiosyncrasy that does not constitute an arbitrary exception to a productive rule is consistent with the criteria-governed productivity solution to Baker's Paradox. One example of this type is the use of *rather* with sentential complements, for example:

(3) I would rather (that) you stayed.

This sense of *rather*, meaning "prefer," otherwise behaves as a comparative adverb, ending in *-er*, taking *than* (e.g. *I would rather stay than go*), and appearing before a bare verb phrase (e.g. *I would rather go*). The ability of *rather* to license sentential complements does not constitute a counterexample to the claim that arbitrary exceptions do not exist because this behavior is an *ability* to behave in a certain way (what can be called a "positive exception") rather than an *inability* to behave in a certain way (a negative exception). *Rather* is similar to other "syntactic nuts" discussed by Culicover (1999) in this respect.

6.2.2 Consequences of the range of phenomena

Expanding the range of phenomena helps to distill the essence of the learnability problem: An instance of Baker's Paradox is found wherever there is a productive generalization and there appear to be arbitrary exceptions to it. When the discussion is limited to verb alternations, issues that are not related to the core of the issue take on unmerited importance – for example, the question of whether or not linking rules are innate is an important issue in domains of Baker's Paradox related to verb alternations (Bowerman 1990; Brinkmann 1996; Marcotte 2005), but this issue is not crucially related to Baker's Paradox in general.

Another way in which expanding the range of phenomena helps to distill the essence of the issue is by showing that having an alternative paraphrase is not a necessary property of an instance of Baker's Paradox. Verb alternations provide two alternative paraphrases (for example, the double object and prepositional dative forms of the dative alternation), but not all Baker's Paradox phenomena have this property. For example, the prenominal use of adjectives has no paraphrase as similar to it in meaning as the alternants of the dative alternation are to each other. Although *an*

angry person can be paraphrased, *a person who is angry,* the latter differs from the former in quite a few ways: it is much longer, it contains a tensed clause, etc. Because this type of case exists and falls under the scope of the problem, the solution must extend to such cases.

The model proposed by Schütze (1997) is an example of a solution that relies on an alternation between roughly equivalent paraphrases. Schütze describes a connectionist network whose output layer contains one node representing the double object construction and one node representing the prepositional dative construction. The input layer contains information about the verb and its arguments. The network as a whole functions as a probabilistic model of choice between alternants of the dative alternation (and in this regard is similar to the model described by Bresnan et al. (2007)). Schütze counts among the virtues of this model the fact that it predicts both productivity and arbitrariness: It can generalize to new verbs, but arbitrary exceptions can also be represented, as direct pathways from the input layer to the output layer. Of course, the main thesis of this dissertation is that predicting arbitrariness is not a virtue of any model, but Schütze's basic model could trivially be altered to rule out arbitrary exceptions. A non-trivial drawback of this model, however, is that its architecture relies on alternative paraphrases. Constraints on the double object construction are learned solely through repeated witnessing of prepositional dative constructions. Such a model would not work for learning constraints on prenominal adjectives, because there is no alternative construction that could serve as the "antagonist," as it were, for reasons described above.

A further drawback of a model that relies on the presence of alternative paraphrases is that, assuming that it is intended to explain contrasts of acceptability, the double object construction is predicted always to be as "bad" as the prepositional dative construction is "good," because the two alternants are yoked. For every increase in the strength of the pathway between a given verb and the double object

construction, there is a corresponding decrease in strength between that verb and the prepositional dative construction. Assuming that the strengths of these pathways represent levels of acceptability, this predicts that the prepositional dative construction should be less acceptable with alternating verbs than with non-alternating verbs. For example, *send something to someone* should be worse than *drag something to someone*, because *send* is alternating and *drag* is non-alternating. The results of Experiment 1 from Chapter 3 do not support this notion; prepositional datives were rated comparably among alternating and non-alternating verbs. In defense of Schütze's model, one might counter that it is intended to model probability rather than acceptability, but in that case it is not intended to account for the empirical phenomena that a solution to Baker's Paradox should account for, namely, contrasts in acceptability.

This is not to say that there is no connectionist network model that could provide a solution to the problem at hand. I believe that an appropriate connectionist model could be designed for this purpose if it had a somewhat different architecture. Such a model would encode *constraints on a syntactic position.* For a given syntactic position, such as "predicative adjective," this model would output "yes" (i.e., acceptable) or "no" (i.e., unacceptable) depending on various input features. The input features would include semantic features of the item to be placed in the position, such as semantic predicativity. In general, any *model of acceptability* for a syntactic position would avoid the problem of being reliant on alternative paraphrases.[1]

Where would negative evidence come from under this view? Negative evidence can come from explicit or implicit signals. Explicit negative evidence, by definition, arises in situations where an utterance or part of an utterance is explicitly said to be ungrammatical, or where caretakers offer corrective feedback. If the corrected utterance contains an instance of a given syntactic construction, then the learner can

[1]This type of model may be equivalent in some respects to models of part-of-speech tag induction (see Klein 2005 and references therein).

hypothesize that the utterance contained an abuse of this syntactic construction. For example, if the utterance contains a prenominal use of an adjective, and is explicitly said to be ungrammatical, then the learner can hypothesize that the sentence contained a faulty use of the prenominal adjective position. As Chouinard and Clark (2003) point out, corrective feedback contains information about the locus of the problem, so this information is available at least on some occasions. Once the locus of the problem has been identified, the learner can identify possible sources of the difficulty. Semantic features of the word used in that position can be hypothesized as the causes of the difficulty.

For example, recall the corrective feedback given in response to an overgeneralization of *alive* to prenominal position, shown at the end of Chapter 5 on page 192. In this instance, the caretaker provides the child with corrective feedback allowing him to infer that there was something ill-formed about his use of *alive* in *alive monsters* (*Live monsters? What are some live monsters?*). As Chouinard and Clark (2003) would point out regarding this example, the contrast between *live* and *alive* tells Mark the locus of the error, and moreover what can be done to fix it. After several experiences of this type, Mark may begin to hypothesize that words that begin with *a-* are not acceptable in prenominal position (or he may hypothesize different explanations, or disregard the evidence entirely).

Implicit negative evidence would work similarly, although it would also require a mechanism for generating expectations, because implicit negative evidence arises when expectations are violated. Models of acceptability do not provide expectations; they merely discriminate acceptable from unacceptable uses of a given syntactic position. However, assuming there is a mechanism for generating expectations, implicit negative evidence could arise when those expectations are violated. For example, if a prenominal use of an adjective is expected in a given instance, and does not occur, then the learner may hypothesize that the expected sentence might contain an

abuse of the prenominal adjective position. The learner can then hypothesize potential causes of difficulty in the same way that he or she can when explicit negative evidence is available.

In summary, the breadth of the phenomena surveyed in this dissertation helps to distill the nature of the problem, and shows that models that rely on the presence of alternative paraphrases are not sufficient to capture the full range of phenomena. This does not imply that no connectionist or gradient model will suffice; a connectionist or gradient model of acceptability would avoid the problem of being reliant on alternative paraphrases.

6.2.3 Types of criteria governing productivity

The primary goal of the empirical studies has been, of course, to show that there are no arbitrary exceptions in these domains. In other words, it has been to deny the empirical premise of Baker's Paradox, "Arbitrariness." This type of approach can be called the "criteria-governed productivity" approach.

The type of analyses that have served as "criteria-governed productivity" solutions to Baker's Paradox have typically fallen in the realm of semantics. Indeed, the approach that aims "to look for semantic or perceptual characteristics that correlate with the syntactic distributions and to propose that these play a significant role in acquisition" is labelled the "semantics approach" by Wonnacott et al. (2008:167). The majority of the constraints and generalizations that I have given as explanations for the behavior of putative arbitrary exceptions have in fact posited relationships between syntax and semantics. For example, the linking rule given in Chapter 2 ("The Causative Alternation") constrains the relationship between the meaning of a verb and the syntactic expression of its arguments. In Chapter 3 ("The Dative Alternation"), some semantic constraints on the ditransitive use of verbs are invoked, such as the requirement that ditransitive verbs describe a transfer of possession. In

Chapter 4 ("Odd Prepositions"), restrictions on stranding are argued to follow from a constraint on the complexity of the event spanned in a long-distance dependency. The Predicativity Principle in Chapter 5 ("Adjectives") establishes semantic restrictions on the type of adjectives that can be used syntactically as predicative adjectives.

However, the "semantics approach" would be a misnomer for the present approach, because meaning is not a critical feature of a solution under this approach. Some purely syntactic constraints have served as solutions. The fact that specifiers precede heads in English, for example, is used in Chapter 4 to explain the ordering of *ago* with respect to its complement. Also in Chapter 4, the fact that *ago* does not strand is argued to follow from the constraint against the extraction of specifiers also seen in "Left Branch Condition" violations such as *Whose did you read _ book?* Explanations based on morphology have also been considered, and supported to some extent: Chapter 3 shows that the hypothesis that morphological complexity governs the productivity of the ditransitive use of verbs remains the most viable of the available hypotheses, and Chapter 5 argues that the inability of *a-* adjectives to function prenominally was due to a morphological property. What this implies is that my claim is weaker than the claim that "everything is predictable from semantics," as it were. Such a claim is opposed to the "Arbitrariness" premise, but is stronger than the one I am advocating.

6.2.4 Domain-specific findings

A benefit of taking a stance against the "Arbitrariness" premise is that one has the opportunity to develop a richer understanding of the phenomena that lie in the scope of Baker's Paradox. The need to evaluate the existence of arbitrary exceptions motivates one to reevaluate both the criteria governing productivity in these domains, and how to establish whether or not these criteria have been met. Although there is more work to be done to fully understand these domains, I hope I have been able to

provide some additional light on the phenomena I have investigated.

The causative alternation is perhaps the best-understood of the phenomena I address, and the criteria for undergoing the causative alternation laid out by Levin and Rappaport Hovav (1995) are supported by the empirical investigations in Chapter 2 ("The Causative Alternation"). Levin and Rappaport Hovav's (1995) theory is often described with respect to the distinction between *internally* and *externally* caused verbs, leaving out a third class, which I have labelled *non-caused*. This third class is important, because two of the most frequently cited verbs in the language acquisition literature on the causative alternation – *come* and *disappear* – fall into this class. The investigations in Chapter 2 also show how the distinction between internal and external causation applies within the class of verbs of cyclical motion. I argue that the non-alternating verbs in this class (*totter, revolve*) describe internally caused eventualities, and the alternating verbs (*spin, rotate*) describe externally caused eventualities. The argument is based on an observation regarding direction of force, which leads to the statement of a principle that can be used for distinguishing internal and external causation, namely, the Direction of Force Principle.

Chapter 3 addresses criteria governing the productivity of the double object construction having to do with form as well as meaning, focusing primarily on form. The experimental results reported in this chapter help to narrow down the set of possible ways of stating Gropen et al.'s (1989) "morphophonological constraint" on the dative alternation, a constraint that has been described in terms of prosodic weight (in terms of metrical feet), etymological origin, formality, and morphological complexity. All of these hypotheses except the morphological complexity hypothesis were tested directly, and none of the hypotheses tested were fully supported. This brings up an obvious question for future work, namely, whether or not the morphophonological constraint is in fact a morphological constraint.

Chapter 4 responds to Culicover's (1999) case study on the behavior of prepositions, primarily focusing on their ability to undergo pied-piping and stranding. Whereas Culicover argues for a highly stipulative view of preposition behavior, this chapter shows that all of the facts that Culicover describes in this case study fall into general classes of phenomena. The analyses offered in this chapter build on the insights of previous authors in many cases, but some new ideas arose from this investigation as well. For instance, I argued for the existence of the Marking Generalization as part of an explanation for the inability of prepositions like *off* and *out* to pied-pipe, and this was shown to be supported by the behavior of other prepositions as well.

Chapter 5 develops several generalizations regarding the predicative and prenominal uses of adjectives. The part of the chapter dealing with constraints on the predicative use of adjectives includes a detailed descriptive classification of the adjectives that fail to occur predicatively. Although this set is somewhat diverse, its members all share the semantic feature of being non-predicative. On the basis of this generalization, I propose the Predicativity Principle, which accounts for a wide range of restrictions on the predicative use of adjectives. The part of the chapter dealing with constraints on the prenominal use of adjectives argues that *a-* adjectives (*asleep*, *abuzz*, etc.) are ruled out from prenominal position because they synchronically contain the morpheme *a-*, which is a productive prefix. Crucially, new forms containing this prefix are unacceptable prenominally, so *a-* adjectives are subject to a general constraint on the productivity of the prenominal adjective position.

Investigating these putative exceptions is thus useful not only for evaluating the existence of arbitrary exceptions, but also fruitful for the study of language, contributing to a deeper understanding of the phenomena themselves.

6.3 Logical foundations of Baker's Paradox

The other major goal in this dissertation has been to analyze the logical foundations of Baker's Paradox. The purpose of doing so is not only to solve the paradox, but also to characterize the possible mechanisms by which restrictions on productive patterns are acquired, as discussed in Chapter 1. In this section, I will review the main points of Chapter 1, and go on to discuss possible explanations for my thesis of non-arbitrariness.

6.3.1 The theoretical landscape

In Chapter 1, I argued for the existence of two independent dichotomies, one between arbitrariness and criteria-governed productivity, and one between conservatism and negative evidence. These two dichotomies are orthogonal in the landscape of possible theories of how restrictions on productivity are acquired. Thus, it is possible to advocate criteria-governed productivity while at the same time assuming that negative evidence is used in language acquisition, as I do.

Although these dichotomies are orthogonal, they are not unrelated; the question of arbitrariness affects the type of negative evidence that it would be necessary for learners to use. The type of negative evidence that would be necessary for learning arbitrary exceptions would pertain to the use of individual words in individual constructions; if arbitrary exceptions do not exist, as I claim, then the learner may use negative evidence pertaining to general properties that words may have.

Chapter 1 also addressed how Baker's Paradox relates to Culicover's (1999) idea of the "Conservative Attentive Learner." According to Culicover (1999), the existence of arbitrary exceptions implies that the learner is conservative and attentive. As I argued in Chapter 1, conservatism is the flip-side of negative evidence. This means that the relationship between arbitrariness and conservatism, just like the relationship

between arbitrariness and negative evidence, is orthogonal, contrary to Culicover's (1999) claim.

However, I agree with Culicover that attentiveness follows from arbitrariness. If there are arbitrary exceptions, then the learner must be attentive to the use of individual words in individual constructions, and furthermore encode and store such information. Contrapositively, non-attentiveness implies non-arbitrariness. This means that non-attentiveness is one possible explanation for non-arbitrariness. However, there are other possible explanations for non-arbitrariness, which will be discussed in the next section.

6.3.2 Explanations for non-arbitrariness

Supposing that arbitrary exceptions do not, in fact, exist, what would this linguistic situation mean for learning? Several possible consequences of arbitrariness have been identified above (negative evidence, attentiveness), but what follows from its negation? The consequences of the negation are more indeterminate than the consequences of the assertion, but there are several possible theoretical views that would explain it: (i) limitations on the architecture of grammar, preventing word-specific constraints from being expressible; (ii) limitations on the nature of learning, in particular, non-attentiveness; (iii) a preference on the part of the learner for general explanations over stipulations, which I call *the explanation-seeking learner*. As I will discuss in more detail below, evidence from psychology suggests that humans are capable of learning word-specific constraints, i.e., attentiveness. I therefore posit the explanation-seeking learner as a way of understanding why arbitrary exceptions are so few and far between.

6.3.2.1 Architectural limitations

To explain the non-existence of arbitrary exceptions, one might imagine that the architecture of the grammar is such that it cannot accommodate the presence of arbitrary exceptions. I do not believe that there is one single architectural assumption about the grammar that could be used to rule out the possibility of all arbitrary exceptions. The phenomena falling under the scope of Baker's Paradox are heterogeneous in such a way that a variety of assumptions would be necessary to grammatically rule out the possibility of arbitrary exceptions.

For example, suppose that the inability of a preposition to strand constitutes a requirement that the preposition's complement be overt. This restriction could be encoded in Head-Driven Phrase Structure Grammar (Pollard and Sag 1994) as a constraint imposed by a preposition that the SYNSEM of its complement be of type *canonical*, as opposed to *empty*. Requiring that a grammar be *incapable* of specifying a constraint like this would amount to requiring the grammar to be such that a preposition cannot impose constraints on the overtness of its complement.

The prohibition of constraint specifications could take on a very different form in the domain of diathesis alternations such as the dative alternation. How exactly this would work depends on how the problem is framed. If constructions are made part of the theoretical apparatus (Goldberg 1995), then the problem can be seen as identifying the constraints on what elements can fill the position of *V* in the *V NP NP* construction. To rule out the possibility of arbitrary exceptions under this view, it is necessary to rule out the possibility of stating restrictions on the ability of individual verbs to appear in the head position of a particular construction.

On the other hand, if argument realization is determined by individually linking partipants to positions in argument structure such as "external argument," and linking those positions in argument structure to surface positions such as "subject," "indirect object," and "direct object" or their configurational equivalents, then the

items whose positions are in question are the arguments of the verb. In this case, one could rule out arbitrary lexical exceptions by preventing argument linking from being sensitive to the identity of the verb. If argument linking were an encapsulated process that could not be influenced by factors other than semantic features of the eventuality and its participants, then non-arbitrariness would follow as a consequence. This view may be too strong, however, in light of the fact that there seems to be a morphophonological constraint on the dative alternation.

Although the nature of the putative word-specific constraints would vary from domain to domain if they were grammatically ruled out, I believe that there is an abstract architectural assumption that could rule out word-specific constraints on acceptability. Recall the connectionist architecture described above, in which the output was "yes" (acceptable) or "no" (unacceptable) depending on various input features, for a specific syntactic position. In such a framework, word-specific constraints on a particular position can be modelled as hard-wired pathways from inputs representing particular words to the "no" output, if the identity of the word is a possible input feature. This type of system is a model of acceptability, which could be implemented using a regression-type model as well. The dependent variable would be "yes" or "no" (or perhaps a gradient acceptability scale), and the predictor variables might include components of the semantics of the item and the other participants, the type of interaction involved (formal vs. informal, for example), the nature of the situation being described, etc.

Word-specific constraints could be seen as fixed effects in such a regression model. Arbitrary exceptions could be ruled out by assuming that language learners and speakers are not guilty of the "language as fixed effect fallacy" (Clark 1973; Raaijmakers 2003). That is, learners assume that anything they may discover about the constraints on the syntactic position in question will be generalizable to new "items" (i.e., words).

6.3.2.2 Non-Attentiveness

A somewhat less restrictive explanation for the lack of arbitrariness would involve constraints not on representation, but on learning. Of course, if there were no way even to represent information about individual words in individual constructions because of the architecture of the grammar, then non-attentiveness of the learner would follow as a consequence, but one need not assume that word-specific constraints are unrepresentable in order to imagine that the learner is non-attentive. Another theory that would derive non-arbitrariness as a consequence is that the learner is simply not attentive to the use of individual words in individual constructions. If the learner does not attend to this information, the information is not encoded and stored.

There are several arguments against this idea, i.e., in favor of learners' attentiveness to the use of individual words in individual constructions. One body of evidence comes from research in sentence comprehension, which has reliably found effects of verb bias on parsing. For example, in an offline forced-choice interpretation study on preposition attachment, Ford et al. (1982) found that the locus of attachment for the prepositional phrase *on the beach* in the sentence *The women discussed/kept the dogs on the beach* followed the subcategorization preference of the verb: The verb *keep*, which prefers locative complements, yields "low" attachment (treating the prepositional phrase as a complement), and the verb *discuss*, with the opposite subcategorization preference, yields "high" attachment (treating the prepositional phrase as a VP or sentential modifier).

In a well-known eye-tracking study, Trueswell et al. (1993) found syntactic misanalysis effects ("garden path effects") in sentential complements following verbs that typically take direct object complements, like *find*, but not following verbs that take sentential complements, like *claim*. Garnsey et al. (1997) found the same effect in an experiment manipulating direct object plausibility. Trueswell and Kim (1998) used priming to find the same effect: the syntactic bias (sentential complement vs. direct

object) of a prime word had a significant impact on the magnitude of garden path effects for following sentences.

Studies such as these support a model of processing such as that of Jurafsky (1996) or Manning (2003), where verb-specific subcategorization probabilities, which can be estimated from corpora and which are presumably derived through exposure, are mentally stored and deployed in parsing. These models are in tension with the central claim of this thesis ("no arbitrary exceptions"), because they do claim that subcategorization preferences are to some extent arbitrary, although these models do not stipulate arbitrary "exceptions" in the grammatical sense; there are no statements of the form, "Verb X is ungrammatical in Construction Y."

The results from sentence comprehension are not the strongest possible indications in favor of arbitrariness, however. Apparently verb-specific subcategorization probabilities could in principle be driven by semantic factors. Hare et al. (2003) and Hare et al. (2004) show that verb subcategorization preference effects in comprehension are conditioned by verb sense. For example, the verb *find* has a direct object preference in its 'locate' sense, and not in its 'realize' sense. When context is used to promote one or the other sense, temporary ambiguities are interpreted in a manner consistent with the sense-determined preference. This implies that if verb-specific subcategorization probabilities are stored, there is a parameter for each separate word sense, rather than for each word form. As Wonnacott et al. (2008:170) point out, "[Hare et al.'s] findings at least raise the possibility that structural preferences may be entirely driven by verb semantics." They continue: "The strong correlation between verb distribution and verb semantics, which hold in each natural language, make it impossible to determine whether verb biases are a result of the verb's own distributional history or of its membership in some more general semantic classes." This quotation eloquently captures why verb-bias effects in comprehension are not necessarily even in tension with the central claim of this thesis.The verb senses that prefer direct

object complements, such as the 'locate' sense of *find*, may share a semantic feature that is itself the cause of the parsing expection for a direct object. In that case, the human parser would not need to store a parameter for each individual sense of each individual verb, but only for each relevant semantic feature.

In order to eliminate the confound between distribution and semantic class, Wonnacott et al. (2008) use an artificial language, in which the distribution and semantics of each word can be controlled. This artificial language had an alternation between two constructions: *Verb Agent Patient* (VAP) and *Verb Patient Agent* followed by the "particle" *ka* (VPA_ka). There were three verb classes in this language: VAP-only, VPA_ka-only, and alternating. Participants witnessed the verbs in non-alternating classes only in their respective constructions, and the alternating verbs were shown in both. Each verb was assigned to an action, such as PUSH, STROKE, TICKLE, etc. and the assignment of verbs to actions was different for every participant. This counterbalancing method was used to ensure that the acquired subcategorization preferences would not be attributable to the verbs' semantics. After training, participants demonstrated sensitivity to these class distinctions in their grammaticality judgments, production, and on-line comprehension. These results strongly suggest that verb-specific constraints are learnable (and learned). This militates against an explanation for non-arbitrariness on the basis of inattentiveness on the part of the learner.[2]

6.3.2.3 The explanation-seeking learner

An even more modest explanation is possible. It is not necessary to assume that arbitrary exceptions cannot exist, or cannot be learned. An *explanation-seeking learner* would potentially be capable of memorizing arbitrary exceptions, but would avoid

[2]It must be kept in mind that the participants in these experiments were adults, however.

doing so if possible. According to this explanation, learners prefer explanatory generalizations to stipulative facts (just as linguists do). This idea could be called the "theory theory" for language development (see Gopnik and Meltzoff 1997 for an articulation and defense of the "theory theory" for child development more generally, and many additional references): Learners develop causal explanations for linguistic evidence, which help them to understand their linguistic environment. Under this view, the child collects evidence about the acceptability of sentences and attempts to develop a theory of which sentences are acceptable and which are not.[3,4]

Like scientific theories, a child's theory can make predictions, which can be contradicted by (positive or negative) evidence. When the predictions of the theory are violated, the child has the opportunity to revise the theory to make it more accurate. For example, imagine that a child has a theory of the causative alternation according to which transitive *break*, for example, is formed from intransitive *break*, via a productive causativization process that applies to any verb describing any eventuality that can be conceptualized as being caused. This predicts, erroneously, that *fall* should be usable as a transitive verb. Corrective feedback would provide an opportunity for revision of this theory. For example, suppose a child utters, *Don't fall me down!* and hears *Don't worry, I won't drop you* in response from his or her caretaker. The child has received a bit of negative evidence, and can infer that for some reason, *drop* is more appropriate than *fall* in this situation. That the evidence tells the learner a fact about a specific word does not imply that the learner must posit a word-specific explanation.

What could this reason be? To account for this piece of linguistic evidence, a

[3]Such a theory could be understood either a theory of the language itself, or a theory of the caregiver's linguistic competence. See Marcotte (2005) for further discussion of this issue.

[4]The reader might be reminded of Chomsky's (1965) "Language Acquisition Device," but the idea here is not the same. Here, "hypotheses" take the form of constraints on syntactic positions, whereas under Chomsky's view, a "hypothesis" corresponds to a grammar. More importantly, Chomsky makes very specific assumptions about the nature of grammars and the innateness of linguistic knowledge, which are unnecessary here.

variety of explanations could be given. One explanation lies in the meaning of *fall*: perhaps it is not really a verb that can be conceptualized as being caused. Alternatively, the child might posit that there is a constraint on the use of lexical causatives that the verb *fall* violates – under this scenario, the meaning is right, but the constraints are wrong. An arbitrary lexical stipulation is another viable explanation for this particular data point. In particular, perhaps *drop* is a suppletive form for causative *fall*.

This learner is not necessarily attentive: this learner may not encode and store every aspect of the data, but only the potentially explanatory aspects, e.g., *fall* started with the sound /f/, it was a short word, it had to do with downward motion, etc.. In other words, attention and encoding could be affected by the type of explanations the child is currently capable of making.

Recent support for the idea that learners are not *always* attentive in language learning comes from another result of Wonnacott et al. (2008). Item-specific learning was not found across the board. Recall that their participants were taught an artificial language with an alternation between two constructions: Verb Agent Patient (VAP) and Verb Patient Agent followed by the "particle" *ka* (VPA_*ka*). All of the participants were taught a version of the language in which some verbs alternated between the two constructions and some were limited in their distribution to one construction or the other, but the proportion of verbs in the lexicon of the artificial language differed across participants. For some participants, there were four alternating verbs, four VAP-only verbs, and four VAP_*ka*-only verbs (12 in total). For other participants, there were *eight* alternating verbs (out of 12 in total), and 2 VAP-only verbs and 2 VAP_*ka*-only verbs. Only in the former condition (with a 4:4:4 ratio of verb types) did the participants acquire word-specific constraints; in the latter condition (with an 8:2:2 ratio), participants did not acquire verb-specific constraints. Thus, whether or not verb specific constraints are acquired depends on the overall

frequency distribution of verb types in the language. Granted, the participants in this experiment were adults rather than children, but it is possible that the same effect may emerge with child participants as well.

A possible interpretation of this result is that learners need to become sensitized to the possibility of non-alternating verbs in order to learn that a given verb does not alternate. Sensitivity to such a feature would allow learners to generate hypotheses about the alternating or non-alternating status of a given verb, which can in turn generate predictions that can be falsified. If a learner hypothesizes that such-and-such a verb is an alternating verb, evidence against this hypothesis could come from the relative rarity of one construction. If no such hypothesis is ever developed, then no such predictions can be generated and falsified – failure to generate such predictions may be what occurs in the case where non-alternating verbs are rare.

The idea that learning, as theory development, depends on the prior establishment of an earlier theory is put forth for child development more generally by Karmiloff-Smith and Inhelder (1974). In their experiments, children played with blocks and learned strategies for building towers. Based on their observations, they concluded (op. cit.: 203–204):

> Frequent counterexamples do not alone induce a change in the child's behavior. If they did, then progress could be achieved by simply providing a large number of counterexamples. The child must first form a unifying rule based on regular patterns he has observed .... Only when this theory is really consolidated and generalized, is he ready to recognize some form of unifying principle for the counterexamples which he earlier rejected as mere exceptions.

The fact that negative evidence is sometimes ignored by children fits nicely into this framework. Children may not always be ready to absorb contradictory evidence and revise their explanations, as shown by the following dialogue from Braine (1971):

(4) Child: Want other one spoon, Daddy.
Father: You mean, you want *the other spoon*.
Child: Yes, I want other one spoon, please, Daddy.
Father: Can you say 'the other spoon'?
Child: Other ... one ... spoon.
Father: Say ... 'other'.
Child: Other
Father: Spoon
Child: Spoon
Father: Other ... spoon
Child: Other ... spoon. Now give me the other one spoon.

This dialogue has been taken to show that learners simply do not make use of negative evidence, but a less radical interpretation is possible. The dialogue in (4) can be taken as support for the view that evidence is not automatically and mechanically absorbed, but rather is taken into account only when old explanations are questioned and new explanations are sought.

Moreover, children may develop different explanations for the same data. This partially explains the finding by Hudson Kam and Newport (2005) that children regularize inconsistent input in different ways. Whereas adults in their experiments produced determiners alongside nouns in an artificial language at a rate proportional to the rate at which nouns were presented with determiners, children tended to regularize the artificial language, producing determiners either completely consistently or never. One child participant found an even more creative strategy, marking nouns in transitive sentences but not intransitive sentences (Hudson Kam and Newport 2005:182). This result can be understood under the view that language learning is not blind number-crunching, but a process of theory formation and revision.

In the realm of scientific theory formation, there is evidence that children prefer explanations that are not *ad hoc*. In an investigation of the kinds of explanations that children seek, Samarapungavan (1992) presented children with pairs of alternative theories and asked them which of the two theories they preferred and why. One

experiment asked children to choose between theories in the domain of chemistry. In this experiment, there were five buckets: two with blue liquid labelled "Cold," two with red liquid labelled "Hot," and one with clear liquid labelled "Cold." The buckets labelled "Cold" (two blue and one clear) were alkaline, and litmus paper turned blue when dipped in them. The two red buckets, both labelled "Hot," were acidic and turned litmus paper red. Children were asked to choose between two theories of why the paper turned color: (i) the paper turns the color of the liquid in the bucket, except sometimes the liquid gets old (the *ad hoc* theory); (ii) the paper turns color based on temperature (the non-*ad hoc* theory). Thus, both accounted for the "data" accurately but one required an *ad hoc* stipulation to account for all of the evidence. Samarapungavan found that children prefer the non-*ad hoc* theory to the *ad hoc* theory across several domains, although this preference was not found at all age levels: Third and fifth graders dispreferred *ad hoc* theories, although first graders did not. This result suggests that, in general, humans prefer explanatory generalizations to arbitrary stipulations, although this preference may take some time to develop.

Finally, the idea that humans possess an innate drive to find explanations is found frequently in writing in philosophy and psychology, perhaps beginning with Thomas Hobbes (*Leviathan*, Part I, Chapter 6, 1651):

> Desire, to know why, and how, CURIOSITY; such as is in no living creature but Man; so that Man is distinguished, not only by his Reason; but also by this singular Passion from other Animals; in whom the appetite of food, and other pleasures of Sense, by predominance, take away the care of knowing causes; which is a Lust of the mind, that by a perseverance of delight in the continual and indefatigable generation of Knowledge, exceedeth the short vehemence of any carnal Pleasure.

The drive for explanation is ubiquitous, as shown by the fact that explanations are found in every human culture, regardless of how scientifically "advanced" it is (Sperber et al. 1995). Gopnik (1998, 2000) even argues that "explanation is to theory

formation as orgasm is to reproduction: the phenomenological mark of the fulfillment of an evolutionarily determined drive" (Gopnik 2000:300).

These findings and observations support the idea that there is a human drive to find explanations that are not *ad hoc*. This drive may play a role in language acquisition as well, which would help to explain why arbitrary exceptions should be so difficult to find in language. Under this view, language learners do not blindly and passively absorb data, but implicitly develop, test, and refine linguistic theories.

6.4 In a nutshell

Where does all this leave Baker's Paradox? As I argued in the first chapter, Baker's Paradox is a paradox in the sense that one cannot maintain all of the premises simultaneously, but it is not a paradox in the stronger sense that all of these premises appear to be true. The "No Negative Evidence" premise can easily be denied. The interesting question is, therefore, not how to solve the paradox, but how many of its premises are true, and what follows from those conclusions about how restrictions on productivity are learned. My principal claim is that the "Arbitrariness" premise is not well-founded, and I have laid out a possible view of learning that explains why this might be.

In a nutshell, my conclusion is this: For the scientist interested in language and language development, it is a fruitful strategy to seek general explanations for apparently idiosyncratic facts of language, as such explanations can be found. I propose that this orderliness comes about because the language learner, likewise, seeks to develop models of linguistic acceptability with general explanatory power.