Cognition 240 (2023) 105563

Contents lists available at ScienceDirect

Cognition

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

The discourse functions of grammatical constructions explain an enduring syntactic puzzle

Nicole Cuneo, Adele E. Goldberg *

Princeton University, Princeton, NJ 08540, USA

ARTICLE INFO

Keywords:
Constructions
Communication
Island effects
Discourse function

ABSTRACT

Each grammatical construction serves a function, such as conveying that part an utterance is at-issue or is backgrounded. When multiple constructions combine to produce an utterance, their functions must be compatible. This preregistered study ($N = 680$) addresses the enigmatic case of "syntactic island constraints": Long-distance dependency constructions (LDDs) do not combine equally well with all base constructions. While widely presumed to require unlearned syntactic constraints, we test the idea that it is infelicitous to make an element both prominent (via an LDD construction) and backgrounded (via the base construction). Using 10 base constructions of English (144 base stimuli), results confirm two independent measures of backgroundedness strongly correlate with acceptability ratings on each of three LDD constructions. Results indicate that "island" constraints arise from a clash between the functions of the constructions being combined.

Each time we speak or sign, we combine words and grammatical constructions to express our intended message. Canonical English clauses license a particular order of constituents: in (1a), the subject argument (*The Library of Congress*) appears before the verb (*lent*), while the direct object (*a crystal flute*) appears immediately after. Certain constructions, including questions (1b) and relative clause constructions (1c), position a constituent a potentially far distance away from its canonical position. That is, to query a particular constituent, a *wh*-phrase can appear initially as in (1b); to modify a certain constituent with a relative clause as in (1c), the constituent again can be positioned before the clause in which it would canonically appear. Because the actual and canonical positions of a constituent may be separated by a long-distance, both *wh*-questions and relative clauses are often referred to as long-distance dependency (LDD) constructions (Kaplan & Zaenen, 1989).[1] LDD constructions are sometimes described as if a constituent (e.g., *what* or *which flute* in 1b) were "extracted" from its canonical position (indicated by underscores) and "moved" to its new position earlier in the utterance, although most researchers today acknowledge that any movement is purely metaphorical (Fodor, Bever, & Garrett, 1974).

(1) a. The Library of Congress lent a crystal flute to Lizzo.
b. What/Which flute did [the Library of Congress lend _ to Lizzo]?
c. People admired the crystal flute [the Library of Congress lent _to Lizzo].

In the current work, we refer to canonical sentences such as (1a), as base sentences, and sentences created by combining a base construction with a question construction (1b) or relative clause construction (1c) as LDD stimuli. Ross (1967) was the first to observe that the constituents of certain constructions do not felicitously combine with LDD constructions. Alluding to the idea that the constituents within such constructions are unable to "move" from their canonical position, he dubbed such constructions "islands," as if the constituents were unable to move off the island.

In what follows, we do not presuppose which constructions are islands or to what extent. Instead, island status is determined empirically. While a great deal of early work was based on introspective judgments, empirical work on islands has become increasing common (e.g., Christensen, Kizach, & Nyvad, 2013; Hofmeister, Casasanto, Sag, Sprouse, & Hornstein, 2013; Kluender & Kutas, 1993; Kush, Lohndal, & Sprouse, 2019; Sprouse, Wagers, & Phillips, 2012). Yet acceptability ratings are commonly collected only on one or two base constructions at a time. For example, there exists work on main verbs with clausal complements (e.g., Ambridge & Goldberg, 2008; Liu, Ryskin, Futrell, & Gibson, 2022; Liu, Winckel, Abeillé, Hemforth, & Gibson, 2022), and separate work on adjuncts (e.g., Bondevik, Kush, & Lohndal, 2021), but

* Corresponding author.
E-mail addresses: nicole.cuneo@princeton.edu (N. Cuneo), adele@princeton.edu (A.E. Goldberg).

[1] We use the term Long-Distance Dependency constructions (LDDs); alternative names include Filler-Gap constructions (e.g., Liu et al., 2022) or Unbounded Dependency constructions (Chaves & Putnam, 2021).

https://doi.org/10.1016/j.cognition.2023.105563
Received 8 February 2023; Received in revised form 31 May 2023; Accepted 11 July 2023
Available online 5 August 2023

few studies use the same methods to compare a broad range of base constructions simultaneously. Exceptions include Sprouse, Caponigro, Greco, and Cecchetto (2016), which compares judgments on four base constructions and Kush et al. (2019), which tests five base constructions in Norwegian. Critically, neither of these larger comparisons investigate the potential role of the functions of the base constructions.[2]

In fact, few studies include any empirical test for semantic or discourse factors which have been proposed as explanations for island effects (but see Ambridge & Goldberg, 2008; Cuneo & Goldberg, 2022; Liu, Ryskin, et al., 2022; Liu, Winckel, et al., 2022; Namboordiripad et al., 2022). The relative lack of empirical work examining the role of the functions of the base constructions means we do not know whether acceptability judgments on combinations of LDDs and base constructions depend on the functions of the constructions. The current work is aimed squarely at this question.

In addition, the majority of empirical work has investigated a single long-distance dependency construction, most commonly *wh*-questions (e.g., Ambridge & Goldberg, 2008; Namboordiripad et al., 2022) or relative clauses (e.g., Christensen & Nyvad, 2022). The few studies that tested judgments on both types of LDDs were designed so that judgments were not directly comparable (Abeillé, Hemforth, Winckel, & Gibson, 2020; Kobzeva et al., 2022; Sprouse et al., 2016). Thus, it remains unclear as to whether acceptability judgments on *wh*-question LDDs and relative clause LDDs can be explained by the same factors.

The current study addresses the role of the base constructions' function and compares *wh*-question and relative clause LDD constructions in a large-scale preregistered crowd-sourced study. Acceptability judgments are collected on instances of 10 base construction types, on their own and in combination with the LDD constructions. Acceptability ratings are then correlated with two separate and independent measures of a discourse factor that has been proposed to predict island status in prior work (e.g., Goldberg, 2006), namely the extent to which each base construction *backgrounds* its content.

Prior literature suggests that the shaded rows in Table 1 will be less island-like than the unshaded rows, but again, we do not rely on an a priori classification of stimuli into islands and non-islands. We instead ask whether ratings on long-distance dependency stimuli (e.g., *wh*-question examples in the rightmost column) follow directly from the acceptability of the base sentences (e.g., examples in the middle column) together with the added complexity of the long-distance dependency, or whether instead, the discourse measure of backgroundedness predicts additional variance. Kluender and Kutas (1993) first introduced this type of methodology as a way to detect islands: for a construction to be considered an island, judgments on long-distance dependency extractions must be judged less acceptable than what would be predicted simply by taking the acceptability of the base sentences without LDDs and the added complexity of the LDD into account (see also Sprouse et al., 2012).

By investigating the functions of the constructions involved, the current work aims to explain *why* certain constructions resist combination with long-distance dependency constructions. We hypothesize that unacceptability arises because of a clash between the functions of the grammatical constructions being combined: Long-distance dependency constructions foreground a certain constituent in their domain, while constructions that are islands relegate their content, to varying degrees, to backgrounded status (see also Goldberg, 2006, 2013). This hypothesis claims that it is infelicitous for a speaker to choose to both foreground an element (via a long-distance dependency) and background it (via the choice of a base construction that backgrounds its content).

The clash-of-functions explanation of islands can be illustrated with examples. In (2a), the sentence asserts the information in its verb phrase (*saved a child*). This information is prominent in the discourse (or "at-issue" in the terminology of Potts, 2004). Therefore, the information can be queried, as in (2b), by the question word, *who*, which is also prominent in the discourse.

(2) a. [The lifeguard [who worked in New Jersey]]$_{RC}$ saved a child.
b. Who did [the lifeguard [who worked in New Jersey]]$_{RC}$ save _?

Similar propositional content, as is expressed in (2a), is likewise conveyed by (3a), although the content is "packaged" differently in the two sentences. In (3a), the speaker conveys the heroic act within a relative clause (RC), which generally expresses backgrounded information. We suggest therefore that (3b) involves a clash of functions: the constituent, *Who,* is at-issue, and yet it corresponds to information within a relative clause, which is backgrounded. Data reported below confirm both the backgrounded status of relative clauses and the fact that examples like (3b) are judged to be quite odd.

(3) a. [The lifeguard [who saved a child]]$_{RC}$ worked in New Jersey.
b.?? Who did [the lifeguard [who saved _]]$_{RC}$ work in New Jersey?

Specifically, the current results demonstrate that the *discourse properties* of base sentences, like those in (2a) and (3a)—which do not contain any island violations—predict the acceptability of long-distance dependencies like those in (2b) and (3b), more than they predict the acceptability ratings of (2a) and (3a) themselves. The proposed explanation of island constraints is characterized by the BCI constraint stated in (A) (see Ambridge & Goldberg, 2008; Cuneo & Goldberg, 2022; Goldberg, 2006, 2013; Namboordiripad et al., 2022):

(A) **BCI: Backgrounded constructions are islands**: Constructions are islands to long-distance dependency constructions to the extent that their content is backgrounded within the domain of the long-distance dependency construction.

The BCI essentially claims that it is infelicitous for a speaker to both foreground a constituent (via a long-distance dependency construction) *and* background it (via the base construction that the constituent semantically depends on). It is worth clarifying the following aspects of the BCI constraint:

BCI.1. The terms *prominence, foregrounding*, and being *at-issue* are intended to be interchangeable and will be operationalized here in two ways (via Negation and Discourse tasks).

BCI.2. Acceptability is gradient and island status (the degree of unacceptability not attributable to the acceptability of the base construction nor to the complexity of the LDD) is predicted to vary with the *degree* of backgroundedness.

BCI.3. The BCI holds of the domain targeted by the long-distance dependency construction. In *wh*-questions, the domain is the entire question: the *wh*-word is prominent with respect to the *wh*-question. In relative clauses, the domain of the long-distance dependency is the Noun Phrase: the head noun is prominent within the NP.

A quite different explanation of island status, widely taken for granted until recently, is formulated purely in syntactic terms. This perspective claims that a constituent cannot "move" past two "bounding nodes," where bounding nodes are defined as clauses or noun phrases (i.e., "IPs" or "DPs") (e.g., Abels, 2017; Chomsky, 1973; Sprouse et al., 2016; see Boeckx, 2012 for review). To the extent that it is valid, this generalization offers a tidy abstract description. But it cannot be said to provide an *explanation* as it does not address *why* dependencies that hold across multiple clauses or noun phrases should be problematic. Moreover, the structural generalization does not predict any systematic variation in judgments among sentences with the same structure

[2] Sprouse et al. (2012) recognize this omission explicitly, stating, "here and throughout we assume the standard linguistic position that island constraints are the result of grammatical constraints, admit that this [an explanation of island effects] is an interesting question in its own right." pg 8.

Table 1
Construction types included as stimuli, along with example base sentences and corresponding *wh*-questions stimuli.

Construction types	Sample base sentence stimuli	Sample *wh*-question stimuli ('gap' site underlined for illustration)
Main Clause	The woman who called Uber for a ride lost her glasses.	What did the woman who called Uber for a ride lose_?
Relative Clause	The woman who lost her glasses called Uber for a ride.	What did the woman who lost _call Uber for a ride?
Causal Adjuncts	He researched it by comparing prices.	What did he research the question by comparing _?
Temporal Adjuncts	He researched it after comparing prices.	What did he research the question after comparing _?
Verb Complements (potentially "bridge")	Bill said that Skyler recited a poem.	What did he say that Skyler recited_?
Verb Complements (unlikely "bridge")	Bill discovered that Skyler recited a poem.	What did he discover that Skyler recited_?
Double Object recipient	Daisy showed him an insurance policy.	Who did Daisy sell _ an insurance policy?
Prepositional paraphrase recipient	Daisy showed an insurance policy to him.	Who did Daisy sell an insurance policy to_ ?

(Francis, 2021). The current study includes two such cases. The first involves LDDs that target an element within the complement clause of a main verb such as *think, mumble, claim, complain,* etc. There is no non-circular way to predict judgment differences on these sentences without appealing to meaning, discourse and/or frequency since the form is held constant (Ambridge & Goldberg, 2008; cf. Baltin, 1982). The second relevant case involves non-finite adjuncts, which vary in their semantic relationship to the main clause, while they all involve dependencies across the same "bounding nodes." For these two cases in particular, any difference in acceptability demands a non-structural explanation (see also Namboordiripad et al., 2022). More generally, structural descriptions do not predict that acceptability judgments on long-distance dependencies should correlate with any particular discourse property that holds of sentences that do not contain any violations themselves.[3]

Nonetheless, in an exploratory analysis we test an additional factor, the distinction between whether the LDD targets a main clause (crossing only one clause boundary) or a subordinate clause (crossing two clause boundaries). This factor, characterizable in structural terms, turns out to account for additional variance. At the same time, for reasons to be discussed, we argue that this factor also follows from the BCI proposal.

The BCI hypothesis is related to the hypothesis proposed by Erteschik-Shir and Lappin (1979) as stated in (B):

(B) **Dominance:** An NP can only be extracted out of clauses which may be interpreted as dominant or out of phrases in which the NP may itself be regarded as dominant (Erteschik-Shir & Lappin, 1979: 51).

A constituent was defined as "dominant" if and only if the speaker intended the utterance to direct the listener's attention to the constituent (1979: 43; see also Deane, 1991; Kuno, 1987; Polinsky, 1998; Richter & Chaves, 2020; Takami, 1989; Van Valin, 1998). While the early definition of dominance relied on speaker's intentions, the authors recognized that grammatical forms are associated with discourse-functions, as the BCI makes explicit (see also Goldberg, 2006, 2013; Newmeyer, 2016). The BCI is also explicitly gradient while Dominance was intended to be binary (see also Ambridge & Goldberg, 2008; Chaves

[3] Many other types of examples that defy structural descriptions are recognized to exist. These include presentational relative clauses, presupposed conjuncts, and so-called "picture noun phrases" (e.g., Chaves, 2013; Deane, 1991; Lakoff, 1986; McCawley 1981; Ross, 1967). The current stimuli do not include any of these low-hanging fruit, although pilot testing indicates, as expected, that they too are explicable in terms of the BCI.

& Putnam, 2021; Hofmeister & Sag, 2010; Richter & Chaves, 2020).

A recent related proposal is the Focus-Backgrounded Constraint (FBC), stated below (Abeillé et al., 2020). Like the BCI and Dominance, the FBC is based on the idea that islands result from a clash between the discourse-functions of the constructions being combined.

FBC (Focus proposal): A focused element should not be part of a backgrounded constituent.

Note that if "focused" is replaced with "prominent," "foregrounded," or "at issue," the FBC is indistinguishable from the BCI, particularly if we grant backgroundedness a gradient interpretation as is intended (Abeillé et al., 2020). However, the authors intend the FBC to differ from the BCI in one particular way. We refer to the FBC, as articulated by Abeillé et al. (2020), as the Focus proposal to help clarify this and minimize the use of acronyms. Like the BCI, the Focus proposal recognizes that the *wh*-word in a question is a focus, and both proposals predict that *wh*-questions cannot target elements within backgrounded constructions. However, the Focus proposal is explicitly *not* intended to apply to relative clause LDDs. Instead, constraints on relative clause extractions are argued to be due factors other than a clash of functions; specifically, constraints on relative clauses are argued to be due to the general preference to minimize the distance between filler and "gap." Abeillé et al. (2020) support this idea by demonstrating that subject-extracted relative clauses, which minimize the distance between filler and gap, are judged more acceptable than object-extracted RCs.

In agreement with the Focus proposal, we recognize a distinction between the discourse functions of questions and relative clauses. In fact, we take it as axiomatic that each distinct grammatical construction serves a distinguishable range of functions (see also Bolinger, 1968; Clark, 1987; Givón, 1995; Goldberg, 1995; Langacker, 1987). Therefore, we do not presuppose that all long-distance dependencies behave alike (see also Postal, 1994; Ross, 1987; Sag, 2010). Yet the BCI proposal only requires that long-distance dependencies require the "extracted" constituent be prominent within the domain characterized by the long-distance dependency. The earlier and non-canonical positioning of the head noun of a relative clause indicates its status as prominent within the domain of the LDD construction. That is, the *wh*-word in a question corresponds to a discourse-prominent element with respect to the question construction, and the head noun in a relative clause corresponds to a discourse-prominent element with respect to the relative clause construction. In summary, the BCI predicts that constraints on relative clauses and questions should correlate with one another, while the Focus proposal predicts that judgments on question LDDs and relative clause LDDs should diverge since they are hypothesized to be due to unrelated factors.

In the current study, acceptability ratings on both *wh*-question LDDs and relative clause LDDs are compared to acceptability ratings on the same set of base stimuli (recall Table 1). This allows us to compare the predictions of the BCI and Focus proposals. We also included a set of "discourse-linked" *wh*-questions, which include a head noun along with the *wh*-word (e.g., *Which boxes*; *What time*), because relative clauses include a contentful head noun, while simple *wh*-questions do not, and we wish to address this potential confound (see also, Abeillé et al., 2020; Sprouse et al., 2016). A sample set of stimuli, distributed across four independent surveys, is provided in (6):

(6) Example main clause with relative clause stimulus set (extraction from main clause)
Base: The door that was blue led to the garage.
Wh-Q: Where did the door that was blue lead to _?
RC: They found the garage that the door that was blue led to _.
D-Q: What room did the door that was blue lead to _?

Before we turn to the current study, let us consider a final type of explanation for certain island effects, namely that unacceptability is due to insufficient familiarity with the relevant constructions (Bybee, 2007). There are two types of accounts that appeal to frequencies in the input. One proposal is that the frequencies of particular combinations of base construction + long-distance dependency construction directly predict acceptability judgments (Dabrowska, Rowland, & Theakston, 2009; Verhagen, 2005). According to this proposal, for instance, questions of the type in (7) are more acceptable than those of the type in (8) because instances of (7) are far more frequent than instances of (8), as indicated here by frequencies provided from the Contemporary Corpus of American English (Davies, 2008):

(7) *Wh*- DO <pronoun> THINK <pronoun>? 1758 tokens
(8) *Wh*- DO <pronoun> FORGET <pronoun>? 0 tokens

Indeed, >90% of all *wh*-questions that "extract" a constituent within a clausal complement involve the main verb *think* or *say* in English (Dabrowska et al., 2009), and also in Dutch (Verhagen, 2005). As predicted by these accounts, *wh*-questions involving *think* and *say* are judged highly acceptable. Yet the proposal does not explain why these two verbs should occur so frequently in *wh*-questions in the first place, in both English and Dutch. It also does not predict that long-distance dependency constructions other than *wh*-questions (e.g., relative clauses), should favor "extraction" from the same verbs' complements, a prediction to be tested in the current work.

A different frequency-based proposal suggests that acceptability ratings on long-distance dependencies result from the additive effect of two simple factors: the complexity of the long-distance dependency constructions plus the input frequency of the base type of sentence (Kothari, 2008). In support of this idea, Liu, Ryskin, et al. (2022) report that the acceptability ratings of both declaratives and *wh*-questions are directly predicted by how frequently the main verb has been witnessed with a clausal complement. This proposal implies that clausal complements of verbs are not technically islands because no *super-additive* effect was found: i.e., no decrement in acceptability beyond what is expected from the base sentences and the additional complexity of the long-distance dependency construction (Sprouse et al., 2012). We investigate this claim about verbal complements in the current data as well. Importantly, in what follows, we offer two independent ways of operationalizing backgroundedness.

1. Testing the BCI account of islands

To test the BCI proposal, it is critical to provide an independent test of backgroundedness, although this step is often overlooked. Here we operationalize backgroundedness in two different ways: via a Negation task, adapted from prior work (Ambridge & Goldberg, 2008; Liu, Ryskin, et al., 2022; Liu, Winckel, et al., 2022; Namboordiripad et al., 2022) and via a new Discourse task introduced here. Our goal is to determine whether the degree of island status is predicted by one or both of these measures, which each depend on constructions' functions rather than their syntactic forms. If the backgroundedness measures predict acceptability ratings on long-distance dependencies more than base sentences, it will support the claim that the more the content of a construction is backgrounded, the less available it is for a long-distance dependency, providing positive evidence that islands depend on the discourse functions of constructions.

We therefore preregistered the hypothesis that results would reveal an interaction: the independent measures of backgroundedness should predict acceptability ratings on the long-distance dependency stimuli better than they predicted ratings on the base sentences (SI.1). This finding would be striking, since both measures of backgroundedness, as detailed in the following section, include base sentences and not the long-distance dependency stimuli: i.e., neither method of operationalizing backgroundedness includes any stimuli involving illicit "extraction." The relative acceptability of base sentences was confirmed in a separate survey, and ratings on the base sentences are included in all analyses.

A set of 144 base sentences was constructed to include 24 sentence stimuli for each of the six pairs of constructions compared in the discourse task, namely:

(1) 24 main clauses w/ temporal adjuncts (extraction from main clause or verbal adjunct)
(2) 24 main clauses with relative clauses[4] (extraction from main clause or relative clause)
(3) 12 sentences with causal adjuncts +12 sentences temporally adjuncts (extraction from adjunct)
(4) 12 double object sentences +12 matched prepositional phrase sentences (extraction of the recipient)
(5) 12 sentences with co-referential arguments (potential parasitic gaps) + 12 sentences with non-coreferential arguments
(6) 24 sentences with clausal complements (extraction from the clausal complement)

To summarize, accounts based on the syntactic properties of long-distance dependencies do not predict a correlation with degree of backgroundedness nor any particular property related to the way information is packaged in sentences that do not include any illicit LDDs. Evidence for the current hypothesis would also present a challenge to a general account of island effects based on the familiarity or frequency of particular constructions, because familiarity can be expected to predict acceptability ratings on the base sentences and LDD stimuli equally (Liu, Ryskin et al., 2022; Robenalt & Goldberg, 2015; Sprouse et al., 2012). Finally, if acceptability ratings on *wh*-questions, discourse-linked *wh*-questions, and sentences containing relative clauses are found to correlate, it would undermine the Focus proposal, which claims that relative clauses are uninfluenced by backgroundedness. Immediately below, we operationalize the discourse function of backgroundedness in two ways, via a Negation task and a new Discourse task, described in turn.

2. Operationalizing backgroundedness

2.1. Negation task

To say that a construction *backgrounds* its content implies that the content is taken for granted, or not at-issue. The Negation task, first employed experimentally by Ambridge and Goldberg (2008), capitalizes on the fact that negating a sentence does not negate content that is presupposed or taken for granted (Karttunen, 1973). While presuppositions are often assumed to be binary, Ambridge and Goldberg (2008) consider backgroundedness to be gradient. They therefore provided participants with a 5-point Likert scale, which we use here as well (see also Liu, Ryskin et al., 2022; Namboordiripad et al., 2022). In particular, we asked participants to rate the extent to which negated versions of base sentences imply the content of the target construction. Two examples are provided in Table 2. According to this task, constructions are backgrounded to the extent that they are impervious to sentence negation.

2.2. Discourse task

We also introduce a new Discourse task based on the claim that backgrounded constructions tend to be ill-suited for providing direct, cooperative responses to prompts for information (Goldberg, 2006, 2013; Morgan, 1975). That is, if a speaker wishes to supply information directly, the information should be made prominent in the reply rather than backgrounded. In this task, each trial prompts participants for certain information (e.g., *Tell me why the puppy is so happy*). Participants are then asked to select which of two sentences is the "more direct and cooperative" response. The two options are both base sentences and both provide the requested information, but they differ in *how* the requested information is conveyed. One response provides the requested information within a construction hypothesized to be backgrounded while the other provides the requested information within a construction hypothesized to be non-backgrounded. Response pairs are minimally different from one another, as they controlled for overall complexity, syntactic form and, to the extent possible, propositional content. For instance, to test whether adjunct clauses serve as a felicitous construction in which to convey requested information directly, both response options contain a main clause and an adjunct clause. The only difference was whether the requested information is provided within the main or adjunct clause.[5] Four sample stimuli used in the Discourse task are provided in Table 3.

Importantly, none of the response options contained any island violations, and all were intended to be reasonably acceptable and natural when presented in a neutral context. Ratings on these, the base sentences, are included in all analyses. If supplying requested information within an "island" construction is considered less direct and cooperative than supplying the same information in a non-island construction, it will support the claim that island constraints are the result of the discourse function of the constructions involved. More generally, we hypothesize that the extent to which information expressed in a target construction is construed to be backgrounded in discourse will predict the extent to which the target construction is an island for combination with a long-distance dependency construction. Specifically, as described in detail below, we predict that the acceptability of sentences containing the long-distance dependency constructions — relative clauses and two types of *wh*-questions — with a "gap" in a target construction will be predicted by one or the other backgrounded measure. We predict an interaction: the backgroundedness measures will predict acceptability ratings on the long-distance dependency sentences more strongly than they predict acceptability ratings on the base sentences.

To foreshadow results, the current data demonstrate that the degree to which constructions background their content in base sentences predicts judgments on *wh*-questions, discourse-linked questions, and relative clauses, across a wide range of constructions. These include various non-finite adjuncts and clausal complements of a variety of main verbs, undermining a purely structural account. In the case of main-verb + complement clauses, we find that log-frequencies correlate with the backgroundedness measures, adding further support to the BCI. We find that acceptability ratings on *wh*-questions and relative clause correlate highly, indicating that both types of long-distance dependencies are subject to the same constraints, counter to the predictions of the Focus proposal. Finally, in an exploratory analysis, we find that whether LDDs target a main or subordinate clauses plays a role, and this distinction also correlates with the foreground/background distinction. Thus, the current results are consistent with the BCI proposal: Put simply, it is infelicitous for a speaker to simultaneously choose to foreground and background the same element.

3. Experiment

Preregistrations and Data Availability The number of participants, exclusion criteria, stopping rule, and analyses were preregistered

[4] Note that the restrictive relative clause construction is used in two ways: as a type of LDD construction, and also as one of the 10 target constructions. Restrictive relative clauses are generally considered islands, meaning that they cannot be combined, for instance, with another relative clause or *wh*-question. They are also a type of LDD construction in that the head noun of a relative clause appears in a non-canonical position before the clause which licenses it.

[5] We considered asking participants how direct and cooperative each statement was on its own as a response to the request for information, rather than a choice between two options, but an initial pilot study found that people struggled to understand the task. They seemed to rate responses randomly or on the basis of plausibility rather than grammar. By offering the chance to compare two responses that contained similar propositional content, we enabled participants to recognize that we were interested in how information was packaged rather than in the information content itself.

Table 2
Two example stimuli used in the Negation task. Responses tending toward NO are considered less backgrounded, more at issue.

Please think about what the following sentence means:	
His rowing club that meets at the lake doesn't start at 6:00 am.	*(Querying Main Clause)*
Does his rowing club start at 6:00?	
■No ■Probably Not ■Can't tell ■Probably ■Yes	
Please think about what the following sentence means:	
The owner did not get Fido outside while giving him treats.	*(Querying Adjunct)*
Did the owner give Fido treats?	
■No ■Probably Not ■Can't tell ■Probably ■Yes	

Table 3
Four example English stimuli used in the Discourse Task: response options that are chosen a higher % of the time are deemed to encode the requested information in a less backgrounded, more at-issue construction. (Information in italics included here only for exposition).

Tell me why Ali got up so early.		
Response options	His rowing club that meets at the lake starts at 6:00. *(Main Clause)*	His rowing club that starts at 6:00 meets at the lake. *(Relative Clause)*
Tell me why that puppy is so happy.		
Response options:	The owner got Fido outside by giving him treats. *(Causal Adjunct)*	The owner got Fido outside while giving him treats. *(Temporal Adjunct)*
Tell me what you did in the garden.		
Response options:	I planted a tree without watering it. *(Parasitic coreference)*	I planted a tree without watering the flowers. *(Non-parasitic reference)*
Tell me why Iris took time off from school.		
Response options:	Dan heard that she wasn't feeling well. *(Clausal complement of a clausal verbal complement of* hear*)*	Dan hated that she wasn't feeling well. *(Clausal complement of the clausal verbal complement of* hate*)*

(see SI.1). We planned and performed the study in the following parts, which were each preregistered before data collection and are combined here: https://researchbox.org/1014.

a) Using the Discourse task to predict judgments on *wh*-questions and base stimuli
b) Using the Negation task to predict judgments on *wh*–questions and base stimuli

c) Gathering judgments on *wh*-questions, relative clauses, and "discourse-linked" *wh*-questions)

3.1. Participants

As preregistered, a total of 680 participants took part in the experiment. For the acceptability judgments, 480 participants divided into four unique groups of 120 participants each completed an acceptability survey on either 1) base sentences, 2) *wh*-questions based on base sentences, 3) sentences that include restrictive relative clauses created from base sentences, or 4) "discourse-linked" *wh*-questions created from the base sentences. A separate group of 120 participants took part in the Negation task, and a final group of 80 participants took part in the Discourse task. All participants were recruited via Prolific. Based on demographics provided, mean age = 37.32; 268 participants self-identified as female, 275 as male, 15 as nonbinary, and 122 participants preferred not to say. As preregistered, subject quality was ensured by requiring participants to respond appropriately on at least 75% of catch trials to be included in the analysis. Participants were recruited until the target number was reached.

The study used a 2 × 1 design, crossing SENTENCE TYPE (base vs. long-distance dependency sentences) with DEGREE OF BACKGROUNDEDNESS, measured and analyzed independently by the Negation task or the Discourse task. Ordinal analyses test for an interaction between sentence type and degree of backgroundedness in predicting acceptability ratings. Included were random intercepts and slopes for item, and random intercepts for base constructions and subjects (measures were all collected between-subjects). We report exploratory analyses of individual constructions in more depth after the main analyses.

3.2. Acceptability ratings

Acceptability judgments were provided on a 7-point Likert scale, with 1 being "very unnatural" and 7 being "very natural" for the four sets of stimuli below:

3.3. Base stimuli acceptability survey

1. 144 base sentences were created by constructing 12 sentence stimuli for each of 10 English constructions illustrated in Table 1. Stimuli were created in pairs to be used in the Discourse task so that key information to be queried was packaged differently, while the propositional content was the same or minimally distinct if required by the constructions being compared.

3.4. Long-distance dependency stimuli acceptability surveys

Three additional surveys were created to collect judgments on the following three types of long-distance dependency constructions, created based on the base stimuli.

2. Wh-Question (Wh-Q) survey: 144 *wh*-questions query a constituent within each base construction
3. Relative Clause (RC) survey: 144 transitive sentences containing a restrictive relative clause formed with its head noun licensed by each target construction
4. "Discourse-Linked Questions" (D-Qs): 144 *wh*-questions with a head noun in the *wh*-phrase, querying a constituent within each target construction.

All stimuli are available at https://researchbox.org/1014.

Thus, acceptability ratings were collected on a total of 576 stimuli (= 144 stimuli of 4 types). Each of the 4 surveys was divided into 4 lists which included 36 target items each, quasi-randomly assigned:In order to minimize explicit comparisons, no participant saw any items that were highly similar semantically: i.e., the two sentences of a given response-pair in the Discourse task were assigned to distinct lists. In order to avoid strategic responding, each participant saw a single type of stimuli: base sentences, *wh*-Qs, RCs, or D-Qs. In order to avoid potential satiation effects which dull participant judgments after repeated exposure to the same type of stimulus (Chaves & Dery, 2019), no participant witnessed >4 instances of any of the 10 target constructions. In order to avoid fatigue effects, each participant was asked to rate only 36 stimuli and 12 filler sentences. Each participant judged a single list, in which the order of presentation was randomized for each participant.

3.5. Preregistered analysis

As preregistered, we predict an interaction, namely that responses on the Negation task and the Discourse task will each independently predict acceptability ratings on the corresponding long-distance dependency stimuli more than acceptability of the base sentences. Analyses used the clmm function from the ordinal package (Christensen & Christensen, 2015) because acceptability was rated on a 7-point Likert scale. The LDD variable was centered, and random intercepts and slopes (base vs long-distance dependency) were included for items, random intercepts were included for subject (unique groups of subjects judged each sentence type) and for construction type.

3.6. Replication

Reviewers of the initial submission identified several problematic stimuli, so we revised those items and reran the entire study with a new group of 680 participants, now on Prolific, rather than the Cloud Research platform as a front end on Mechanical Turk (Litman, Robinson, & Abberbock, 2017). Results reported below replicate those of the original version.

4. Results

Recall that we do not rely on an a priori classification of constructions into islands and non-islands. We are instead interested in whether acceptability of long-distance dependency stimuli correlates with performance on one or both independent measures of backgroundedness better than the base sentences do. As detailed below, results confirm that both measures of backgroundedness show the predicted interaction: backgroundedness (inversely) predicts acceptability judgments on long-distance dependency constructions more than base stimuli across the 10 target constructions tested.

4.1. Negation task and ratings

Since the Negation task indicates the extent to which information conveyed by each target construction is construed to be negated (at-issue), higher scores on the Negation task indicate that the target construction is less backgrounded. The predicted interaction is evident in Fig. 1. Individual acceptability ratings (zscored, for visualization) on base sentences are represented in blue and are close to horizontal, indicating little relationship between negation scores and acceptability ratings on these sentences. As is visible, scores on the Negation task (x-axis) predict acceptability ratings on each of the LDDs, as acceptability increases as the target content is judged to be less backgrounded (i.e., more negated by main clause negation).

The preregistered ordinal model confirms that scores on the Negation task predict acceptability ratings on the LDDs more than they predict acceptability ratings on the base sentences: i.e., the hypothesized interaction is significant ($\beta = 0.80$, $z = 7.62$, $p < .00001$). Full results of the preregistered model are provided in Table 4. As is to be expected, the long-distance dependency stimuli were judged less acceptable than their base sentence counterparts ($\beta = -4.88$, $z = -14.19$, $p < .00001$). There is also a correlation between acceptability ratings and negation scores across all stimuli ($\beta = 0.51$, $z = 7.5$, $p < .00001$). We tested whether

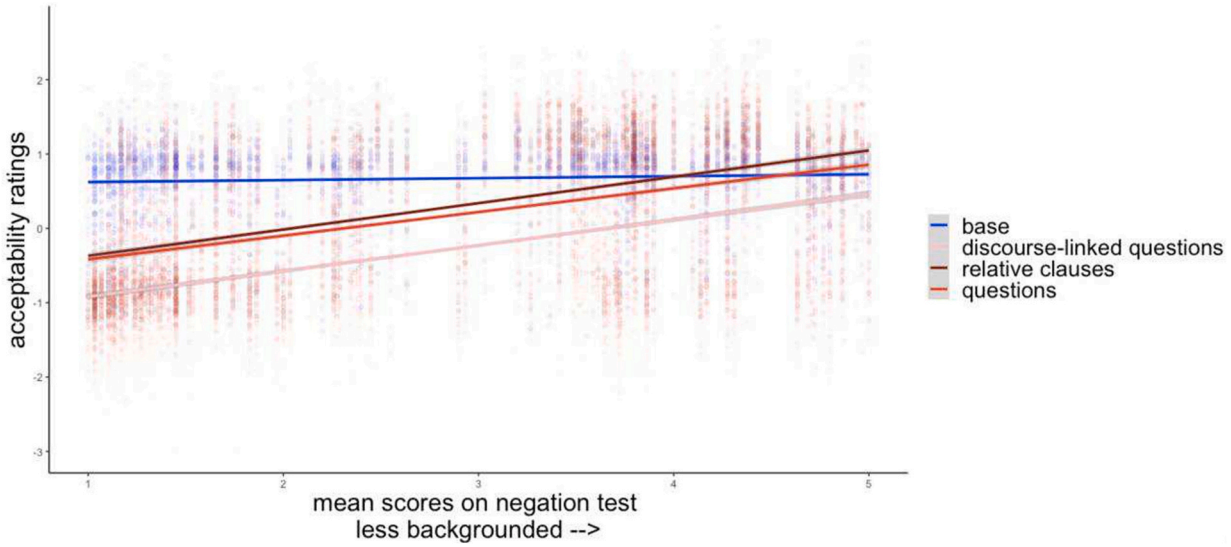


Fig. 1. Judgments on the Negation task of backgroundedness (x-axis) predict acceptability ratings (y-axis) on the long-distance dependency stimuli (shades of red), but not the base sentences (in blue) which are all judged to be relatively acceptable. Acceptability scores are (only) displayed using zscores to address individuals' different use of the scale. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Results of ordinal model confirm an interaction (boldfaced) between scores on the Negation task and acceptability ratings of LDDs vs base sentences.

clmm(raw_response ~ negation_mean*BaseorLDD + (1|subject) + (1 + BaseorLDD|item) + (1|Cx_Cx), data)

Random effects:	Groups	Name	Variance	SD	Corr
	subject	(Intercept)	0.95	0.98	
	item	(Intercept)	0.94	0.97	
		BaseorLDD	2.46	1.57	0.38
	Cx_Cx	(Intercept)	0.16	0.41	
Fixed effect Coefficients	Estimate	SE	z value	Pr(>|z|)	
Negation mean score	0.51	0.06	7.50	< 0.00001***	
Base vs LDD	−4.88	0.34	−14.19	< 0.00001***	
Negation Mean score: Base or LDD	**0.80**	**0.11**	**7.62**	**< 0.00001*** **	

scores on the negation task predicted acceptability ratings on just the base sentences, using a simple ordinal model with random intercepts for subjects, items, and constructions; we find no significant influence of negation scores on the acceptability of base sentences ($\beta = 0.13$, $z = -1.59$, $p = .112$), confirming that the blue correlation line in Fig. 1 does not significantly vary with negation scores. The interaction supports the BCI claim that backgrounded constructions are islands.

We additionally ran the same model on each of the three long-distance dependency constructions separately (along with the base stimuli) and find a significant interaction with comparable effect sizes and p values for each (see SI.4 for full models).

4.2. Discourse task and ratings

Recall the second way that backgroundedness was operationalized is based on the idea that constructions that provide information more directly should be more available for *wh*-extraction (i.e., less "island" like). The Discourse task measured how direct and cooperative responses were in a two-alternative forced-choice task, which prompted participants to supply information conveyed in one of two target constructions. The dependent measure was the proportion of times participants chose each base sentence as the preferred response.

The same preregistered model used for the Negation task measure was used to test the discourse measure of backgroundedness: an ordinal mixed model was fit to determine if acceptability ratings were predicted by Sentence Type (base vs. long-distance dependency), the discourse-based response preference, and their predicted interaction, with random intercepts and slopes for items, and random intercepts for subjects and construction type. Results confirm that responses on the Discourse task predict the acceptability ratings of *wh*-questions more than the acceptability of the base responses themselves ($\beta = 2.23$, $z = 4.32$, $p < .00001$). That is, as is evident in Fig. 2, the probability of a base sentence being selected as a direct and cooperative response (x-axis) predicted judgments on long-distance dependency stimuli (y-axis, shades of red), more than the acceptability of the responses themselves which were all base sentences (in blue).

Full results are provided in Table 5. As is to be expected, the long-distance dependency stimuli were judged less acceptable overall than their base sentence counterparts ($\beta = -3.73$, $z = -11.75$, $p < .00001$). There is an effect of scores on the Discourse task across sentence types ($\beta = 2.29$, $z = 8.43$, $p < .00001$). When we test whether scores on the Discourse predict acceptability ratings on just the base sentences in a simple ordinal model with random intercepts for subjects, items, and constructions, we find a significant effect remains ($\beta = 1.33$, $z = 4.15$, $p < .00001$). This suggests that sentences which are considered more acceptable overall are considered better responses. Crucially, the predicted interaction remains significant regardless.

As was done for the Negation task, we reran the same model on each of the three long-distance dependency constructions independently, and again find a significant interaction with the Discourse task scores, with comparable effect sizes for each (see SI.4 for full models).

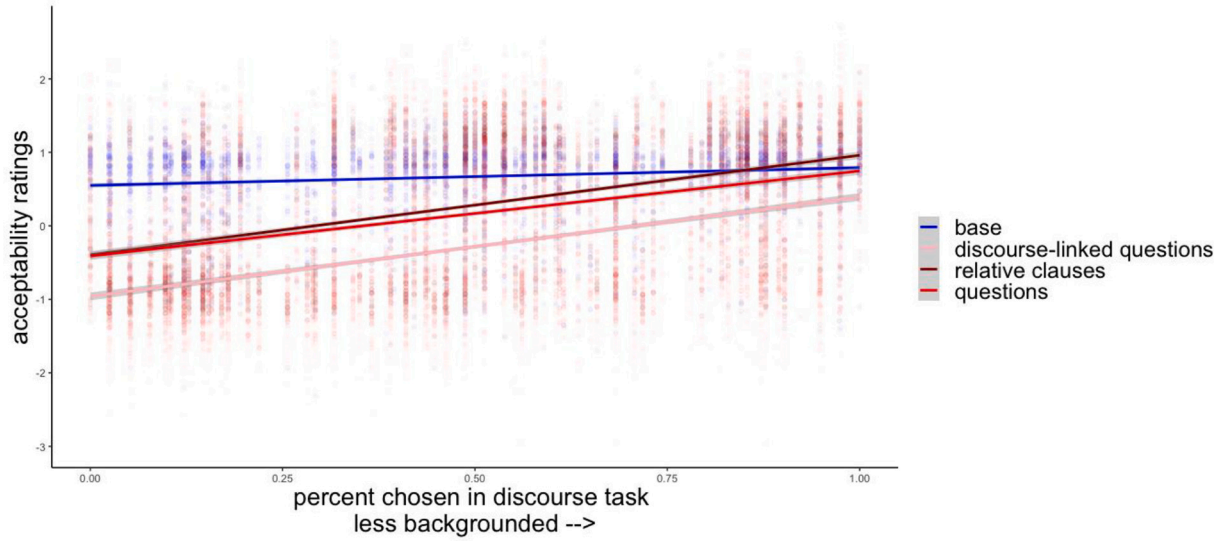


Fig. 2. Proportion of choices of one construction over another in the discourse measure of backgroundedness (*x*-axis) predicts acceptability ratings (*y*-axis) on the long-distance dependency stimuli (shades of red) more than the base sentences (in blue) which are judged to be relatively acceptable. $N = 680$ participants across 4 acceptability surveys and the independent Discourse task. Acceptability scores are shown in z-scores to account for different participants' use of the scale. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
Results of ordinal model confirm an interaction (boldfaced) between scores on the Discourse task and acceptability ratings of LDDs vs. base sentences.

clmm(raw_response ~ percent_chosen*BaseorLDD + (1|subject) + (1 + BaseorLDD | item) + (1|Cx_Cx), data)

Random effects:	Groups	Name	Variance	SD	Corr
	subject	(Intercept)	0.96	0.98	
	item	(Intercept)	0.86	0.93	
		BaseorLDD	3.11	1.76	0.47
	Cx	(Intercept)			
Coefficients:	Estimate	SE	z value		Pr(>|z|)
Percent chosen	2.29	0.27	8.43		< 0.00001***
Base vs LDD	−3.73	0.32	−11.75		< 0.00001***
Percent chosen: Base vs. LDD	**2.23**	**0.52**	**4.32**		**< 0.00001***

Table 6
Correlations between acceptability ratings on 144 items in *wh*-questions, relative clauses and discourse-linked questions. Each long-distance dependency type was rated by a separate group of participants.

	Wh-questions	Relative Clauses	Discourse-linked *wh*–questions
Wh-questions	1		
Relative Clauses	0.80	1	
Discourse-linked *wh*–questions	0.88	0.90	1

5. Correlations among the 3 long-distance-dependency sets of stimuli

As mentioned, and detailed in the SI, both measures of Backgroundedness predict the acceptability of each long-distance dependency construction considered separately, more strongly than they predict acceptability of the base sentences. Since long-distance dependency constructions serve a variety of functions (e.g., Abeillé et al., 2020; Birner & Ward, 1998; Gregory & Michaelis, 2001; Sag, 2010), we also test whether acceptability ratings on the 3 types of long-distance dependency stimuli correlate with one another, since the BCI predicts they should, while the Focus proposal predicts RCs are not subject to the same constraints. Table 6 shows the pairwise correlations for each pair of long-distance dependency constructions. Recall that separate groups of participants judged each LDD so that judgments on one could not influence the judgments on others directly.

Recall that the set of discourse-linked questions were included in order to compare judgments on *wh*-questions and relative clauses, since both discourse-linked questions and relative clauses provide a lexical head noun. Unsurprisingly, the two types of questions—*wh*-Qs and discourse-linked—correlate highly with one another ($r = 0.88$, $t = 22.15$, $p < .00001$). Strikingly, acceptability ratings on the long-distance dependency sentences involving relative clauses also correlated highly with those of both discourse-linked questions ($r = 0.90$, $t = 24.10$, $p < .00001$) and bare *wh*-questions ($r = 0.80$, $t = 16.10$, $p < .00001$). The strong correlations support the idea that *wh*-questions and restrictive relative clauses are subject to highly similar constraints and challenge the claim of the Focus proposal (the FBC) that relative clauses should not be influenced by measures of backgroundedness. In both questions and relative clauses, it is infelicitous for the construction that licenses the

prominent element to background the same content.

Since we used two measures of backgroundedness in the current study, we can ask whether they reliably correlate with one another. They do ($r = 0.46$, $p < .00001$). This indicates that both measures index related factors, although the moderate correlation confirms that the two measures are not identical.

We have seen that acceptability ratings on the three long-distance dependency constructions tested, which are highly correlated with one another, are predicted by each backgroundedness measure, across the 10 base constructions. Another way to visualize the results is to consider how the target constructions compare with one another. We explore this by using difference scores—the difference in acceptability between mean ratings per target construction on declarative stimuli and ratings on the long-distance dependency constructions.

6. Comparisons across construction types

The BCI proposal predicts that constructions that are less island-like, in terms of showing a smaller difference in acceptability rating between base and long-distance dependency stimuli, should be less backgrounded according to the Negation and Discourse tasks. Fig. 3 provides difference scores on the *wh*-questions for each of the 10 target constructions (*y*-axis) and each construction's average score on each measure of backgroundedness (*x*-axes). The correlations are displayed in Fig. 3 between ratings on *wh*-questions and Negation task scores ($r = -0.73$) in the left panel, and *wh*-questions and Discourse task scores ($r = -0.67$) on the right. Correlations are similarly high between ratings on the relative clause stimuli and each task (Negation task scores: $r = -0.77$; Discourse task scores: $r = 0.60$); and the discourse-linked questions and each task (Negation task scores: $r = -0.73$; Discourse task scores: $r = .64$). Given that there are only 10 datapoints available for these analyses, the correlations are quite striking.

It is clear in Fig. 3 that some constructions fall more closely along the regression line than others. In the following section we examine pairs of cases more closely.

7. Exploratory analyses of individual construction pairs

Recall that the base stimuli were created such that each stimulus was paired with another stimulus which was highly semantically similar but differed in how its information was packaged, in order to provide two highly similar response options in the Discourse task. This pairing allows us to meaningfully compare results across pairs of base constructions, while controlling for complexity and propositional meaning to the extent possible.

The same preregistered ordinal models used in the planned analyses are applied to each pair of constructions, except that construction type was excluded as a factor, since only one pair was included in each analysis. Specifically, raw acceptability ratings were the outcome variable in an ordinal model that included as fixed factors, BaseorLDD, a measure of backgroundedness, and their interaction. Random effect structure included random slopes and intercepts for items, and random intercepts for subjects (since each type of judgment was collected across participants, no slopes for subjects are included). Again, we hypothesized an interaction: backgroundedness measures should predict acceptability ratings on the long-distance dependency sentences more than they predict acceptability ratings on the base sentences.

Each comparison includes only 12 pairs of items (24 base stimuli sentences), so results must be interpreted cautiously. Table 7 provides figures for each comparison. Significance levels of the predicted interaction shown in each panel are based on the ordinal models: '***': $p < .0001$; '**: $p < .01$, '*' $p < .05$. Results of the models are discussed in turn, and the predicted interaction term, with significant level indicated in Table 7, is boldfaced in each Table of results. As described below, in the case of clausal complements of verbs, log frequency was additionally included as a factor in order to address the frequency-based account outlined in the introduction.

7.1. Main clauses with temporal adjuncts

A set of 24 base sentences all included both a main clause and a temporal adjunct. For half of the stimuli, the content that was targeted by the Negation task and Discourse task was conveyed by the temporal adjunct, and in the other half the targeted content was conveyed by the main clause. The long-distance dependency constructions likewise targeted an element from the main clause in half of the items and an element within the temporal adjunct in the other half. As expected, sentences with long-distance dependencies were rated less acceptable overall than base sentences, presumably due to their complexity. As predicted, both backgroundedness measures predict acceptability ratings of the long-distance dependency stimuli better than they predict acceptability ratings of the base sentences, as indicated by the boldfaced interaction terms in Table 8.

7.2. Main clauses with relative clauses

In this comparison, 24 base stimulus sentences all included both a relative clause and a main clause. In half of the stimuli, the content that was targeted in the Negation task and Discourse task was conveyed by the relative clause, and in the other half, by the main clause. Results

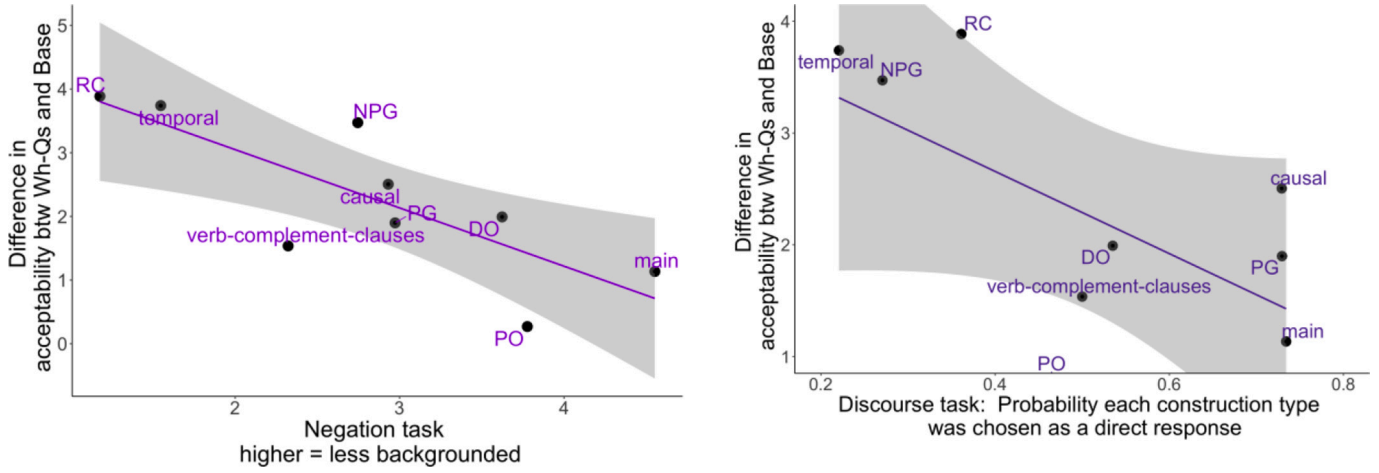


Fig. 3. Correlations between the difference in acceptability between *wh*-questions and declarative stimuli (y-axes) and the responses on the Negation task ($r = -0.73$; x-axis on Left) and on the Discourse task ($r = -0.67$, x-axis on Right), by construction type.

Table 7
Each of the 12 panels is based on 12 minimal pairs (= 24 base stimuli), which differed in how information was packaged. BCI predicts that backgrounded measures should predict ratings better on Long-Distance Dependency stimuli (in shades of red) more than on the base sentences (in blue). For display purposes only, acceptability ratings were normed by participant (z-scores).

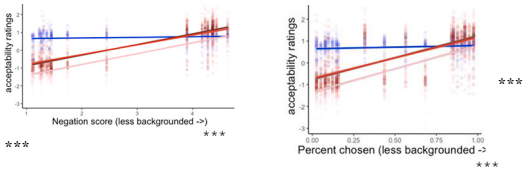
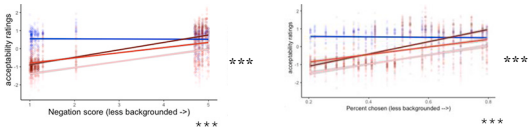
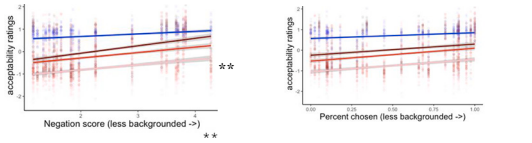
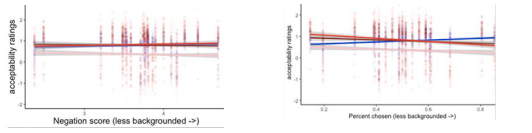
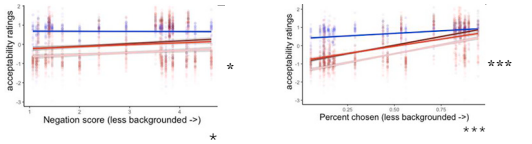
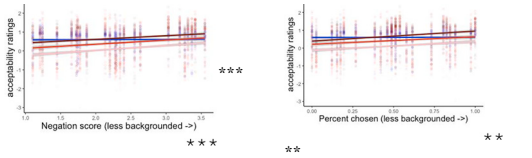
Target Constructions (extraction site)	Negation Scores x Acceptability	Discourse Scores x Acceptability
Main clauses with temporal adjuncts (extraction from main or adjunct)		
Main clauses with relative clauses (extraction from main or relative clause)		
Adjuncts (non-finite)		
Double object constructions or prepositional dative constructions (extraction of recipient argument)		
Co-referential arguments (parasitic gaps) and non-coreferential arguments (non-parasitic gaps) (extraction of second NP)		
Verbs + clausal complements (extraction from complements) (frequency included as a factor in models)		

Table 8
Model results for the negation (top) and discourse task (bottom) from the subset of data containing 12 minimal pairs with LDDs targeting the main clause or temporal adjunct.

Negation task: clmm(raw_response ~ negation_mean*BaseorLDD +(1|subject) + (1+ BaseorLDD| item), datatempmain)				
	Estimate	*SE*	*z value*	*Pr(>|z|)*
Negation Mean score	0.88	0.03	24.52	< 0.00001
Base or LDD	−6.88	0.32	−21.69	< 0.00001
Negation Mean: Base or LDD	**1.46**	**0.07**	**21.10**	**< 0.00001*** **
Discourse Task: clmm(raw_response ~ percent_chosen*BaseorLDD +(1|subject) + (1+ BaseorLDD| item), datatempmain)				
	Estimate	SE	z value	Pr(>|z|)
Percent Chosen	3.02	0.13	22.61	< 0.00001
Base or LDD	−4.50	0.22	−20.32	< 0.00001
Percent chosen: Base or LDD	**4.06**	**0.26**	**15.78**	**< 0.00001*** **

shown in Table 9 confirm that both backgroundedness measures predict ratings more for the long-distance dependency stimuli than for the base sentences, even in this small subset of the data.

7.3. Sentences containing non-finite adjuncts (semantically related temporally or causally)

Each of the 24 base sentences analyzed in this subset of the data

Table 9
Model results for the negation (top) and discourse task (bottom) from the subset of data containing 12 minimal pairs, with extraction from the main clause or relative clause.

Negation task: clmm(raw_response ~ negation_mean*BaseorLDD +(1|subject) + (1+ BaseorLDD| item), data_rcmain)				
	Estimate	SE	z value	Pr(>|z|)
Negation Mean	0.45	0.02	18.92	< 0.00001***
Base or LDD	−5.98	0.25	−23.49	0.50
Negation Mean: Base or LDD	**0.97**	**0.05**	**20.10**	**< 0.00001*** **
Discourse Task: clmm(raw_response ~ percent_chosen*BaseorLDD +(1|subject) + (1+ BaseorLDD| item), data_rcmain)				
	Estimate	SE	z value	Pr(>|z|)
Percent Chosen	3.52	0.25	14.27	< 0.00001
Base or LDD	−6.30	0.32	−20.03	< 0.00001***
Percent chosen: Base or LDD	**7.08**	**0.49**	**14.49**	**< 0.00001*** **

Table 10
Model results for the negation (top) and discourse task (bottom), with the subset of data containing 12 minimal pairs, with extractions from temporal or causal non-finite verbal adjuncts.

Negation task clmm(raw_response ~ negation_mean*BaseorLDD +(1|subject) + (1+ BaseorLDD| item), data_adjuncts)				
	Estimate	SE	z value	Pr(>|z|)
Negation Mean	0.42	0.05	8.17	< 0.00001***
Base or LDD	−4.86	0.31	−15.24	< 0.00001***
Negation Mean:Base or LDD	**0.32**	**0.10**	**−3.17**	**0.002****
Discourse task clmm(raw_response ~ percent_chosen*BaseorLDD +(1|subject) + (1+ BaseorLDD| item), data_adjuncts)				
	Estimate	SE	z value	Pr(>|z|)
Percent Chosen	1.25	0.14	9.07	< 0.00001***
Base or LDD	−4.4	0.27	−15.94	< 0.00001***
Percent chosen: Base or LDD	**0.44**	**0.27**	**1.64**	**0.102**

included either a temporal non-finite adjunct (headed by *while, before, after*) or a causal non-finite adjunct (headed by *by, from,* or *to*). The Negative and Discourse tasks always targeted the content of an adjunct. The long-distance dependency constructions likewise all targeted a noun phrase from within the adjunct clause. Clausal adjuncts are often taken to be islands uniformly. Yet Truswell (2007) and others have argued that the island status of adjuncts varies on the basis of semantics. Recent work by Namboordiripad et al. (2022) compared different adjuncts systematically in two studies and found the same Negation task used here predicted acceptability ratings better on long-distance dependencies than the Base stimuli, in two experiments. The current data replicates this finding with the Negation task, where we once again find the predicted interaction: the negation scores predict judgments on the LDD stimuli more than they predict scores on the Base sentences. Results on the Discourse task do not reach significance in this comparison (See Table 10).

7.4. Double-object or Prepositional paraphrase constructions

Of all the constructions tested, only the double-object (DO)/ prepositional paraphrase (PO) cases fail to pattern as predicted. Perhaps tellingly, the two constructions are in fact rarely viewed as differing in island status (but Erteschik-Shir, 1979; Goldberg, 2006, 2013). Yet it has been observed that the recipient of a PO is more amenable to long-distance dependencies than the recipient argument of a DO (Jackendoff & Culicover, 1971). The difference in predicted judgments is indicated by the ">" in (9) and (10):

(9) PO: Who did they give the award to _? >

DO: Who did they give _ the award?

(10) PO: Peter knew the Ukrainians who they gave the award to _. >

DO: Peter knew the Ukrainians who they gave _the award.

The difference in judgments is confirmed in the current study: LDDs involving POs are significantly more acceptable than those involving DOs ($\beta = 2.81, z = 13.36; p < .00001$), in a model with maximal random effect structure (See SI.5 for details). However, the dispreference is not predicted by either measure of backgroundedness. The negation task yields null results, and the Discourse task reveals a significant effect in the *opposite* direction: participants indicated that double-object responses were more direct and cooperative responses to queries about the recipient argument. For instance, when prompted with *Tell me who saw the new painting,* participants were more likely to choose *She showed her the portrait* than *She showed the portrait to her.* This suggests that the relative ill-formedness of LDDs targeting the recipient argument of the DO construction (e.g.,?*Who did she give a book?*) requires an explanation other than the BCI. One potential explanation is that speakers tend to interpret the written DO examples as involving a narrow focus on the recipient (*She showed HER the portrait*), which would require stress on the recipient argument as indicated by the capitalized font. We plan to retest the DO-PO stimuli using auditory stimuli, but for now, we leave this single case unresolved. Fig. 3 suggests it is the POs rather than the DOs that are outliers in the extent to which the backgroundedness measures predicts acceptability ratings. That is, th PO stimuli is essentially judged more acceptable than is expected from either discourse measure insofar as there is minimal difference in acceptability between judgments on their LDD stimuli and their base versions of the same stimuli (Table 11).

Table 11
Model results for the negation (top) and discourse task (bottom) from the subset of data containing 12 minimal pairs, with extraction from recipient of the double object construction or the prepositional phrase (to/for-dative).

Negation task: clmm(raw_response ~ negation_mean*BaseorLDD +(1|subject) + (1+ BaseorLDD| item), data_dopo)				
	Estimate	SE	z value	Pr(>|z|)
Negation Mean	−0.21	0.56	0.37	0.71
Base or LDD	−0.70	0.68	1.02	0.31
Negation Mean: BaseorLDD	**−0.21**	**0.18**	**−1.16**	**0.25**
Discourse Task clmm(raw_response ~ percent_chosen*BaseorLDD +(1|subject) + (1+ BaseorLDD| item), datadopo)				
	Estimate	SE	z value	Pr(>|z|)
Percent Chosen	−0.45	0.31	−1.43	0.15
Base or LDD	0.17	0.43	0.39	0.69
Percent chosen: Base or LDD	**−2.76**	**0.63**	**−4.41**	**< 0.00001*****

Table 12
Model results for the negation (top) and discourse task (bottom) from the subset of data containing 12 minimal pairs, with extraction from a parasitic gap (involving both main clause and adjunct) or non-parasitic gap (only adjunct).

Negation Task clmm(raw_response ~ negation_mean*BaseorLDD +(1|subject) + (1+ BaseorLDD | item), datanpgpg)				
	Estimate	SE	z value	Pr(>|z|)
Negation Mean	0.06	0.07	−0.90	0.37
Base or LDD	−2.75	0.35	−7.76	< 0.0001***
Negation Mean:Base or LDD	**0.24**	**0.11**	**2.18**	**0.03***
Discourse task clmm(raw_response ~ percent_chosen*BaseorLDD +(1|subject) + (1+ BaseorLDD| item), datanpgpg)				
	Estimate	SE	z value	Pr(>|z|)
Percent Chosen	3.21	0.15	21.39	< 0.00001***
Base or LDD	−3.80	0.26	−14.76	< 0.00001
Percent chosen:Base or LDD	**2.11**	**0.29**	**7.39**	**< 0.00001*****

7.5. Sentences containing parasitic gaps or non-parasitic gaps

"Parasitic gaps" are so named because they involve a "gap" that is parasitic on another "gap" for its interpretation. An example of such a case is in (12). Long-distance dependencies with parasitic gaps are recognized to be more acceptable than comparable stimuli without parasitic gaps (e.g., 13).

(12) Which book did Monique review _ without reading _? >

(13) Which book did Monique review a paper without reading _?

There is a great deal of work on parasitic gaps, which we do not attempt to review here (see Culicover & Winkler, 2019; Engdahl, 1983; Postal, 1993; Ross, 1967). Because untensed adjuncts were already included as stimuli in the comparison between temporal vs causal adjuncts, we chose to include tensed adjuncts for half of the parasitic/non-parasitic pairs; we include "without" phrases in the other half of parasitic/non-parasitic pairs.

Few have suggested a functional explanation as to why parasitic gaps are more acceptable than non-parasitic gaps (but see Chaves & Dery, 2019: 9–10), and no one that we know of has empirically tested a functional explanation. Yet results once again reveal the predicted interaction, both for the negation task and the discourse task: i.e., Scores on each of the backgroundedness measures predicted acceptability ratings of long-distance dependencies significantly better than base sentences (see Table 12).

Table 13
Twenty main verbs used in 24 stimuli with clausal complements.

announce	*discover*	*grumble*	*say*
believe	*hear*	*guess*	*see*
bet	*fear*	*imagine*	*shout*
claim	*feel*	*realize*	*whisper*
complain	*forget*	*remember*	*yell*

7.6. Clausal verb complements (potential "bridge" verbs)

The final comparison included in the current study involves clausal complements of the 20 verbs in Table 13. Using the same ordinal models, we once again find that the predicted interaction is significant when the Negation task is used to predict judgments ($\beta = 48$, $z = -3.85$, $p = .0001$) and when the Discourse task is used ($\beta = 0.73$, $z = 2.44$ $p = .01$). That is, both measures of Backgroundness predict ratings of the long-distance dependency stimuli more than they predict ratings of the base stimuli.

However, recall that Liu, Ryskin, et al. (2022) suggest an alternative explanation. They argued that the complement clauses of main verbs may not resist long-distance dependencies any more than is expected based on the familiarity of the base sentences and the complexity added by the long-distance dependency construction. In particular, they found that the log frequencies of the verb + clausal complements, as found in corpus data, predicted the acceptability of *wh*-questions and declarative sentence equally well.

Given the claim that frequencies of verb + clausal complement combinations should predict acceptability ratings of base sentences and long-distance dependency stimuli *equally well,* we calculated the log frequencies for each verb + complement combination, based on searches of: VERB *that* PRON in the multi-billion-word COCA corpus (Davies, 2008), for each of the verbs in Table 13 (see SI.3). We first test the role of frequency without including either measure of backgroundedness (Table 14). Consistent with Liu, Ryskin, et al. (2022), we find that log frequencies influence acceptability across the board ($\beta = 0.55$, $z = 6.47$,

Table 14
Log frequencies of individual verb + complement combinations predict ratings on long-distance dependency stimuli better than they predict ratings on base stimuli.

clmm(raw_response ~ log.frequency*BaseorLDD + (1|subject) + (1 + BaseorLDD | item), databridge)

	Estimate	SE	z value	Pr(>|z|)
log.frequency	0.55	0.08	6.47	< 0.00001***
Base or LDD	−2.99	0.67	−4.48	< 0.00001***
log.frequency:Base or LDD	**0.45**	**0.17**	**2.68**	**< 0.01** **

Table 15
Model results for the negation (top) and discourse task (bottom) from the subset of data containing 12 minimal pairs with extraction from clausal complements of 20 unique main verbs.

Negation task
clmm(raw_response ~ negation_mean*BaseorLDD +log.frequency + (1|subject) + (1 + BaseorLDD | item), data_verbcomp)

	Estimate	SE	z value	Pr(>|z|)
Negation Mean	0.13	0.07	1.79	0.07
Base or LDD	−2.55	0.33	−7.78	< 0.0001
log.frequency	0.61	0.07	8.74	< 0.00001
Negation Mean: Base or LDD	**0.52**	**0.13**	**4.10**	**< 0.00001**

Discourse Task
clmm(raw_response ~ percent_chosen *BaseorLDD +log.frequency + (1|subject) + (1 + BaseorLDD | item), data_verbcomp)

	Estimate	SE	z value	Pr(>|z|)
Percent chosen	0.10	0.20	0.50	0.62
Base or LDD	−1.87	0.36	−5.14	< 0.00001***
log.frequency	0.55	0.09	5.85	< 0.00001***
Percent chosen: Base or LDD	**0.84**	**0.31**	**2.69**	**0.01****

$p < .00001$). Yet unexpectedly, we additionally find a significant interaction: log-frequencies predict ratings on the long-distance dependency stimuli more than the base stimuli ($\beta = 0.45$, $z = 2.68$, $p = .01$). There is no theoretical reason to expect frequencies of verb + complement clauses to predict acceptability of long-distance dependencies more than base sentences.

An explanation may stem from the fact that log-frequencies turn out to correlate significantly with each discourse measure. That is, log-frequencies of the verb + complement correlate with results on the Discourse task ($r = 0.45$, $p < .00001$) and with judgments on the Negation task ($r = 0.069$, $p = .0002$). These correlations indicate that verb + complement clauses that occur more frequently tend to be those that make the content of their complement clauses more at-issue (less backgrounded). If verbs with similar meanings have similar frequencies across languages, the BCI would offer an explanation: people are more likely to produce complement clauses of verbs when those complement clauses are at-issue in discourse (see also Richter & Chaves, 2020 for cogent discussion).

We tested ordinal models on the predicted interaction of each measure of backgroundedness, with log frequency included as a (non-interacting) factor additional factor. Results, shown in Table 15, reveal that each measure of backgroundedness influences acceptability judgments of LDDs more than the Base sentences, even when frequency is included as a separate factor. This lends strong support to the BCI hypothesis that backgrounded constructions are islands (see also Ambridge & Goldberg, 2008).

Table 16
LDDs targeted several types of main clause and subordinate clause constituents.

Main clause	Subordinate clause
Transitive main clauses	Temporal verbal adjuncts
Double object construction	Causal verbal adjuncts
Prepositional object construction	Non-parasitic gaps
	Clausal complements of verbs
	Relative clauses

8. Including an additional factor: main clause vs. subordinate clause

In final exploratory analyses, we test an additional binary factor, namely whether the element targeted by the LDD is licensed within a main or subordinate clause. The main / subordinate clause distinction can be related to a simplified version of the structural constraint outlined in the introduction: LDDs have been claimed not to relate elements separated by more than a single "bounding node" (Chomsky, 1973). Clause boundaries are generally considered to be bounding nodes. The current target constructions include several types of main and subordinate clauses (see Table 16). Because parasitic gaps target information from both the main and subordinate clause, they were omitted from analysis, although results do not change if they are included in either category.

New models were created that included the main-clause factor and its interaction with Base or LDDs (Table 17). We additionally included each of measures of backgroundedness and its interaction with Base or LDDs. We did not test for a three-way interaction because none was predicted, but all factors were also included as simple fixed effects. Results show that LDDs are significantly more acceptable when the extraction site targets a main clause compared to a subordinate clause. The predicted interactions involving the backgroundedness measures remain significant in the case of the negation task (top), and marginal in the case of the discourse task.

Critically, the main clause vs. subordinate clause distinction is related to the notion of backgroundedness: Each measure of backgroundness predicts the main_clause / subordinate clause distinction in separate linear models: Negation task ($\beta = 0.27$ $t = 139.70$ $p < .000001$); Discourse task ($\beta = 0.59$ $t = 48.07$, $p < .000001$).

9. General discussion

The current work addresses the long-standing question as to *why* certain combinations of constructions are judged to be systematically less acceptable than others. The fact that long-distance dependency constructions resist combination with certain base constructions, dubbed "islands," has been widely presumed to depend on syntactic

Table 17
Model results for the negation (top) and discourse task (bottom), including a new factor, main_clause which divided the target constructions according to whether the LDD targeted a main clause or subordinate clause.

Negation task
clmm(raw_response ~ main_clause + BaseorLDD + negation_mean + negation_mean:BaseorLDD + main_clause:BaseorLDD + (1|subject) + (1 + BaseorLDD|item) + (1|Cx_Cx), data_plus)

	Estimate	SE	z value	Pr(>|z|)
main_clause	1.07	0.39	2.74	< 0.01**
BaseorLDD	−3.66	0.52	−7.03	**< 0.001***
negation_mean	0.28	0.10	2.86	0.004**
BaseorLDD:negation_mean	**0.42**	**0.16**	**2.66**	**0.007****
main_clause:BaseorLDD	1.45	0.43	3.36	**< 0.001**

Discourse Task
clmm(raw_response ~ main_clause + BaseorLDD + percent_chosen + percent_chosen:BaseorLDD + main_clause:BaseorLDD + (1|subject) + (1 + BaseorLDD|item) + (1|Cx_Cx), data_plus)

	Estimate	SE	z value	Pr(>|z|)
main_clause	1.34	0.27	4.94	< 0.00001***
BaseorLDD	−2.88	0.32	−9.02	< 0.00001***
percent_chosen	1.27	0.30	4.24	< 0.00001***
BaseorLDD: percent_chosen	**1.02**	**0.52**	**1.96**	**0.0502.**
main_clause:BaseorLDD	2.09	0.31	6.68	< 0.00001***

constraints, while the possibility that the functions of the constructions predict acceptability has primarily relied on introspective judgments (Deane, 1991; Erteschik-Shir & Lappin, 1979; Goldberg, 2006). The current work is based on a specific hypothesis, namely that it is infelicitous for a speaker to simultaneously choose to foreground an element (via a long-distance dependency construction) and background the same element (via a base construction). That is, we hypothesized that when an element is made prominent by a long-distance dependency construction, the question or sentence is infelicitous to the extent that the same element is backgrounded by the base construction it depends on for its interpretation. This is the Backgrounded Constructions are Islands (BCI) hypothesis.

The experimental work that had tested the function-based account had, to date, included only one or two base constructions at a time and had only collected acceptability ratings on a single type of LDD (*wh*-questions) (Ambridge & Goldberg, 2008; Cuneo & Goldberg, 2022; Liu, Ryskin, et al., 2022; Liu, Winckel, et al., 2022; Namboordiripad et al., 2022). Here we tested 3 sets of LDD constructions on 10 base constructions. Importantly, we operationalized the gradient notion of backgroundedness in two distinct ways: via a Negation task and a new Discourse task. As hypothesized, each measure predicted acceptability ratings on long-distance dependency stimuli more than base sentences. These results are striking since each measure of backgroundedness probes only base sentences: The Negation task probes inferences on negated base sentences, and the Discourse task required a choice between two base-sentence responses. Nonetheless, results on each task predict the acceptability of questions and sentences containing long-distance dependencies better than they predicted the acceptability of the base sentences themselves, across the 10 target base constructions. The between-subjects design prevented participants from comparing base sentence stimuli with related sentences containing a long-distance dependency construction and prevented them from comparing the long-distance dependency constructions themselves. It also prevented participants in the Negation task or the Discourse task from explicitly considering acceptability when performing judgments. At the same time, the fact that each participant judged sentences from a variety of base construction types allowed us to avoid satiation effects (Chaves & Dery, 2019).

Current results demonstrate that *wh*-questions and relative clauses pattern alike with respect to acceptability judgments, at least in the current stimuli. In fact, acceptability ratings on all three sets of long-distance dependency stimuli (including discourse-linked *wh*-questions) were highly correlated with one another ($0.80 < r < 0.90$). Thus, although relative clauses and questions undoubtably have distinct functions, results indicate the head noun is prominent within the domain of the RC construction (the Noun Phrase) just as the *wh*-word is prominent within the domain of the question construction. This result is unexpected by the Focus proposal, which predicted that the acceptability of *wh*-questions should depend on backgroundedness, but relative clauses should not.

We had not predicted that subsets of only 24 items would provide sufficient power to show the predicted interaction. Yet exploratory comparisons between pairs of constructions were strikingly consistent with the BCI hypothesis. Results on the Negation task predicted the island status of long-distance dependencies for all subsets with the exception of the PO-DO comparison which showed a null effect. Main clauses with relative clauses, main clauses with temporal adjuncts, a variety of non-finite adjuncts, clausal complements of 20 different main verbs, and sentences with and without parasitic gaps each showed the hypothesized interaction. Finally, even with an exploratory factor distinguishing main from subordinate clauses included, the predicted interaction with the negation task results remained significant.

The discourse task — the proportion of times participants chose one base sentence over another as a direct, cooperative response — yielded somewhat less clean results compared to the negation task, but it was significant in the overall preregistered model. The Discourse task also predicted judgments on certain subsets of the data independently: for main clauses with relative clauses, main clauses with temporal adjuncts, variation among main verbs with complement clauses, and for sentences with and without parasitic gaps. The discourse task results did not show a significant interaction in the case of non-finite adjuncts and was significant in the unexpected direction when the DO-PO case was examined. Finally, the extent to which it accounted for variation beyond the influence of the main clause / subordinate clause distinction fell on the razor's edge of significance. We had anticipated that participants would find it challenging to decide how "direct and cooperative" a response was on its own and we employed the two-alternative forced choice task for this reason (Note 6). Participants showed some preference for choosing base sentences that were simply judged to be more *acceptable* by other speakers, which was not intended. Moreover, the task yielded fewer independent data points: if one response was selected *P* proportion of the time, it entailed that the other member of the pair was selected *1-P* proportion of the time. For these reasons, the negation task is a more reliable test of backgroundedness than the discourse task. Nonetheless,

both the negation and the discourse tasks reveal the predicted interaction for each LDD considered separately (SI.4), as well as for the combined stimuli (Table 5).

Since the current study included 10 base construction types, it was possible to test for a correlation in the extent to which each measure predicted backgroundedness across constructions. For this, we calculated mean backgroundedness scores for each construction type and correlated those with the mean difference in judgments between long-distance dependencies and the corresponding base stimuli. This analysis confirmed a significant correlation for both backgroundedness measures: the extent to which a construction was, on average, backgrounded (inversely) predicted the extent to which it was judged acceptable to extract an element from the construction (Fig. 3).

The outlier in Fig. 3 is the prepositional object construction, which was judged somewhat more acceptable in LDDs than would be expected on the basis of either measure of backgroundedness. A closer look at the stimuli that included the double-object construction and prepositional dative echoed the fact that this set of stimuli did not follow the same pattern as the rest. We speculated that the unexpected results may be due to the fact that the stimuli were written and so allowed for a narrow focus on the double-object recipient (e.g., *She told HIM the story*). Additional work is required to understand why double-object recipients tend to resist LDDs, an observation that was confirmed in the current study, since neither measure of backgroundedness patterned as predicted.

Current results do shed light on the so-called distinction between "bridge" and "non-bridge" verbs. Without presupposing which verbs ought to be considered "bridge" verbs, results demonstrated that each measure of backgroundedness predicted judgments on the LDDs more than on the base sentences, as predicted by the BCI. The interaction remained significant even when the log frequencies of the main verbs + complement, as determined by corpus data, was included as an additional factor, demonstrating that complement clauses are more amenable to LDDs the more foregrounded (less backgrounded) they are (see also Ambridge & Goldberg, 2008; Dabrowska et al., 2009). In fact, the current data indicates that log frequencies are indirectly related to backgroundedness: first, log frequencies of the main verbs + complement clauses unexpectedly displayed a stronger effect on LDDs than on base sentences. In an attempt to explain this, we probed whether the log frequencies correlated with each measure of backgroundedness and found that they did. That is, verbs that are more likely to occur with complement clauses are also more likely to foreground their complement clauses: the functions of verb + complement combinations influence how frequently they are used (and/or vice versa).

In final exploratory analyses, we divided stimuli according to whether the LDDs targeted an element in the main or subordinate clause and found that main clauses were more amenable to long-distance dependencies than subordinate clauses. This is consistent with a simplified structural description of the phenomenon. It is of course possible that both discourse and structural factors play a role (e.g., Boeckx, 2012; Goodall, 2015). But as was the case with log frequencies of verb + clausal complements, the distinction between main and subordinate clauses itself evokes the notion of backgroundedness. As their monikers suggest, *main* clauses tend to express the main content, while *subordinate* clauses tend to express information that is subordinate. More concretely, each measure of backgroundedness turned out to predict main vs. subordinate clause status. The general idea that the main – subordinate clause distinction correlates with backgroundedness is in fact far from new. Tomlin (1985) performed an empirical comparison of the clause types used in free descriptions of an animated cartoon and an independent analysis of which events in the cartoon were more central to the storyline. As he reports, "independent clauses code foreground and pivotal information; dependent clauses code background information."

Interpreting the distinction between main and subordinate clause as codifying a distinction in backgroundedness offers the possibility of accounting for certain other cases, not directly tested here. For example, relative clauses (e.g., 11) that express the main content of the sentence are formally subordinate but are compatible with at least certain LDDs (e.g., 12) (Duffield & Michaelis, 2011; Kluender, 1998):

(11) There are few people who will finish this paper.

(12) This is a paper there are few people who will finish _.

The alternative would be to state that LDDs that target main clauses are simply more acceptable than those targeting a subordinate clause for unexplained reasons. It would then be a coincidence that the distinction happens to significantly correlate with each measure of backgroundedness. Critically, we cannot easily assimilate backgroundedness itself to a distinction between main and subordinate clauses, particularly in the case of clausal complements or various adjuncts. In both cases, the structure is held constant, while backgroundedness predicts differences in judgments (see also Ambridge & Goldberg, 2008; Namboordiripad et al., 2022). Parasitic gaps, too, would require some additional machinery since they appear in subordinate clauses and yet are judged relatively acceptable.

The work here leaves much more to do. Cross-linguistic experimental work on island constructions has been increasing (e.g., Christensen & Nyvad, 2014; Kush et al., 2019; Müller & Eggers, 2022; Sprouse et al., 2016; Stepanov, Mušič, & Stateva, 2018), yet the current measures of backgroundedness have not yet been applied in languages other than English. Doing so will allow us to explore whether apparently analogous constructions are backgrounded similarly in different languages, and the extent to which judgments on long-distance dependencies pattern accordingly. The current work tested multiple types of long-distance dependencies: *wh*-questions, "discourse-linked" questions, and restrictive relative clauses and found them to correlate highly. Other long-distance dependency constructions including topicalizations and clefts are worth comparing as well, to determine whether all LDDs make the targeted element prominent to the same degree. While the current work has tested a range of potential island constructions, a number of other constructions remain to be investigated using empirical measures of backgroundedness. Of particular interest are subjects (Chaves, 2013; Chaves & Dery, 2014; Kluender, 1998; Phillips, 2006; Polinsky et al., 2013), presentational relative clauses (Nyvad, Müller, & Christensen, 2022), and certain types of conjunctions (e.g., Lakoff, 1986; Ross, 1967). Finally, languages that rely less on LDDs and instead position question words in their canonical positions are important to investigate (e.g., Lu, Thompson, & Yoshida, 2020), as are the pragmatically restricted range of in-situ *wh*-questions in English (Chaves & Putnam, 2021; Ginzburg & Sag, 2000).

10. Conclusions

Language systems are highly complex, and certain combinations of constructions are systematically judged less felicitous for reasons that have long eluded a satisfactory explanation. A central such case involves long-distance dependency constructions, such as *wh*-questions and relative clauses, which position a constituent some distance away from the clause that the constituent depends on for its semantic interpretation. The current results support the idea that infelicity arises to the extent that the constructions being combined have incompatible functions. Specifically, it is anomalous for a speaker to simultaneously foreground an element (by targeting it for a long-distance dependency) and background it (via choice of base construction).

The current paper tested three long-distance dependency constructions on ten base constructions and provided two independent tests of backgroundedness. This method allowed us to investigate to what extent and *why* base constructions can felicitously be combined with long-distance dependency constructions. The two measures of backgroundedness were: a) the extent to which sentence negation is construed as negating the content expressed by the target construction (Negation task) and b) the likelihood of participants deciding that the

target construction is a preferred way to convey information directly and cooperatively in the two alternative forced choice task (Discourse task). Results from the Negation task demonstrate that participants treat the content of base constructions that are easier to extract from as less presupposed and more influenced by main clause negation. Separate results from the Discourse task confirm that when asked to choose the more direct and cooperative response to a prompt for information, participants favor responses in which the requested information is conveyed by a construction from which it would be easier to "extract." These findings are consistent with the BCI claim that constructions are islands that resist extraction to the extent that they are backgrounded in discourse.

The current findings challenge the claim that island effects are the result of illicit syntactic movement, since judgments are predicted by backgroundedness, a discourse factor, and none of the stimuli used to operationalize backgroundedness included any island violations. It might be possible to add discourse functions to trees as inaudible nodes, thereby treating backgroundedness as if it were a syntactic category (Erteschik-Shir, 2007). However, aside from being ad hoc, it is unclear how this approach could predict the gradient nature of the effects (see also, Culicover & Winkler, 2019; Kluender, 1998; Kluender & Kutas, 1993; Liu, Ryskin, et al., 2022; Liu, Winckel, et al., 2022; Namboordiripad et al., 2022). Finally, we note that it is self-evident that the functions of constructions are learned, insofar as speakers use and interpret the functions appropriately; it is far from clear whether and why syntactic counterparts, presumably in the form of inaudible nodes in trees, would be or could be learned.

Current results also undermine a general explanation in terms of frequency or familiarity. If familiarity with base constructions were primarily responsible for acceptability ratings, we would expect ratings on the base sentences to parallel ratings on the long-distance dependency stimuli. There would be no reason to expect backgroundedness measures to explain additional variance in ratings on the long-distance dependency stimuli. It may be possible to appeal to the familiarity of the *combination* of long-distance dependency and base construction to predict acceptability ratings but given the high correlation between acceptability ratings on questions and relative clauses, a frequency-based account would need to assume that speakers happen to both question and add relative clauses to the same type of constituents, with the same relative frequencies. Finally, if frequencies are responsible for acceptability judgments, it would beg the question as to why frequencies pattern as they do, particularly if similar patterns are found cross-linguistically.

In order to explain the extent to which long-distance dependency constructions can be felicitously combined with other constructions, the current results indicate that it is critical to attend to the functions of the constructions involved. The Backgrounded Constructions Are Islands (BCI) hypothesis is straightforward: it should be infelicitous for a speaker to simultaneously choose to foreground and background the same element. Results are consistent with this claim, whether the data is analyzed across all datapoints at once, for the base constructions and each type of LDD separately, in subsets of the data for pairs of constructions, or as a correlation of average responses across the 10 construction types tested. Both independent ways of operationalizing backgroundedness predict the degree of island-status.

To put the current results in a simpler context, it is worth bearing in mind that to produce language, we typically need to combine multiple constructions to express our message. The current results support the claim that it is infelicitous for a speaker to combine a construction that makes an element semantically prominent (an LDD construction) with a construction that backgrounds the same element (in the base construction). To do so would be like trying to put crystal flute on display by wrapping it in a brown paper bag.

Funding

Funding for this work was provided by the Data Driven Social Science Initiative at [University].

IRB approval was granted by [University] #4951.

Credit author statement

Cuneo contributed to the Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing-review and editing. Goldberg contributed to the Conceptualization, Data curation, Formal analysis, Methodology, Recourses, Visualization, Writing- original draft.

Data availability

The data, stimuli and code are all publicly available at the link provided

Acknowledgments

We are grateful to Arielle Belluck for careful editing and help with data collection, to Anne Abeillé, David Adger, Ted Gibson, Ray Jackendoff, Shahar Shirtz, and Elodie Winckel for discussions and to three anonymous reviewers for very helpful feedback on an earlier version of the manuscript. We are grateful to the Data Driven Social Science Initiatve at Princeton for funding.

Appendix A. Supplementary data

Supplementary Information regarding preregistration, data, items, code, and full models not included in main text. https://doi.org/10.1016/j.cognition.2023.105563

References

Abeillé, A., Hemforth, B., Winckel, E., & Gibson, E. (2020). Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition, 204*, Article 104293.

Abels, K. (2017). Movement and islands. In *The Oxford handbook of ellipsis*.

Ambridge, B., & Goldberg, A. E. (2008). The island status of clausal complements: Evidence in favor of an information structure explanation. *Cognitive Linguistics, 19*(3), 349–381.

Baltin, M. R. (1982). A landing site for movement rules. *Linguistic Inquiry, 13*, 1–38.

Birner, B. J., & Ward, G. L. (1998). *Information status and noncanonical word order in English*. John Benjamins Publishing.

Boeckx, C. (2012). *Syntactic Islands*. Cambridge: Cambridge University Press.

Bolinger, D. (1968). *Aspects of language*. NY: Harcourt: Brace & World, New York.

Bondevik, I., Kush, D., & Lohndal, T. (2021). Variation in adjunct islands: The case of Norwegian. *Nordic Journal of Linguistics, 44*(3), 223–254.

Bybee, J. (2007). Review of constructions at work by AE Goldberg. *Journal of Child Language, 34*(3), 692–297.

Chaves, R. P. (2013). An expectation-based account of subject islands and parasitism. *Journal of Linguistics, 49*(2), 285–327.

Chaves, R. P., & Dery, J. E. (2014). Which subject islands will the acceptability of improve with repeated exposure? In R. E. Santana-LaBarge (Ed.), *31st west coast conference on formal linguistics* (pp. 96–106). Somerville, MA: Cascadilla Proceedings Project.

Chaves, R. P., & Dery, J. E. (2019). Frequency effects in subject islands. *Journal of Linguistics, 55*(3), 475–521.

Chaves, R. P., & Putnam, M. T. (2021). *Unbounded dependency constructions : Theoretical and experimental perspectives*. Oxford University Press.

Chomsky, N. (1973). In S. Anderson, & P. Kiparsky (Eds.), *Conditions on transformations. In a festschrift for Morris Halle* (pp. 232–286). Holt, Rinehart & Winston: NYC.

Christensen, K. R., Kizach, J., & Nyvad, A. M. (2013). The processing of syntactic islands–an fMRI study. *Journal of Neurolinguistics, 26*(2), 239–251.

Christensen, K. R., & Nyvad, A. M. (2014). On the nature of escapable relative islands. *Nordic Journal of Linguistics, 37*(1), 29–45.

Christensen, K. R., & Nyvad, A. M. (2022). *The island is still there: Experimental evidence for the inescapability of relative clauses in English*. Studia Linguistica.

Christensen, R. H. B., & Christensen, M. R. H. B. (2015). Package 'ordinal'. *Stand, 19*(2016).

Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition 1987* (pp. 1–33).

Culicover, P. W., & Winkler, S. (2019). Parasitic gaps aren't parasitic, or the case of the uninvited guest. *The Linguistic Review, 39*(1), 1–35.

Cuneo, N., & Goldberg, A. E. (2022). Islands effects without extraction: The discourse function of constructions predicts island status. In *Proceedings of the 42nd annual conference of the cognitive science society*.
Dabrowska, E., Rowland, C., & Theakston, A. (2009). The acquisition of questions with long-distance dependencies. *Cognitive Linguistics, 20*, 571–597.
Davies. (2008). The corpus of contemporary American English 1990–present. Online: http://corpus.byu.edu/coca.
Deane, P. (1991). Limits to attention: a cognitive theory of island phenomena. *Cognitive Linguistics, 2*(1), 63.
Duffield, C. J., & Michaelis, L. A. (2011). Why subject relatives prevail: Constraints versus constructional licensing. *Language and Cognition, 3*(2), 171–208.
Engdahl, E. (1983). Parasitic gaps. *Linguistics and philosophy*, 5–34.
Erteschik-Shir, N. (1979). Discourse constraints on dative movement. In S. Laberge, & G. Sankoff (Eds.), *Syntax and semantics* (pp. 441–467). New York: Academic Press.
Erteschik-Shir, N. (2007). *Information structure: The syntax-discourse interface* (Vol. 3). OUP Oxford.
Erteschik-Shir, N., & Lappin, S. (1979). *Dominance and the functional explanation of island phenomena. 6(1–3)* pp. 41–86).
Fodor, J., Bever, T. G., & Garrett, M. (1974). *The psychology of language: An introduction to psycholinguistics and generative grammar*. McGraw-Hill.
Francis, E. (2021). *Gradient acceptability and linguistic theory*. Oxford University Press.
Ginzburg, J., & Sag, I. (2000). *Interrogative investigations*. Stanford: CSLI Publications.
Givón, T. (1995). *Functionalism and grammar*. Philadelphia: J. Benjamins.
Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press.
Goldberg, A. E. (2013). Backgrounded constituents cannot be "extracted". In J. Sprouse (Ed.), *Syntactic Islands*. Oxford: Oxford University Press, 221–238.2013b.
Goodall, G. (2015). The D-linking effect on extraction from islands and non-islands. *Frontiers in Psychology, 5*, 1493.
Gregory, M. L., & Michaelis, L. A. (2001). Topicalization and left-dislocation: A functional opposition revisited. *Journal of pragmatics, 33*(11), 1665–1706.
Hofmeister, P., Casasanto, L. S., Sag, I. A., Sprouse, J., & Hornstein, N. (2013). *Islands in the grammar? Standards of evidence. Experimental syntax and island effects* (p. 42).
Hofmeister, P., & Sag, I. A. (2010). Cognitive constraints and island effects. *Language, 86*, 366–415.
Jackendoff, R. S., & Culicover, P. (1971). A reconsideration of dative movements (with Peter Culicover). *Foundations of Language, 7*(3), 397–412.
Kaplan, R. M., & Zaenen, A. (1989). Long-distance dependencies, constituent structure, and functional uncertainty. *Alternative Conceptions of Phrase Structure, 17*, 42.
Karttunen, L. (1973). Presuppositions of compound sentences. *Linguistic Inquiry, 4*(2), 169–193.
Kluender, R. (1998). On the distinction between strong and weak islands: A processing perspective. *Syntax and Semantics, 29*, 241–279.
Kluender, R., & Kutas, M. (1993). Bridging the gap - evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience, 5*, 196–214.
Kobzeva, A., Sant, C., Robbins, P. T., Vos, M., Lohndal, T., & Kush, D. (2022). Comparing island effects for different dependency types in Norwegian. *Languages*, 1–25.
Kothari, A. (2008). Frequency-based expectations and context influence bridge quality. In M. Grosvald, & D. Soares (Eds.), *WECOL: Western conference on linguistics* (pp. 136–149). California State University.
Kuno, S. (1987). *Functional syntax: Anaphora, discourse and empathy*. Chicago: University of Chicago Press.
Kush, D., Lohndal, T., & Sprouse, J. (2019). On the island sensitivity of topicalization in Norwegian: An experimental investigation. *Language, 95*(3), 393–420.
Lakoff, G. (1986). *Frame semantic control of the coordinate structure constraint*. Chicago: Parasession on Pragmatics and Grammatical Theory CLS.
Langacker, R. W. (1987). *Foundations of cognitive grammar: Volume I: Theoretical prerequisites, (Vol. 1)*.. Stanford University Press.
Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods, 49*(2), 433–442.
Liu, Y., Ryskin, R., Futrell, R., & Gibson, E. (2022). A verb-frame frequency account of constraints on long-distance dependencies in English. *Cognition, 222*, Article 104902.
Liu, Y., Winckel, E., Abeillé, A., Hemforth, B., & Gibson, E. (2022). Structural, functional, and processing perspectives on linguistic island effects. *Annual Review of Linguistics, 8*, 495–525.
Lu, J., Thompson, C. K., & Yoshida, M. (2020). Chinese wh-in-situ and islands: A formal judgment study. *Linguistic Inquiry, 51*(3), 611–623.
Morgan, J. L. (1975). In P. Gole, & J. L. Morgan (Eds.), *Vol. 3. Some interactions of syntax and pragmatics. Syntax and semantics: Speech acts*. NYC: Academic Press.
Müller, C., & Eggers, C. U. (2022). Island extractions in the wild: A corpus study of adjunct and relative clause islands in Danish and English. *Languages, 7*(2). https://doi.org/10.3390/languages7020125
Namboordiripad, S., Cuneo, N., Kramer, M. A., Sedarous, Y., Bisnath, F., Sugimoto, Y., & Goldberg, A. E. (2022). *Backgroundedness predicts island status of non-finite adjuncts in English*. Cognitive Science Society Proceedings.
Newmeyer, F. J. (2016). Nonsyntactic explanations of island constraints. *Annual Review of Linguistics, 2*, 187–210.
Nyvad, A. M., Müller, C., & Christensen, K. R. (2022). Too true to be good? The non-uniformity of extraction from adjunct clauses in English. *Languages, 7*(4), 244. https://doi.org/10.3390/languages7040244
Phillips, C. (2006). The real-time status of island phenomena. *Language*, 795–823.
Polinsky, M. (1998). A non-syntactic account of some asymmetries in the double object construction. In J.-P. Koenig (Ed.), *Conceptual structure and language: Bridging the gap*. Stanford: CSLI.
Polinsky, M., Gallo, C. G., Graff, P., Kravtchenko, E., Morgan, A. M., & Sturgeon, A. (2013). Subject islands are different. In J Sprouse, & N Hornstein (Eds.), *Experimental syntax and island effects* (pp. 286–309). Cambridge University Press.
Postal, P. M. (1993). Parasitic gaps and the across-the-board phenomenon. *Linguistic Inquiry, 24*(4), 735–754.
Postal, P. M. (1994). Contrasting extraction types. *Journal of Linguistics, 30*(1), 159–186.
Potts, C. (2004). *The logic of conventional implicatures*. OUP Oxford.
Richter, S. N., & Chaves, R. P. (2020). *Investigating the role of verb frequency in factive and manner-of-speaking islands*. Cognitive Science Proceedings.
Robenalt, C., & Goldberg, A. E. (2015). Judgment evidence for statistical preemption: It is relatively better to vanish than to disappear a rabbit, but a lifeguard can equally well backstroke or swim children to shore. *Cognitive Linguistics, 26*(3), 467–503.
Ross, J. R. (1967). *Constraints on variables in syntax*. Ph.D. Thesis. Cambridge: Massachusetts Institute of Technology.
Ross, J. R. (1987). Islands and syntactic prototypes. In B. Need, E. Schiller, & A. B. Bosch (Eds.), *Papers from 23rd CLS* (pp. 309–320). Chicago: Chicago Linguistic Society.
Sag, I. A. (2010). English filler-gap constructions. *Language, 86*, 486–545.
Sprouse, J., Caponigro, I., Greco, C., & Cecchetto, C. (2016). Experimental syntax and the variation of island effects in English and Italian. *Natural Language & Linguistic Theory, 34*, 307–344.
Sprouse, J., Wagers, M., & Phillips, C. (2012). Working-memory capacity and island effects: A reminder of the issues and the facts. *Language, 88*(2), 401–407.
Stepanov, A., Mušič, M., & Stateva, P. (2018). Two (non) islands in Slovenian: A study in experimental syntax. *Linguistics, 56*(3), 435–476.
Takami, K. I. (1989). Preposition stranding: Arguments against syntactic analyses and an alternative functional explanation. *Lingua, 76*, 299–335.
Tomlin, R. S. (1985). Foreground-background information and the syntax of subordination. *Text-Interdisciplinary Journal for the Study of Discourse, 5*(1–2), 85–122.
Van Valin, R. D. (1998). The Acquisition of Wh-Questions and the mechanisms of language acquisition. In M. Tomasello (Ed.), *The new psychology of language*. Hillsdale, NJ: Lawrence Erlbaum.
Verhagen, A. (2005). In A. Athanasiadou, C. Canakis, & B. Cornillie (Eds.), *On subjectivity and "long distance Wh-movement". Subjectification: Various paths to subjectivity* (pp. 323–346). New York: Mouton de Gruyter.