Referential garden path effects in modified Haddock descriptions

(word count: 12,227 words)

Helena Aparicio, Roger Levy, and Elizabeth Coppock

Abstract

As Haddock (1987) points out, *the rabbit in the hat* is perfectly felicitous in a scenario with multiple hats. The uniqueness requirement of the lower definite in these *nested (definite) descriptions* pertains to rich descriptive content: not *hat* (absolute reading), but *rabbit-containing hat* (relative reading). **The incrementalist view**: For Haddock, the loosening of the uniqueness requirement is due to an incremental procedure by which constraints are successively imposed on discourse referents. Haddock envisions a constraint satisfaction problem that can be formalized as a set of open formulas like $\mathsf{rabbit}(x)$, $\mathsf{in}(x, y)$, and $\mathsf{hat}(y)$, applied in sequence. The uniqueness requirement of the inner definite is satisfied if there is only one satisfier of y left, or in other words, if the *candidate set* has only one element. **The scope-based view**: Like Haddock, Bumford (2017) assumes that *the* imposes a cardinality-one check on the candidate set for the relevant discourse referent, but it is scope that determines what constraints have applied prior to this check. Both views predict relative-like readings for nested definite descriptions containing gradable modifiers like *the rabbit in the big/bigger bag*, but only the scope view predicts absolute readings. In this paper, we give experimental evidence for the scope-based approach, by showing that participants access absolute interpretations. The evidence comes from **referential garden paths**, where a dynamic constraint evaluation process temporarily settles on the wrong referent before eventually failing. It is exactly when the scope-based theory predicts referential garden paths—including absolute readings—that we observe penalties.

1 Introduction

Definite descriptions such as *the hat* are generally taken to presuppose uniqueness with respect to the descriptive content (*hat*), in some sense. If there is more than one hat around, then *the hat* is infelicitous. Whatever this uniqueness requirement amounts to exactly, it seems to be lifted when the description is syntactically embedded inside another DP, as in *the rabbit in the hat*. As Haddock (1987) points out, such a description is perfectly felicitous in a scenario with multiple hats such as the one pictured in Figure 1, taken from Haddock's paper. We refer to descriptions in which one definite is embedded in another as *nested (definite) descriptions*. Speaking pre-theoretically, the uniqueness requirement of the lower definite seems to be calculated over descriptive content that is richer than the plain descriptive content of the inner definite: *the rabbit in the hat* requires not that there be a unique hat, but a unique *rabbit-containing* hat. We describe cases in which the

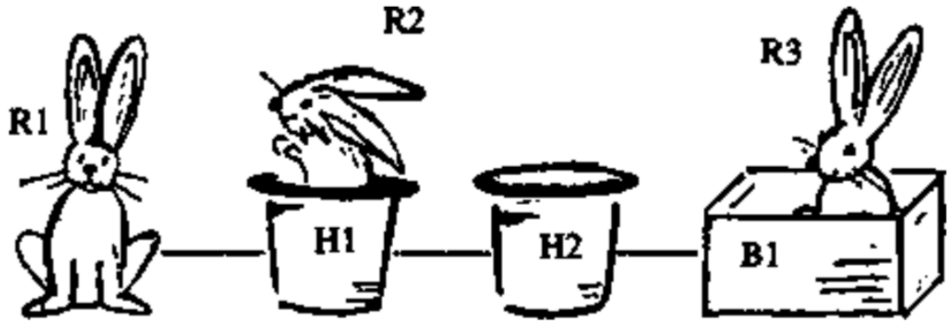


Figure 1: Haddock's scenario with rabbits, hats and a box.

uniqueness check of the inner definite in nested definite description is relativized to descriptive content from the higher NP as *Haddock readings*.

According to Haddock, the reason the uniqueness requirement is loosened is that the uniqueness check for the inner definite article applies to a discourse referent that has been constrained not only by the descriptive content of the inner nominal, but also by some descriptive content contributed by the outer nominal, thanks to an incremental processing procedure by which constraints are successively imposed on discourse referents. Haddock envisions a constraint satisfaction problem that can be formalized as a set of open formulas with free variables, like $\mathsf{rabbit}(x)$, $\mathsf{in}(x, y)$, and $\mathsf{hat}(y)$. The uniqueness requirement of the inner definite article is satisfied if there is only one possible satisfier of y left after previous constraints have applied, or in other words, if the so-called *candidate set* has only one element. Borrowing Haddock's own terminology, we label this view an *incrementalist* one.

An alternative view, which we refer to as *scope-based*, is provided by Bumford (2017). According to Bumford, the loosening of the uniqueness constraint is due to the possibility of the definite article taking higher-than-surface scope. Like Haddock, Bumford assumes that the uniqueness requirement for a definite determiner is sensitive to the size of the candidate set for a discourse referent (requiring it to be a singleton), but it is scope rather than word order that determines what constraints have applied before the definite article's uniqueness check occurs.

In this paper, we give experimental evidence for the scope-based approach using nested definite descriptions containing gradable modifiers, like *the rabbit in the big/bigger bag*. Contrary to the predictions of the strong incrementalist hypothesis, modifiers in the inner nominal of a nested description are not always constrained by the information that has been placed on the relevant discourse referent by the linearly preceding words. In particular, in a nested description like *the rabbit in the big bag*, the positive adjective *big* can be interpreted relative to a comparison class that is not restricted to bags that have rabbits in them.

Consider a scenario with three bags of different sizes, where the largest of the three contains a frog and the smaller two contain rabbits. Clearly, *the big bag* picks out the bag with the frog in it. Yet *the rabbit in the big bag* clearly picks out the rabbit in the *medium* bag. Hence an absolute reading is not available for the nested description in this scenario. Both Bumford's and Haddock's theories converge on the correct referent for both expressions, as we discuss in more detail in the next section. But the two theories make available different sets of non-viable interpretations. Essentially, Haddock predicts only a relative reading, where the adjective and the definite determiner are evaluated relative to a more constrained comparison class (rabbit-containing bags), a reading that succeeds. Bumford predicts, in addition, an absolute reading for the modifier (*big*), on which it is interpreted relative to the content corresponding to its complement noun (bags). If the threshold for *big* is set too high, so that only the biggest bag counts as *big*, then the lower noun phrase will refer to the bag with the frog in it, and the nested description as a whole will fail to refer because it doesn't contain a rabbit.

In some cases, as we will show, non-viable interpretations are just ignored and make no difference to the process of identifying a referent. But in other types of cases, we can observe what we call a REFERENTIAL GARDEN PATH effect, where the interpretation converges on a certain referent before ultimately crashing, over the course of dynamic evaluation of a constraint sequence. Using partially-masked auditory stimuli that can be resolved in one of two ways given a visual scene, we show that listeners disprefer resolutions associated with interpretations that yield a referential garden path effect.

2 Theoretical predictions

In this section, we spell out our research question and the possible answers to it in detail. We begin by explicating the core contrast between Haddock's and Bumford's theories of Haddock descriptions in section 2.1. In section 2.2, we present ways of extending the two theories to handle positive and comparative adjectives that we take to be in the spirit of the original proposals. We will see that the two extended theories make different predictions, which we will test experimentally in section 3.

2.1 Solutions to the Haddock Puzzle

In the literature on Haddock descriptions (Haddock 1987; van Eijck 1993; Champollion and Sauerland 2010; Bumford 2017; Grudzińska and Zawadowski 2019), a common idea keeps recurring: that the interpretation of a noun phrase involves a sequential application of constraints on variables in a way that can be modeled in a dynamic framework. For Haddock (1987), the ordering of the constraints is determined for the most part by the order of the words, in concert with a shift-reduce parsing algorithm by which the sentence is processed incrementally from left to right. Bumford's (2017) account, on the other hand, is cast in a compositional dynamic semantics in which the constraint ordering is determined not by word order, but rather by scope. In what follows, we characterize the constraint sequences involved under both theories.

Let us begin with Haddock, for whom, again, the sequence is determined for the most part by the order of the words, with the exception that the definite article's uniqueness requirement is applied *after* its complement NP's constraints are incorporated. Each constraint in the sequence has the potential to narrow down on the set of possible values for variables. Setting aside certain details about how variables are unified with each other, the sequence of constraints involved in the interpretation of *the rabbit in the hat* runs as follows, where open logical formulas like $\mathsf{rabbit}(x)$ express constraints. Relative to the scenario depicted in 1, the possible values for x and y are updated at each step as indicated below each constraint.

(1)	Step	1	2	3	4	5
		$\mathsf{rabbit}(x)$	$\mathsf{in}(x,y)$	$\mathsf{hat}(y)$	$\mathsf{unique}(y)$	$\mathsf{unique}(x)$
	x	R1, R2, R3	R2, R3	R2	R2	R2
	y	(any)	H1, B1	H1	H1	H1

As reflected in this example, the application of the uniqueness constraint contributed by each definite determiner is delayed until the corresponding NP is "syntactically closed" (Haddock 1987: 662). The uniqueness requirement of the definite is met if, at the time of application, there is only one remaining possible value for the indicated variable.

Although Haddock does not provide an explicit semantics for his formal language, it is clear how the process he describes could be implemented in a dynamic semantic framework where contexts determine a set of possible variable assignments, and formulas determine updates to such contexts. To do so, we use a version of dynamic semantics in which states are sets of assignments, and formulas determine updates on states so construed.[1] In the language we will define, the following formula captures the sequence of constraints envisioned by Haddock:[2]

(2) $\mathsf{rabbit}(x) + \mathsf{in}(x,y) + \mathsf{hat}(y) + \mathsf{uniq}(y) + \mathsf{uniq}(x)$

We define the semantics of formulas like this in a variant of Dynamic Predicate Logic (DPL; Groenendijk and Stokhof 1989) in the update semantics style presented in Groenendijk and Stokhof 1990. The formal language is based on the language of first-order predicate logic, and the semantic value of a formula, given a model, is an update to a given input state. More precisely:

- A model $M = \langle D, I\rangle$ is an ordinary first-order model consisting of a domain of individuals D and an interpretation function I, mapping each non-logical constant (name, predicate, or relation) to its extension in the model.
- An assignment g is a function from variables to elements of D.
- For any term t (individual-denoting variable or constant), $|t|^{M,g} = \begin{cases} g(t) & \text{if } t \text{ is a variable} \\ I(t) & \text{if } t \text{ is a constant} \end{cases}$
- A state σ is a set of assignments.

For a formula ϕ, we write $\sigma[\![\phi]\!]^M$ to denote the result of updating state σ with ϕ with respect to model M. The superscript M is dropped for readability. The nature of the update depends on the shape of ϕ. As usual in DPL, if ϕ is an atomic formula consisting of a predicate π applied to a sequence of terms $t_1, ...t_n$, then the update keeps assignments that map the terms to tuples in the denotation of π:

- For any predicate π, $\sigma[\![\pi(t_1,...,t_n)]\!] = \sigma \cap \{g \mid \langle |t_1|^g, ...|t_n|^g\rangle \in I(\pi)\}$

[1] As we will see below, it is is important that states are sets of assignments rather than single assignments because the uniq predicate imposes a constraint on the full set of assignments under consideration. For this reason, we find it curious that van Eijck (1993), though he explicitly discussed Haddock descriptions in a dynamic framework, did not opt for this approach.

[2] We are glossing over the issue of whether x and y are 'fresh' variables, although presumably, they must be, and Bumford's (2017) analysis ensures this via the existential component of the definite article. As we are dealing with referring expressions in isolation, none of the data we consider here is impacted by this simplification.

We distinguish between two different types of conjunction: dynamic (or 'sequential') and simultaneous (or 'parallel'). Dynamic conjunction, in which the left conjunct applies before the right conjunct, is written as $+$. For simultaneous update, we use the standard conjunction symbol $\wedge$:

- $\sigma[\![[\phi + \psi]]\!] = \sigma[\![\phi]\!][\![\psi]\!]$
- $\sigma[\![[\phi \wedge \psi]]\!] = \sigma[\![\phi]\!] \cap \sigma[\![\psi]\!]$

We will often omit brackets; please read $+$ as left-associative. With simultaneous updates, both formulas update the initial context directly, and the results are merged by intersection. It follows that the order doesn't matter; $[\phi \wedge \psi]$ determines the same update as $[\psi \wedge \phi]$.[3]

Haddock's 'unique' predicate can only be evaluated relative to the full set of assignments under consideration, because unique expresses a "meta-constraint" (p. 662); in other words, it is a predicate that is evaluated over the whole set of entities to which the relevant variable can be assigned. In dynamic semantics terms, the same idea can be expressed by saying that the update is not 'distributive', as it does not apply pointwise to each of the assignments under consideration, but to the whole set of assignments being considered at once. With this in mind, the semantics of Haddock's unique can be defined as follows (dropping the final two letters for typographical reasons).[4]

- For any variable v,
$\sigma[\![\mathsf{uniq}(v)]\!] = \begin{cases} \sigma & \text{if there is exactly one } k \text{ for which there is a } g \in \sigma \text{ s.t. } g(v) = k \\ \emptyset & \text{otherwise} \end{cases}$

With these definitions, it can be shown that each conjunct of 2 corresponds to a step in the sequence of updates envisioned by Haddock; the set of possible referents for x and y will be narrowed down in exactly the manner he describes.

The Bumford (2017) analysis of Haddock descriptions also involves sequential application of constraints. Bumford's theory is explicitly formalized in a (compositional) dynamic semantic framework in which meanings can update the set of constraints currently imposed on a given variable. The two theories share the central insight that Haddock readings obtain when the uniqueness check of the inner definite applies after semantic constraints introduced by descriptive material in the outer description. For Bumford, however, the (virtual, non-temporal) delay in the application of the uniqueness check is achieved by allowing the (uniqueness component of the) definite article to take wider-than-surface scope.

Bumford assumes that the meaning of a definite determiner is split between an existential component, which is always interpreted *in situ*, and a uniqueness check, which can either be enforced

[3] To complete the language, we can make the following assumptions:

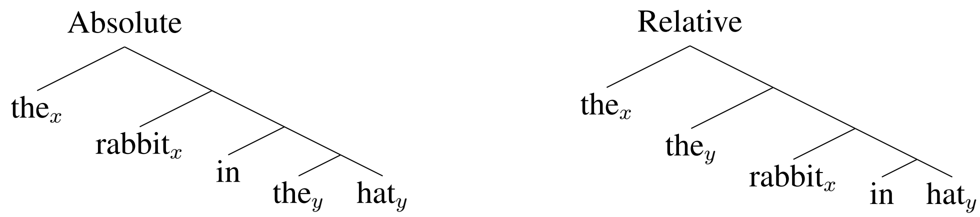
- $\sigma[\![t_1 = t_2]\!] = \sigma \cap \{g \mid [\![t_1]\!]^g = [\![t_2]\!]^g\}$
- $\sigma[\![\neg\phi]\!] = \sigma - \downarrow \phi$,
where $\downarrow \phi = \{g \mid \{g\}[\![\phi]\!] \neq \emptyset\}$
- $\sigma[\![\exists x \phi]\!] = \sigma[\![\approx_x]\!][\![\phi]\!]$
where $\sigma[\![\approx_x]\!] = \bigcup_{g \in \sigma}\{h \mid g \approx_h x\}$

For any two assignments g and h: $g \approx_x h$ means that g and h differ at most with respect to the value they assign to x.

[4] Cf. the **1**-operator of Bumford (2017), which is quite similar in meaning to this uniq.

in situ or higher up in the structure of the derivation. The former results in an ordinary interpretation of the description (see structure on the left in 3), as it presupposes that there is exactly one hat, while the latter results in a Haddock interpretation (see structure on the right), as it only presupposes that there is exactly one hat *with a rabbit in it*. The reading on which the definite article has high scope – on which the uniqueness constraint is delayed – can also be called a 'relative' reading, and the ordinary reading can be called an 'absolute' reading, because on Bumford's theory, the ambiguity between these two readings of the definite article parallels the ambiguity that superlatives exhibit between absolute and relative readings, as we will discuss in the next subsection.

(3)



These structures yield different constraint sequences.[5] Bumford's analysis does not imply a total ordering on the constraints involved, so to represent his theory faithfully we use both simultanous conjunction ($\wedge$), which is not order-sensitive, and dynamic conjunction ($+$), which is sensitive to order. The absolute reading involves the sequence of constraints in 4; the relative reading involves the sequence in 5:

(4) $[[\mathsf{hat}(y) + \mathsf{uniq}(y)] \wedge \mathsf{in}(x, y) \wedge \mathsf{rabbit}(x)] + \mathsf{uniq}(x)$ (Absolute)

(5) $[\mathsf{hat}(y) \wedge \mathsf{in}(x, y) \wedge \mathsf{rabbit}(x)] + \mathsf{uniq}(y) + \mathsf{uniq}(x)$ (Relative)

In both cases, each definite article imposes a uniqueness requirement vis-a-vis a given discourse referent relative to all of the constraints that are placed on it within the syntactic scope of the definite article at LF. Notice that in the case of the absolute reading, by the time $\mathsf{uniq}(y)$ applies, the only constraint that has been placed on y is that it is a hat. Hence there must be no more than one hat total. On the relative reading, uniqueness is checked relative to a richer set of constraints, so in this case there may be more than one hat, as long as there is only one rabbit-containing hat.

Both absolute and relative interpretations are available on the scope-based view, then, whereas Haddock's analysis does not predict an ambiguity, and only sanctions a relative-like reading, where uniqueness is checked relative to richer descriptive content. The debate between these two views, then, boils down to the question of whether absolute readings are available for nested descriptions (in addition to relative readings). This paper argues that of these two views, the scope-based view is superior, because we do find evidence for absolute readings. Our data comes from modified Haddock descriptions like *the rabbit in the big bag*. The presence of absolute readings can be seen through 'referential garden path' effects: a phenomenon where the dynamic constraint evaluation process temporarily settles on a unique referent before crashing. As we will show, it is exactly when the scope-based theory predicts referential garden paths that we observe penalties.

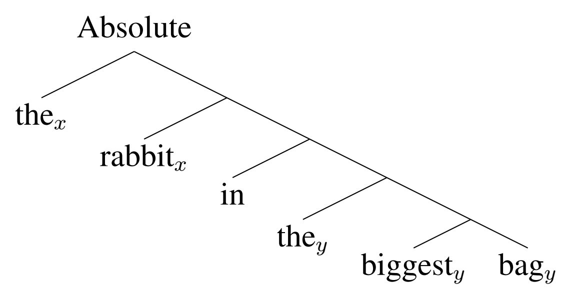
[5] See Bumford (2017) for the full picture; here we are glossing over many details of Bumford's analysis, in order to bring out the contrast of interest between his account and Haddock's.

2.2 Bringing modifiers into the picture

To spell out the predictions that the scope theory makes with respect to modified Haddock descriptions, we begin with cases involving superlatives, which Bumford (2017) specifically addressed. Bumford uses the same tools to account for both Haddock descriptions and so-called 'relative readings' of superlatives, building on the insight that both involve a computation over richer semantic content. For example, on a relative interpretation, *George is wearing the biggest hat* says that George is wearing the hat that is bigger than any hat *worn by any salient alternative to George*, not that George is wearing the hat that's bigger than any other hat, *simpliciter*; the latter would be the absolute interpretation. Similarly, *the rabbit in the biggest hat* on a relative interpretation picks out the rabbit that is in a bigger hat than any other hat *that a rabbit is in*. In this section, we will review how Bumford's account works with superlatives, and then extend it to the types of modifiers that we use in our experimental materials.

A nested definite description like *the rabbit in the biggest bag* can have (at least) two interpretations: an absolute interpretation, in which both the definite determiner and the superlative take scope in their surface positions, and a relative interpretation, in which both the definite determiner and the superlative adjective have high scope.[6]

(6)



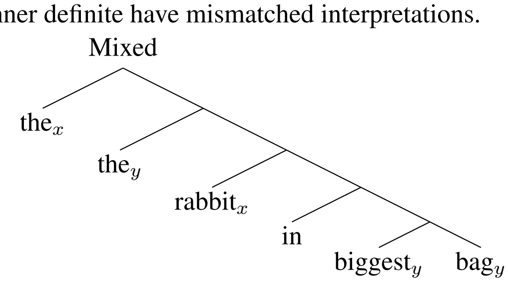
Given the scenario in Figure 2, the relative reading would pick out the rabbit in the middle bag. The absolute reading would fail to refer, because the container that one would describe as the biggest bag has no rabbit in it.

Let us see how this works in more detail. Loosely following Bumford (2017), we adopt the following denotation for superlatives:

(7) For any variable v,
$\sigma[\![\textsf{sup}(v, A)]\!] = \{g \in \sigma \mid \text{for all } h \in \sigma : h(v) = g(v) \text{ or } g(v) >_A h(v)\}$

[6] A third possible reading is represented by the pseudo LF in (i), a 'mixed reading', because the adjectival modifier and the inner definite have mismatched interpretations.

(i)



Bumford did not discuss mixed readings, but as far as we can see, they are compatible with his approach, at least in spirit. See Section 4.6.2 for further discussion of mixed readings.

Figure 2: A scenario supporting a relative (or Haddock) reading of the complex DP *the rabbit in the biggest bag* but no absolute reading for it.

Here A is a gradable predicate; if $x >_A y$ then the degree to which x is A is greater than the degree to which y is. More formally, we assume that the interpretation function I maps gradable predicates A to individual-degree pairs, so $I(A)$ is a set of pairs $\langle k, d\rangle$ such that individual k has adjectival property A to degree d. Then $x >_A y$ can be defined as: 'the greatest degree d such that $\langle x, d\rangle \in I(A)$ exceeds the greatest degree d such that $\langle y, d\rangle \in I(A)$'. Thus superlatives filter out assignments to v that are not maximal with respect to A. This treatment implements Bumford's idea that the comparison class for the superlative is the set of possible assignment-values for the relevant discourse referent. Note that we are not attempting a compositional semantics here; our goal is only to illustrate the sequence of constraints involved. Bumford's work provides an example of how a compositional semantics for these kinds of dynamic meanings could be defined.

The sequence of constraints is determined by scope, moving from most embedded to least embedded. According to Bumford (2017), the absolute and relative readings involve the following sequences of constraints, respectively:

(8) $[[\mathsf{bag}(y) + \mathsf{sup}(y, \mathsf{big}) + \mathsf{uniq}(y)] \wedge \mathsf{in}(x, y) \wedge \mathsf{rabbit}(x)] + \mathsf{uniq}(x)$ (Absolute)

(9) $[\mathsf{bag}(y) \wedge \mathsf{in}(x, y) \wedge \mathsf{rabbit}(x)] + \mathsf{sup}(y, \mathsf{big}) + \mathsf{uniq}(y) + \mathsf{uniq}(x)$ (Relative)

The absolute reading finds the biggest bag, and then the unique rabbit in it. The relative reading looks for bags with rabbits in them, finds the biggest one, and checks for uniqueness of both the bag and the rabbit at this point.

Observe that the superlative modifier is applied after the head noun in both cases, even with the absolute reading. Assuming that a superlative modifier is applied after the noun it modifies accounts for the fact that *the biggest bag* is biggest among bags, and not biggest in general—a fact that Heim 1999 captured by assuming an otherwise-unmotivated LF movement within the nominal. Throughout this paper, we assume that modifiers are always applied after the noun they modify, in accordance with the the *Head Primacy Principle* of Kamp and Partee (1995: 161): "In a modifier-head structure, the head is interpreted relative to the context of the whole constituent, and the modifier is interpreted relative to the local context created from the former context by the interpretation of the head." The kind of example Kamp and Partee use to motivate their Head Primacy Principle is the contrast between *giant miniature* and *miniature giant*; the first is large for a miniature; the latter is small for a giant. As we have just seen, it has welcome consequences for superlatives as well. It also conveniently cuts in half the number of derivations to consider, and does not affect our qualitative predictions.

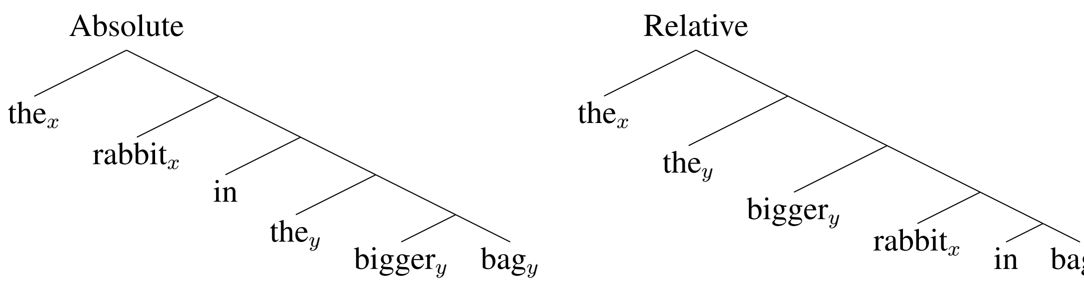
Positive and comparative modifiers will be our focus here. As we argue in detail in Appendix A, comparatives exhibit the same kind of relative readings that superlatives have. For a brief illustration, observe that relative to the three-bag scenario depicted in Figure 2, *the bigger bag* on its own is infelicitous, having no identifiable referent. Yet *the rabbit in the bigger bag* is perfectly felicitous, and it picks out the rabbit in the middle bag. This observation is a reflection of the fact that there is no felicitous absolute reading for the *the rabbit in the bigger bag* in this context, but there is a relative reading. This can be accounted for by treating comparatives analogously to superlatives:

(10) For any variable v,
$\sigma[\![\mathsf{cmpr}(v, A)]\!] = \{g \,|\, \text{there is an } h \in \sigma : g(v) >_A h(v)\}$

This lexical entry says that a comparative like *bigger*, interpreted relative to a discourse referent v, acts as a filter that keeps only the assignments to v for which there is another assignment in the current state that maps v to something smaller. In a definite comparative like *the bigger bag*, the uniqueness requirement will join forces with the semantics of the comparative to ensure that there are exactly two candidates remaining for the variable in question. This correctly predicts that *the bigger bag* is most felicitous in contexts where there are exactly two bags.

As with superlatives, Bumford's account predicts both absolute and relative readings for comparatives, while Haddock's predicts only a relative-like reading. Schematically, they look as follows:

(11)



The sequence of constraints corresponding to the relative reading is as follows. Below each constraint we depict the set of possible referents remaining for the variable y, relative to the scenario in Figure 2, after the constraint has applied (B1 = the smallest bag; B2 = the middle bag; B3 = the biggest bag):

(12)

	$\mathsf{bag}(y)$	+	$\mathsf{in}(x,y) \wedge \mathsf{rabbit}(x)$	+	$\mathsf{cmpr}(y, \mathsf{big})$	+	$\mathsf{uniq}(y)$	+	$\mathsf{uniq}(x)$
y:	B1, B2, B3		B1, B2		B2		B2		B2

By the time the comparative comes along, the biggest bag has already been ruled out from consideration. The smallest bag is then ruled out by the comparative, as it is not *bigger* among the candidates. The uniqueness test for y succeeds, then, and the reader may verify that the uniqueness test for x succeeds as well (the candidates for x are not shown). The middle bag (B2) is the sole candidate for y left by the time all of the constraints have applied, and the sole candidate left for x is the rabbit in the middle bag, in accordance with clear native speaker intuitions about what *the rabbit in the bigger bag* refers to.

The sequence of constraints corresponding to the absolute reading, on the other hand, does not converge on a referent for y (or x):

(13)

	$\mathsf{bag}(y)$	+	$\mathsf{cmpr}(y, \mathsf{big})$	+	$\mathsf{uniq}(y)$	+	$\mathsf{in}(x,y) \wedge \mathsf{rabbit}(x)$	+	$\mathsf{uniq}(x)$
y:	B1, B2, B3		B2, B3		$\emptyset$		$\emptyset$		$\emptyset$

We use the empty set symbol ($\emptyset$) to signify that there are no candidates left. The problem is that when the uniqueness constraint applies, there are still two candidates left (the two that count as *bigger*), so the uniqueness constraint fails, and the derivation never recovers. Thus, under the assumptions we have made, relative to the scenario in Figure 2, there is no successful absolute reading for the comparative, although there is a successful relative reading for the comparative.

Let us turn now to positive-form gradable adjectives like *big*. For positive-form gradable adjectives, we assume that the context provides a threshold $\theta(A)$ for each gradable adjective A.

(14) For any variable v,
$\sigma[\![A(v)]\!]^{\theta} = \{g \in \sigma \mid \text{there is a } d \geqslant \theta(A) :\ \langle |v|^{g}, d\rangle \in I(A)\}$
if θ is a non-vacuous threshold for A relative to v in σ; $\emptyset$ otherwise

Informally speaking, θ is a 'non-vacuous threshold' for A relative to v in σ if θ distinguishes among members of the candidate set so that some count as A and some do not. Hence, for the comparison class in Figure 2, two thresholds are possible: one separating the smallest from the two larger bags, and one separating the small and the medium bags from the largest. We call this constraint on thresholds the 'Non-vacuity Principle', borrowing Kamp and Partee's (1995) label for the same idea. The Non-vacuity Principle is defined by Kamp and Partee (1995: 161) as follows: "In any given context, try to interpret any predicate so that both its positive and negative extension are non-empty." This principle can be implemented as a requirement that the threshold for a gradable adjective be set in a way that it discriminates among the current candidates for the discourse referent in question. We define non-vacuity formally as follows:

(15) Non-vacuity
θ is a non-vacuous threshold for A relative to v in σ if and only if there exist g and g' in σ such that $g \in \sigma[\![A(v)]\!]^{\theta}$ and $g'(v) \notin \sigma[\![A(v)]\!]^{\theta}$.

Hence the semantics of the gradable adjective is sensitive to the makeup of the current candidate set. In that sense, the current candidate set constitutes the 'comparison class' for the gradable adjective.[7]

Now, moving on to the description *the rabbit in the big bag*, let us consider the scenario in Figure 2. Here, two comparison classes are possible: (i) the set of rabbit-containing bags, so that only the lower threshold is possible; or (ii) the set of all bags, so that either the lower or the upper threshold could be in play. Which comparison class we have depends on the order in which the constraints apply.

[7] The reader may worry that in conjunction with the Head Primacy Principle, the Head Non-vacuity Principle makes overly strict predictions about how modifiers can function, requiring them always to be used informatively, in the strict sense that they eliminate candidate referents. While listeners generally expect adjectival modifiers to eliminate candidate referents and be informative in that sense (Sedivy et al. 1999), there is evidence that modifiers can be used non-eliminatively, as in *the big blue pin* in reference to the only blue pin Degen et al. (2019). We suggest that such "over-informative" uses of adjectives involve non-restrictive modification (cf. Sproat and Shih's (1988) distinction between direct and indirect modification), and that non-restrictive modification involves simultaneous (non-sequential) update between the head and the modifier. Thus, in *Whoa! Look at that big bear!* regarding a bear that is not particularly big for a bear, but still somehow big, we have a non-restrictive use of *big*, and simultaneous update with both *big* and *bear*, so that the comparison class is *not* restricted to bears by the time *big* applies. A restrictive use involves update first with *bear*, and then with *big*, so that the comparison class is restricted to bears. In both cases, the adjective *big* applies non-vacuously to the current comparison class, so the Non-vacuity constraint is satisfied, even if the Head Primacy Principle is not. We take restrictive modification to be an unmarked, default choice, so the Head Primacy Principle is typically, but not always respected.

Let us begin with relative readings. Assume that the smallest bag in Figure 2 is size 1, the medium bag is size 2, and the largest bag is size 3. There is a relative reading that succeeds, so long as $\theta = 2$. If $\theta = 3$, then the relative reading fails.

(16)
$$\begin{array}{lcccccc} & \mathsf{bag}(y) & + \ \mathsf{in}(x,y) \wedge \mathsf{rabbit}(x) & + \ \mathsf{big}_{\theta=2}(y) & + \ \mathsf{uniq}(y) & + \ \mathsf{uniq}(x) & \checkmark \\ y: & \text{B1, B2, B3} & \text{B1, B2} & \text{B2} & \text{B2} & \text{B2} & \end{array}$$

(17)
$$\begin{array}{lcccccc} & \mathsf{bag}(y) & + \ \mathsf{in}(x,y) \wedge \mathsf{rabbit}(x) & + \ \mathsf{big}_{\theta=3}(y) & + \ \mathsf{uniq}(y) & + \ \mathsf{uniq}(x) & ✗ \\ y: & \text{B1, B2, B3} & \text{B1, B2} & \emptyset & \emptyset & \emptyset & \end{array}$$

As the reader can see, when $\theta = 3$, a failure occurs when the gradable adjective applies, because once the set of candidates for y has been narrowed down to bags containing a rabbit, there is no candidate left that is *big* relative to that threshold.

On the absolute reading, the comparison class is just the set of bags, since the only constraint imposed on y by the time the modifier comes along is that it be a bag. Absolute readings fail in the Figure 2 scenario regardless of whether $\theta = 2$ or $\theta = 3$. They fail in different ways, though. When $\theta = 2$, the uniqueness check for y fails. In the case where $\theta = 3$, the derivation converges on a referent for y early on, and the uniqueness check succeeds, but ultimately the derivation crashes. This is the type of situation we refer to as a REFERENTIAL GARDEN PATH (notated 💣).

(18)
$$\begin{array}{lcccccc} & \mathsf{bag}(y) & + \ \mathsf{big}_{\theta=2}(y) & + \ \mathsf{uniq}(y) & + \ \mathsf{in}(x,y) \wedge \mathsf{rabbit}(x) & + \ \mathsf{uniq}(x) & ✗ \\ y: & \text{B1, B2, B3} & \text{B2, B3} & \emptyset & \emptyset & \emptyset & \end{array}$$

(19)
$$\begin{array}{lcccccc} & \mathsf{bag}(y) & + \ \mathsf{big}_{\theta=3}(y) & + \ \mathsf{uniq}(y) & + \ \mathsf{in}(x,y) \wedge \mathsf{rabbit}(x) & + \ \mathsf{uniq}(x) & 💣 \\ y: & \text{B1, B2, B3} & \text{B3} & \text{B3} & \emptyset & \emptyset & \end{array}$$

In general:

(20) Relative to a given scene, a linguistic expression gives rise to a REFERENTIAL GARDEN PATH if it has an interpretation on which a unique referent is assigned to some variable at one stage in the dynamic process, but there are no candidates left by the end.

An interpretation (dynamic sequence) assigns a unique referent to a given variable at some stage if there is only one candidate left at that stage, i.e., if the candidate set is a singleton. Our hypothesis is that this 'early' success in identifying a referent (in an abstract, mathematical, rather than temporal sense) leads listeners astray, causing a dispreference.

The term *referential garden path* is inspired by the notion of garden path familiar from the parsing literature (Bever 1970; Ferreira and Henderson 1991; Frazier and Rayner 1982; Garnsey et al. 1997; Trueswell et al. 1993: a.o.). Referential and parsing garden paths differ in at least one obvious way: while the latter involve alternative syntactic parses, referential garden-paths involve alternative dynamic-semantic interpretations. Put differently, comprehenders get *referentially* garden-pathed because they commit to the wrong referent, not the wrong parse. In the current work, the relevant connection between these two types of garden paths is that they both involve a local, not global, ambiguity (structural in the case of parsing garden paths and semantic in the case of referential garden paths), and that the resolution of this temporary ambiguity incurs a cost that yields a dispreference.

Let us take stock. The scope-based, Bumford-style view allows for both absolute and relative interpretations of the modifiers. A strongly incrementalist, Haddock-style view predicts that the

Table 1: Summary of outcomes for modified Haddock descriptions under absolute vs. relative readings under the *bag* resolution, relative to the scenario in Figure 2. ✓=success; ✗=failure; 💣=referential garden path.

	Absolute	Relative
big ($\theta = 2$)	✗	✓
big ($\theta = 3$)	💣	✗
bigger	✗	✓

comparison classes for all three types of modifiers should be constrained by descriptive content from the outer nominal (i.e., relative-like interpretations). In the context of Figure 2, a referential garden path is predicted for the positive adjective *big* only under an absolute reading (with $\theta = 3$). Referential garden paths are never predicted for *big* under relative readings or for *bigger*. The outcomes for the various readings are summarized in Table 1. If referential garden paths are problematic for listeners, and *the rabbit in the big bag* has an absolute reading, then we would expect a dispreference for this string in a scenario like the one in Figure 2. Only on the scope view does it have an absolute reading, so only under the scope view is this penalty predicted. Our main experiment is designed to test this prediction. But first, we will establish experimentally that the mere existence of alternative interpretations that fail is not the problem.

3 Experiment 1: Unmodified nested definites

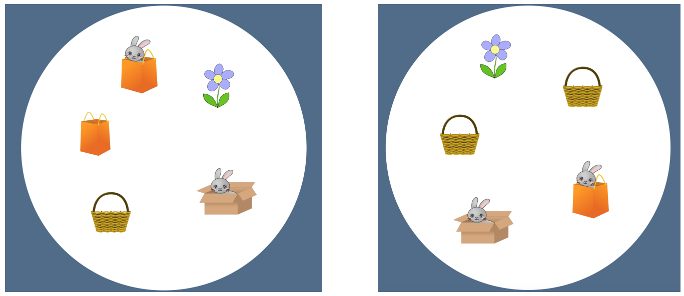
Before getting to the main experiment, let us warm up with a simpler one involving unmodified nested definites, like *the rabbit in the bag*. The goal of this experiment is to establish whether there is a general dispreference for situations in which an absolute reading fails to yield a referent for some linguistic expression, and only a relative reading is available. We will see that there is no such dispreference. This finding sets up an important point of contrast with respect to Experiment 2, which shows that *some* alternative readings that fail *do* yield a penalty, namely those that involve referential garden paths. The presentation of this experiment will also serve to introduce the methodological approach that we use for both experiments, and help to make vivid why addressing our research question demands that we include modifiers in the embedded noun phrase, as we do in Experiment 2.

3.1 Design & Materials

3.1.1 Experimental stimuli

Experiment 1 consisted of a reference resolution task in which participants were presented with visual scenes containing five possible referents. Each scene was paired with an auditory instruction that asked participants to click on one of the objects in the display. In experimental trials, the auditory instruction contained a nested definite description that was manipulated such that the embedded noun was masked with low-amplitude static noise (*Click on the rabbit in the* ***). The masked string was compatible with exactly two possible resolutions, i.e., two possible underlying

Instruction: *Click on the rabbit in the ***.*



TWO-BAGS condition   ONE-BAG condition

Figure 3: Example scenes for Experiment 1. Left panel represents the TWO-BAGS condition, containing the following five objects: **Target 1:** rabbit in a bag; **Target 2:** rabbit in a box; **Competitor:** empty bag; **Distractor 1:** empty basket; **Distractor 2:** flower. Right panel represents the ONE-BAG condition and differs from the TWO-BAGS condition only in that the Competitor object is substituted by **Distractor 3:** a second empty basket.

word strings, e.g. *rabbit in the bag* or *rabbit in the box*. As seen in Figure 3, this was achieved by ensuring that all experimental trial displays contained two potential referents that were compatible with the first noun in the instruction (e.g., *rabbit*), but crucially were associated with two different resolutions of the ambiguous masked instruction (e.g., a rabbit inside a bag (Target 1), and a rabbit inside box (Target 2). Thus, participants' task was to determine which embedded noun (e.g., *bag* or *box*) was more likely, given the linguistic information available in the instruction and the properties of the visual display.

Twelve items were constructed; see Table 2 in Appendix B for a full listing. In experimental trials, the instruction was always of the form [*the* OUTER-NOUN PREPOSITION *the* INNER-NOUN], where INNER-NOUN could be realized in two different ways (e.g. *bag/box* for one item, or *pillow/paintbrush* for another). The two inner nouns shared an onset (e.g., *bag* and *box*). Each of the twelve items had two VERSIONS, A and B, which differed only in which of the two alternative possibilities for the inner noun was the 'primary' one. The 'primary' inner noun determined properties of both the auditory instruction and the visual display.

The auditory stimuli were recorded by a male native speaker of American English in a soundproof booth. Care was taken to ensure that speech rate, volume, and pitch were as consistent as possible. This was done in order to control for coarticulation effects that could provide participants with phonetic cues regarding the identity of the masked noun. For each of the twelve items, two versions were recorded (Version A and Version B), differing in the choice of inner noun. The inner nouns were then masked by replacing the sound with low-amplitude static noise. To our ears, the masked Version A and Version B stimuli were indistinguishable, but still we paired the audio with the displays according to version, so that Version A displays were paired with masked Version A audio, and similarly for Version B.

Visual scenes were constructed to accompany the instruction. See Figure 3 for a example. There were two target referents in each scene, Target 1 and Target 2, both matching the outer noun in the instruction (e.g. *rabbit*), and realizing one or the other alternative for the inner noun. The preposition in the instruction (*in* or *with*) determined the spatial relation between the two nouns. Displays varied according to whether or not an absolute reading was available for the possible resolutions (*bag* or *box*). The main variable of interest was SCENE, which had two levels, TWO-BAGS and ONE-BAG. In the ONE-BAG condition, the two potential target referents were compatible with a absolute *and* a relative interpretation of the description. This is because absolute readings entail relative ones; whenever there is exactly one bag, and that bag has a rabbit in it, there is exactly one satisfier of 'bag with a rabbit in it'. In the TWO-BAGS condition, a relative reading for one of the resolutions was enforced by including a third referent, the Competitor, which matched the primary inner-noun alternative (e.g. *bag*). Thus, in the TWO-BAGS condition, Target 2 (e.g., a rabbit in a box) was compatible with an absolute interpretation of the description (i.e., box resolution in the left panel of Figure 3), whereas Target 1 (e.g., a rabbit in a bag) was only compatible with a relative interpretation (i.e., left panel of Figure 3). Finally, depending on the condition (TWO-BAGS or ONE-BAG), two or three additional referents were included as distractors. Distractors could not be described by either the outer noun or the inner noun (e.g. a basket and a flower, or car and a boat).

Like the auditory instructions, the visual displays had two VERSIONS, differing in whether the Competitor was, for example, an empty bag or an empty box. This was done in order to compensate for any biases that participants may have had toward one specific type of object in the display over the other, independent of our experimental manipulations.

We employed a Latin Square design with four lists ensuring that no participant saw an item in more than one condition, and across the four lists, each item appeared in all four conditions.

3.1.2 Fillers

A total of 24 fillers were included. Filler trials were of four different types (see Table 5 in Appendix B for the exhaustive list of fillers). The first class of fillers (6 filler trials) consisted of unambiguous instructions where none of the nouns were masked (e.g. *Click on the man in the truck*). The PPs in these 6 trials were either helpfully informative (in the sense that the PP helped to eliminate at least one candidate referent), or over-informative (i.e., the PP was not needed to identify the intended referent). Furthermore, a subset of these trials necessitated a relative interpretation; others licensed an absolute interpretation of the description. The second type of fillers resembled target items in that the second NP in the auditory instruction was masked (5 fillers). However, unlike experimental trials, the PP was over-informative, and in some cases the Haddock interpretation was not available. The third class of fillers consisted of complex descriptions where the first noun was masked (e.g. *Click on the *** with the fish*). PPs in this class varied according to informativity and whether the PP could receive a relative interpretation or not. Finally, the fourth type of filler trials (6 fillers) consisted of descriptions that contained a single NP, where none of the material had been masked (e.g. *Click on the bird*).

3.2 Procedure

Subjects participated in the experiment remotely, and were subjected to three practice trials before beginning the experiment, two with masked auditory stimuli and one without any masking (*Click on the square with stripes*; *Click on the triangle with* ***; *Click on the red* ***). On each trial, participants were presented with a display showing five images arranged in a circle, equidistant from each other and from an audio button in the center. The order of the images around the circle on the display was randomly generated for each participant on each trial. In order to trigger the auditory instruction, participants clicked on an audio button in the center of the display. They proceeded to the next trial when they clicked on one of the five images. Clicks to one of the five potential referents triggered the next trial only after the auditory instruction was over. This prevented participants from skipping to the next trial without having heard the full auditory instruction. The order of the experimental trials and the fillers was randomized, and experimental trials were pseudo-randomly interspersed with filler trials, ensuring that no more than two fillers were presented in a row.

3.3 Participants

A total of 52 participants were recruited through Amazon Mechanical Turk and Prolific. Results from 8 participants were discarded either because they failed to pass our attention check (giving more than three unexpected responses on filler items), or because they did not self-report being native speakers of English (3 participants). All analyses reported below are based on data belonging to the remaining 41 participants.

3.4 Predictions

By comparing target selection rates to the two potential referents, it is possible to gauge participants' interpretive preferences as a function of the available linguistic input and the properties of the visual display. Recall that in the TWO-BAGS condition no absolute reading is available for the *bag* resolution. Only a relative reading is available, since in this condition there are two bags, so the uniqueness check fails for the inner noun. If the failure of an absolute reading causes a dispreference for a given resolution, then it is predicted that there should be fewer clicks to the target object corresponding to the *bag* resolution in the TWO-BAGS condition compared to the ONE-BAG condition. Similarly, if absolute readings are preferred over relative ones, we also predict fewer clicks to the referent corresponding to the *bag* resolution in the TWO-BAGS condition compared to the ONE-BAG condition. As we will see, participants' decisions about how to resolve the ambiguous string did not match either of these predictions.

3.5 Results & Discussion

Figure 4 plots the rate at which participants selected the target object corresponding to the *bag* resolution in the TWO-BAGS and ONE-BAG conditions, respectively. Participants overwhelmingly selected one of the two potential targets, with the exception of one single trial. Clicks to the target object corresponding to the *bag* resolution (the resolution with the primary inner noun) were submitted to a logistic Bayesian mixed effects regression model using SCENE as a fixed effect predictor, and participants and items as mixed effects. Unless otherwise noted, all model specifications for random effects reported in this paper are maximal. Analyses were performed using the brms package in R (Bürkner 2017). Results did not reveal any significant effects of SCENE ($\beta = -0.2$, $CI = [-0.49, 0.53]$).

Each of the two theories of nested definites under consideration accounts for this null effect in different ways. From Bumford's perspective, on which nested definites are ambiguous between a Haddock (relative) reading and an absolute reading, the current results suggest that there is no penalty (at least in this paradigm) for resolutions of the string for which an absolute reading fails as a result of a violation of the inner definite's uniqueness presupposition. Similarly, the current results do not point at a preference for absolute readings over relative ones. Under Haddock's

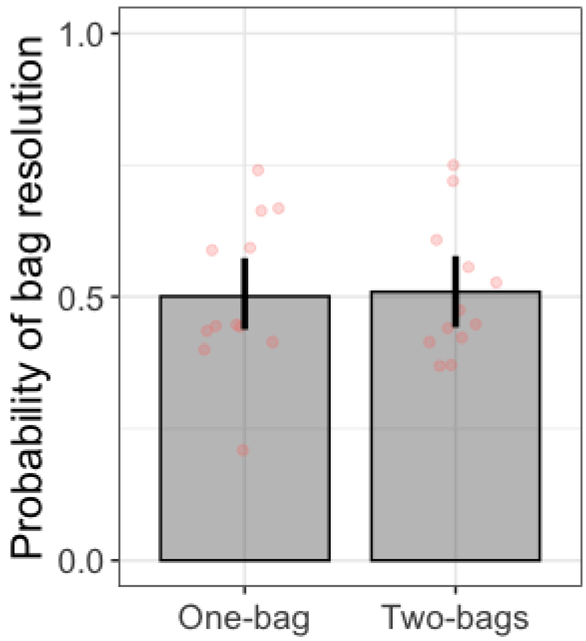


Figure 4: Results of Experiment 1. The error bars show a 95% bootstrapped confidence intervals of the mean. Floating dots represent item means.

account, which sanctions only relative readings, the lack of a difference between the two conditions tested is expected, because there is no absolute reading to begin with.

Experiment 1 has established that a nested definite description such as *the rabbit in the bag* is equally acceptable in contexts that support both relative and absolute interpretations and contexts where only a relative interpretation is available. What this shows is that if absolute readings are licensed by the grammar, there is no penalty for resolutions that happen to lack them. But it does not tell us whether or not absolute readings are licensed by the grammar; these results are consistent with both Bumford's and Haddock's views. In the next experiment, we will see evidence that absolute readings are indeed licensed by the grammar as Bumford predicts. We see a trace of them, even in cases where they do not ultimately succeed, in situations where the listener can posit a unique referent for the embedded noun phrase, leading to a referential garden path effect. For the next experiment, we use embedded noun phrases containing context-sensitive gradable predicates in order to construct situations where a unique referent can be posited for the embedded noun phrase even though ultimately only the relative reading succeeds.

4 Experiment 2: Modified nested definites

Experiment 2 tests whether absolute readings for nested descriptions are available, as predicted by Bumford's scope-based theory. The experiment uses the same paradigm as Experiment 1. Participants performed a reference resolution task in which Haddock descriptions were presented auditorily. Each instruction was paired with a visual display containing five potential referents. As in Experiment 1, in target trials, the inner noun of the definite description was masked with

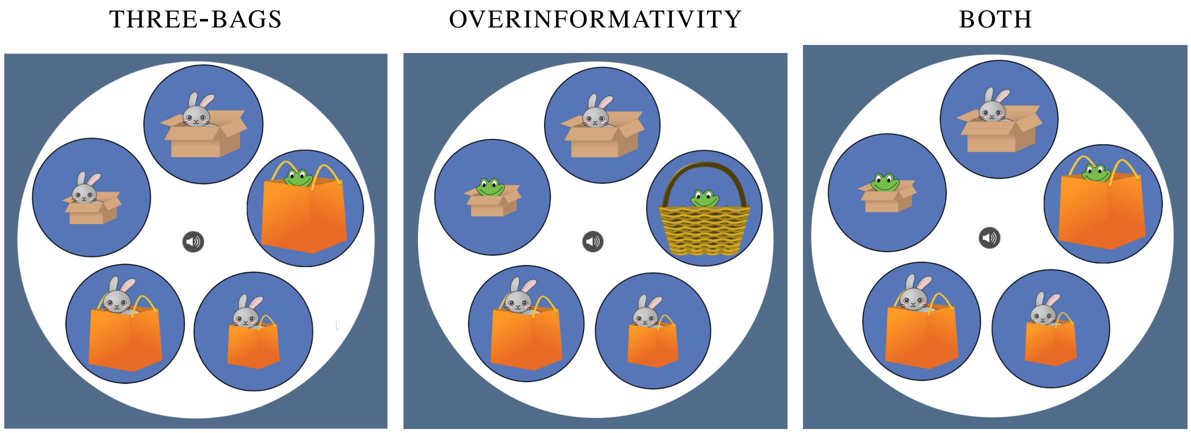


Figure 5: Visual Displays tested in the Main Experiment. Clicking on the audio button in the center of the screen would launch an auditory instruction of the form, "Click on the rabbit in the big/ger ***"

quiet static noise. Visual displays were crafted in such a way that the truncated description was compatible with two possible referents (e.g. a rabbit in a bag and a rabbit in a box). The crucial difference between Experiment 1 and 2 is that the descriptions contained an adjectival modifier in the embedded noun phrase, which was either in the positive (*big*) or the comparative (*bigger*) form (e.g. *the rabbit in the big/ger* ***). The presence of this modifier caused referential garden paths on absolute readings with respect to some scenes, allowing us to detect them.

4.1 Materials

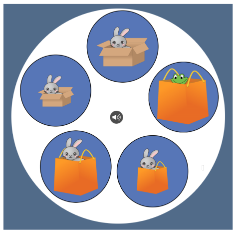
4.1.1 Experimental items

In experimental trials, the instruction was of the form [*the* OUTER-NOUN PREPOSITION *the* MODIFIER INNER-NOUN], where INNER-NOUN could be realized in two different ways (e.g. *bag/box* for one item, or *pillow/paintbrush* for another). As in Experiment 1, the two inner nouns shared an onset (e.g., *bag* and *box*). The modifier was either positive (*big*) or comparative (*bigger*).

Such instructions were tested against three types of visual SCENEs. In the first scene type, only one of the possible resolutions was compatible with an absolute interpretation. This was accomplished by manipulating the number and sizes of the inner-noun objects, as well as the type of outer-noun entities associated with each inner-noun object in the display. As an example, consider the THREE-BAGS scene in Figure 4. Here, Target 1 is a rabbit in a bag, while Target 2 is a rabbit in a box. The scene contains two other rabbits that are inside a smaller bag and a smaller box respectively. This ensured that the bag and the box associated with the two targets could be judged as *big* or *bigger*, depending on the instruction. Whether each of the two possible resolutions was compatible with an absolute interpretation was achieved by including a third referent that could not be described using the outer noun in the description, e.g. a frog. This referent was always associated with the same inner-noun object type as Target 1: a bag, in this case. This bag was biggest among the bags.

As seen in derivations (21) and (22), repeated from (18) and (19), the absolute interpretation of the *bag* resolution fails in the THREE-BAGS scene, regardless of the threshold value adopted.

THREE-BAGS scene



	Absolute	Relative
big ($\theta = 2$)	✗	✓
big ($\theta = 3$)	💣	✗
bigger	✗	✓

	Absolute	Relative
big ($\theta = 2$)	✓	✓
big ($\theta = 3$)	✗	✗
bigger	✓	✓

Figure 6: Summary of outcomes for modified Haddock descriptions under absolute vs. relative readings in the THREE-BAGS condition. The left panel contains outcomes for the *bag* resolution (Target 1), while the right panel contains outcomes for the *box* resolution (Target 2). ✓=success; ✗=failure; 💣=failure with referential garden path.

(21)
$$\begin{array}{rcccccccccl} & \mathsf{bag}(y) & + & \mathsf{big}_{\theta=2}(y) & + & \mathsf{uniq}(y) & + & \mathsf{in}(x,y) \wedge \mathsf{rabbit}(x) & + & \mathsf{uniq}(x) & ✗ \\ y\text{:} & \text{B1, B2, B3} & & \text{B2, B3} & & \emptyset & & \emptyset & & \emptyset & \end{array}$$

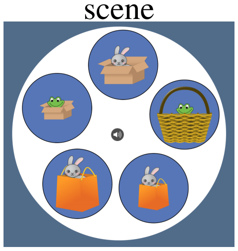
(22)
$$\begin{array}{rcccccccccl} & \mathsf{bag}(y) & + & \mathsf{big}_{\theta=3}(y) & + & \mathsf{uniq}(y) & + & \mathsf{in}(x,y) \wedge \mathsf{rabbit}(x) & + & \mathsf{uniq}(x) & 💣 \\ y\text{:} & \text{B1, B2, B3} & & \text{B3} & & \text{B3} & & \emptyset & & \emptyset & \end{array}$$

In (21), which assumes $\theta = 2$, the derivation crashes once the uniqueness check on bags applies. At that point, the candidate set contains two possible referents (B2 and B3), and so the uniqueness check fails. When $\theta = 3$, as in (22), the derivation converges on a single referent for y, i.e. B3. However, upon the application of subsequent filters the derivation crashes. Therefore, in (22) the failed derivation gives rise to a referential garden path. Given that no value of the threshold variable results in a defined description for the absolute reading of the *bag* resolution, the only possible reading is a relative one. On the other hand, under the *box* resolution, both the relative and the absolute readings are defined. The full range of readings for the THREE-BAGS condition is spelled out in Figure 6. We refer the reader to Appendix C for the full set of derivations.

The second type of manipulation involved the informativity of the modifier. In the THREE-BAGS scene, the inclusion of the prenominal modifier was required for successful reference, regardless of whether the masked portion of the string was resolved as *bag* or *box*; an equivalent but unmodified instruction such as *the rabbit in the bag/box* would have failed to refer in the THREE-BAGS scene. This is due to the fact that an unmodified instruction like this could not distinguish between the rabbit in the smallest bag/box and the rabbit in the medium bag/box. Therefore, in the THREE-BAGS condition the adjective was always helpfully informative. In the OVERINFORMATIVITY scene (middle panel Figure 4), the visual display was designed such that the adjectival modifier in the instruction was helpfully informative under the *bag* resolution (Target 1), but redundant/overinformative under the *box* resolution (Target 2). This was accomplished by substituting the rabbit in the smallest box with a frog. We refer to this type of display as the OVERINFORMATIVITY condition. Finally, in order to independently assess the effect of the informativity manipulation independently from the number-of-bags manipulation, in the OVERINFORMATIVITY scene, the third bag was substituted with a basket.

We now turn to the referential garden paths predicted to arise in the OVERINFORMATIVITY scene. The informativity manipulation causes a referential garden path to arise for the *box* resolution of the comparatively modified description (*the rabbit in the bigger box*) under a relative interpretation (23). Recall that the comparative meaning discussed in (10) is a filter that keeps only

OVERINFORMATIVITY scene



	Absolute	Relative
big ($\theta = 2$)	✓	✓
big ($\theta = 3$)	✗	✗
bigger	✓	✓

	Absolute	Relative
big ($\theta = 2$)	✓	✓
big ($\theta = 3$)	✗	✗
bigger	✓	💣

Figure 7: Summary of outcomes for modified Haddock descriptions under absolute vs. relative readings in the OVERINFORMATIVITY scene. The left panel contains outcomes for the *bag* resolution (Target 1), while the right panel contains outcomes for the *box* resolution (Target 2). ✓=success; ✗=failure; 💣=failure with referential garden path.

assignments to v for which there is another assignment that maps v to something smaller. Therefore, once the outer noun filter $\textsf{rabbit}(x)$ has applied, the output candidate set contains a single referent (namely the rabbit in the medium bag). However, this referent is later incompatible with the meaning of the comparative, which requires multiple candidates in its input state. When the comparative applies, nothing passes the test. The derivation thus yields a referential garden path because it settles temporarily on one referent and then ultimately fails.

(23)

	$\textsf{rabbit}(x)$	+	$\textsf{in}(x,y) \wedge \textsf{box}(y)$	+	$\textsf{cmpr}(y, \textsf{big})$	+	$\textsf{uniq}(y)$	+	$\textsf{uniq}(x)$	💣
x:	R1, R2, R3		R2		R2		R2		R2	
y:	(any)		B2		$\emptyset$		$\emptyset$		$\emptyset$	

The full range of (un)available readings associated with the OVERINFORMATIVITY scene are spelled out in Figure 7. We refer the reader to Appendix C for the full set of derivations.

Finally, the third scene type, called BOTH, combines the number of bags manipulation and the informativity manipulation (see right panel Figure 5). For this scene, the predicted readings for the *bag* resolution are just the same as in the THREE-BAGS scene, and the predicted readings for the *box* resolution are just the same as in the OVERINFORMATIVITY scene.

As in Experiment 1, 12 experimental items were constructed for Experiment 2. The two potential resolutions and distractors all began with the same phoneme; for example, *box*, *bag*, and *basket* all start with a bilabial voiced stop /b/. Experimental items are listed in Table 3 in Appendix B. As in Experiment 1, the auditory stimuli were recorded by a male native speaker of English in a soundproof booth, and care was taken to ensure that speech rate, volume, and pitch were as consistent as possible. For each item/modifier pair, there were two versions, Version A and Version B, depending on whether the first or second inner noun was employed. This inner noun was then masked, resulting again in a situation where the Version A stimuli were pairwise-indistinguishable from the Version B stimuli. Still we paired the masked Version A stimuli with Version A displays, and similarly for Version B. As in Experiment 1, we constructed two versions of each scene. Version A is as just described; Version B had inner noun 1 and inner noun 2 switched, to remove associated biases.

We thus had three within-items variables: SCENE (3 levels), ADJECTIVE TYPE (positive vs. comparative), and VERSION (A vs. B). We employed a Latin Square design with $3 \times 2 \times 2 = 12$ lists ensuring that no participant saw an item in more than one condition, and across the 12 lists, each item appeared in all 12 conditions. Each participant saw all 12 items and all 24 fillers.

4.1.2 Fillers

Filler items ($n = 24$) consisted of an instruction paired with five images arranged in a circle, constituting a visual display. Fillers were always unambiguous such that the referent could always be inferred from the information available from the auditory instruction given the visual display. The auditory instructions could consist of complex DPs but were never adjectivally modified.

There were several types of fillers. For the first type of filler, the instruction consisted of a complex description where none of the nouns had been masked (e.g. *Click on the girl with the grapes*). Among this type, there was variation in whether the description in the instruction was compatible with a relative interpretation and whether the PP was required for target identification (i.e. whether the PP was informative or not). The second type of filler consisted of complex DPs where the inner noun was masked (e.g. *Click on the girl with the* ***). As with the first type of fillers, some of the fillers of this type allowed for relative readings. The PP was always overinformative among fillers of this type, however, since reference identification would otherwise had not been possible since the inner noun was masked. In the third type of filler, the first noun was masked (e.g., *Click on the *** in the bathtub*). Some of these allowed for relative interpretations and in some cases the PP was overinformative. Finally, the fourth type of filler consisted of auditory descriptions containing a simple DP the noun was not masked (e.g., *Click on the horse*).[8]

4.2 Procedure

The procedure was the same as in Experiment 1.

4.3 Participants

We collected data from 242 native speakers of English, recruited through the crowd-sourcing platforms Amazon Mechanical Turk and Prolific. Data from 19 participants was removed from data analysis due to a failure to pass attention checks, giving more than 3 unexpected responses in the filler trials. Data from three additional participants was removed since they took the experiment twice, resulting in a total of 217 participants.

4.4 Predictions

We expected that our participants' behavior would be modulated by two factors: 1) whether a particular resolution gives rise to an informativity violation; and 2) whether a particular resolution is associated with a referential garden path. The particular situations in which we expect a referential garden path depends on which readings are sanctioned by the grammar. If the scope theory is right, and the grammar sanctions absolute readings, then we expect referential garden paths under the *bag* resolution of the string with the positive form adjective in both of the scenes involving three bags (the THREE-BAGS scene and the BOTH scene). Independently of whether the grammar sanctions absolute readings, the referential garden path hypothesis predicts a penalty for the *box* resolution of the string with the comparative modifier *bigger* in the OVERINFORMATIVITY and

[8] A full list of the fillers used in the experiment can be found in Table 5 in Appendix B.

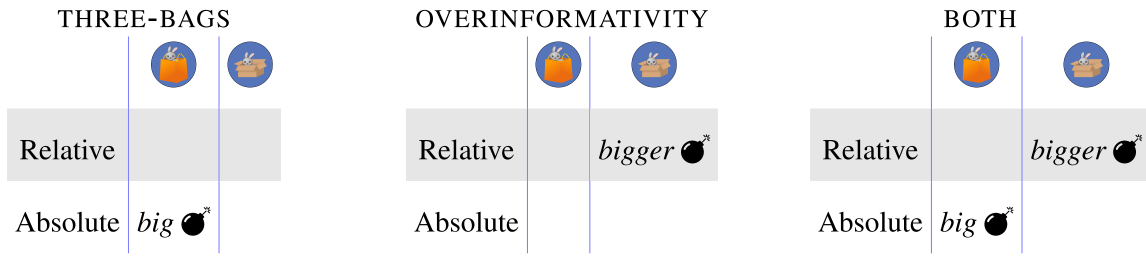


Figure 8: Referential garden paths predicted by the twelve conditions tested in Experiment 2.

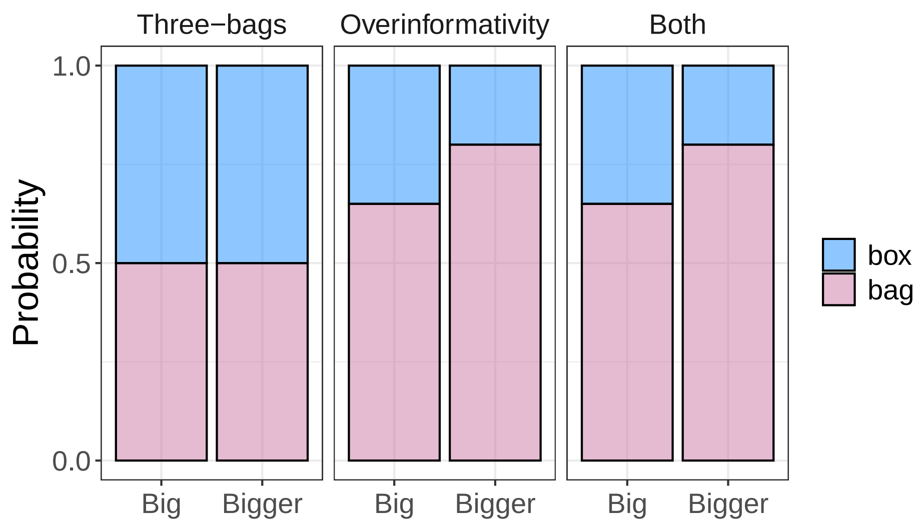


Figure 9: Predictions under the incrementalist view: Participants should be at chance with both *big* and *bigger* in the THREE-BAGS scene (Scene 1), and there should be no difference between the OVERINFORMATIVITY and BOTH scenes (Scenes 2 and 3).

BOTH scenes. Figure 8 summarizes all the referential garden paths associated with each possible resolution in the six experimental conditions.

Our qualitative predictions for each of the conditions tested in our experiment are summarized in Figures 9 and 10. In the THREE-BAGS scene, the use of the adjectival modifier is required for successful target identification in both the bag and the *box* resolution. Therefore, informativity should not play a role in this scene's results. The only modulator of target selection rates in the THREE-BAGS scene should be the referential garden path triggered by the absolute interpretation of the the *bag* resolution when the description contains a positive adjective (see left panel Figure 8). If, as predicted by the scope view, absolute readings are considered by our participants, we should observe a dispreference for the *bag* resolution over the *box* resolution when the description contains a positive form adjective (*big*), such that the target selection rate for the *bag* resolution should be significantly below 50% in this display (see left panel Fig. 10). On the other hand, if *only* relative readings are accessed by our participants (as predicted by the incrementalist view), both resolutions should be equally probable and participants should be at chance in deciding among them (see left panel Fig. 9). Finally, if the description contains a comparative adjective, we predict that the *bag* and the *box* resolutions should be equally probable, since neither involves an overinformative use

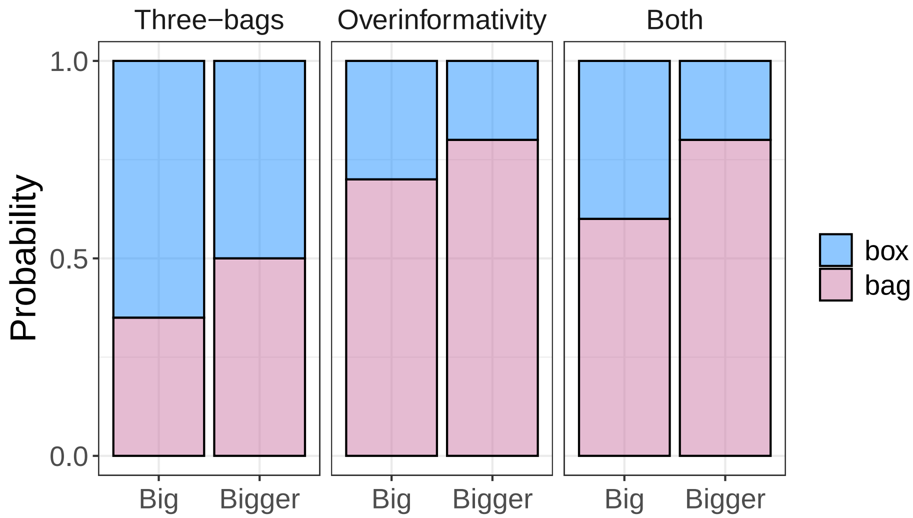


Figure 10: Predictions under the scope view: For *big*, but not *bigger*, there should be a dispreference for the *bag* resolution in the THREE-BAGS scene (Scene 1), and there should be a difference between the OVERINFORMATIVITY and BOTH scenes (Scenes 2 and 3).

of the adjective or a referential garden path. Therefore, the two theories under consideration make the same predictions with respect to this condition (see left panel of Figures 9 and 10).

In the OVERINFORMATIVITY scene, the inclusion of a modifier in the instruction is required for successful target identification under the *bag* resolution. However, the use of the adjective is redundant under the alternative *box* resolution. This should lead to higher-than-chance selection rates for the *bag* resolution regardless of whether the instruction contains a positive or a comparative adjective, a prediction that is shared by both the incrementalist and the scope based view. Furthermore, given that the *box* resolution involves a referential garden path under the relative reading with a comparative adjective (see middle panel Figure 8), it is predicted that participants should display an even higher preference for the *bag* resolution in the comparative condition, above and beyond the informativity effect, compared to the positive form condition, which does not involve a referential garden path (see middle panel Figures 9 and 10).

Finally, let us consider the predictions for the BOTH scene. The makeup of the BOTH scene is a mixture of the THREE-BAGS scene and the OVERINFORMATIVITY scene. Therefore, we expect our results to display both informativity and referential path effects effects (see right panel Figure 8). The strong incrementalist account makes the same exact predictions in the OVERINFORMATIVITY and BOTH scenes, since the presence of the third bag is only expected to give rise to a referential garden path under the absolute reading of the *bag* resolution, and that is a reading that is not predicted to be generated in the first place. The scope-based account, on the other hand, makes different predictions for the BOTH scene compared to the OVERINFORMATIVITY scene. While the results for the comparative are not predicted to be different between the those two scenes, the instruction with the positive form should yield a compounded effect of informativity *and* a referential garden path. On top of the informativity penalty for the *box* resolution (common to the OVERINFORMATIVITY and BOTH scenes), we should observe an effect of the referential garden path produced by the absolute reading of the string with the positive adjective under the *bag*

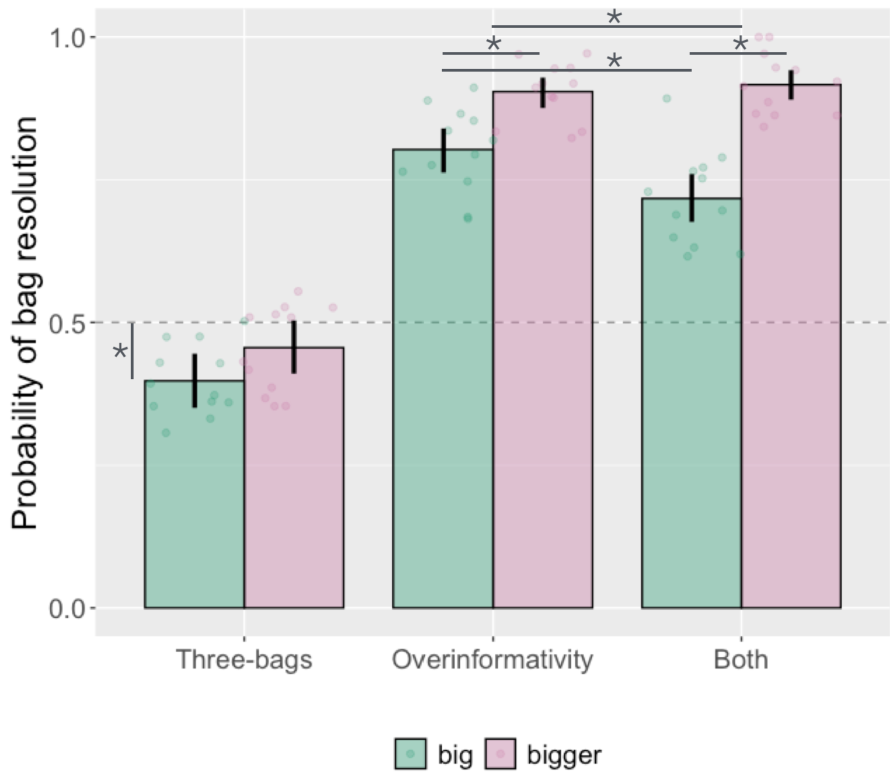


Figure 11: Results for Experiment 2. The Y-axis plots the proportion of responses corresponding to the *bag* resolution in each of the six conditions tested. The error bars represent Bootstrap 95% confidence intervals.

resolution. Therefore, with the positive form *big*, we expect that target selection rates for the *bag* resolution should be significantly lower in the BOTH scene compared to the OVERINFORMATIVITY scene and higher than the *bag*-selection rates observed in the those observed in the THREE-BAGS (see right panel Figure 10).[9]

4.5 Results

Proportions for the *bag* resolution (where the participant selected Target 1) are shown in Figure 11 in the six conditions tested. In order to assess whether the predictions just described were borne out, we conducted a series of Bayesian mixed-effects logistic regressions.[10] We start with the predictions pertaining to the THREE-BAGS scene. In order to assess whether selection rates for one target or the other significantly differed from chance in the THREE-BAGS scene, we fit two separate intercept-only models to the data, one for each adjective. Random intercepts by

[9]We predict qualitative comparisons between conditions, rather than specific quantitative values, so the visualization in Figure 10 is more specific than our actual predictions. In particular, we make no prediction about whether the Overinformativity *big* condition will be above or below chance; we predict only that target selection rates for the *bag* resolution should be between those for the BOTH scene and the OVERINFORMATIVITY scene.

[10]All mixed-effects analyses were conducted using with the brms package in R. Simple effects were calculated with the function `hypothesis`. All models reported in this section included the maximal random effect structure justified by the analysis and the experimental design. All models assumed uninformative priors.

subjects and items were also included. Selection rates for Target 1 (corresponding to the *bag* resolution) with the positive form adjective *big* were significantly below chance ($\beta = -0.44$, $SE = 0.12$, $CI = [-0.69, -0.21]$).[11] In other words, there was a significant preference for the *box* resolution. No significant preference for one resolution or the other was observed with the comparative adjective *bigger* ($\beta = -0.20$, $SE = 0.13$, $CI = [-0.45, 0.05]$). The main effect of ADJECTIVE did not reach significance, however ($\beta = 0.23$, $SE = 0.17$, $CI = [-0.09, 0.56]$).[12]

Next, we discuss the results pertaining to the two scenes for which the *box* resolution carried an informativity violation (OVERINFORMATIVITY and BOTH). In order to assess the effects of the two referential garden paths predicted by the scope-based hypothesis (see §4.4), data from these scenes was submitted to a model predicting selection rates for Target 1 (corresponding to the *bag* resolution) from fixed effects ADJECTIVE and SCENE and their interaction. The model contained random by-speaker and by-item intercepts and slopes.[13] A significant ADJECTIVE $\times$ SCENE interaction was detected ($\beta = 0.91$, $SE = 0.47$, $CI = [0.08, 1.94]$). Simple effects comparisons revealed that the interaction was driven by a decrease in target selection rates in the positive form conditions, such that the *bag* resolution was dispreferred in the BOTH scene compared to the OVERINFORMATIVITY scene ($\beta = -0.53$, $SE = 0.23$, $CI = [-0.98, -0.08]$). No equivalent effect was found for the comparative conditions ($\beta = 0.38$, $SE = 0.46$, $CI = [-0.42, 1.38]$). Finally, *bag* selection rates were significantly higher in the comparative conditions compared to the positive ones both in the OVERINFORMATIVITY scene ($\beta = 0.99$, $SE = 0.28$, $CI = [0.45, 1.57]$) and the BOTH scene ($\beta = 1.9$, $SE = 0.43$, $CI = [1.19, 2.86]$).

4.6 Discussion

4.6.1 Absolute readings

We began this paper by considering two alternative theories of Haddock descriptions, a scope-based view, which sanctions both absolute readings and relative readings, and an incrementalist view, which sanctions only relative-like readings. Of these two, the data support the former. One data point supporting this conclusion is the significant dispreference for the *bag* resolution in the THREE-BAGS scene, where the presence of the third, largest bag tempts the listener into an interpretation where the inner definite refers to it. As spelled out in the predictions section, this pattern of results is to be expected if participants encounter the referential garden path associated with the absolute interpretation of the positive form condition under the *bag* resolution. In the comparative condition, no such dispreference was detected, as expected given that comparative form adjectives are not theorized to give rise to referential garden paths under any reading in this scene.

The second data point supporting our conclusion is the significant ADJECTIVE $\times$ SCENE interaction found in the OVERINFORMATIVITY and BOTH scenes. As discussed in the previous section, this interaction was driven by significantly lower *bag* resolution rates in the positive form conditions in the BOTH scene compared to the OVERINFORMATIVITY scene. Again, this pattern of

[11]The model specification was as follows: TARGET $\sim 1 + (1 \mid \text{SUBJECT}) + (1 \mid \text{ITEM})$. Note that in this model the estimate corresponds to the intercept.

[12]The model specification was as follows: TARGET $\sim$ ADJECTIVE $+ (1 +$ ADJECTIVE $\mid$ SUBJECT$) + (1 +$ ADJECTIVE $\mid$ ITEM$)$.

[13]The model specification was as follows: TARGET $\sim$ ADJECTIVE $\times$ SCENE $+ (1 +$ ADJECTIVE $\times$ SCENE $\mid$ SUBJECT$) + (1 +$ ADJECTIVE $\times$ SCENE $\mid$ ITEM$)$.

results is consistent with the hypothesis that the absolute interpretation of the description containing the positive form leads to a referential garden path under the *bag* resolution in the BOTH scene but not in the OVERINFORMATIVITY scene. With the comparative modifier *bigger*, on the other hand, the preference for *bag* did not decrease when a third bag was introduced to the scene, and this is as expected because the *bag* resolution is not associated with a referential garden path for either the absolute or relative reading of the comparative description in these two scenes.

The current patterns of results cannot be accommodated by strong incrementalist theories, which by design do not predict the existence of absolute readings in the first place. Rather, the current results support the scope-based view, where Haddock descriptions are truly ambiguous between a relative and an absolute reading, over the incrementalist view.

4.6.2 Mixed readings

In a modified Haddock description like *the rabbit in the big/ger bag*, there are two elements in the inner noun phrase that could in principle take either low or high scope: the definite determiner and the adjectival predicate. Leaving aside inverse scope configurations, where scope-taking operators fail to preserve the ordering of the *in situ* configuration, this allows for three possible situations:

- Pure absolute readings: both the inner determiner and the inner adjective scope low
- Mixed readings: the inner adjective scopes low and the inner determiner scopes high
- Relative readings: both the inner adjective and the inner determiner scope high

A mixed reading is exemplified in the pseudo-LF in (24), where the determiners scope high and the adjectival predicate scopes low.

(24)

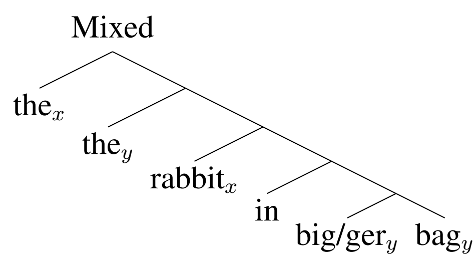


Table 12 details the order of application of the different semantic filters given by absolute, relative and mixed readings. As can be appreciated in the table, mixed readings share features with both absolute and relative readings. As far as we can see, nothing about Bumford's (2017) scope-based view rules out mixed readings.

Absolute	$\mathsf{bag}(y)$	+	$\mathsf{big}_{\theta=2}(y)$	+	$\mathsf{uniq}(y)$	+	$\mathsf{in}(x,y) \wedge \mathsf{rabbit}(x)$	+	$\mathsf{uniq}(x)$
Relative	$\mathsf{bag}(y)$	+	$\mathsf{in}(x,y) \wedge \mathsf{rabbit}(x)$	+	$\mathsf{big}_{\theta=2}(y)$	+	$\mathsf{uniq}(y)$	+	$\mathsf{uniq}(x)$
Mixed	$\mathsf{bag}(y)$	+	$\mathsf{big}_{\theta=2}(y)$	+	$\mathsf{in}(x,y) \wedge \mathsf{rabbit}(x)$	+	$\mathsf{uniq}(y)$	+	$\mathsf{uniq}(x)$

Figure 12: Sequence of constraints under different scopal configurations. Blue marks in situ scope; Green marks scoped out filter.

So far, we have left mixed readings out of the discussion, but they are important to bring into the picture as we carefully consider the significance of our experimental findings. As it happens, mixed readings yield referential garden paths in the same exact scenarios as absolute readings. In particular, referential garden paths arise with mixed readings under the *bag* resolution of descriptions containing a positive form adjective in the THREE-BAGS and BOTH scenes:

(25)
$$\begin{array}{rccccccccccl} & \mathsf{bag}(y) & + & \mathsf{big}_{\theta=2}(y) & + & \mathsf{in}(x,y) \wedge \mathsf{rabbit}(x) & + & \mathsf{uniq}(y) & + & \mathsf{uniq}(x) & ✗ \\ y\text{:} & \text{B1, B2, B3} & & \text{B2, B3} & & \text{B2} & & \text{B2} & & \text{B2} & \end{array}$$

(26)
$$\begin{array}{rccccccccccl} & \mathsf{bag}(y) & + & \mathsf{big}_{\theta=3}(y) & + & \mathsf{in}(x,y) \wedge \mathsf{rabbit}(x) & + & \mathsf{uniq}(y) & + & \mathsf{uniq}(x) & 💣 \\ y\text{:} & \text{B1, B2, B3} & & \text{B3} & & \emptyset & & \emptyset & & \emptyset & \end{array}$$

The fact that absolute and mixed readings converge with respect to where they give rise to referential garden paths shows that the referential garden path effect we observed with positive form adjectives is really due to the absolute (low/early) interpretation of the positive form adjective. So when it comes to positive form adjectives, what our experimental results show is that the listener considers a low-scope interpretation of them in nested descriptions, whether or not the inner determiner scopes high. A strongly incrementalist account of Haddock effects does not make any such interpretation available, so our findings indeed support the scope view, even with this added wrinkle that mixed readings are a possibility.

4.6.3 Referential garden paths

Finally, we would like to highlight the remarkable accuracy with which the notion of referential garden path captures the results reported in this paper. Referential garden paths not only revealed absolute readings of nested descriptions with positive form gradable adjectives, they also captured subtle effects with comparatives. In particular, a referential garden path was predicted to arise for the relative interpretation of the *box* resolution of the string with a comparative modifier in the OVERINFORMATIVITY and BOTH scenes. No corresponding effect was predicted for the positive form. So the *box* resolution was predicted to be dispreferred in these scenes for comparatives for two reasons—the informativity violation, and the referential garden path—whereas the positive form was only subject to the informativity violation. And indeed, bearing out this prediction, we found that in these two scenes, the baseline dispreference for the *box* resolution was greater for comparatives, compared to positives, as shown by the main effect of ADJECTIVE found in the OVERINFORMATIVITY and BOTH scenes. Taken together, these findings point at the robustness of referential garden paths as a new behavioral signature of semantic processing, opening up exciting possibilities for the study of referential communication.

5 Conclusion

In this paper we have examined two alternative theories of Haddock descriptions that aim to account for the apparent obviation of the uniqueness requirement displayed by embedded definites: the incrementalist account proposed by Haddock (1987), and the scope-based account by Bumford (2017). The two theories differ in the range of readings that they generate; while the incrementalist account predicts that Haddock descriptions should only involve relative readings, the scope-based approach is able to generate both relative and absolute readings, among possibly others. Determining whether the grammar generates absolute readings is a difficult question to answer on the basis of introspection; absolute readings entail relative ones, and it is therefore in principle possible that all attested readings of these complex definites are indeed generated through relative semantic derivations.

The experimental approach adopted in this paper has allowed us to circumvent this limitation. In two reference resolution tasks, we examine the derivations under which adjectivally modified Haddock descriptions are hypothesized to give rise to *referential garden paths*, a failing semantic derivation that converges on a single referent before eventually crashing. Remarkably, we were able to detect the expected behavioral effects (i.e., a dispreference for the resolution that involved a referential garden path) in precisely all the circumstances where referential garden paths were theoretically predicted to arise in our experimental design. These circumstances involved different visual scenes, different readings and different types of adjectivally modified descriptions, suggesting that referential garden paths are a general feature of semantic processing. To our knowledge, the current work is the first to point out their existence and we are hopeful that this novel semantic processing effect will aid in the study of referential communication more broadly.

Second, our results allow us to adjudicate between the two theories under consideration, as we were able to detect referential garden path effects associated with absolute readings for an embedded modifier. The current results militate against Haddock's incrementalist account, and suggest that the disappearance of the uniqueness constraint in Haddock descriptions is due to the fact that modifiers and determiners are scopally mobile within nested descriptions.

As far as we know, the idea of a referential garden path has not been introduced in previous literature. There is much remaining to be understood about how they work, and the extent to which they are analogous to the better studied garden paths discussed in the parsing literature. It is at this point an open question whether we should think of the mechanism underlying the penalty for readings incurring a referential garden path in parallel way. In particular, it is an open question whether the strength of a referential garden path effect is modulated by the likelihood or strength of commitment to the reading that gives rise to it. Our experimental results discourage us from identifying absolute interpretations (or *in situ* scope interpretations) as the default reading in general, since referential garden paths associated with relative readings were also detected in contexts where the absolute interpretation was in fact licensed. Rather, it seems that comprehenders entertain *all* possible derivations licensed by the grammar. But these derivations may still be entertained to varying degrees. Our hope is that referential garden paths will become better understood in future work and prove useful to the psycholinguistics community as a behavioral signature that can be used to probe referential processing.

Figure 13: Left: A scenario supporting a relative reading of *the rabbit in the biggest bag*. Right: A scenario supporting a relative reading of *the rabbit in the bigger bag*.

A Relative readings: superlative & comparative adjectives

It is well-known that superlative adjectives like *biggest* can have relative readings (Szabolcsi 1986; Heim 1999; Bumford 2017: i.a.). One way of seeing this is via the definedness of the description in the lefthand panel of Figure 13, where the only available interpretation of the superlative requires the exclusion of bags that do not contain a rabbit, even if they are bigger. The goal of this section is to show that comparative adjectives like *bigger* can also give rise to relative readings. The fact that *the rabbit in the bigger bag* is well-defined in the scenario depicted in the righthand side of Figure 13 is one way of seeing this. Furthermore, as we will show, relative readings of comparatives display all the previously discussed hallmarks of such readings; just like relative superlatives, relative readings of comparatives obviate definiteness effects, are blocked by possessives and by non-modal infinitival clauses, and give rise to similar ambiguities when the adjective modifies the head of a relative clause interpreted in the scope of an attitudinal propositional verb.

As Szabolcsi (1986: i.a.) pointed out, superlatives obviate definiteness effects. In (27a), where *have* takes a complement headed by the relational noun *sister*, the variant with the definite article is quite strange. But there is nothing out of the ordinary about (27b), in which a superlative modifier is added. As Szabolcsi points out (see esp. sec. 4), examples like (27b) require focus on the subject (here *Bernie*), and involve a relative reading of the superlative, in that the comparison is among the subject's focus alternatives. Comparatives, too, obviate definiteness effects when they modify a relational noun in object position of a *have*-sentence.

(27) a. Bernie has a/??the sister.
b. Bernie has the nicest sister.
c. Bernie has the nicer sister.

Just as in the case of the superlative; focus on *Bernie* is required in (27c). Like (27b), it could be used in answer to the question, *Who has the nicer/nicest sister?* but not *What is Bernie's family like?* (setting aside the elative/intensifier use of *nicest* in the case of (27b)). Thus comparatives behave exactly like relative superlatives with respect to this type of definiteness effect.

Second, as discussed by Bumford (2017), relative readings of superlatives are blocked by prenominal possessives:

(28) a. Who has read the longest play by Shakespeare?
b. Who has read Shakespeare's longest play?
≡ Who has read Hamlet?
Absolute Reading: Who has read the play by Shakespeare that is longer than any other play by Shakespeare?
Missing Relative Reading in (28b): Who has read a longer play by Shakespeare than anyone else has read?

In contrast to (28a), example (28b) can only be interpreted as a question about *Hamlet*, the longest play ever written by Shakespeare. It cannot be construed as asking who read a longer Shakespeare play than anyone else read.

Now let us consider the comparative versions:

(29) a. Who has read the longer play by Shakespeare?

b. #Who has read Shakespeare's longer play?
$\rightsquigarrow$ Shakespeare wrote two plays.
Absolute Reading: Of the two plays written by Shakespeare, who has read the longer one?
Missing Relative Reading (29b): Of the two contextually salient plays written by Shakespeare, who has red the longer one?

Example (29b) implies that Shakespeare only wrote a total of two plays. This is due to a general fact about comparatives in definite descriptions: they tend to be felicitous only when the comparison class contains exactly two elements. Since Shakespeare wrote more than two plays in his lifetime, the absolute reading leads to presupposition failure. In (29a), a relative reading is available, so the sentence is fine; in (29b), only an absolute reading is available due to the presence of the possessive, so the sentence is forced to carry a false presupposition.

A third parallel between superlative and comparative descriptions can be observed through a phenomenon discussed by Bhatt (2006). Bhatt points out that superlative descriptions associate with focus, giving rise to truth-conditionally different readings depending on the placement of the focused constituent (30a-b). Even though Bhatt does not cast the discussion in terms of the relative vs. absolute distinction, we point out that focus-sensitivity effects are contingent on a relative interpretation of the superlative. With an absolute interpretation, the different information structures exemplified in (30a-b) result in non-equivalent truth-conditions, as shown in the associated paraphrases.

(30) a. Joan$_F$ gave Mary the most expensive telescope.

b. Joan gave Mary$_F$ the most expensive telescope.

Relative reading for (30a): Some people gave Mary telescopes. Of all those telescopes, the telescope that was given by Joan was the most expensive one.
Relative reading for (30b): Joan gave some people telescopes. Of all those telescopes, the telescope Joan gave to Mary was the most expensive one.
Absolute reading: Joan gave Mary something, and that thing was the most expensive telescope among all the telescopes.

Bhatt observes that association with focus effects are blocked by non-modal infinitival relative clauses (31).

(31) Joan$_F$ gave Mary the most expensive telescope to be built in the 9th century.
Absolute reading: Of all the telescopes built in the 9th century,
Joan gave Mary the most expensive one.
Missing relative reading: Some people gave Mary telescopes built in the 9th century.
Of all those telescopes, the telescope that was given by Joan was the most expensive.

Bhatt's observation can be taken as evidence that non-modal infinitival relatives block relative readings of superlatives.

Examples (32)-(33) show that comparative adjectives pattern with superlatives in this respect.

(32) a. Joan$_F$ gave Mary the more expensive telescope.
b. Joan gave Mary$_F$ the more expensive telescope.
Relative reading for (32a): Two people gave Mary telescopes. Of all those telescopes, the telescope that was given by Joan was the more expensive one.
Relative reading for (32b): Joan gave two people telescopes. Of those telescopes, the telescope Joan gave to Mary was the more expensive one.
Absolute reading: Joan gave Mary a telescope. It was the more expensive of the two telescopes.

(33) Joan$_F$ gave Mary the more expensive telescope to be built in the 9th century.
Absolute reading: Of the two telescopes built in the 9th century, Joan gave Mary the most expensive one.
Missing relative reading: Some people gave Mary telescopes built in the 9th century. Of the two 9th century telescopes she received, the telescope that was given by Joan was the more expensive one.

A final piece of evidence comes from ambiguities such as (34b), again due to Bhatt (2002). Bhatt observes that superlatives modifying a noun heading a relative clause can be interpreted within the scope of a propositional attitude verb inside the relative clause. Example (34c) shows that comparative adjectives exhibit the same ambiguity.

(34) a. the long book that John said Tolstoy had written
b. the longest book that John said Tolstoy had written
c. the longer book that John said Tolstoy had written
High reading: of the books John said Tolstoy wrote, the long/longer/longest one
Low reading: the book John said was long/longer/longest among the ones written by Tolstoy.

On the high reading (34), both the superlative and the comparative adjective are interpreted against the comparison class of the books mentioned by John, whereas on its low interpretation, the relevant comparison class comprises the books written by Tolstoy. The ordinary positive-form gradable adjective *long* has only a high reading, whereas the comparative and superlative adjectives have low readings as well.

All of these diagnostics show that comparatives have relative readings just like superlatives. If we a Bumford-style analysis of relative readings is right for superlatives, then the analogous kind of analysis should be right for comparatives. In particular, we assume that comparatives can act as filters whose application is delayed in the dynamic sequence, effectively operating at a higher scope position.

B Experimental materials

The target items for Experiments 1 and 2 are listed in Tables 2 and 3 respectively.

The fillers for Experiments 1 and 2 are given in Tables 4 and 5 respectively.

Table 2: Target items for Experiment 1.

outer noun	prep.	inner noun 1	inner noun 2	distractor 1	distractor 2
rabbit	in	box	bag	basket	flower
frog	in	bathtub	bucket	boat	duck
cat	in	truck	tree	tower	sandwich
boy	with	pillow	paintbrush	pen	car
lady	with	fan	fish	flower	glue
girl	with	dog	doll	duck	cherries
man	with	socks	scarf	sandwich	basket
bird	in	can	cup	car	boat
elephant	with	glasses	grapes	glue	tower
farmer	with	cheese	chair	chessboard	pen
policewoman	with	vacuum	violin	vase	lamp
horse	with	ladder	letter	lamp	vase

Table 3: Target items for Experiment 2.

itemnum	outer1	outer2	prep	inner1	inner2	distractor
1	rabbit	frog	in	box	bag	basket
2	frog	bird	in	bathtub	bucket	boat
3	cat	bird	in	truck	tree	tower
4	boy	girl	with	pillow	paintbrush	pen
5	lady	man	with	fan	fish	flower
6	girl	man	with	dog	duck	doll
7	man	girl	with	sock	scarf	sandwich
8	bird	rabbit	in	can	cup	car
9	monkey	frog	with	glue	grapes	glasses
10	farmer	lady	with	cheese	chair	cherries
11	policewoman	boy	with	carrot	cookie	cake
12	horse	cat	with	ladder	lizard	lamp

Table 4: Fillers in Experiment 1.

number	type	phrase	image1	image2	image3	image4	image5
1a	*bleep* P the N1	*bleep* in the bathtub	rabbit-bathtub	cat-bag	basket	basket	flower
1b	*bleep* P the N1	*bleep* in the box	frog-box	bag	boat	boat	flower
2a	*bleep* P the N1	*bleep* in the truck	cat-truck	tower	truck	tree	sandwich
2b	N2 P the *bleep*	girl with the *bleep*	boy-pillow	girl-pillow	pillow	paintbrush	car
3a	*bleep* P the N3	*bleep* with the fish	elephant	lady	lady-fish	farmer	man
3b	*bleep* P the N1	*bleep* in the tree	girl-tree	girl-car	girl-truck	tower	tree
4a	N1 P the N1	man in the truck	man-truck	girl-truck	cat-truck	truck	truck
4b	N2	bird	dog	bird	cat	boat	bag
5a	N1 P the N1	elephant with the ladder	elephant-ladder	ladder	ladder	elephant-letter	letter
5b	N2	farmer	man	farmer	lady	girl	fish
6a	N1 P the N1	policewoman in the car	policewoman-car	policewoman-cat	man-car	farmer-car	policewoman-can
6b	N1 P the N1	horse with the scarf	horse-scarf	elephant-box	cat-bathtub	fish	rabbit-car
7a	N1 P the *bleep*	rabbit in the *bleep*	rabbit-cup	cat-cup	fish	rabbit	dog-cup
7b	N3	frog	fish	fan	frog	flower	farmer
8a	N1 P the *bleep*	cat in the *bleep*	cat-car	can	cat	car	cup
8b	N3	boy	bathtub	boat	boy	box	basket
9a	N3 P the N3	lady with the ladder	boy-box	farmer-fish	lady-ladder	girl-grapes	cat-car
9b	N3 P the N3	girl with the grapes	girl-glue	girl	girl-grapes	girl-glasses	grapes
10a	N2 P the *bleep*	man with the *bleep*	girl-tree	man-vase	lady-car	cat-boat	dog-boat
10b	N2 P the *bleep*	bird in the *bleep*	dog-boat	bird-boat	cat-boat	frog-boat	cat-boat
11a	*bleep* P the N3	*bleep* with the fan	horse-scarf	elephant-box	elephant-fan	policewoman-can	man-can
11b	*bleep* P the N3	*bleep* with the vacuum	farmer-violin	farmer-fish	farmer-vacuum	farmer-vase	farmer-glue
12a	N3	policewoman	lady	girl	policewoman	dog	pen
12b	N3	horse	duck	bird	horse	elephant	dog

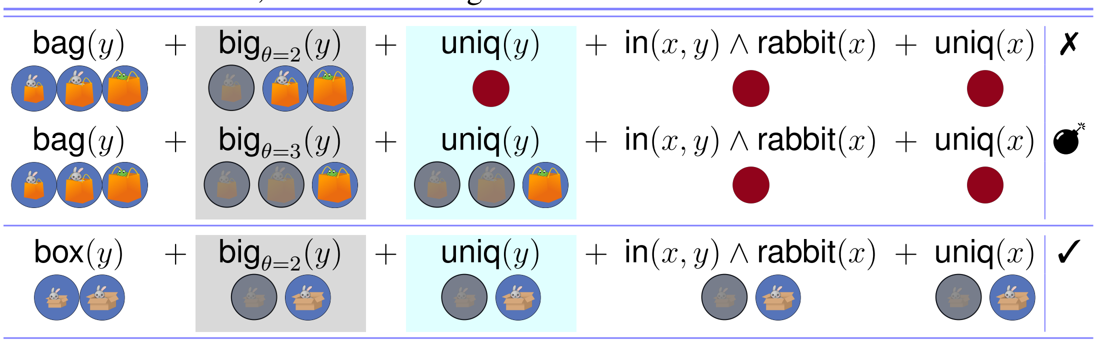
Table 5: Fillers in Experiment 2.

number	phrase	image1	image2	image3	image4	image5
1	the glasses	big_glasses	rabbit	big_sandwich	farmer-med_cheese	med_bathtub
2	the green ***	green_frog	brown_frog	violin	cat-med_truck	cat-small_truck
3	the blue fish in the ***	med_fish	red_fish	med_chair	farmer-med_cheese	gray_cat
4	the yellow ***	yellow_bird	girl-grapes	elephant	big_cherries	boy-box
5	the gray *** in the car	gray_cat	rabbit-big_car	cat-med_tree	big_doll	boy-med_paintbrush
6	the cheese	med_cheese	big_tower	med_tree	dog-big_cup	dog-small_cup
7	the tall ***	tall_man	short_man	big_glasses	fork	med_grapes
8	the tall tower with the ***	tall_tower-flag	short_tower-flag	elephant-glasses	bird-med_can	elephant-fan
9	the cat with the tall ***	cat-big_ladder	small_ladder	girl-big_dog	girl-med_dog	man-vase
10	the tall *** with the man	tall_policewoman-man	tall_policewoman	med_truck	med_sock	rabbit-big_car
11	the girl	girl	rabbit-big_box	rabbit-small_box	frog-big_bag	cat-cup
12	the short ***	short_woman	tall_woman	med_pillow	big_sandwich	violin
13	the short pencil with the ***	short_pencil-notepad	long_pencil-notepad	farmer-med_chair	farmer-big_cheese	farmer-small_cheese
14	the dog with the short ***	dog-short_lamp	big_lamp	big_flower	short_flower	man-vacuum
15	the short *** with the glass	short_pitcher-glass	tall_pitcher	man-big_truck	man-small_truck	med_truck
16	the car	big_car	elephant-box	girl-big_pillow	girl-truck	lady-med_fan
17	the taller ***	big_tree	small_tree	big_cherries	fork	med_fish
18	the taller building with the ***	tall_building-antenna	short_building-antenna	man-vacuum	man-med_scarf	horse
19	the car with the tall ***	car-tall_trafficlight	short_trafficlight	frog-small_bathtub	frog-big_bathtub	frog-med_bucket
20	the tall *** with the dog	tall_boy-dog	short_boy	big_ladder	girl-grapes	big_basket
21	the shorter ***	short_flower	big_flower	big_cherries	med_fish	chessboard
22	the short glass with the ***	short_glass-plate	tall_glass-plate	short_pencil-notepad	long_pencil-notepad	small_ladder
23	the cat with the shorter ***	cat-short_bottle	tall_bottle	tall_boy-dog	short_boy	girl-grapes
24	the shorter *** with the horse	short_farmer-horse	tall_farmer-horse	tall_tower-flag	short_tower-flag	big_tower

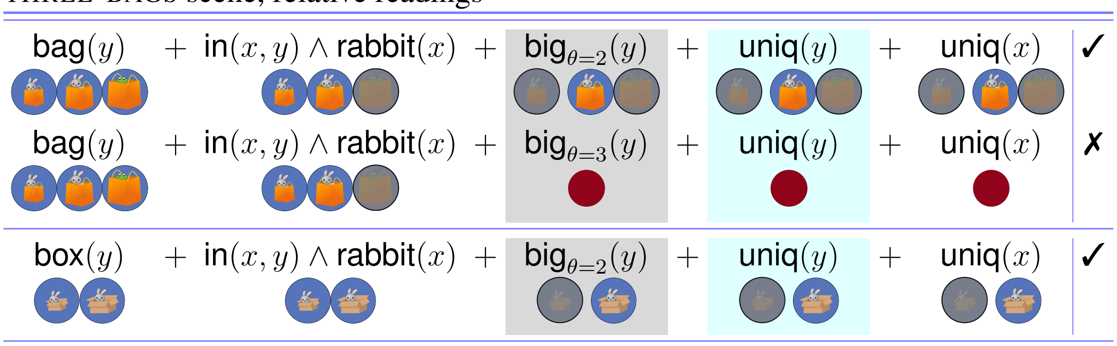
C Full set of derivations

C.1 Positive

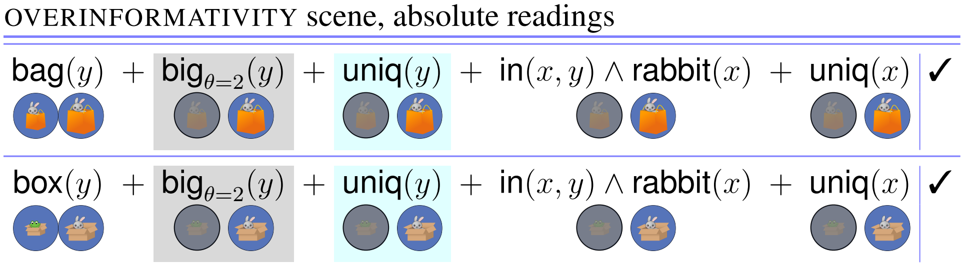
THREE-BAGS scene, absolute readings



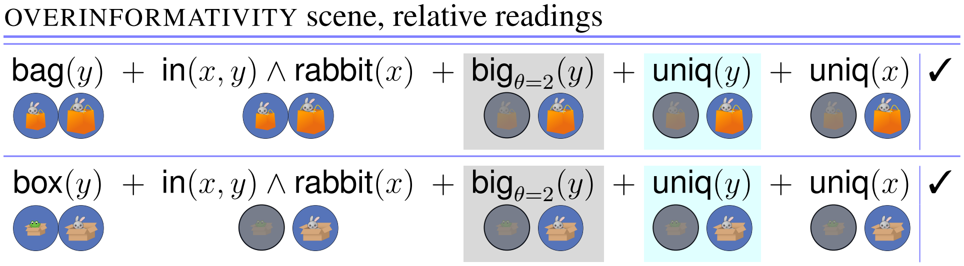
THREE-BAGS scene, relative readings



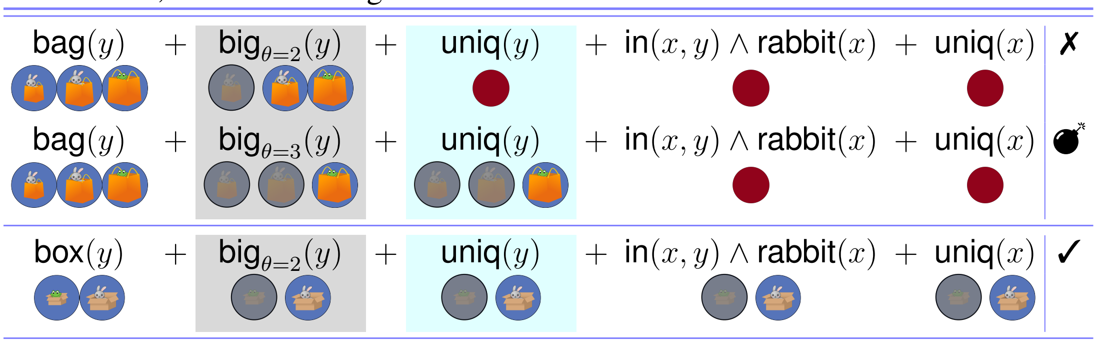
OVERINFORMATIVITY scene, absolute readings



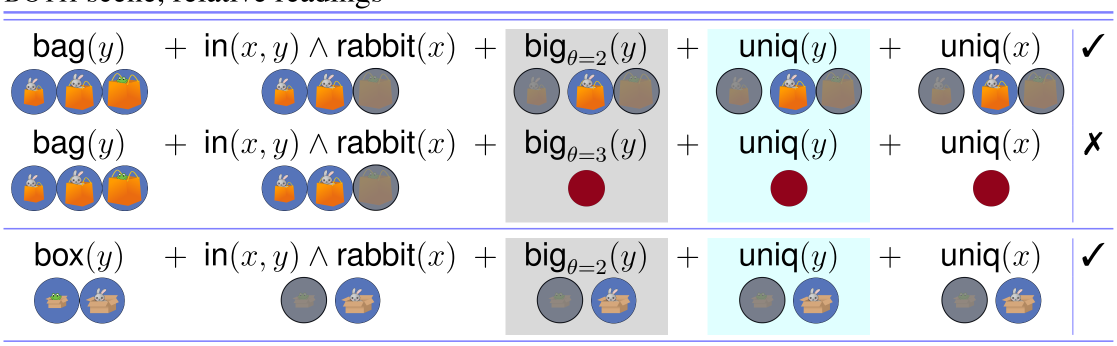
OVERINFORMATIVITY scene, relative readings



BOTH scene, absolute readings

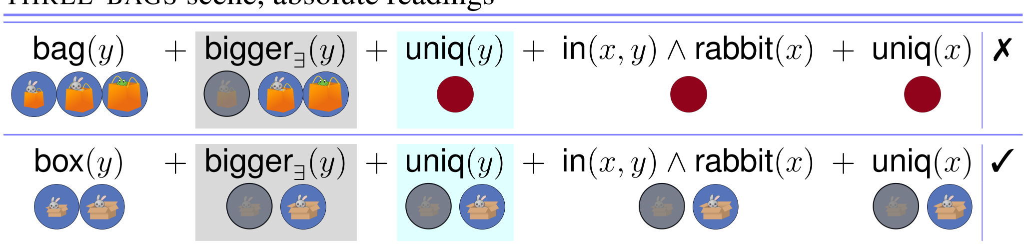


BOTH scene, relative readings

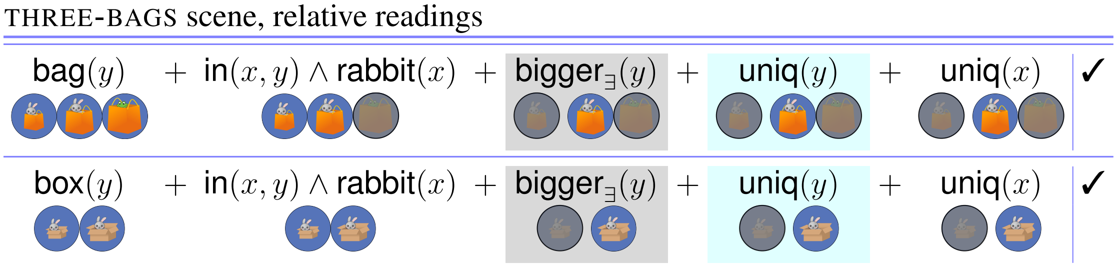


C.2 Comparatives

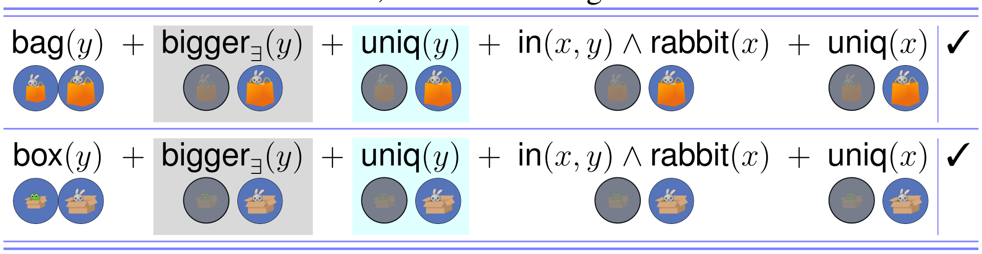
THREE-BAGS scene, absolute readings



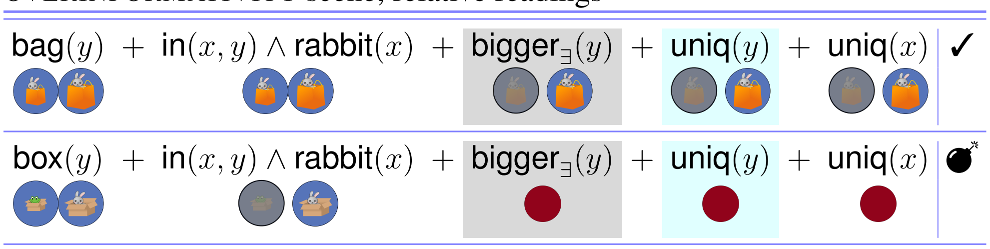
THREE-BAGS scene, relative readings



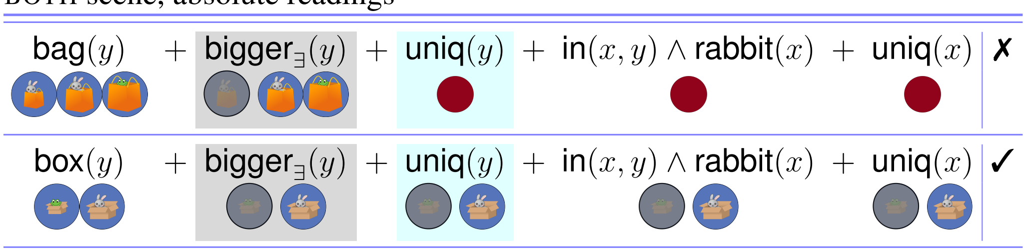
OVERINFORMATIVITY scene, absolute readings



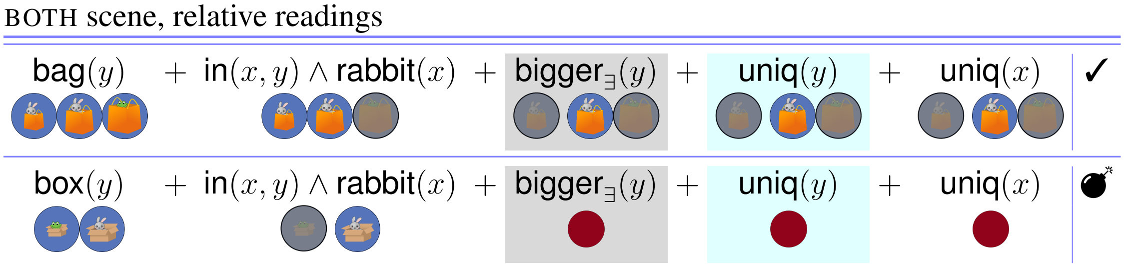
OVERINFORMATIVITY scene, relative readings



BOTH scene, absolute readings



BOTH scene, relative readings



Data Accessibility Statement: The data and analysis scripts, as well as the auditory and visual stimuli corresponding to Experiments 1 & 2 can be accessed at the following anonymized OSF repository: https://osf.io/asbnj/?view_only=c49c658e64eb4f9a8d1d153f3c692898.

References

Bever, T. G. (1970). The cognitive basis for linguistic structures. In Hayes, J. R., editor, *Cognition and the Development of Language*, pages 279–362. John Wiley & Sons.

Bhatt, R. (2002). The raising analysis of relative clauses: Evidence from adjectival modification. *Natural Language Semantics*, 10:43–90.

Bhatt, R. (2006). *Covert Modality in Non-finite Contexts*. Mouton de Gruyter, Berlin.

Bumford, D. (2017). Split-scope definites: Relative superlatives and Haddock descriptions. *Linguistics and Philosophy*, 40(6):549–593.

Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1).

Champollion, L. and Sauerland, U. (2010). Move and accommodate: A solution to Haddock's puzzle. In Bonami, O. and Hofherr, P. C., editors, *Empirical Issues in Syntax and Semantics 8*, pages 27–52.

Degen, J., Hawkins, R. X. D., Graf, C., Kreiss, E., and Goodman, N. D. (2019). When redundancy is rational: A bayesian approach to 'overinformative' referring expressions. *CoRR*.

Ferreira, F. and Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, 30(6):725–745.

Frazier, L. and Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14:178–210.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., and Lotocky, M. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37:58–93.

Groenendijk, J. and Stokhof, M. (1989). Type-shifting rules and the semantics of interrogatives. In Partee, B. H. and Turner, R., editors, *Properties, Types and Meanings: Semantic Issues*, volume 2, pages 21–68. Kluwer.

Groenendijk, J. and Stokhof, M. (1990). Two theories of dynamic semantics. In *JELIA 1990: Logics in AI*, pages 55–64. European Workshop on Logics in Artificial Intelligence.

Grudzińska, J. and Zawadowski, M. (2019). Inverse linking, possessive weak definites and Haddock descriptions: A unified dependent type account. *Journal of Logic, Language and Information*, 28:239–260.

Haddock, N. J. (1987). Incremental interpretation and Combinatory Categorial Grammar. In *Proceedings of the 10 International Joint Conference on Artificial Intelligence*, volume 2, pages 661–663. Morgan Kaufmann.

Heim, I. (1999). Notes on superlatives. Manuscript, MIT.

Kamp, H. and Partee, B. (1995). Prototype theory and compositionality. *Cognition*, 57:129–191.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., and Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, pages 109–147.

Sproat, R. and Shih, C. (1988). Prenominal adjectival ordering in English and Mandarin. In Blevins, J. and Carter, J., editors, *Proceedings of NELS 18*, pages 465–489. GSLA, University of Massachusetts, Amherst.

Szabolcsi, A. (1986). Comparative superlatives. In Fukui, N., Rapoport, T., and Sagey, E., editors, *Papers in Theoretical Linguistics*, pages 245–265. MITWPL, Cambridge, MA.

Trueswell, J. C., Tanenhaus, M. K., and Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(3):528–553.

van Eijck, J. (1993). The dynamics of description. *Journal of Semantics*, 10(3):239–267.