Article

Extraction from Present Participle Adjuncts: The Relevance of the Corresponding Declaratives

Andreas Kehl

English Department, Universität Tübingen, Wilhelmstraße 50, 72074 Tübingen, Germany; andreas.kehl@uni-tuebingen.de

Abstract: In this article, I will argue that many of the theoretical approaches to extraction from participle adjunct islands suffer from the fact that the focus of investigation lies on perceived grammaticality differences in interrogative structures. Following approaches which make an explicit connection between extraction asymmetries and properties of the underlying proposition, I will argue that there is good evidence for the existence of similar differences in declarative adjunct constructions which can explain most of the grammaticality patterns observed for interrogatives. A crucial distinction to the majority of previous theories is the focus on acceptability rather than grammaticality, and the assumption that acceptability in declaratives is determined by a variety of semantic and syntactic complexity factors which do not influence how strongly extraction degrades the structure. This line of argumentation is more compatible with approaches to island phenomena that explain the low acceptability of some extractions by independent effects such as processing complexity and discourse function instead of syntactic principles blocking the extraction. I will also discuss a partially weighted, multifactorial model for the acceptability of declarative and interrogative participle adjunct constructions, which explains the judgment patterns in the literature without the need for additional, complex licensing conditions for extraction.

Keywords: adjunct islands; *wh*-extraction; locality; present participle; gradient acceptability; acceptability model

1. Introduction

Since the formulation of the Condition on Extraction Domain (CED, Huang 1982), more and more apparent counterexamples to this strict locality condition have surfaced, including extractions from subjects and adjuncts that are judged as grammatical. Compare the ungrammatical extraction from an adverbial clause in (1a) with the extraction from an adjunct headed by a present participle in (1b), which is considered grammatical; the (participial) adjunct predicate is shown in square brackets in most of the examples used in this article. An acceptable extraction from subject is shown in (2).

(1) a. *Who did Mary cry [after John hit t]?   (Huang 1982, p. 503)

     b. What$_i$ did John arrive [whistling t$_i$]?   (Borgonovo and Neeleman 2000, p. 200)

(2) What did [$_{IP}$ [$_{NP}$ the attempt to find __] end in failure]?

                                    (Hofmeister and Sag 2010, p. 370)

Attested examples of extraction from participle adjuncts, as in (1b), are often found in the form of relativization, as in (3) from Santorini (2019) and (4) from a news article; in these cases, it is a nominal element that is associated with a gap site in the complement position of a participle adjunct instead of a *wh*-pronoun.

(3) a. The magazine I spend most of my days [reading __].

 b. "This is the game I grew up [watching __]," Wilson added.

<div align="right">(Santorini 2019)</div>

(4) Already uncomfortable with the policy—the USD 1.3 trillion legislation includes massive spending hikes that contradict prior GOP complaints about the debt—many Republicans were left dumbfounded by [a process]$_i$ that looked a lot like one they had won office [criticizing __$_i$]. (Washington Post, 22 March 2018)

I follow Truswell (2011, p. 30) in referring to adjunct constructions, such as (1b), (3), and (4) as Bare Present Participle Adjuncts (BPPA); they are characterized by an untensed present participle as the head of the adjunct predicate, as well as the absence of an explicitly encoded subject or subordinators.

However, not all BPPA constructions allow extraction as easily as (1b): minimally different examples, such as (5), are reported to block the extraction.

(5) *What$_i$ did John dance [imagining t$_i$]? (Borgonovo and Neeleman 2000, p. 199)

There are, thus, two problems to be addressed: (i) apparently grammatical extractions from adjuncts that should be excluded by the CED (1a vs. 1b), and (ii) variation within an adjunct type where some extractions are allowed while others are not (1b vs. 5). The first problem has been addressed in the Minimalist literature in two ways: first, abandoning or modifying the original formulation of the CED to accommodate such cases,[1] and second, to reconsider the adjunct status of apparent counterexamples to the CED (e.g., Graf 2015). These approaches still assume that there is a syntactic principle at work which determines when extraction is possible. A major alternative to this syntactic perspective is taken by approaches which do not assume a syntactic principle behind extraction asymmetries, but rather more general principles. This includes approaches based on processing (e.g., Sag et al. 2008; Hofmeister and Sag 2010), information structure (Goldberg 2006, 2013), pragmatics (Chaves and Putnam 2020), and discourse functions (Abeillé et al. 2020). Such approaches line up with the Radical Unacceptability Hypothesis proposed in Culicover et al. (2022), to which I return at the end of this article.

I will focus on the second problem in the rest of this article.[2] For reasons of space, the discussion focuses on examples with *wh*-extraction, but it should be kept in mind that there is growing evidence that different types of filler–gap dependencies yield different effects, so that so-called island constraints do not appear to be cross-constructionally active (Liu et al. 2022); see also Sprouse et al. (2016) and Abeillé et al. (2020) for findings and discussion, as well as Kehl (2021, experiment 1) for a comparison between declarative, interrogative, and relativized BPPA constructions. Differences between types of extractions become all the more relevant since much of the existing literature focuses on *wh*-extraction, whereas many attested examples are instances of relativization (Chaves and Putnam 2020; Santorini 2019). I will briefly address other dependency types at the end of Section 5.

The variation in the extraction behavior of interrogative BPPA constructions has resulted in several approaches that try to find an explanation for such patterns; the influential theoretical approaches in Borgonovo and Neeleman (2000) and Truswell (2007, 2011) propose licensing conditions to accommodate this island-internal variation. In this article, I will follow the discussion in Brown (2017) and Kehl (2021), agruing that the *grammaticality* patterns observed for extraction from BPPA constructions are actually a reflection of different degrees of *acceptability* which are already observable in the declarative counterparts and evoke the impression of grammaticality differences in interrogatives. Thus, I assume that the acceptability difference between (1b) and (5) is equal to that between the declaratives in (6):

(6) a. John arrived [whistling the Blue Danube].

    b. John danced [imagining the Gobi Desert].

<div align="right">(Borgonovo and Neeleman 2000, pp. 199–200)</div>

By extension, I also assume that the same acceptability contrast between verbs such as *arrive* and *dance* is visible in other dependency types, such as the relative clauses in (7), which are more similar to the attested data in (3) and (4) above:

(7) a. This is [the song]$_i$ that John arrived [whistling ___$_i$].

    b. This is [something]$_i$ that John danced [imagining ___$_i$].

The basic idea behind this assumption, which I will argue for in the remainder of this article, is that differences between the two main verbs *arrive* and *work* result in different degrees of acceptability independently of whether the sentence form is declarative, relativization, or interrogative. In other words, the relative acceptability of the declaratives are good predictors for relative acceptability in different sentence forms; see also Chaves and King (2019), who find a relation between judgments of relevance and acceptability of subextraction from subjects. This line of research shifts the focus of attention to the semantic and/or pragmatic factors which affect acceptability in the underlying declarative structures. This comparison of extraction from island constructions to possible differences in the underlying declaratives ties into the growing body of research that does not focus on extraction constructions alone (among others, Abeillé et al. 2020; Brown 2017; Chaves and King 2019; Chaves and Putnam 2020). The relevance of drawing on more subtle differences in declaratives to explain differences at the fringes of grammaticality in extraction structures goes back to at least Kuno (1987), an idea that is picked up prominently in the pragmatic approach to extraction asymmetries in Chaves and Putnam (2020), but also the discussion of complexity differences in Culicover and Winkler (2022).

The discussion in this article centers around the question whether extraction asymmetries observed for BPPAs need to be captured by a grammatical principle, as proposed in Borgonovo and Neeleman (2000) and Truswell (2007, 2011), or whether these asymmetries can be explained independently. I argue for the second position and discuss the possibility of capturing judgment differences, such as (1b) vs. (5); the underlying idea is that the semantic compatibility between the two predicates in this construction affects acceptability both in the presence and absence of a dependency such as *wh*-extractionextraction. The discussion of this narrow set of examples is closely related to the more general proposal in Culicover and Winkler (2018, 2022) and Culicover et al. (2022) that many instances of such judgment differences in extraction phenomena can be accounted for without the need to introduce grammatical principles.

This article is structured as follows: I will first provide a short summary of the grammaticality patterns reported in Borgonovo and Neeleman (2000) as well as Truswell (2007) in Section 2 as a basis for the remainder of the discussion; in Section 3, I discuss the relations between the concepts of grammaticality and acceptability, as well as the potential mapping problems between gradient and binary judgments; I will then suggest a factorial design for the detection of island-internal variation that allows for an experimental validation of factors that are assumed to influence how strong extraction affects different types of declaratives; Section 4 examines previous experimental studies which compare declarative and interrogative BPPA constructions and whether their results speak for or against the conclusions in the theoretical literature; Section 5 then discusses factors which influence the acceptability of declarative BPPA constructions independently of extraction and combine these factors into an acceptability model for declarative and interrogative BPPA constructions; in Section 6, I take a brief look at evidence from related phenomena that also come to the conclusion that differences in declaratives have an impact on theory development; Section 7 concludes this article.

2. Reported Grammaticality Patterns

In this section, I will summarize the reported grammaticality patterns for extraction from participle adjuncts in two influential accounts: Borgonovo and Neeleman (2000) and Truswell (2007).[3] Both accounts share the intuition that different grammaticality patterns exist in interrogatives which are not present in declaratives; this leads them to propose additional licensing mechanisms for extraction to accommodate these interrogative patterns. I will not go into the technical details of these accounts for reasons of space and because the focus of this article is on the relation between declaratives and interrogatives instead of the licensing mechanisms they propose. As I will show in Section 3, such a comparison uncovers problematic aspects of these accounts.

2.1. Transparency Depends on Verb Types

Borgonovo and Neeleman (2000) report on a grammaticality pattern that allows extraction from participial adjuncts modifying unaccusative and reflexive transitive main verbs, as in (8a) and (8b); in contrast, extraction from adjuncts modifying unergative and non-reflexive transitives, as in (8c) and (8d), results in ungrammaticality.

(8) a. What$_i$ did John arrive [whistling t$_i$]? [unaccusative]

    b. What$_i$ did John hurt himself [trying to fix t$_i$]? [reflexive transitive]

    c. *What$_i$ did John dance [imagining t$_i$]? [unergative]

    d. *What$_i$ did John hurt Bill [trying to fix t$_i$]? [non-reflexive transitive]

    (Borgonovo and Neeleman 2000, pp. 199–200)

The main proposal resulting from this pattern is that some verb types are able to L-mark adjuncts by means of a syntactic reflexivity relation where the internal argument DP binds both the θ-roles of the adjunct predicate and the main verb. The adjunct will then count as L-marked and obeys the CED because it is properly governed. Only unaccusatives and reflexive transitives are able to L-mark the adjunct because the right structural configuration is only possible with an internal argument that is also the external argument of the adjunct predicate. Unergatives fail to L-mark the adjunct because they do not have an internal argument and do not project the necessary V′-layer (Borgonovo and Neeleman 2000, pp. 212–13); L-marking is not possible for non-reflexive transitives because the external argument of the adjunct is also the external argument of the main verb. In both cases, extraction is banned by the CED because the adjunct is not L-marked.

Crucially, L-marking is a condition that is specific to the licensing of extraction: it does not have an effect in declaratives because it is irrelevant there. Therefore, the declarative sentences in (9) underlying the interrogatives in (8c) and (8d) are completely unmarked.

(9) a. John danced [imagining the Gobi Desert].

    b. John hurt Bill [trying to fix the roof].

    (Borgonovo and Neeleman 2000, pp. 199–200)

Because declarative BPPA constructions are unconstrained in terms of grammaticality differences, the source of ungrammaticality in interrogatives is caused by the extraction operation itself, which fails to be licensed if L-marking cannot be established for unergatives and non-reflexive transitives. The required adjustments to subjacency-based locality theory are modest and can be expressed in core-syntactic terms, even if the theory requires ternary branching to establish syntactic reflexivity between the verb, its internal argument, and the adjunct predicate. Still, a major problem with this account is that it does not consider any potential variation in the declarative counterparts and exclusively relies on extraction-related factors to explain the pattern in interrogative BPPA constructions.

2.2. Transparency Depends on Telicity

A slightly different pattern is described in Truswell (2007), who focuses on the event structure of the BPPA construction. The key proposal is that extraction from an adjunct predicate is only licensed in the grammar if the adjunct fills an open or underspecified event position in the event structure of the matrix predicate. This means two things: (i) the matrix predicate needs to encode at least two subevents, and (ii) one of these is underspecified by the lexical semantics of the matrix predicate. The two event types that encode more than one subevent are achievements and accomplishments in terms of Vendler (1957); they are composed of a culmination point and a durative subevent leading up to this endpoint, which is optional for achievements; see Rothstein (2004). States and activities, on the other hand, either encode no event at all (states) or only a single subevent (activities). In case the adjunct can be interpreted as supplying more information about the underspecified subevent, the two predicates describe facets of a single event, mirroring the lexical semantics of a maximally complex verb (Truswell 2007, p. 1369). This amounts to the generalization that extraction from the adjunct is only possible if the matrix predicate is telic; this derives the predictions for the contrast in (10) with the atelic verb *work* (10a) and the telic *arrive* (10b):

(10) a. *What did John work whistling __ ?   [atelic matrix predicate]

     b. What did John arrive whistling __ ?   [telic matrix predicate]
                                                (see Truswell 2007, p. 1369)

These predictions are similar to those in Borgonovo and Neeleman (2000), but formulated in event-semantic terms, which are not exclusively tied to extraction. In addition to achievement matrix predicates, such as (10b), extraction is also possible from structures with accomplishment main verbs, such as in (11), provided that the adjunct can describe the cause of the matrix predicate:

(11) What did John drive Mary crazy [trying to fix t]?   (Truswell 2007, p. 1356)

Like Borgonovo and Neeleman (2000), Truswell (2007) concludes that the corresponding declaratives in (12) do not show a similar pattern and that the grammaticality pattern in interrogatives is the result of extraction. Both accounts do consider declarative counterparts with respect to their grammaticality, but do not observe significant differences in acceptability.[4]

(12) a. John worked [whistling a song].

     b. John arrived [whistling a song].
                                         (see Truswell 2007, pp. 1369, 1373)

This means that the syntactic extraction operation needs to be sensitive to the distinctions between different event types, but also to the lexical semantics of the two predicates, as well as potential causal chains between them. Unless information about the aspectual type and causality are directly encoded syntactically, as, for example, in Borer (2005) and Ramchand (2008), this extraction pattern is impossible to explain in core syntactic terms. It is not an immediate problem that Truswell (2007) considers both sentences in (12) grammatical, but this focus on grammaticality requires the formulation of extraction conditions in event-semantic terms (or a post-syntactic event-semantic output filter, as suggested in Truswell 2011).

Both accounts sketched in this section agree that declarative BPPA constructions are relatively unconstrained with respect to grammaticality differences and that the pattern in interrogatives is a direct result of failures in the licensing mechanism for extraction. I will argue in the following section that this perspective overestimates the reported grammaticality differences in interrogatives, and at the same time underestimates potential differences

in the declarative counterparts. The main reason for these problematic aspects is rooted in the distinction between the concepts of grammaticality and acceptability, as well as the relation between gradient and binary judgments.

3. Grammaticality, Acceptability, and the Relation between Declaratives and Interrogatives

In this section, I will discuss problematic aspects of the exclusive focus on *grammaticality* differences in interrogatives without also considering potential *acceptability* differences in the declarative counterparts. The problem is one of mapping relations between binary grammaticality judgments and gradient acceptability judgments, because sentences that receive the same binary grammaticality marking may still show significant differences in acceptability that are not properly represented in all grammaticality judgments. For example, it is reasonable to consider both examples in (13) grammatical, but experimental evidence suggests that (13a) is less acceptable than (13b). Among others, the lower acceptability and negative impact on online sentence processing of additional arguments is shown in Jurka (2010, 2013), Polinsky et al. (2013), Brown (2017), and Culicover and Winkler (2022). An additional issue in (13a) is that there is a degree of ambiguity whether the adjunct refers to John or Bill. In connection with syntactic dependencies, the greater processing cost and, thus, reduced acceptability is predicted by Dependency Locality Theory (Gibson 1998, 2000); see also Section 4.[5]

(13) a. John hurt Bill [trying to fix the roof].

b. John arrived [whistling the Blue Danube].

(Borgonovo and Neeleman 2000, p. 200)

Subsection 3.1 describes the contrast between binary grammaticality and gradient acceptability judgments, as well as their relation in more detail; the focus here is on which conclusions can be drawn from these two measurements and the risk of not distinguishing between them properly. Subsection 3.2 proposes an adapted factorial experiment design that allows for the investigation of island-internal variation which includes a comparison to the declarative base position. Subsection 3.3 emphasizes the usefulness of including standardized reference fillers in acceptability judgment tasks for conceptual and methodological reasons.

3.1. Gradient and Binary Judgments

One of the core issues in the evaluation of the theoretical approaches in Borgonovo and Neeleman (2000) and Truswell (2007, 2011) lies in the distinction between the concepts of *grammaticality* and *acceptability* discussed in Chomsky (1965). Chomsky (1965) models this distinction as one between *competence* and *performance*: the former refers to those aspects of language that are part of a speaker's grammar, whereas the latter reflects the use of language that is also affected by other factors. Grammaticality is seen as a measure of whether a sentence is licensed by the grammar; this evaluation has often been considered to be a categorical distinction, even though Chomsky (1965, p. 11) already notes that it is probably "a matter of degree". Acceptability as a measure of naturalness and comprehensibility does not solely depend on grammaticality, but grammaticality is one of the factors that determine acceptability: a sentence that is considered grammatical can still show low acceptability because they are semantically or pragmatically anomalous, or because they are difficult to process (Chomsky 1965, p. 11). Ungrammaticality refers to the fact that a given structure cannot be computed by the grammar, or runs afoul at the interfaces, for example because not all uninterpretable features are checked and deleted during the derivation. Acceptability is partially fed by grammaticality, but also affected by additional factors that are independent of grammaticality: as is well known, there are sentences which can be generated by the grammar but can be anomalous semantically and/ or pragmatically, or pose processing difficulties that impact acceptability judgments. On the other hand,

there are sentences which are grammatically ill-formed but appear intuitively acceptable, a phenomenon called 'illusions of grammaticality' in Phillips (2013, p. 106).

There is, thus, a mapping problem between grammaticality and acceptability because not all sentences that are considered grammatical are necessarily equally acceptable, also noted in Chomsky (1965, p. 11). Especially problematic are cases where acceptability is on the borderline or threshold of grammaticality: minimally different acceptable sentences run the risk of being assigned opposite grammaticality judgments, even if the relative distance in acceptability between them is smaller than the distance between two fully grammatical or ungrammatical sentences. I will elaborate on this problem in the remainder of this subsection.

Consider the two declarative BPPA constructions in (14), with an unaccusative (14a) and an unergative (14b) matrix predicate. The predictions of Truswell (2007) and Borgonovo and Neeleman (2000) agree on the fact that extraction from the adjunct in (14a) will be grammatical, whereas extraction from (14b) will not.

(14)   a.   John arrived whistling a funny song.

       b.   John worked whistling a funny song.

Let us assume a gradient Likert-type judgment scale with seven discrete points, and a binary categorization into grammatical and ungrammatical sentences. Let us also assume that gradient judgments on or above the middle of the gradient scale, i.e., every gradient judgment ≥ 4, will be mapped to the binary judgment 'grammatical', and that gradient judgments < 4 will be mapped to 'ungrammatical'. Thus, if (14a) is assigned a gradient judgment of 7 and (14b) a judgment of 5, both will be mapped onto a grammatical binary judgment; this is shown in (15).

(15)   a.   (14a) → gradient judgment: 7 → binary judgment: grammatical

       b.   (14b) → gradient judgment: 5 → binary judgment: grammatical

This is, in essence, what Borgonovo and Neeleman (2000) and Truswell (2007, 2011) assume about declarative BPPA constructions, with a focus on the outcome of the binary grammaticality judgment. For now, it is not immediately relevant why (14b) should be less acceptable on a gradient scale compared to (14a). The data reported in Brown (2017) and Kehl (2021) support the assumption that there is a statistically significant acceptability difference between the two, even if this difference might not be as pronounced as in this hypothetical example.

When extraction takes place from the adjunct, the gradient judgment will decrease for both structures because the formation and resolution of filler–gap dependencies is a cognitively costly operation and because interrogatives are semantically more complex than declaratives (Chaves and Putnam 2020; Hofmeister and Sag 2010; Wagers 2013). Since the extraction domain is an adjunct, this judgment decrease will probably be larger compared to extraction from a subcategorized complement, as predicted by the CED.[6] The interrogative counterparts of (15) are shown in (16), without judgment marks.

(16)   a.   What did John arrive whistling?

       b.   What did John work whistling?

A final assumption made here, again supported by the experimental evidence in Brown (2017) and Kehl (2021), is that both structures are affected to the same degree by extraction, meaning that the decrease in the gradient judgment is identical; the gradient judgment for (16b) will then fall below the threshold in the middle of the scale, resulting in an ungrammatical binary judgment. For (16a), the gradient judgment remains on or above the threshold, yielding a grammatical binary judgment; this is shown schematically in (17).

(17)　a.　(16a) → gradient judgment: 5 → binary judgment: grammatical

　　　b.　(16b) → gradient judgment: 3 → binary judgment: *ungrammatical

On the surface, this results in exactly the grammaticality patterns constructed in Borgonovo and Neeleman (2000) and Truswell (2007). However, what we are mostly interested in is whether the relative differences in gradient judgments between the two sentence pairs are identical or whether one is larger than the other. To check for this, we subtract the two declarative judgments from one another and compare this to the same difference between the interrogative counterparts. If the difference pairs are equal to each other (or at least not significantly different), then there is no need for additional licensing mechanisms for extraction because the gradient judgment differences in interrogatives can be predicted from the differences in declaratives in a linear additive way. This is shown in (18a) and is the simpler case because then the only explanation required is what causes the differences in declaratives plus the independent decrease caused by extraction. If, on the other hand, the relative differences are of different magnitudes, as shown in (18b) and (18c), then this requires an explanation for this additional difference that cannot be predicted from the gradient contrasts in declaratives. These patterns can be referred to as superadditive and subadditive. Depending on whether the difference between declaratives is smaller than that for interrogatives, as in (18b), or the other way round, as in (18c), this leads to the need either for additional licensing mechanisms or a repair mechanism, respectively.

(18)　a.　$\frac{\text{differences between declaratives}}{\text{differences between interrogatives}} = 1$　　　　　　　[no licensing mechanism required]

　　　b.　$\frac{\text{differences between declaratives}}{\text{differences between interrogatives}} < 1$　　　　　　　[licensing mechanism required]

　　　c.　$\frac{\text{differences between declaratives}}{\text{differences between interrogatives}} > 1$　　　　　　　[repair mechanism required]

This metric is similar to the differences-in-differences score employed in Sprouse et al. (2012, 2013), which isolates the effect sizes of individual factors and evaluates whether the combination of two factors negatively impact acceptability scores to a greater (or lesser) degree than the two individual factors.

Figure 1 illustrates the first two possibilities in (18): the left panel corresponds to (18a) where the gradient judgment differences are identical for both types of matrix predicate; the right panel shows the pattern where the decrease caused by extraction is larger for atelic matrix predicates than that for telic predicates (18b). I omit the case of (18c) for expository purposes. The shaded area in Figure 1 shows the range of the gradient scale that will be mapped onto an ungrammatical binary judgment. The experimental results in both Brown (2017) and Kehl (2021, experiment 2) correspond more closely to the pattern on the left rather than the one on the right, showing that the strength of the acceptability decrease in interrogatives is not influenced by the other factors they investigate. I will return to this discussion in Section 4 below.
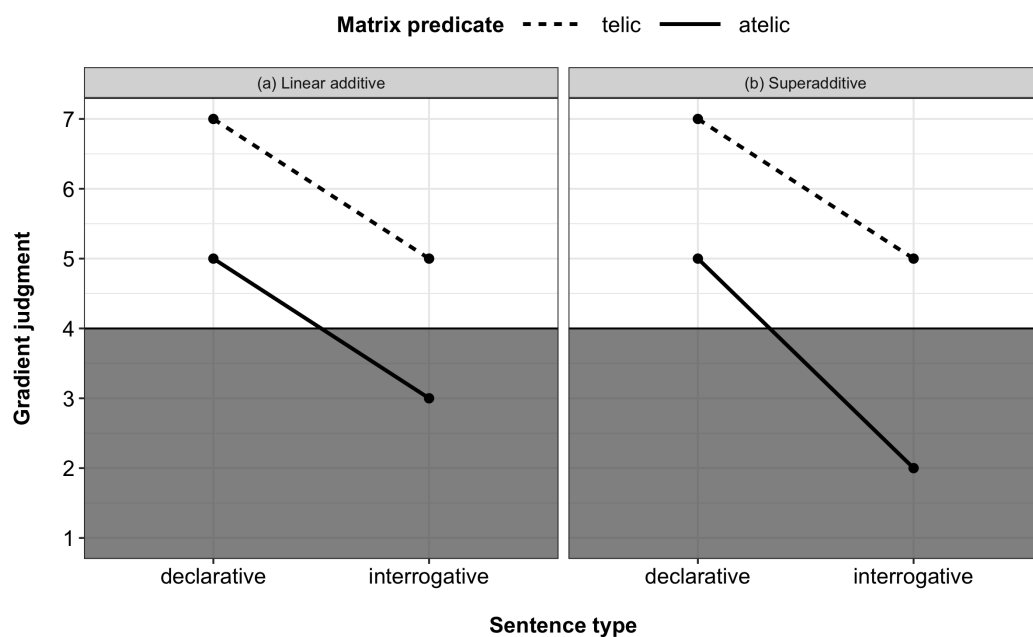
Figure 1. Schematic illustration of the linear additive and superadditive gradient judgment patterns in (18); the shaded area shows the part of the scale that will be mapped to ungrammatical judgments.

The difference between these patterns is obscured if the sole focus is on the binary judgment because this ignores potential differences in gradient judgments for declaratives; in a sense, some information is lost in the mapping between gradient and binary judgments. On its own, this is not problematic, but it becomes so if used as a basis for the postulation of licensing mechanisms for extraction. One possibility to avoid the potential pitfalls of binary judgments is to broaden the data pool and gather binary judgments from multiple informants, which can then be converted to a gradient scale similar to Likert-scales by calculating the ratio of grammatical-to-ungrammatical responses (Bader and Häussler 2010, 2019). This method has been shown to result in similar patterns as judgments on discrete or continuous scales.

Taking a step back, the binary judgment differences in Borgonovo and Neeleman (2000) and Truswell (2007) can be converted to acceptability measures, meaning that the grammatical extractions are more acceptable than the ungrammatical ones. However, the formulation as grammaticality judgments runs the risk of leading to proposals about the architecture of the syntactic component and its interfaces with semantics and pragmatics. Therefore, I think that it is advisable to focus on acceptability first and then reason about the model of grammar that best fits with these results.

3.2. A Factorial Design for Island-Internal Variation

The procedure described in the previous section represents a modification of the factorial design for island effects in Sprouse et al. (2012, 2013) and Kush et al. (2018, 2019). The original design compares conditions in a way that allows to isolate the individual effects of two factors: the difference between extraction from matrix clauses vs. embedded domains and between extraction from non-island vs. island domains. See (19) for an illustration of the design:

(19) a. Who __ thinks that John bought a car? [NON-ISLAND/MATRIX]

   b. What do you think that John bought __ ? [NON-ISLAND/EMBEDDED]

   c. Who __ wonders whether John bought a car? [ISLAND/MATRIX]

   d. *What do you wonder whether John bought __ ? [ISLAND/EMBEDDED]
      (Sprouse et al. 2013, p. 25)

This design allows quantifying three sets of contrasts and the respective effects they have on acceptability: (i) the contrast between (19a)–(19b) isolates a possible effect between extraction from the matrix clause vs. the embedded clause; (ii) the contrast between (19a)–(19c) detects whether the presence or absence of an island domain, in this case a *wh*-island introduced by *whether*, affects acceptability; and (iii) the contrast between (19b)–(19d) compares the cost of extraction from a non-island vs. from an island domain (see Sprouse et al. 2013, p. 25). Often, theoretical approaches will focus on the contrast between (19b) vs. (19d) and conclude that *whether*-clauses are islands if this extraction feels less acceptable than the non-island. However, this leaves unaccounted the potential effect that the presence of a *whether*-clause has on acceptability independently of extraction.

To solve this, Sprouse et al. (2012, 2013) include this effect in the calculation of potential island effects: if the acceptability judgment for the 'worst' condition (19d) compared to the unmarked reference condition (19a) cannot be predicted from the differences between (19a) and (19b) and (19a)–(19c), then this additional acceptability decrease is called an 'island effect' which needs to be accounted for theoretically.

The same reasoning can be applied to investigate the validity of theoretical approaches such as those in Borgonovo and Neeleman (2000) or Truswell (2007): instead of comparing an island construction with a non-island, two instances of the same island type are tested in declarative and interrogative conditions. They differ minimally in one of the factors isolated in the literature, such as event structure or the verb type of the matrix predicate. This allows to examine whether such factors determine how strongly extraction degrades acceptability, as well as whether there are acceptability differences in the declaratives that are the source of the reported differences in interrogatives. An example of such a design is given in (20), based on the example sentences discussed in the previous section. This relatively simple 2 × 2 design manipulates the matrix predicate as telic or atelic, as well as the difference between declarative and interrogative sentences. The manipulation of other factors, also ones with more levels, is of course also possible; for a more complex 2 × 2 × 2 design that crosses the factors TELICITY, TRANSITIVITY, and EXTRACTION, see Brown (2017). For example, the simple comparison between declarative and interrogative sentence forms can be augmented to also include relative clauses and topicalization.

(20)   a.   John arrived whistling a funny song.   [telic/−wh]

        b.   What did John arrive whistling?   [telic/+wh]

        c.   John worked whistling a funny song.   [atelic/−wh]

        d.   What did John work whistling?   [atelic/+wh]

The statistical analysis will then compare the effects of the two factors, in this case telicity and extraction, as well as the interaction between them. The absence of a significant interaction indicates that the strength of extraction is not influenced by the factor that distinguishes the declarative conditions. If there is a significant interaction, additional licensing or repair mechanisms are called for, as explained in the previous section. Like the detection of island effects in the original factorial design in Sprouse et al. (2012, 2013), the question whether extraction from a 'suboptimal' adjunct island configuration leads to drops in acceptability that cannot be explained independently of extraction would lead to additional licensing requirements. Determining this need for licensing mechanisms should be at the core of investigations into island-internal variation and should be backed up with experimental data in addition to initial, intuitive judgments.

3.3. The Use of Standardized Fillers

The results of gradient judgment studies can sometimes be difficult to interpret. Typically, the experimental conditions are compared to each other in terms of significant differences between conditions in the data pool, or in terms of effect structures in the case of factorial designs. Although this is the main interest of an experimental study, i.e., to test hypotheses about acceptability contrasts and the influence of specific factors, it is also

of interest to compare where the experimental conditions are located on the continuum of gradient acceptability, regardless of whether this continuum is expressed in discrete Likert-type scales or truly continuous judgments as in Magnitude Estimation (Bard et al. 1996) or Thermometer judgments (Featherston 2020). One possibility is to add control conditions that are closely related to the construction under investigation, as implemented in Abeillé et al. (2020) with grammatical and ungrammatical controls.[7] In her experiment on extraction from adjuncts in English that is closely related to the design in (20), Brown (2017) includes grammatical and ungrammatical controls as in (21); as extraction from tensed adjuncts as in (21b) is not always considered unacceptable, extraction from a conjunct as in (22a) can also be used for unacceptable controls because there is general agreement in the literature that such extractions are ungrammatical (Liu et al. 2022). A resumptive pronoun at the gap site, as shown in (22b), can also be used to construct ungrammatical control conditions that are close to the design implemented (Chaves and Putnam 2020, pp. 218–19).

(21)   a.   Which ice cream did Mary eat before she saw the celebrity?

[grammatical control]

        b. *Which celebrity did Mary eat an ice cream before she saw?

[ungrammatical control]
(Brown 2017, p. 120)

(22)   a. *What did Mary go to work and whistle?

        b. *What did Mary arrive at the office whistling it?

The set of standardized reference fillers developed for English in Gerbrich et al. (2019) are designed to provide anchor points along gradient or discrete judgment scales, ranging from a high level of acceptability to a low level; the idea of providing a standardized scale for acceptability is also found in Featherston (2009), who develops a set of German reference fillers.

The goal of the standardized fillers is to provide anchor points on the extremes of the rating scale with highly acceptable and highly degraded sentences, as well as a range of acceptability in between; ideally, this results in a reference scale with equal distances between the individual levels, so that the experimental items can be assigned a relative level of normed acceptability. The choice of very general levels of well-formedness along the spectrum of acceptability which is not limited to control items that are related to the construction has the advantage that the fillers can be re-used across multiple experiments and, thus, allows a more grounded discussion of acceptability across experiments. It is of course possible to include both the standard fillers and construction-specific control conditions in an experiment. A sample of the reference fillers is given in (23); the assignment of more traditional graded grammaticality judgment marks, such as '?' or '*', are adapted from Gerbrich et al. (2019, p. 310).

(23)   a. $\sqrt{}\sqrt{}$A: The patient fooled the dentist by pretending to be in pain.

        b. $\sqrt{}$B: Before every lesson the teacher must prepare their materials.

        c.   ?C: Hannah hates but Linda loves eating popcorn in the cinema.

        d.   ??D: Who did he whisper that had unfairly condemned the prisoner?

        e.   *E: Historians wondering what cause is disappear civilization.

(see Gerbrich et al. 2019, p. 315)

The two best levels A and B are usually not marked in such judgment schemes, and are both considered fully grammatical; still, Gerbrich et al. (2019) suggest that there are significant acceptability differences between these grammatical levels, which are difficult to detect in judgments with limited conventionalized markings. Judging from their experimental results with the standardized fillers, Gerbrich et al. (2019, p. 309) conclude that

there may be even more distinguishable levels of well-formedness. Note that the E-level is still interpretable, but highly unnatural; it is possible to add a further level with low interpretability, as for example in the adaptation of the standard fillers in Brown et al. (2021). Brown et al. (2021, p. 10) refer to this as a "clearly ungrammatical level" with examplessuch as *The ink was for spilled* that are considered both unacceptable and uninterpretable.

Figure 2 illustrates the expected distribution of the five sets of standardized fillers on a 7-point scale of acceptability; see also the discussion in Featherston (2020, pp. 168–72) showing a similar distribution in *z*-scores based on an actual experiment. The exact values may vary from experiment to experiment, and it may not always be the case that the distance between the levels is evenly distributed, especially if target conditions fall between two of the levels (Gerbrich et al. 2019, pp. 315–16). From these predicted values and the judgment marks in Gerbrich et al. (2019), it becomes apparent that the binary ungrammaticality marking may be limited to a rather small gradient acceptability area, unlike the assumption above that the threshold for binary grammaticality judgments lies in the middle of the gradient scale. I leave this point open for discussion here.
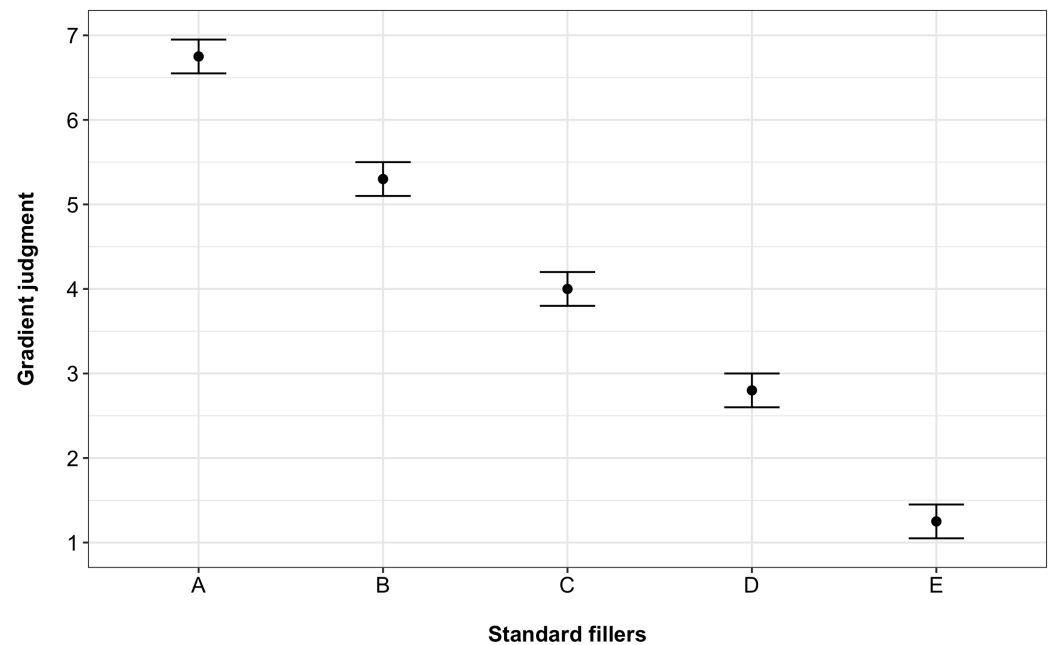

Figure 2. Expected rating distribution for the A–E levels of the standard fillers in (23) on a gradient judgment scale.

By comparing the experimental items in declarative and interrogative conditions relative to their location on the gradient acceptability continuum established by the reference fillers, more reliable conclusions about the relative acceptability of BPPA constructions can be made. Since the fillers leave enough room in the upper half of the scale (A–C) for highly acceptable to slightly marked levels of acceptability, even subtle differences in declarative BPPA constructions that are obscured in binary intuitive judgments can be detected. With respect to interrogative BPPA constructions, it is of interest whether they decline all the way to the bottom of the scale in suboptimal conditions and how large the difference is to conditions that the literature considers to be grammatical.

The use of the standardized fillers in an experimental setting also has two more mundane, methodological benefits: (i) a plausibility check for the target items, and (ii) a plausibility check for participant responses.

In a typical experiment, it is advisable to construct target items that avoid the extreme points of a closed scale to prevent ceiling and floor effects. It is also advisable to exclude target items that have no unique structural representation (*word salad*) because the researcher cannot determine which structural parse is being judged (Gerbrich et al. 2019, pp. 310–11).

The E-standards are marked in several clearly determinable ways but still have a unique structural representation, whereas the A-standards contain no structural or semantic faults. This means that researchers should become skeptical if their target items fall outside the range of the standard fillers, i.e., if there are items averaging significantly better than the A-standards or significantly below the E-standards. There can be, of course, good reasons for such situations, but the results should be scrutinized closely. Target items that fall somewhere between the ranges of the standard fillers can be more clearly evaluated for their overall gradient acceptability.

The second point concerns the reliability of participant judgments. As these judgments are collected in an anonymous fashion and there are no negative consequences for incoherent or blatantly random judgments, data quality needs to be ensured at some point. Especially experiments that are carried out with compensation of some kind, be it monetary or for course credit, may create an environment where participants are not really engaged with the task and do the experiment half-heartedly. Large crowdsourcing platforms, such as Amazon's Mechanical Turk or others, have on the one hand been shown to provide usable data (Gibson et al. 2011; Sprouse 2011), but on the other hand it can always happen that participants try to complete as many tasks as possible for maximum compensation. An ethical amount of payment is a first step to avoid this, but does not guarantee accurate data. To solve this issue, the standard fillers can be evaluated for individual participants to see whether they reproduce the expected decline in mean acceptability from the A-standards to the E-standards. If the E-level averages higher than the C-level, for instance, this is a good indication that the experiment was not carried out diligently, providing a principled reason to exclude this participant from the statistical evaluation. Although the judgments for the standard fillers may not always exactly follow the expected pattern, as shown in Featherston (2020, p. 170), it is still possible to distinguish completely random judgments from those that are slightly off.

These two methodological points have shown that the standardized fillers have a valid use in experimental judgment studies in addition to the better comparability with stable levels of acceptability. They provide a more fine-grained scale of well-formedness compared to binary judgments, and also allow for a more principled conversion to traditional judgment marks, such as the question mark or the asterisk.

3.4. Interim Conclusion

In this section, I have discussed three issues that should be considered in the analysis of island-internal variation, exemplified with an evaluation of theoretical approaches to BPPA constructions. First, the relation between grammaticality and acceptability and how this relation can become problematic for theoretical conclusions about locality operations, such as *wh*-extraction. I have argued that there is nothing wrong in considering the two declarative sentences in (14) grammatical; it is, however, problematic to ignore potentially interesting differences in acceptability. Second, I have described the use of a factorial design to better describe island-internal variation in relation to the variation that is independent of extraction from the island. This design avoids potential confounds that arise if too much emphasis is placed on variation in interrogatives. Third, I have discussed how acceptability judgment tasks can benefit from the use of the standardized fillers in Gerbrich et al. (2019), from both conceptual and methodological perspectives. In combination with a factorial design that includes declarative base-structures, this allows for a principled analysis of the effects operating in specific types of islands.

In the following section, I consider existing experimental work on the acceptability of declarative and interrogative BPPA constructions and how these results compare to the issue of gradient acceptability and ramifications for the construction of licensing mechanisms for extraction.

4. Previous Experimental Investigations

The idea that not all declarative BPPA constructions are equally acceptable because the adjunct predicate is not semantically licensed in all configurations was first proposed in Brown (2015, 2016, 2017). She argues that only low-merged VP adjuncts are in the right structural configuration to allow extraction, whereas high-merged *v*P adjuncts resist extraction.[8] By hypothesis, not all types of participle adjunct predicates qualify as low-merging adjuncts to all types of matrix predicate. This means that some participle adjuncts fail to be licensed in the configuration that would allow extraction, which leads to reduced acceptability that is independent of whether extraction takes place or not. Brown (2017) formulates this as a distinction between the semantic licensing conditions on the low-merging adjunct and the syntactic licensing conditions for extraction. For the semantic licensing conditions of low-merging adjuncts, she suggests that the temporal interval of the matrix predicate should be properly included in that of the adjunct predicate, which works best if the adjunct predicate is atelic and the matrix predicate telic; this is essentially the generalization formulated in Truswell (2007). Kehl (2021) goes in a similar direction by proposing a set of semantic compatibility and syntactic complexity criteria that determine the acceptability level of the declarative BPPA construction, taking into account the properties of the host predicate. Both approaches share the common assumption that there is a principled relation between acceptability differences in interrogatives and the corresponding declaratives.

Brown (2017) shows experimentally that there are significant effects of transitivity in declarative BPPA constructions, and that this effect does not interact with the presence vs. absence of a gap: thus, the relative acceptability difference between the intransitive (24a) and the less acceptable transitive (24b) is the same as that between the corresponding declaratives in (25a) and (25b).

(24)   a.   Which tune did Monica arrive whistling?   > more acceptable

          b.   Which tune did Julia pick the candidates whistling?

                                                       (Brown 2017, p. 119)

(25)   a.   Lucy arrived whistling the national anthem.   > more acceptable

          b.   Mary picked the candidates whistling the national anthem.

                                                         (Brown 2017, p. 119)

What this means is that transitivity shows an effect on acceptability but does not determine how strongly extraction affects acceptability. This result is unexpected in the framework proposed by Borgonovo and Neeleman (2000).

In addition, Brown (2017) shows that telicity only has a significant effect for intransitive matrix predicates, i.e., for unergatives and unaccusatives. Transitive atelic activities and transitive telic accomplishments do not show a similar sensitivity. This is not predicted by the event-semantic account in Truswell (2007). For example, the telic transitive sentence in (26a) is equally acceptable as the atelic transitive in (26b), but the telic intransitive in (27a) is more acceptable than the atelic intransitive in (27b); the same obtains for the corresponding interrogatives. Similar observations are found in Kehl (2021).

(26)   a.   Mary picked the candidates whistling the national anthem.

          b.   Sophie finished sketches whistling the national anthem.

                                                       (Brown 2017, p. 119)

(27)   a.   Lucy arrived whistling the national anthem.   > more acceptable

          b.   Lucy shivered whistling the national anthem.

                                                       (Brown 2017, p. 119)

Brown (2017) concludes that transitivity is a key factor in determining the acceptability of declarative and interrogative BPPA constructions;[9] she also concludes that the relation between acceptability contrasts in declaratives and interrogatives should be taken seriously. These results fit her two-component model with independent licensing conditions for the adjunct and extraction operations. The complex acceptability pattern observed for interrogative BPPA constructions in the literature can be traced back to similar differences in declaratives, obviating the need for additional licensing mechanisms that are tied to extraction.

Similarly, Kehl (2021) reports that telic matrix predicates have an advantage over atelic ones (experiments 1 and 2) and that unaccusative matrix predicates are judged as more acceptable compared to unergatives and transitives (experiment 4); in none of the experiments, however, do these factors interact with extraction, so that the acceptability differences in interrogatives can be reliably predicted from identical contrasts in declaratives. These results obviate the requirement for additional syntactic or semantic licensing conditions for extraction as postulated in Borgonovo and Neeleman (2000) and Truswell (2007). For example, there are already significant differences between declarative conditions with telic and atelic matrix predicates, respectively, seen in (28). To be precise, the relative difference is exactly the same as in the interrogatives in (29), as the telicity of the matrix predicate does not interact with the presence or absence of extraction.

(28) a. John arrived whistling a funny song. [telic matrix predicate]

b. John worked whistling a funny song. [atelic matrix predicate]

(29) a. What did John arrive whistling? [telic matrix predicate]

b. What did John work whistling? [atelic matrix predicate]

Additionally, the same contrasts are obtained in relativizations such as (30), which are closer in form to the attested examples in Santorini (2019). A comparison of declarative, interrogative, and relativization BPPA constructions shows that the effect of telicity remains the same across these sentence types, but the overall acceptability is shifted: declarative BPPA constructions are generally more acceptable than relativizations, which, in turn, are more acceptable than interrogative BPPA constructions. This points towards the fact that different types of long-distance dependencies require different degrees of processing effort.

(30) a. This is the song that John arrived whistling. [telic matrix predicate]

b. This is the song that John worked whistling. [atelic matrix predicate]

Similar results are obtained for the distinction between unaccusative, unergative, and transitive matrix predicates; this points towards the fact that the proposals in Borgonovo and Neeleman (2000) and Truswell (2007) are not related to extraction from the adjunct. From an architectural perspective, it is easier to include a condition on the possibility of L-marking along the lines of Borgonovo and Neeleman (2000) instead of making core syntactic operations sensitive to semantic factors; whether an event-semantic approach to acceptability differences in declaratives fares better than one based on the grammatical verb type of the matrix predicate remains to be seen, but both are most likely related to how complex the resulting BPPA construction is for the parser to interpret and how plausible the complex event described there is; see also Chaves and Putnam (2020) for similar points. It is probably the case that Truswell (2007) is on the right track concerning the influence of event structure, even if this factor does not seem to depend on the presence or absence of extraction.

Several experiments in Kehl (2021) also show that there are considerable differences between declarative conditions, which are not directly predicted in Borgonovo and Neeleman (2000) or Truswell (2007), again pointing to the importance of considering the relative acceptability of the underlying declaratives instead of only their grammaticality. These

differences in declaratives can be captured in the comparison with the standardized reference fillers from Gerbrich et al. (2019): in most of the reported experiments, there is a contrast between the more acceptable declarative conditions, which are located between the A- and the B-level of the reference fillers, and the less acceptable declarative conditions with judgments clearly below the B-level and sometimes closer to the C-level. This shows that these differences are not too subtle to be irrelevant, or "unremarkable" as Truswell (2007, p. 1373) puts it.

The declarative counterparts of BPPA constructions are also compared to interrogatives in Kohrt et al. (2018), who do not find evidence for the theoretical claims about the factor agentivity in Truswell (2011), but crucially also no interaction of their factor ±extractable with extraction (their experiment 1). Against the predictions from Truswell (2007) and Truswell (2011), they do not find significant effects of verb type distinctions between extractable *arrive*-type verbs and non-extractable *work*-type verbs; see the example items in (31a). The only significant effect they find is between declaratives (31a) and interrogatives (31b), which is the predicted negative effect of extraction on acceptability.

(31)   a.   John wondered whether his best friend {worked/arrived} at the office drinking some coffee late this afternoon.

   b.   John wondered which coffee his best friend {worked/arrived} at the office drinking __ late this afternoon.

(Kohrt et al. 2018)

The lack of a significant effect of whether the matrix predicate is a suitable predicate for extraction may partially be caused by their assignment of event types to either extractable or non-extractable conditions: they include states in the extractable category and accomplishments in the non-extractable category, which is in line with the claims about agentivity in Truswell (2011), but is problematic from the observations about telicity in Truswell (2007) and the possibility for accomplishments to allow extraction when the adjunct specifies the causal component of the accomplishment, which is explicitly acknowledged in Truswell (2011).

The experimental evidence provided by Brown (2017) and Kehl (2021) supports the hypothesis that the factors identified in the literature do not influence the strength of extraction from the adjunct; there is no need to postulate additional licensing mechanisms to evade the CED. Both find that there are systematic acceptability differences in declaratives that are carried over to the interrogative structures without additional effects requiring an explanation.

5. A Model for the Acceptability of Participle Adjuncts

Once the focus of interest is shifted to a principled comparison between declarative base positions and *wh*-interrogatives, as well as the underlying acceptability differences in declaratives, the question is what causes these acceptability differences found in Brown (2017) and Kehl (2021). In this section, I will first discuss factors which influence the acceptability of (declarative) participle adjuncts; some, but not all of these factors have been discussed in the previous literature. At the end of this section, I will combine the factors into a partially weighted model for predicting the acceptability of declarative and interrogative participle adjunct constructions. This model will be conceptually based on graded and multifactorial models of acceptability such as the Decathlon Model (Featherston 2008, 2019) and the Cumulative Effect Hypothesis discussed in Haegeman et al. (2014) and Greco et al. (2017).[10] In these types of model, the violation of individual constraints show negative effects on acceptability; these constraint violations are cumulative, so that the violation of each additional constraint further decreases acceptability. I will argue that extraction from the adjunct is simply one additional negative effect that is added to the combined effects of the factors which influence acceptability in declarative BPPA constructions; crucially, the size of the extraction effect does not depend on whether other effects apply in the

declarative or not.[11] This is precisely the fundamental assumption made in Brown (2017) and Kehl (2021), which differentiates these accounts from previous approaches to extraction from adjuncts.

5.1. Transitivity: Multiple Referents Incur Independent Processing Costs

One of the factors that determines whether a BPPA construction is highly acceptable in declaratives is transitivity, i.e., whether the matrix predicate selects one or more arguments. Brown (2017) finds that transitivity is a relevant factor because it determines whether telicity has an effect at all, shown by an interaction of the two factors in her experiments. For transitive predicates, it is not important whether it is an atelic activity or a telic accomplishment, but intransitives are sensitive to the unergative–unaccusative distinction, with unaccusative achievements being more acceptable than unergative activities. This result is also found in Kehl (2021, experiment 4), where unaccusatives have a general advantage over unergatives and transitives, which are not differentiated between telic and atelic.

An additional observation made in Kehl (2021), based on the discussion in Borgonovo and Neeleman (2000), is that the nature of the second argument is important: reflexive objects as in (32a) and subjects of resultative constructions as in (32b) behave differently than prototypical transitive predicates with two distinct discourse referents, as in (32c).

(32) a. John hurt himself [trying to fix the roof].   [reflexive]

   b. John drove Mary crazy [talking about his sacking].   [resultative]

   c. John cut Bill [carving the turkey].   [transitive]
                                              (Borgonovo and Neeleman 2000, p. 211)

Here I will not go into a detailed discussion why resultative constructions differ from transitives; see Winkler (1997), Rothstein (2017), and Hu (2018) for discussion of how the subject of the resultative is assigned its θ-role. Incidentally, Borgonovo and Neeleman (2000, p. 212) observe that extraction from the adjunct in (32b) is ungrammatical, whereas Truswell (2007, 2011) considers this a prime example of transparent accomplishments; this emphasizes the need to investigate this type of matrix predicate in more detail.

In more general terms, a second argument increases complexity in the BPPA construction, also because potential control conflicts of the adjunct predicate need to be resolved: in a transitive sentence, the adjunct can be controlled by both the subject and the object of the matrix clause, which increases the amount of processing to resolve this ambiguity. Some event types show restrictions in their control possibilities (Rapoport 2019; Simpson 2005), but then the parsing of the wrong control orientation should lead to even lower acceptability.[12]

The observation that transitivity in general incurs drops in acceptability independently of extraction operations is also made in Jurka (2010, 2013), Polinsky et al. (2013), and Konietzko (2021); they all find that predicates which select a second argument are slightly less acceptable than intransitives (unergatives and unaccusatives) in declarative structures. Polinsky et al. (2013, p. 296) refer to this as a 'transitivity penalty', which is probably caused by the processing effort to parse the second argument. Similar effects of transitivity are also discussed in relation to extraction in Dependency Locality Theory (Gibson 1998, 2000), which also offers an explanation for the behavior of transitives; I follow Polinsky et al. (2013) in assuming that the effects of transitivity are not exclusive to sentences with extraction.

The negative effects of transitivity make the prediction that the more arguments are selected by the matrix predicate, the higher the processing effort required of the parser, with at least some effect on acceptability. Thus, I predict a relative decline in the acceptability of the sentences in (33), even if all structures might receive a grammatical binary judgment:

(33) a. John arrived singing an obscene song.   [intransitive]

   b. John offended Mary singing an obscene song.   [transitive]

   c. John gave Mary a letter singing an obscene song.   [ditransitive]

The full paradigm of transitivity thus ranges from purely intransitive to reflexive transitive, resultative, transitive, and, finally, ditransitive. It is also possible that not only the number of arguments, but also other factors play a role; this could be formulated in terms of the multi-faceted definition of the transitivity continuum in Hopper and Thompson (1980). An additional problem that arises in ditransitives is that there is a potential orientation ambiguity for the participle adjunct depending on its lexical content: the orientation can be shifted towards the direct object, as in (34), and is sometimes the preferred interpretation.

(34) $John_i$ gave $Mary_j$ a $letter_k$ [lying on the $table]_k$.

In the interrogatives corresponding to (33), the contrast between the intransitive and the (di-)transitive structures is noticeable, but the ditransitive is even worse than the transitive. This is not directly reflected in the binary judgments in (35), but should be visible in a judgment study. The low acceptability of the ditransitive structure (35c) carries over to the alternative ordering in the double object construction in (35d).

(35) a. What did John arrive singing __ ? [intransitive]

b. *What did John offend Mary singing __ ? [transitive]

c. *What did John give Mary a letter singing __ ? [ditransitive I]

d. *What did John give a letter to Mary whistling __ ? [ditransitive II]

Chaves and Putnam (2020, p. 15) point to the fact that optional transitivity may confound the intended interpretation of interrogative BPPA constructions because the *wh*-phrase may be linked to a gap in complement position of an optionally transitive matrix predicate instead of the complement position of the adjunct; see also Staub (2007) and Ness and Meltzer-Asscher (2019).[13] This ambiguity is shown in (36), where potential optional gap sites are indicated by underscores in parentheses.

(36) $What_{i/j}$ did John walk ($\_\_i$) whistling ($\_\_j$)?

a. John walked the dog whistling.

b. John walked whistling a funny song.

An ambiguous parse with gap position after the matrix predicate can be avoided by restricting adjunct predicates to obligatorily transitive predicates, such as *proclaiming*, as in (37). Here the gap site after the main verb would trigger ungrammaticality because the gap after the adjunct is obligatory, here indicated by the lack of parentheses around the gap site following the adjunct predicate. This means that the *wh*-pronoun cannot associate with the optional potential gap site in the matrix clause. A parasitic gap reading is also possible here if the filler can be the object of both predicates; I do not discuss this possibility further here.

(37) $What_{i/j}$ did John walk ($\_\_i$) proclaiming $\_\_j$?

a. *John walked the dog proclaiming.

b. John walked proclaiming his love for Pam.

Yet another way to reduce gap site ambiguity is if a motion verb like *walk* is augmented with a directional phrase, as in (38); it is still possible that John walks his dog to the park, but this parse becomes less likely than in (37).

(38) What did John walk to the park whistling __ ?

To sum up, transitivity, even if it is optional, increases the overall complexity of the BPPA construction and thus gradually builds up hurdles for extraction. Unambiguously intransitive predicates are predicted to have an advantage over potentially transitive and

unambiguously transitive predicates; reflexive and resultative predicates occupy the middle ground because on the one hand they include a second argument, but this argument is either not directly selected by the main verb (resultatives) or is co-referential with the main verb's subject (reflexives).

5.2. Event Structure: Durativity Instead of Telicity

Another factor which has an effect on the acceptability of declarative and interrogative BPPA constructions is based on the observation that not all types of matrix predicate can be felicitously modified by an adjunct predicate. The restrictions on BPPA constructions resemble those that operate in depictive secondary predication, where likewise not all types of main verb accept depictives to the same degree (Rapoport 2019; Simpson 2005). There is an ongoing discussion whether complex adjuncts, such as BPPAs, can be analyzed as depictives, but I will assume this for the present discussion; see also Rothstein (2017, p. 3874). For example, permanent statives, as in (39a), are odd with a BPPA, whereas temporary statives, as in (39b), are more acceptable.

(39) a. ? John was blond [wearing his new sunglasses]. [permanent state]

   b. John lay in bed [wearing his new sunglasses]. [temporary state]

The difference between these types of states is that temporary states have an event variable, which permanent states lack (Rapoport 1993, p. 173). Permanent states are property ascriptions whereas temporary states are predicated of the subject for a temporal interval that allows delimitation. This distinction also shows up in the corresponding interrogatives in (40):

(40) a. *What was John blond [wearing __]?

   b. What did John lie in bed [wearing __]?

Since both permanent and temporary states are atelic, these acceptability differences are problematic for the telicity-based account in Truswell (2007) and Brown (2017), as well as for the reflexivity account in Borgonovo and Neeleman (2000).

A telicity requirement is also problematic for purely punctual achievements like *appear*, which should be ideal candidates for a temporal inclusion relation in Brown (2017); still, these predicates are degraded in interrogatives, as seen in (41):

(41) a. John appeared wearing a beautiful bespoke suit.

   b. *What did John appear [wearing t]?

(Truswell 2007, p. 1374)

Similar observations can be made for verbs such as *notice* and other perception verbs. The question is whether this carries over to the declarative counterparts; as far as I am aware, this has not been directly tested in a controlled experiment. What permanent states and purely punctual achievements have in common is that both fail to felicitously appear in the progressive, as seen in (42a) and (42b). Crucially, temporary states are fine with the progressive, shown in (42c).

(42) a. ? John is being blond.

   b. ? John is appearing.

   c. John is lying in bed.

In terms of Rothstein (2004), punctual achievements and many perception verbs such as *notice* fail to appear in the progressive because the progressive cannot target an interval preceding the culmination point. The situation is different in cases similar to *arrive*, where the preceding interval can be conceptualized as the path component that leads up to the

arrival. With *appear*, the perspective is different: it is inherently external to the appearing entity, whereas *arrive* allows a conceptualization from the perspective of the arriving entity. This is a first indication that telicity alone makes the wrong predictions in these cases; rather, it seems that there is a certain correlation between the reported interrogative patterns and the ability to appear in the progressive.

Thus, the generalization about telicity in Truswell (2007) needs to be revised to exclude purely punctual achievements and to allow for temporary states. Instead of telicity, I argue that a first step towards a descriptive pattern is to consider the encoding of a durative subevent as relevant for acceptability, which is not the case for permanent states and punctual achievements.

5.3. Incrementality: Themes, Paths, and Properties

An exclusive focus on durativity leads to problems with the experimental results for activity main verbs in Brown (2017) and Kehl (2021): BPPA constructions with activity main verbs are less acceptable than achievements. To further constrain declarative BPPA constructions, a comparison with depictive secondary predicates shows that not all activity main verbs license a depictive, as shown in (43). The pattern is more difficult to capture than that of permanent and temporary states or punctual achievements, but if the BPPA construction can be analyzed as depictive secondary predication, similar effects can be expected there as well. It is also noteworthy that the addition of an object in (43c) ameliorates the modification of *draw* by a depictive.

(43) a.   Jones danced/lectured drunk.

    b.   *Jane laughed/drew drunk.

    c.   Jane drew pictures drunk.

(Rapoport 2019, pp. 434–35)

The distinction between *draw* and *draw pictures* in (43b) and (43c) also shows up in BPPA constructions, where the bare form in (44) is degraded in the interrogative; as noted above, the experimental evidence in Brown (2017) and Kehl (2021) suggest that the declarative counterparts are also less acceptable than sentences with achievement main verbs.

(44) a.   I work listening to music.

    b.   *What do you work [listening to t]?

(Truswell 2007, p. 1373)

The sentences improve in the presence of a direct object, seen in (45). This is contrary to the expectations derived from transitivity in the previous subsection, but suggests that some form of temporal delimitation may be a factor contributing to acceptability, without leading to a telicity requirement.

(45) a.   Mary worked on her thesis drinking coffee.

    b.   What did Mary work on her thesis drinking __ ?

All the acceptable depictive constructions in (43) involve an activity predicate that is in some sense delimited, but still atelic. A specific dance or a lecture have a specified duration, and the drawing of pictures can be measured by the amount of pictures produced, whereas laughing and drawing in the sense of aimlessly doodling are not delimited in the same sense. It could be argued that this type of delimitation is connected to the concept of incremental themes (Dowty 1979): a lecture, pictures, and working on a thesis can be measured against a scale of progress, similar to the incrementality of eating one, two, or three apples. The analogy to incremental themes also extends to the domain of motion, which also come in incremental and non-incremental forms. As noted in Dowty (1979),

Tenny (1995), and Borgonovo and Neeleman (2000), unergative manner of motion verbs like *walk* behave differently when they are followed by a directional PP like *to the station*; this PP introduces a path component that can be measured similar to incremental themes. The effect is shown in (46):

(46)　a. ? Mary walked whistling a funny song.

　　　b.　Mary walked to the station whistling a funny song.

Incrementality also extends to properties, which captures cases such as (47), where the degree of being scared increases with the progress through the movie (the gradual reading of this sentence probably comes from the durative character of the adjunct predicate, but this discussion is outside the scope of this paper).

(47)　John got scared watching a horror movie.

Similar effects of incrementality are seen with semelfactive main verbs such as *jump* in (48), where a particle inducing iterativity and thus durativity has a positive effect in interrogatives. A possible factor in addition to transitivity and durativity could thus be the potential of the event described by the main verb to be measurable or quantifiable in some sense.

(48)　What did she jump *(around) [singing t]?　　　　　　　　　(Truswell 2007, p. 1361)

Taken together, there is at least some evidence that purely temporal inclusion of the matrix interval within the interval of the adjunct predicate is not able to account for the full data pattern, which casts doubt on the scale amalgamation process suggested in Brown (2017). The overall picture emerging from this discussion is that it is unlikely that there is a single factor which determines whether a given main verb will be highly acceptable with a BPPA. This bears close similarity to the multiple factors which influence performance and acceptability along the lines of Chomsky (1965), suggesting that the acceptability of declarative BPPA constructions is a matter of syntactic and semantic complexity and compatibility criteria instead of strict syntactic licensing requirements.

5.4. Combining the Factors into an Acceptability Model

Based on the theoretical discussion of the relation between acceptability in declarative and interrogative BPPA constructions in Brown (2017) and the evidence supporting it, Kehl (2021) develops a model that captures this relation; this model includes factors that differ from those in Brown (2017) and other approaches. The main focus is on the fact that the factors which operate in interrogatives are also visible in declaratives. Extraction simply acts as an additional factor that is independent of the individual decreases in acceptability resulting from other factors, such as transitivity or durativity. The model can be summarized as follows:

(49)　Model for the acceptability of BPPA constructions:

　　i.　Determine the acceptability of the declarative sentence; factors: transitivity, durativity, incrementality

　　ii.　Determine the acceptability of the interrogative sentence by adding the processing costs of extraction to the result of (i)

In the first stage of the model (49i), the factors discussed above influence the acceptability of the BPPA construction: transitivity will decrease acceptability because more arguments require more processing effort. Durativity and incrementality work similarly: the absence of a durative subevent, i.e., for permanent states and purely punctual achievements, decreases acceptability, as does the absence of a delimited or incremental meaning component. Transitivity is most likely a result of increased processing effort, but durativity

and incrementality are semantic factors which seem more related to the conceptual felicity of the situation described in the sentence. Kehl (2021) collects durativity and incrementality under the term *semantic compatibility*.[14] In contrast to these factors, transitivity can be captured in syntactic terms, but the reason that transitivity matters is more likely to be found in relation to ease of processing and the ambiguity between transitive and intransitive uses of the verb in question.

The second stage of the model (49ii) adds the cognitive cost of establishing a dependency (Wagers 2013); this cost is most likely higher than into other domains, such as subcategorized complements, in line with the CED.[15] As this dependency formation is more demanding than a declarative sentence, this results in decreased acceptability. Crucially, the application of extraction and the resulting decreases in acceptability are independent of the factors which determine acceptability in the declarative: in a sense, extraction is blind to these factors. This is compatible with the independence of syntactic operations from purely semantic properties of the sentence (Brown 2017).

With respect to the relative weight of the factors that affect acceptability in declarative BPPA constructions, the previous experimental work on this construction in Brown (2017), Kohrt et al. (2018), and Kehl (2021) does not directly allow conclusions. The negative effect of transitivity is observed and isolated as a key factor in Brown (2017) and is in agreement with the transitivity penalty discussed in Polinsky et al. (2013). Scalar change and durativity are more complex to evaluate because the previous experimental work has focused on the telic–atelic distinction to check the predictions of Truswell (2007), but this distinction does not directly map to the factors discussed here. The complex interactions of these factors should be addressed in future experimental research. Based on the experimental results from Brown (2017) and Kehl (2021), it is possible to assign a preliminary weighting to this model: the effect of extraction is much stronger than that of durativity, incrementality, or transitivity. This observation connects to the discussion above about the subtle acceptability differences in declarative BPPA constructions, which run the risk of being considered irrelevant, especially if the focus of the approach in question is in grammaticality rather than acceptability. The acceptability model can be graphically represented as in Figure 3, taken from Kehl (2021).
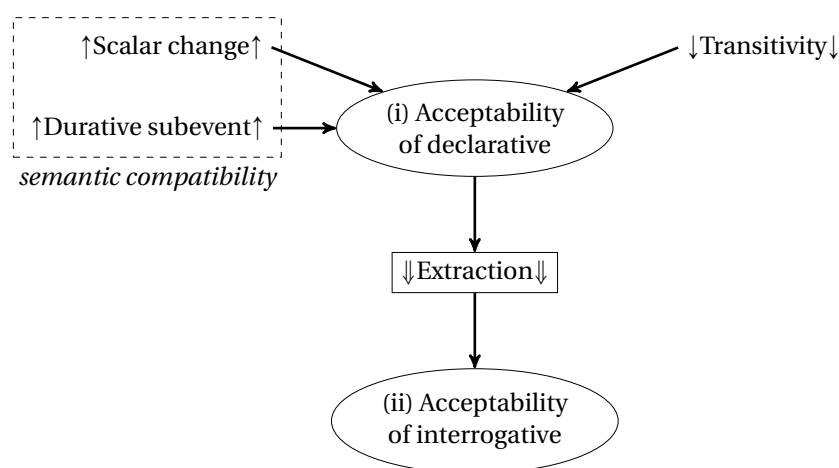

Figure 3. Acceptability model for BPPA constructions from Kehl (2021, p. 284); the factor *scalar change* corresponds to incrementality in this paper. Upwards arrows indicate a positive effect on acceptability, (double) downward arrows a negative effect.

This illustration shows the positive effects of durativity and incrementality with upward arrows, as well as the negative effect of transitivity with downward arrows; double downward arrows on the factor extraction indicate that this effect is stronger than the others. The central characteristic of this model is that it incorporates the relation between declarative and interrogative acceptability as formulated in Brown (2017), which is stated in Kehl (2021) as the independence of extraction from the factors operating in declaratives.

This model accounts for the sometimes subtle acceptability differences in declarative BPPA constructions, as well as the central factors isolated for participle adjunct islands in Borgonovo and Neeleman (2000) and Truswell (2007). At the same time, however, this model is conceptually simpler because the extraction operation remains blind to semantic characteristics of the sentence in question.

The model captures the following judgment differences discussed in the literature: (i) the advantage of telic over atelic matrix predicates due to scalarity (50i), (ii) the oddity of punctual matrix predicates because the latter do not satisfy durativity (50ii), (iii) the improvement with path scales and incremental themes for atelic matrix predicates because they introduce a scalar meaning component (50iii), and (iv) the effect of the number of arguments selected by the matrix predicate as a reflex of transitivity (50iv). If these contrasts can be shown to be observable in declaratives as well as interrogatives alike, this supports the predictions of the factorial acceptability model.

(50)   i.   What did John arrive/*work whistling __ ?   [scalarity]

      ii. *What did John appear/notice whistling __ ?   [durativity]

      iii. What did John work *(on his thesis) whistling __ ?   [scalarity]

      iv. What did John hurt himself/*Bill trying to fix __ ?   [transitivity]

Not all of these contrasts have been tested experimentally in the literature: the contrast in (50i) is the one that most of the existing literature focuses on, e.g., Brown (2017), Kohrt et al. (2018), and Kehl (2021). Likewise, transitivity effects as in (50iv) are to a certain extent explored in these studies, but further studies are required to see where reflexive and resultative matrix predicates lie in relation to intransitive and transitive sentences. The contrasts between purely punctual and extendable achievements in (50ii) as noted in Truswell (2007) and the precise effect of an added scalar meaning in cases like (50iii) also require additional work.

This acceptability model focuses on simple declarative and interrogative BPPA constructions, but it can also be modified to include other sentence forms, such as relativization or topicalization; these sentence forms also encode unbounded dependencies, but are not interrogative (Chaves and Putnam 2020). It can thus be expected that they do not show the same degree of decreased acceptability as the *wh*-interrogatives focused on in this article, which is also indicated in the data reported in Abeillé et al. (2020) and Liu et al. (2022). Compare the declarative BPPA construction in (51) with the different types of dependencies in (51a)–(51c).

(51)   John arrived [whistling an annoying song].   [declarative]

      a.   Which song did John arrived [whistling $\_i$]?   [*wh*-interrogative]

      b.   This annoying song$_i$, John arrived [whistling $\_i$].   [topicalization]

      c.   I hated the song$_i$ (that/which) John arrived [whistling $\_i$].   [relativization]

Initial evidence that relativization leads to a generally smaller decrease in acceptability than bare *wh*-interrogatives is given in (Kehl (2021) [experiment 1]). This might be related to a better match between the information-structural status of the adjunct constituent from which extraction takes place and the discourse function of relativization, as proposed in Abeillé et al. (2020). The visualization of the acceptability model in Figure 3 can be generalized by adding more extraction types than just *wh*-extraction, and by linking these different types of dependency formation to separate acceptability levels; this is shown in Figure 4, where relativization and topicalization are allowed for negative effects on acceptability that are not necessarily identical to that of *wh*-extraction. I will have to leave the relative magnitude of these effects for future experimental research. The underlying hypothesis remains that the contrast between matrix verbs such as *arrive* and *work* can be observed equally across these different dependency types; this assumption follows the

argumentation in Chaves and Putnam (2020) that the pragmatic felicity of the underlying proposition has a strong role to play in island effects and extraction asymmetries.
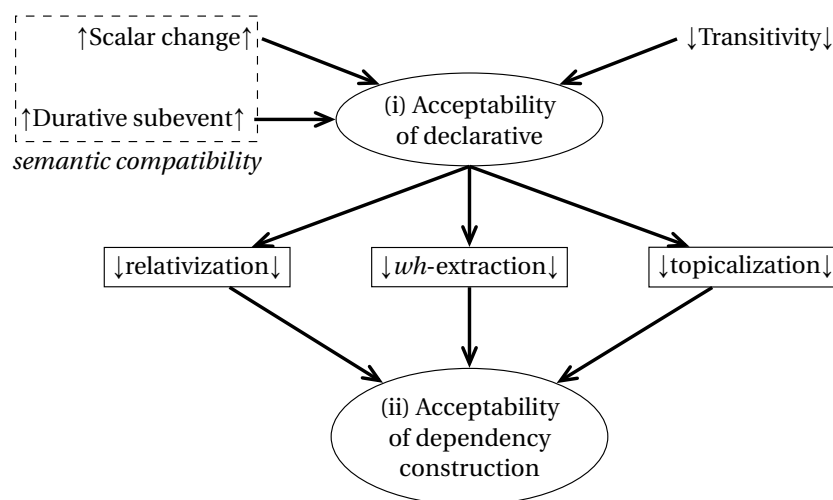

Figure 4. Adapted acceptability model for BPPA constructions in different sentence forms; the acceptability of a dependency construction is modeled as the acceptability of the underlying simple declarative plus the effect of establishing a dependency construction. Upwards arrows indicate a positive effect on acceptability, downward arrows a negative effect.

Another important issue is how strong the factors of the acceptability model are affected by variation in speaker judgments. So far, I am not aware of experimental studies that explicitly take this factor into account. There are studies on the related phenomenon subject islands investigating whether judgments improve depending on presentation order: Chaves and Dery (2019) report that judgments improved if the item was presented later in the experiment, suggesting that there is a satiation effect and that the initially low acceptability judgment improves with repeated exposure as a type of learning effect. If violations of the subject condition can improve over time, it seems plausible that the type of semantic mismatches resulting from scalarity and durativity can also improve with repeated exposure, but this requires further investigation.

In conclusion to the factors related to acceptability in the BPPA construction and the model proposed in Kehl (2021), it seems that Truswell (2007) is not right in his claim that declarative BPPA constructions which do not meet his extraction condition are unremarkable. The exact opposite holds: acceptability differences in declaratives resulting from a variety of different factors are the key determinants of acceptability in interrogative BPPA constructions, and it is not the extraction operation that triggers these differences in interrogatives.

6. Converging Evidence for the Relevance of Acceptability Differences in Declaratives

More recent work agrees about the relevance of potential acceptability differences in declaratives for the acceptability of movement constructions. The proposals diverge slightly in the source of such differences, but the focus has shifted from purely syntactic explanations towards more interface-based ones. Transitivity as a processing-related complexity criterion and event structure as a semantic notion have been the focus of this article.

Similar conclusions about extraction-independent effects of processing-related complexity on acceptability are drawn for the apparent licensing of island-violating extractions in so-called parasitic gap (PG) environments in Culicover and Winkler (2022); they trace back the ameliorating effect ascribed to parasitic gaps to complexity differences between (declarative) PG and non-PG constructions. The former are more acceptable because they are less complex for processing due to the fact that one less referentially distinct argument is encoded. In the contrast in (52), the additional gap in the matrix clause in (52b) leads to the fact that there is only one discourse referent in the sentence, whereas there are two

in (52a). There is, thus, an underlying difference in complexity that pushes (52a) below a threshold for grammaticality, which is not the case in (52b). The parasitic gap is indicated by *pg* in this example.

(52) a. *a person who$_i$ [talking to t_i] about this would prove to the Mayor that there is a problem

    b. a person who$_i$ [talking to pg_i] about this would prove to t_i that there is a problem
<div align="right">(Culicover and Winkler 2022, p. 2)</div>

Although the two corresponding non-extraction sentences in (53) are certainly both grammatical, (53a) is more complex than (53b) from a processing perspective because an additional discourse referent needs to be processed. Whether this results in noticeable acceptability differences is a question that is outside the scope of this paper, but can explain the strong judgment difference reported for (52) by Culicover and Winkler.

(53) a. Talking to *person X* about this would prove to *person Y* that there is a problem.

    b. Talking to *person X* about this would prove to *person X* that there is a problem.

The conclusions in Culicover and Winkler (2022) are very similar to that discussed in this article: there is no requirement for a dedicated licensing or repair mechanism associated with parasitic gaps; a sufficient description of the underlying complexity differences is sufficient to explain why PG constructions are more acceptable than the non-PG construction. Culicover and Winkler (2022) also discuss the important distinction between grammaticality and acceptability that can be used to provide a comprehensive explanation of the patterns detected for parasitic gaps in the literature.

Another set of factors comes from the interface of syntax with pragmatics: Chaves and Putnam (2020) point out that apparent grammaticality contrasts in syntactically marked constructions, such as *wh*-questions, often have their origin in sometimes subtle pragmatic differences that are unrelated to the formation of the marked construction. They propose a largely pragmatic approach to most island domains by arguing that the low acceptability can often be traced back to issues of relevance and salience: if the island domain is not salient or relevant, acceptability contrasts in unmarked constructions can arise and evoke the impression of stronger grammaticality contrasts in marked constructions. This is captured in the Relevance Presupposition Condition (RPC):

(54) RELEVANCE PRESUPPOSITION CONDITION: the referent that is singled out for extraction in a UDC must be highly relevant (e.g., part of the evoked conventionalized world knowledge) relative to the main action that the sentence describes. Otherwise, extraction makes no sense from a Gricean perspective, as there is no reason for the speaker to draw attention to a referent that is irrelevant for the main contribution of the sentence to the discourse.   (Chaves and Putnam 2020, p. 206)

The contrast in (55) is given as an example of this, but the grammaticality difference is unrelated to extraction:

(55) a. What did you read a book about?

    b. *What did you drop a book about?
<div align="right">(Chaves and Putnam 2020, p. 207)</div>

It has been noted as early as Kuno (1987) that the corresponding declaratives already show a noticeable acceptability difference; this is shown in (56):

(56) a. Speaking of Napoleon, I just read a book about him.

   b. ?Speaking of Napoleon, I just dropped a book about him.

<div align="right">(Chaves and Putnam 2020, p. 205)</div>

The reasoning to explain these independent acceptability differences is along the following lines: verbs evoke certain conceptualizations when they are encountered by the parser, and some meaning components are more easily accessible than others. Reading a book evokes the concept of a topic covered by the book, which is relevant information. However, the topic is not as relevant and easily evoked when a book is dropped (Chaves and Putnam 2020, p. 207). This has clear ramifications for acceptability in marked constructions, but may not be as clear in unmarked ones.

For BPPA constructions, the predictions of the RPC predict that adjuncts which supply relevant information invoked by the event described in the matrix predicate can be targeted by extraction. This serves as an explanation for the relative acceptability of cases where the adjunct describes the cause of the matrix predicate, as in (57a). The distinction between non-causal adjuncts discussed in Truswell (2007, 2011), as in (57b) and (57c) is less clear, but it could be argued that telic predicates like *arrive* are informationally light, so that the adjunct can be analyzed as relevant in the sense of Chaves and Putnam (2020); atelic predicates such as *work*, on the other hand, can be argued compete with the adjunct in terms of which information is more relevant, so that the extraction is not licensed by the RPC.

(57) a. What did Peter drive Mary crazy whistling __ ?

   b. What did Peter arrive whistling __ ?

   c. *What did Peter work whistling __ ?

The acceptability model discussed in the previous section is not mutually exclusive with the RPC; the generalizations in the model could be seen as factors that influence the relevance of the adjunct compared to the matrix predicate and hence have an effect on the acceptability of extraction. I agree with Chaves and Putnam (2020, p. 230) that "extraction from such island environments is contingent on the proposition itself, rather than strictly on its syntax". This captures the idea in the model that the factors described by the generalizations show effects that are independent of extraction.

There exists a number of experimental studies that test the relation between declaratives and interrogatives in related phenomena:[16] for example, Chaves and King (2019) find a strong correlation between plausibility ratings for declaratives and acceptability of subextraction from objects, indicating that plausibility ratings act as a predictor of acceptability that is not modulated by extraction. However, Chaves and Putnam (2020) report on another experiment investigating extraction from tensed adverbial clauses, where they do not find a correlation between declarative and interrogative acceptability, meaning that the latter is not reliably predicted by the former. In such cases, it is reasonable to assume that there is another factor which distorts the relation, similar to the factorial definition of island effects in Sprouse and Hornstein (2013). The effects of tensed adjuncts are also discussed from a theoretical perspective in Truswell (2011, pp. 175–79) and experimentally investigated in a cross-linguistic study in Müller (2019). Abeillé et al. (2020) examine relativization from subjects and objects, with the result that extraction from subjects is actually better than extraction from objects, contrary to the predictions of locality constraints such as the CED, which do not discriminate between different types of extractions; this points towards the conclusion that not all extractions function alike, and that the discourse functions of the extraction operation and the extracted element should be included in an analysis.

These brief glances beyond the scope of this paper show that theory development is well advised to take subtle acceptability differences in declaratives seriously in the discussion of licensing mechanisms for movement. Differences in processing complexity, semantic compatibility, and pragmatic characteristics can affect canonical word orders to

such a degree that the application of movement operations invokes the impression of strong grammaticality differences.

7. Conclusions

In this article, I have emphasized the importance of the underlying declarative sentences in the discussion of extraction from participial adjunct islands. Once the distinction between grammaticality and acceptability is taken seriously, it becomes possible to explain the acceptability differences in interrogatives by examining potential acceptability differences in the declarative counterparts. The result is an approach to extraction from adjunct islands that does not require additional and complicated licensing machinery as in the theories presented in Borgonovo and Neeleman (2000) or Truswell (2007, 2011). The approaches in Brown (2017) and Kehl (2021) both emphasize the relevance of acceptability differences in declarative BPPA constructions and propose factors to capture the acceptability variation independently of extraction. I have discussed three factors that are of interest in these accounts: the notion of transitivity, expressed in the number of arguments directly selected by the main verb, the event structure of the main verb, as well as the encoding of an incremental measure scale in the matrix predicate. The effect of transitivity can be described as a processing advantage of verbs with lower transitivity: more arguments to be processed incurs processing costs that can be reflected in acceptability. As far as event structure is concerned, I have argued that a simple telicity requirement, as postulated in Truswell (2007) and Brown (2017) is insufficient to explain the low judgments observed in the literature (e.g. Truswell 2007, p. 1370) for extraction from BPPA constructions with purely punctual matrix predicates, such as *appear* and the relatively acceptable judgments with temporary stative predicates, such as *lie in bed* (Truswell 2011, pp. 158–59). One of the key components isolated in the discussion is durativity instead of telicity, even if further factors need to be taken into account in order to explain the low acceptability with activity matrix predicates. The last factor is that of incrementality, where the progression of the matrix predicate can be measured against an incremental scale, formulated either as paths, incremental themes, or property values. Together, these factors provide a first set of tools to capture the acceptability differences in declarative and interrogative BPPA constructions without the need for additional, complex licensing mechanisms.

A final, more programmatic note about the nature of so-called 'island constraints' such as the CED: there is recent evidence that not all extraction types show the same effects in CED-violating operations, and that the magnitude of the extraction effect also depends on other factors of the island domain. For example, Abeillé et al. (2020) have shown that relativization has a different effect than *wh*-extraction in subject islands, which is hard to explain in pure syntactic terms such as the CED; similar observations are reported in Kehl (2021) for *wh*-extraction and relativization from BPPA constructions, who finds that relativization from BPPAs is more acceptable than *wh*-extraction, and that the aspectual classes of the matrix and adjunct predicates have identical effects, as in declaratives and interrogatives. Additionally, experimental work in Müller (2019) suggests that some adverbial clauses are harder to extract from than others, involving factors such as adverbial clause type and tense-marking. It would appear that the notion of categorical extraction constraints, such as the CED, should be critically evaluated: are such constraints really binary in core syntactic terms, meaning that the grammar can compute the extraction only in one but not in another configuration? Or is this the same type of overgeneralization that has been shown here to be problematic for accounts like Borgonovo and Neeleman (2000) and Truswell (2007)? This is a general problem faced by binary or categorical models of grammar because they are at risk of glossing over subtle acceptability differences in favor of broad general predictions; a graded model of grammar such as the Decathlon Model (Featherston 2008, 2019) has the flexibility of assigning individual decreases in acceptability to different operations from minimally different constructions, so that these effects can be individually quantified and summed up to predict acceptability in a wider range of configurations than the categorical predictions of the CED. The upshot from this brief

discussion is that there are good reasons to assume that extraction from some structural domains is harder than extraction from others, as captured in the original formulation of the CED; whether this is due to derivational or structural factors (competence-based) or the result of increased processing complexity (performance-based) is beyond the scope of this article. I leave the details of such an analysis of island constraints to future research and conclude here that BPPA constructions are an interesting showcase of island-internal variation that can be fruitfully employed to dive deeper into the nature of acceptability and its relation to intuitively observed grammaticality patterns in island constructions.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BPPA	Bare Present Participle Adjunct
CED	Condition on Extraction Domain
DP	Determiner Phrase
PG	Parasitic Gap
PP	Preposition Phrase
RPC	Relevance Presupposition Condition

Notes

[1] See Stepanov 2007 for an overview of developments in the Minimalist tradition, many of which incorporate the concepts of phases (Chomsky 2001, 2008) and cyclicity (Abels 2012). There are by now several approaches which syntactically derive adjunct-internal gaps without violating syntactic principles, such as Narita (2014), den Dikken (2018), and Brown (2017).

[2] I do not address in detail the questions whether the effects of the CED in its original formulation or later adaptations are the result of a grammatical constraint (competence-based in terms of Chomsky 1965) or of a conspiracy of other factors, such as complexity or plausibility (which are rooted in performance). The influence of performance-based factors on extraction operations reflects a growing body of research in this direction; see, among others, Sag et al. (2008), Hofmeister and Sag (2010), and Chaves and Putnam (2020).

[3] Both accounts are based on intuitive author judgments without direct empirical validation; their predictions have since been experimentally tested with mixed results in Brown (2017), Kohrt et al. (2018), and Kehl (2021).

[4] Experimental research on BPPA constructions in Brown (2017) and Kehl (2021) does not confirm this intuition because there are significant acceptability differences between different declarative conditions, which points to the imperfect alignment between grammaticality and acceptability, which I discuss in Section 3. I discuss previous experimental results on extraction from adjuncts in more detail in Section 4. The relevance of considering even small acceptability differences in declaratives and relate them to differences in non-canonical structures is now more prevalent in the literature.

[5] A reviewer wonders whether working memory capacity impacts acceptability in these cases. Experimental studies on the relation of individual working memory and acceptability judgments in island phenomena, such as Sprouse et al. (2012, 2013), have so far not shown a significant correlation between these two measures; as noted by the authors, this is partially due to the method in which memory capacity is measured. Sprouse et al. (2012, p. 116) report on no interaction between the dependency length effect and participant groups with high and low working memory scores when using recall scores as a measure of memory capacity, but a significant interaction if three-back scores are used.

6   In other words, I assume that extraction from adjuncts, such as (16), always shows a superadditive island effect in terms of Sprouse and Hornstein (2013, p. 2); since I focus on a single adjunct construction, the question is whether semantic differences further influence how strong the island effect of extraction is.

7   If these control conditions are part of the experimental design, they can be directly included in the statistics; Abeillé et al. (2020) implement this with sub-models that isolate the effects and interactions of the different factors.

8   This distinction is explained in purely tree-geometric terms in Brown (2017), but it could also be argued that high *v*P adjuncts are inactive from a phase perspective, perhaps because they are introduced by late merger (Stepanov 2007) and are thus entirely opaque for syntactic operations that apply earlier in the derivational cycle.

9   An effect of transitivity is not surprising from the perspective of Dependency Locality Theory (Gibson 1998, 2000) because the dependency crosses over another discourse referent only in the transitive cases. The effect of transitivity is significant independently of extraction (Brown 2017, pp. 124–25), which is compatible with the findings in Jurka (2010, 2013) and Polinsky et al. (2013).

10   A crucial difference between constraint satisfaction models, such as the Cumulative Effect Hypothesis and the Decathlon Model, is that these models work with cross-constructionally active constraints that are part of a speaker's competence; in contrast, the generalizations discussed in the model discussed here are conceived of as factors that contribute to semantic compatibility and syntactic complexity, both with ensuing effects on acceptability. These generalizations need not be part of a speaker's competence grammar, but can be linked to how easily a given sentence is processed, considering that two events plus the temporal and semantic relations between them need to be processed.

11   To see whether there is a difference between extraction from the adjunct and extraction in general, a reviewer suggests that extraction from BPPAs should be compared with extraction from gerundive complements, which are very similar in their surface structure. For example, is the effect of extraction in *What type of cigars did John stop smoking __ last week?* of the same magnitude as in *What type of cigars did John arrive smoking __ last week?* and does the presence or absence of an adjunct have an effect in the corresponding declaratives?

12   A reviewer suggests that there could be a preference to have a simpler situation in the matrix predicate when the adjunct is complex; I agree that this could be a more general explanation for the effects of transitivity with complex adjuncts. The same reviewer also points out that sentences "tend to be subject–verb–complement or subject–verb–adjunct more than subject–verb–complement–adjunct". A discussion of expectation and usage frequency goes beyond the scope of the present article, but provides a fruitful area for future research.

13   The ambiguity between transitive and intransitive alternates is also related to the relative frequency of the two forms: some ambiguous verbs occur primarily in their intransitive uses, others in combination with a prepositional complement, as transitives, and other constructions. This may influence whether a verb is preferably parsed as intransitive or transitive. Roland et al. (2007) analyze the frequencies of such occurrences in large corpus data; for example, the verb *walk* used in (36) has a frequency ordering of PP > intransitive > transitive, whereas the verb *leave* has the frequency ordering transitive > intransitive > PP (both verbs also occur in other configurations, this is just an expository selection). I thank an anonymous reviewer for pointing me towards this corpus data. For the role of frequency data in relation to acceptability from island domains, see for example Chaves and Richter (2020).

14   Whether these semantic factors are actually encoded in the grammar and, thus, part of a speaker's competence is a question (raised by one of the reviewers) that goes beyond the scope of this article; it seems possible that there is a considerable degree of inter-speaker variation in the judgments of these factors, which should be explicitly tested in additional studies.

15   This means that a locality condition such as the CED still has a place in syntactic theory, but the question whether it is a categorical constraint or rather a gradual phenomenon should be investigated more closely. If it can be shown that dependencies into non-complement domains are computationally possible, i.e., locally well-formed (as exemplified for low VP-adjuncts in Brown 2017), the CED could be reduced to a processing phenomenon instead of a grammatical principle. Research in this direction is still ongoing, as in Culicover and Winkler (2018) and Culicover et al. (2022), but an interesting venue for future research.

16   I am grateful to one of the reviewers for pointing out these relevant studies on closely related issues.

References

Abeillé, Anne, Barbara Hemforth, Elodie Winckel, and Edward Gibson. 2020. Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition* 204: 104293. [CrossRef] [PubMed]

Abels, Klaus. 2012. *Phases: An Essay on Cyclicity in Syntax*. Berlin and Boston: de Gruyter. [CrossRef]

Bader, Markus, and Jana Häussler. 2010. Toward a model of grammaticality judgments. *Journal of Linguistics* 46: 273–330. [CrossRef]

Bader, Markus, and Jana Häussler. 2019. How to get from graded intuitions to binary decisions. In *Experiments in Focus: Information Structure and Semantic Processing*. Edited by Sam Featherston, Robin Hörnig, Sophie von Wietersheim and Susanne Winkler. Boston and Berlin: de Gruyter, pp. 183–207. [CrossRef]

Bard, Ellen Gurman, Dan Robertson, and Antonella Sorace. 1996. Magnitude estimation of linguistic acceptability. *Language* 72: 32–68. [CrossRef]

Borer, Hagit. 2005. *Structuring Sense, Vol. 2: The Normal Course of Events*. Oxford: Oxford University Press. [CrossRef]

Borgonovo, Claudia, and Ad Neeleman. 2000. Transparent adjuncts. *Canadian Journal of Linguistics/La Revue Canadienne de Linguistique* 45: 199–224. [CrossRef]

Brown, Jessica M. M. 2015. Apparently semantically-motivated extraction in an autonomous syntax. *Cambridge Occasional Papers in Linguistics* 8: 6–23.

Brown, Jessica M. M. 2016. Blackholes and subextraction from adjuncts in English and Norwegian. In *Proceedings of the Chicago Linguistic Society (CLS)*. Chicago Linguistic Society. Chicago: Chicago Linguistic Society, vol. 51, pp. 67–81.

Brown, Jessica M. M. 2017. Heads and Adjuncts: An Experimental Study of Subextraction from Participials and Coordination in English, German and Norwegian. Ph. D. thesis, University of Cambridge, Cambridge, UK.

Brown, Jessica M. M., Gisbert Fanselow, Rebecca Hall, and Reinhold Kliegl. 2021. Middle ratings rise regardless of grammatical construction: Testing syntactic variability in a repeated exposure paradigm. *PLoS ONE* 16: 1–26. [CrossRef]

Chaves, Rui P., and Adriana King. 2019. A usage-based account of subextraction effects. *Cognitive Linguistics* 30: 719–50. [CrossRef]

Chaves, Rui P., and Jeruen E. Dery. 2019. Frequency effects in subject islands. *Journal of Linguistics* 55: 475–521. [CrossRef]

Chaves, Rui P., and Michael T. Putnam. 2020. *Unbounded Dependency Constructions: Theoretical and Experimental Perspectives*. Oxford: Oxford University Press. [CrossRef]

Chaves, Rui P., and Stephanie Richter. 2020. Investigating the role of verb frequency in factive and manner-of-speaking islands. In *Proceedings of the Cognitive Science Society (COGSCI)*. Austin: Cognitive Science Society, vol . 42, pp. 1771–77.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A Life in Language*. Edited by Michael Kenstowicz. Cambridge: MIT Press, pp. 1–52.

Chomsky, Noam. 2008. On phases. In *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*. Edited by Robert Freidin, Carlos P. Otero and Maria Luisa Zubizarreta. Cambridge and London: MIT Press, pp. 133–66.

Culicover, Peter W., Giuseppe Varaschin, and Susanne Winkler. 2022. The radical unacceptability hypothesis. *Languages* 7: 96. [CrossRef]

Culicover, Peter W., and Susanne Winkler. 2018. Freezing: Between grammar and processing. In *Freezing: Theoretical Approaches and eMpirical Domains*. Edited by Jutta M. Hartmann, Marion Jäger, Andreas Kehl, Andreas Konietzko and Susanne Winkler. Berlin and New York: Mouton de Gruyter, pp. 353–86. [CrossRef]

Culicover, Peter W., and Susanne Winkler. 2022. Parasitic gaps aren't parasitic, or, the case of the uninvited guest. *The Linguistic Review* 39: 1–35. [CrossRef]

den Dikken, Marcel. 2018. *Dependency and Directionality*. Cambridge: Cambridge University Press. [CrossRef]

Dowty, David R. 1979. *Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and in Montague's PTQ*. Dordrecht: Reidel. [CrossRef]

Featherston, Sam. 2008. The decathlon model of empirical syntax. In *Linguistic Evidence: Empirical, Theoretical and Computational Perspectives*. Edited by Stephan Kepser and Marga Reis. Berlin: Mouton de Gruyter, pp. 187–208. [CrossRef]

Featherston, Sam. 2009. A scale for measuring well-formedness: Why syntax needs boiling and freezing points. In *The Fruits of Empirical Linguistics, Vol. 1: Process*. Edited by Sam Featherston and Susanne Winkler. Berlin: de Gruyter, pp. 47–74. [CrossRef]

Featherston, Sam. 2019. The decathlon model. In *Current Approaches to Syntax: A Comparative Handbook*. Edited by András Kertész, Edith Moravcsik and Csilla Rákosi. Berlin and Boston: Mouton de Gruyter, pp. 155–86. [CrossRef]

Featherston, Sam. 2020. Can we build a grammar on the basis of judgments? In *Linguistic Intuitions: Evidence and Method*. Edited by Samuel Schindler, Anna Drożdżowicz and Karen Brøcker. Oxford: Oxford University Press, pp. 165–88. [CrossRef]

Gerbrich, Hannah, Vivian Schreier, and Sam Featherston. 2019. Standard items for English judgment studies: Syntax and semantics. In *Experiments in Focus: Information Structure and Semantic Processing*. Edited by Sam Featherston, Robin Hörnig, Sophie von Wietersheim and Susanne Winkler . Boston & Berlin: de Gruyter, pp. 305–28. [CrossRef]

Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68: 1–76. [CrossRef]

Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. In *Image, Language, Brain*. Edited by Alec Marantz, Yasushi Miyashita and Wayne O'Neil. Cambridge: MIT Press, pp. 95–126. [CrossRef]

Gibson, Edward, Steve Piantadosi, and Kristina Fedorenko. 2011. Using Mechanical Turk to obtain and analyze English acceptability judgments. *Language and Linguistics Compass* 5: 509–24. [CrossRef]

Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Goldberg, Adele E. 2013. Backgrounded constituents cannot be "extracted". In *Experimental Syntax and Island Effects*. Edited by Jon Sprouse and Norbert Hornstein. Cambridge: Cambridge University Press, pp. 221–38. [CrossRef]

Graf, Thomas. 2015. The syntactic algebra of adjuncts. In *Proceedings of the Chicago Linguistic Society (CLS)*. Chicago: Chicago Linguistic Society, vol . 49, pp. 101–13.

Greco, Ciro, Marco Marelli, and Liliane Haegeman. 2017. External syntax and the Cumulative Effect in subject sub-extraction: An experimental evaluation. *The Linguistic Review* 34: 479–531. [CrossRef]

Haegeman, Liliane, Ángel L. Jiménez-Fernández, and Andrew Radford. 2014. Deconstructing the subject condition in terms of cumulative constraint violation. *The Linguistic Review* 31: 73–150. [CrossRef]

Hofmeister, Philip, and Ivan A. Sag. 2010. Cognitive constraints and island effects. *Language* 86: 366–415. [CrossRef]

Hopper, Paul J., and Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56: 251–99. [CrossRef]

Hu, Xuhui. 2018. *Encoding Events: Functional Structure and Variation*. Oxford: Oxford University Press.

Huang, Cheng-Teh James. 1982. Logical relations in Chinese and the Theory of Grammar. Ph. D. thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.

Jurka, Johannes. 2010. The Importance of Being a Complement: CED-Effects Revisited. Ph. D. thesis, University of Maryland, College Park, MD, USA.

Jurka, Johannes. 2013. Subject islands in German revisited. In *Experimental Syntax and Island Effects*. Edited by JJon Sprouse and Norbert Hornstein. Cambridge: Cambridge University Press, pp. 265–85. [CrossRef]

Kehl, Andreas. 2021. Adjunct Islands in English: Theoretical Perspectives and Experimental Evidence. Ph. D. thesis, University of Tübingen, Tübingen, Germany.

Kohrt, Annika, Trey Sorensen, and Dustin A. Chacón. 2018. The real-time status of semantic exceptions to the adjunct island constraint. In *Proceedings of the Chicago Linguistic Society (CLS)*. Chicago: Chicago Linguistic Society, vol . 54, pp. 225–36.

Konietzko, Andreas. 2021. PP Extraction from Subject Islands in German. *Glossa*, accepted for publication.

Kuno, Susumu. 1987. *Functional Syntax: Anaphora, Discourse and Empathy*. Chicago and London: The University of Chicago Press.

Kush, Dave, Terje Lohndal, and Jon Sprouse. 2018. Investigating variation in island effects: A case study of Norwegian wh-extraction. *Natural Language & Linguistic Theory* 36: 743–79. [CrossRef]

Kush, Dave, Terje Lohndal, and Jon Sprouse. 2019. On the island sensitivity of topicalization in Norwegian: An experimental investigation. *Language* 95: 393–420. [CrossRef]

Liu, Yingtong, Elodie Winckel, Anne Abeillé, Barbara Hemforth, and Edward Gibson. 2022. Structural, functional, and processing perspectives on linguistic island effects. *Annual Review of Linguistics* 8: 495–525. [CrossRef]

Müller, Christiane. 2019. *Permeable Islands: A Contrastive Study of Swedish and English Adjunct Clause Extractions*. Lund: Centre for Languages and Literature, Lund University.

Narita, Hiroki. 2014. *Endocentric Structuring of Projection-Free Syntax*. Amsterdam and Philadelphia: John Benjamins.

Ness, Tal, and Aya Meltzer-Asscher. 2019. When is the verb a potential gap site? The influence of filler maintenance on the active search for a gap. *Language, Cognition and Neuroscience* 34: 936–48. [CrossRef]

Phillips, Colin. 2013. On the nature of island constraints I: Language processing and reductionist accounts. In *Experimental Syntax and Island Effects*. Edited by Jon Sprouse and Norbert Hornstein. Cambridge: Cambridge University Press, pp. 64–108. [CrossRef]

Polinsky, Maria, Carlos G. Gallo, Peter Graff, Ekaterina Kravtchenko, Adam Milton Morgan, and Anne Sturgeon. 2013. Subject islands are different. In *Experimental Syntax and Island Effects*. Edited by Jon Sprouse and Norbert Hornstein. Cambridge: Cambridge University Press, pp. 286–309. [CrossRef]

Ramchand, Gillian Catriona. 2008. *Verb Meaning and the Lexicon: A First Phase Syntax*. Cambridge: Cambridge University Press. [CrossRef]

Rapoport, Tova R. 1993. Stage and adjunct predicates: Licensing and structure in secondary predication constructions. In *Knowledge and Language, Vol. 2: Lexical and Conceptual Structure*. Edited by Eric Reuland and Werner Abraham. Dordrecht, Boston and London: Kluwer, pp. 157–82. [CrossRef]

Rapoport, Tova R. 2019. Secondary predication. In *The Oxford Handbook of Event Structure*. Edited by Robert Truswell. Oxford: Oxford University Press, pp. 426–55.

Roland, Douglas, Frederic Dick, and Jeffrey L. Elman. 2007. Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language* 57: 348–79. [CrossRef] [PubMed]

Rothstein, Susan. 2004. *Structuring Events: A Study in the Semantics of Lexical Aspect*. Malden: Blackwell. [CrossRef]

Rothstein, Susan. 2017. Secondary predication. In *The Wiley Blackwell Companion to Syntax*, 2nd ed. Edited by Martin Everaert and Henk van Riemsdijk. Hoboken: Wiley, vol. 6, pp. 3872–901. [CrossRef]

Sag, Ivan A., Philip Hofmeister, and Neal Snider. 2008. Processing complexity in subjacency violations: The complex noun phrase constraint. In *Proceedings of the Chicago Linguistic Society (CLS)*. Chicago: Chicago Linguistic Society, vol . 43, pp. 215–29.

Santorini, Beatrice. 2019. (Un?)expected Movement. Available online: https://www.ling.upenn.edu/~beatrice/examples/movement.html (accessed on 27 March 2022).

Simpson, Jane. 2005. Depictives in English and Walpiri. In *Secondary Predication and Adverbial Modification: The Typology of Depictives*. Edited by Nikolaus P. Himmelmann and Eva Schultze-Berndt. Oxford: Oxford University Press, pp. 69–106. [CrossRef]

Sprouse, Jon. 2011. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods* 43: 155–67. [CrossRef]

Sprouse, Jon, Ivano Caponigro, Ciro Greco, and Carlo Cecchetto. 2016. Experimental syntax and the variation of island effects in English and Italian. *Natural Language & Linguistic Theory* 34: 307–44. [CrossRef]

Sprouse, Jon, and Norbert Hornstein. 2013. Experimental syntax and island effects: Toward a comprehensive theory of islands. In *Experimental Syntax and Island Effects*. Edited by Jon Sprouse and Norbert Hornstein. Cambridge: Cambridge University Press, pp. 1–17. [CrossRef]

Sprouse, Jon, Matthew W. Wagers, and Colin Phillips. 2012. A test of the relation between working-memory capacity and syntactic island effects. *Language* 88: 82–123. [CrossRef]

Sprouse, Jon, Matthew W. Wagers, and Colin Phillips. 2013. Deriving competing predictions from grammatical approaches and reductionist approaches to island effects. In *Experimental Syntax and Island Effects*. Edited by Jon Sprouse and Norbert Hornstein. Cambridge: Cambridge University Press, pp. 21–41. [CrossRef]

Staub, Adrian. 2007. The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology* 33: 550–69. [CrossRef]

Stepanov, Arthur. 2007. The end of CED? Minimalism and extraction domains. *Syntax and Semantics* 10: 80–126. [CrossRef]
Tenny, Carol. 1995. How motion verbs are special: The interaction of semantic and pragmatic information in aspectual verb meanings. *Pragmatics & Cognition* 3: 31–73. [CrossRef]
Truswell, Robert. 2007. Extraction from adjuncts and the structure of events. *Lingua* 117: 1355–77. [CrossRef]
Truswell, Robert. 2011. *Events, Phrases, and Questions*. Oxford: Oxford University Press. [CrossRef]
Vendler, Zeno. 1957. Verbs and times. *The Philosophical Review* 66: 143–60. [CrossRef]
Wagers, Matthew W. 2013. Memory mechanisms for wh-dependency formation and their implications for islandhood. In *Experimental Syntax and Island Effects*. Edited by Jon Sprouse and Norbert Hornstein. Cambridge: Cambridge University Press, pp. 161–85. [CrossRef]
Winkler, Susanne. 1997. *Focus and Secondary Predication*. Berlin and New York: Mouton de Gruyter. [CrossRef]