Covert Movement in Multiple-Wh Questions: Experimental and Theoretical Investigations

Shlonsky, Ur; Villata, Sandra; Franck, Julie

How to cite

SHLONSKY, Ur, VILLATA, Sandra, FRANCK, Julie. Covert Movement in Multiple-Wh Questions: Experimental and Theoretical Investigations. In: Syntax, 2020, vol. 23, n° 2, p. 185–202. doi: 10.1111/synt.12192

This publication URL: https://archive-ouverte.unige.ch/unige:129201
Publication DOI: 10.1111/synt.12192

Syntax
A Journal of Theoretical, Experimental and Interdisciplinary Research

Syntax 2020 DOI: 10.1111/synt.12192

Covert Movement in Multiple-*Wh* Questions: Experimental and Theoretical Investigations

Ur Shlonsky, Sandra Villata, and Julie Franck

Abstract. Results from a new grammaticality-judgment experiment in French confirm the published finding in English that sentences containing a Superiority violation involving a bare extracted element and a lexically restricted intervener (e.g., 'What did which student buy?'), a configuration termed *inverse inclusion*, are more acceptable than those involving a lexically restricted extracted element and a bare intervener (e.g., 'Which book did who buy'), a configuration termed *inclusion*. To account for this pattern, we adopt an explicit implementation of covert movement and propose some modifications in the characterization of the class of interveners. Interestingly, experimental findings on extraction from *wh* islands attest the opposite pattern: there, inclusion is more acceptable than inverse inclusion. We argue that whereas (overt) extraction from *wh* islands is sensitive to the feature content of the extractee and the intervener (i.e., whether or not they are lexically restricted), the degree of (un) acceptability of Superiority violations hinges on the different landing-site options that the features of the extractee and the intervener permit.

1. Introduction

Ross 1967 observes that *wh* movement of an object *wh* out of an indirect *wh* question is degraded in English, as compared with *wh* movement out of a noninterrogative complement:

(1) a. Movement out of a *wh* island
??What do you wonder who bought?
b. Movement out of a noninterrogative complement clause
What do you think John bought __?

Indirect *wh* questions belong to a family of what Ross terms *islands*, that is, domains opaque to *wh* movement.

Over the years, it has become clear that various factors influence the grammatical status of *wh*-island violations. One factor that systematically ameliorates *wh* movement out of a *wh* island is the addition of a lexical restriction to the moved *wh*. Thus, (2) is typically judged as being only mildly deviant, contrasting with (1a).

(2) ?Which book do you wonder who bought?

There are various reasons, however, to view (1a) not just as an island violation, that is, an extraction from an opaque domain, but also as a violation of the principle that restricts the formation of chains across intervening elements: Relativized Minimality

The authors are grateful to Farhad Mirdamadi for his contribution to the research that ultimately led to this paper and to two *Syntax* reviewers for comments on an earlier version.

(Rizzi 1990 and subsequent work). One immediate advantage of Relativized Minimality, specifically of *featural* Relativized Minimality (Rizzi 2004, Starke 2001), is that it provides a relatively straightforward way to express the contrast between (1a) and (2).

From the vantage point of featural Relativized Minimality, the unacceptability of (1a) is due to intervention of the subject *wh* element *who* in the chain linking *what* with its copy, in the object position of *bought*. Since the intervener and the target of *wh* movement have *identical* morphosyntactic features, namely the feature [+Wh], a chain relating the surface position of *what* to its base position is ill-formed, as it violates Relativized Minimality (Rizzi 2004). This configuration is schematized in (3a). The relative acceptability of (2) is due to the fact that the lexically restricted *which book* contains an additional feature, one that is absent on the intervener *who*, a bare *wh* element. Following Grohmann 2006, Richards 2001, and Rizzi 2011, we initially take this additional feature to be [+Top(ic)], expressing the properties of discourse linking (D-linking), namely specificity and presuppositionality, which are typically associated with lexically restricted *wh* expressions. (We motivate a revision of the label in section 4.) Sentence (2) instantiates a case of (feature) *inclusion*, where the feature set of the intervening *wh* has the feature [+Wh] while the feature set of the extracted *wh* has the features [+Wh, +Top]. This configuration is schematized in (3b).

(3) a. What . . . who . . . ⟨what⟩ — Feature identity
[+Wh] [+Wh] [+Wh] — (unacceptable)
b. Which book . . . who . . . ⟨which book⟩ — Feature inclusion
[+Wh, +Top] [+Wh] [+Wh, +Top] — (mildly deviant)

2. A Puzzle

In a series of acceptability-judgment experiments on *wh* islands in French, reported in Villata, Rizzi & Franck 2016, it was observed that feature inclusion improved sentence acceptability as compared with feature identity. The authors also found that when the feature set of the extracted *wh* was included in the feature set of the structurally intervening *wh*, manifesting the configuration dubbed *inverse inclusion*, exemplified in (4), sentences received the same acceptability ratings as sentences that manifested feature identity.

(4) *What do you wonder which student bought __? — Inverse inclusion

Converging evidence has been reported in English (Atkinson et al. 2015) and in Persian (Mirdamadi, Shlonsky & Frank 2016) to the effect that inclusion is more acceptable than inverse inclusion, which receives the same acceptability ratings as identity.

According to Relativized Minimality, what counts is the featural richness of the extracted *wh* object relative to that of the intervening *wh* subject. And indeed, in the inclusion configuration, the feature set of the extracted element is richer than that of

the intervening elements, whereas under identity and inverse inclusion, the feature set of the intervener is either identical to or forms a superset of the feature set of the extractee.

A reverse profile of acceptability, however, has emerged from two studies of Superiority violations, Hofmeister et al. 2007, 2013. The authors observed that inverse inclusion was more acceptable than inclusion, which was judged on a par with identity:

(5) a. ?Mary wondered what which boy read. Inverse inclusion
b. *Mary wondered which book who read. Inclusion
c. *Mary wondered what who read. Identity

The aim of the present paper is to shed light on this puzzle, asking why inverse inclusion is more acceptable than inclusion in Superiority violations—a profile of acceptability that is the opposite of the one reported in *wh* islands.[1]

Since the pattern in question has only been observed in English, we consider it important to first corroborate Hofmeister et al. 2007, 2013's results in another language. In section 3 we report the results of a grammaticality-judgment experiment in French, comparing the acceptability of inverse inclusion in Superiority-violation structures with that of inclusion. The experimental results show that the reversed acceptability profile for the inclusion and inverse-inclusion conditions in Superiority violations extends to French. Section 4 proposes an account for the pattern of acceptability of inclusion and inverse inclusion in Superiority violations, an account in terms of covert movement. Section 5 articulates our solution to the puzzle described in this section, and section 6 concludes the paper.

3. Experiment

3.1. Participants

The participants were 49 native French speakers, all students at the University of Geneva. They received course credits for their participation and were naive as to the purpose of the experiment.

3.2. Materials

The design consisted of a single four-level variable, Configuration, obtained by crossing the factors *Wh* Object Type and *Wh* Subject Type, each having two levels, lexically restricted and bare. This design generated two configurations besides *inverse inclusion* and *inclusion* (see section 2), namely *bare identity*, with two bare *wh* elements (cf. (3a), (5c)), and *complex identity*, with two lexically restricted *wh* elements. We will not discuss the pattern of acceptability of these two additional configurations in this paper. The reason that we included them in this study was to

[1] We should add that the fourth experiment in Hofmeister et al. 2013 attested no difference between inverse inclusion and inclusion. As this pattern of acceptability was only reported in one experiment, we attribute it to lack of power.

have a fully crossed design, manipulating lexical restriction on both *wh* elements. Table 1 gives a sample experimental item consisting of four sentences, one for each of the experimental conditions. We generated 24 such experimental items. The *wh* object was always inanimate and the *wh* subject was always animate. All *wh* elements were masculine and singular. Half of the experimental sentences contained *se demander* 'wonder' as a matrix verb, while the other half contained *savoir* 'know'. All sentences with *se demander* were affirmative, while half of the sentences with *savoir* were negative. We also created 176 fillers, which consisted of 112 sentences involving *wh* islands and 64 sentences in which there was no Superiority violation and the *wh* subject and the *wh* object both appeared in their base positions. Like in the experimental items, we manipulated lexical restriction on both the *wh* subject and the *wh* object in the filler items, to make them as comparable as possible. We created four lists of 68 items by intermixing 24 test sentences (one from each experimental item) with 44 different fillers (one-fourth of the fillers). The items in each list were randomly presented to participants, and participants were asked to judge each of them.

Table 1. Sample experimental item, showing the four conditions[2]

Configuration	Sentence
Bare identity	Je me demande ce que qui a acheté. I me ask this that who has bought 'I wonder what who bought.'
Inverse inclusion	Je me demande ce que quel étudiant a acheté. I me ask this that which student has bought 'I wonder what which student bought.'
Inclusion	Je me demande quel livre qui a acheté. I me ask which book who has bought 'I wonder which book who bought.'
Complex identity	Je me demande quel livre quel étudiant a acheté. I me ask which book which student has bought 'I wonder which book which student bought.'

3.3. Procedure

We programmed the experiment with E-Prime (Schneider, Eschman & Zuccolotto 2012). Sentences were presented one at a time on a computer screen and participants were asked to make acceptability judgments on the basis of a 1–7 Likert scale (with 1 corresponding to a completely unacceptable sentence and 7 to a fully acceptable one). We gave each participant thorough instructions, including three examples illustrating how to use the scale, followed by 10 practice items. There was no time pressure during the experiment.

[2] In French indirect questions, simple *que* 'what' does not occur, and instead one finds *ce que*, literally 'this that'. We assume that *ce que* questions contain a bare *wh* operator, and we leave open the question of why simple *que* cannot be used.

3.4. Data Analyses

We analyzed the data with linear mixed-effect models estimated with the lmerTest package (Bates et al. 2015) in the R software environment (R Development Core Team 2015). We used Configuration as the sole fixed factor,[3] with random intercepts and slopes for both participants and items (we always used the maximal random-effect structure by participant and item).[4] We did not exclude any data point from the data set. In addition, to directly compare the relative acceptability of the four experimental conditions, we ran an additional lmer model in which contrasts were dummy coded, such that the intercept of the model represented the reference level (e.g., bare identity) to which the three other structures were compared. Prior to statistical analysis, acceptability judgments from each participant were *z* scores transformed in order to correct for scale bias. The Satterthwaite approximation for degrees of freedom was used to estimate *p* values and *z* scores; transformed data were calculated subject by subject in order to avoid bias in using the scale and to improve statistical power.

3.5. Results

Figure 1 reports the mean acceptability (in *z* scores) and standard errors for the four experimental conditions. The results attest a significant main effect of Configuration ($F = 57.71$, $p < .001$), revealing that acceptability rates differ across the four structures. Pairwise comparisons attest significantly higher acceptability rates for inverse inclusion ($M = 0.065$) than for inclusion ($M = -0.327$) ($\beta = -0.392$, $SE = 0.119$, $t = -3.293$, $p = .002$). With respect to the other two configurations, namely bare identity and complex identity, the results indicate that inclusion is not significantly different from bare identity ($t < 1$), but complex identity ($M = 0.649$) is higher rated than inverse inclusion ($\beta = -0.584$, $SE = 0.132$, $t = -4.415$, $p < .001$).[5] The observed pattern is thus the following ("=" means "on a par with" and "<" means "lower rated than").

(6) Bare identity = inclusion < inverse inclusion < complex identity

[3] For the sake of comparison with data analyses from Hofmeister et al. 2007, 2013, in an appendix to this article we report the results of a two-factor model, with lexical restriction of the *wh* subject and lexical restriction of the *wh* object as the two factors.

[4] An anonymous reviewer points out that despite the common assumption that Likert scores can be analyzed through a linear-regression model, a Likert scale is actually a set of ordered categories. However, as discussed in Gelman & Hill 2007:123, ordinal data can be analyzed through linear regression if the number of categories is large enough and if a reasonable range of the categories is used by participants. As far as the number of categories is concerned, there is evidence suggesting that when there are five or more ordered categories, there is little harm in treating these variables as continuous (e.g., Johnson & Creech 1983). With respect to the range of categories used by participants, the 1,176 observations were distributed over the seven categories as follows: 246 "1" answers, 151 "2" answers, 219 "3" answers, 213 "4" answers, 197 "5" answers, 91 "6" answers, and 59 "7" answers. Thus, participants used the whole scale.

[5] A *Syntax* editor wonders whether the variable length of the tested sentences could have affected the results. We ran an additional model that included length (word count) as a covariate. Results showed no main effect of length ($t < 1$), attesting that length as such had no effect on the results.

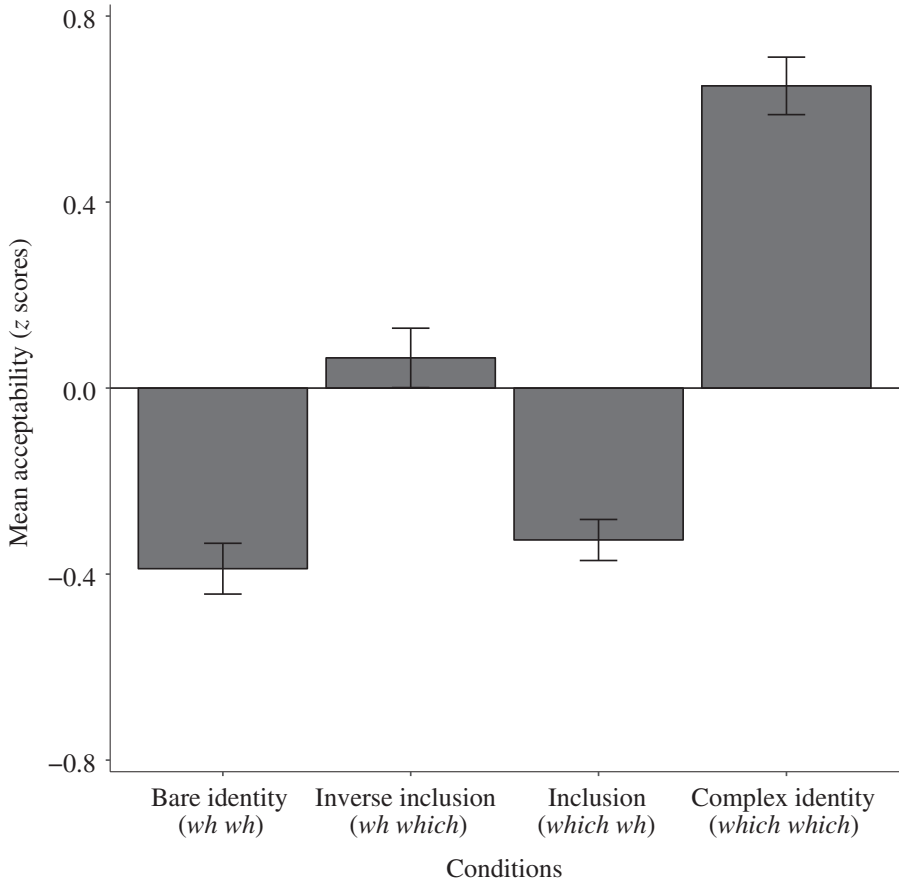


Figure 1. Mean acceptability ratings for the four experimental configurations, after conversion to *z* scores. The zero represents the grand mean, so that a positive *z* score means that the structure is rated above the grand mean, while a negative *z* score means that the structure is rated below the grand mean.

4. Accounting for the Puzzle: The Role of Covert Movement

In section 3 we reported the results of a grammaticality-judgment experiment in French, replicating Hofmeister et al. 2007, 2013's results in English, showing that inverse inclusion is more acceptable than inclusion in Superiority violations. This pattern of acceptability is the opposite of the pattern of acceptability in extraction out of *wh* islands, where inclusion is more acceptable than inverse inclusion (see Mirdamadi 2018 for similar results on Persian islands and Superiority violations).

This reversal in acceptability rates is amenable to a principled analysis if we consider a fundamental syntactic difference between Superiority and *wh*-island

configurations. In the case of extraction out of *wh* islands, the two *wh* elements have scope over two distinct clauses. In (1a), repeated here as (7), *who* scopes over the embedded clause and *what* over the matrix clause. The position of the two *wh* elements corresponds to their scope position.

(7) ??What do you wonder who bought?

In Superiority-violation structures, the *wh* elements have scope over the same clause, but only the "ex-situ" *wh* is pronounced in its scope position. On the assumption that *wh* elements must be in their scope position at the interface with semantic interpretation, there has to be some way for the in-situ *wh* to come to occupy a scope-taking, clause-initial position. A widely known analysis of multiple-*wh* questions posits that the in-situ element undergoes covert movement and comes to occupy its scope position at Logical Form (LF) (Chomsky 1981, Huang 1982; see also, for elaboration, Fox 2003, Hornstein 1995, Huang 1995, among others). Covert movement does not take place in *wh* islands, because each *wh* element is overtly moved to its appropriate scope position.

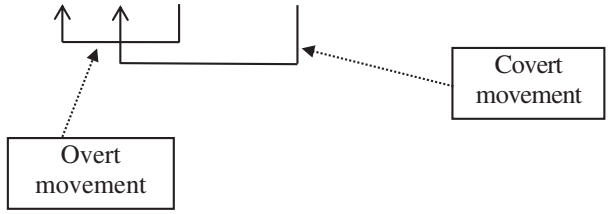
In Villata, Rizzi & Frank 2016: sec. 5.3, the authors suggest that this formal difference between extraction from *wh* islands and Superiority violations lies at the core of the reversal in the acceptability rates of the inclusion and inverse-inclusion configurations. Our discussion in the following paragraphs is inspired by Rizzi 2017's more explicit execution of this insight.

Consider the contrast in (8). The Superiority condition of Chomsky 1973 requires the highest *wh* element, here the subject, to move overtly (see, e.g., Huang 1982). The sentence in (8a) is thus a grammatical multiple-*wh* question with an overtly moved subject and an in-situ object. The sentence in (8b) illustrates an ungrammatical (Superiority-violating) format: the object is overtly moved and the subject remains in situ.

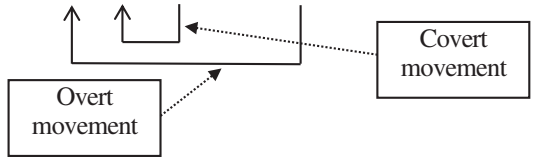
(8) a. (John wonders) who said what?
b. *(John wonders) what who said?

At LF, *wh* in-situ expressions undergo covert movement. In agreement with Richards 2001, we assume that covert movement involves "tucking in": the landing site of the covertly moved expression lies below that of the overtly moved one. Richards argues that tucking in involves movement to a lower specifier of the *wh* probe. A different implementation—consistent with the "one specifier, one head" constraint (see Cinque 1999's argument for the existence of a head position in between any two specifiers)—consists of moving only the *wh* feature of the in-situ phrase directly to the *wh* probe (adapting Roberts 2010's unification of agreement and pronominal cliticization to the *wh* domain). What is crucial here are the representations in (9), showing the LFs for (8a) and (8b) after covert movement applies. (Elements surrounded by ⟨ ⟩ are lower copies.)

(9) a. . . .who what ⟨who⟩ said ⟨what⟩ = (8a)



b. *. . .what who ⟨who⟩ said ⟨what⟩ = (8b)



The two chains, (*what*, ⟨*what*⟩) and (*who*, ⟨*who*⟩), intersect in the grammatical (9a): Movement and tucking in of *what* under *who* only crosses the lower link of the (*who*, ⟨*who*⟩) chain. In the ungrammatical (9b), covert movement of *who* tucks in under the overtly moved *what*, and the result is that the (*who*, ⟨*who*⟩) chain comes to be nested within the (*what*, ⟨*what*⟩) chain.

In order to bring this formal difference between crossing and nesting under the purview of Relativized Minimality, we adopt Krapova & Cinque 2008's proposal to the effect that only full chains count as interveners; individual links of chains do not.[6] Krapova & Cinque's analysis is based on ordering constraints in multiple *wh* fronting in Bulgarian, but it can be generalized and adapted to multiple-*wh* questions in English-like languages. On the basis of their proposal, (9b) violates Relativized Minimality: the entire (*who*, ⟨*who*⟩) chain intervenes, rendering the (*what*, ⟨*what*⟩) chain illicit. The structure in (9a), however, is well-formed with respect to this interpretation of Relativized Minimality, since only one member of the (*what*, ⟨*what*⟩) chain intervenes in the representation of the (*who*, ⟨*who*⟩) chain.

Observe that (9b) formally corresponds to the bare-identity condition that we tested in the experiment we reported in section 3 (see table 1). Recall that it received the lowest acceptability ranking. In light of this discussion, we can attribute its low acceptability to a violation of Relativized Minimality, as that principle applies to the representation of chains at LF.

What about the inclusion and inverse-inclusion conditions? Recall that the former was judged on a par with bare identity, while the latter was ranked slightly (but significantly) higher. A principled grammatical explanation of this variation is available, once we assume that lexically restricted *wh* expressions are *not required* to

[6] This proposal is, in fact, the reverse of Kuno & Robinson 1972's Constraint on Crossing Dependencies and Pesetsky 1982's Path-Containment Condition, which rule out crossing dependencies. Ordering constraints on multiply fronted *wh* words in several Eastern European languages, brought to light in Rudin 1988, falsify these early attempts to account for Superiority effects; see Richards 2001.

tuck in and have the option of moving to a higher position in the left periphery, a position that is inaccessible to bare *wh* expressions (Krapova & Cinque 2008, Soare 2009).

For evidence of a different landing site for lexically restricted *wh*, it is useful to look in multiple-*wh*-fronting languages, which impose a rigid order on the *wh* expressions in the left periphery of interrogatives. Soare 2009, for example, shows that in Romanian multiple-*wh*-fronting constructions that combine a lexically restricted *wh* and a bare *wh*, the lexically restricted one always precedes the bare one. This is particularly striking when it is a subject *wh* and an object *wh* that are combined: while both bare and lexically restricted *wh* subjects must precede a bare *wh* object, as in (10a,b), a bare subject must follow a lexically restricted object, as in (10c).

(10) a. Cine ce a citit?
who what has read
'Who read what?'
b. Care student ce a citit?
which student what has read
'Which student read what?'
c. Pe care carte cine a citit-o?
OBJ which book who has read-it
'Which book did who read?'

In agreement with Krapova & Cinque 2008, which examines Bulgarian, Soare 2009 argues that lexically restricted *wh* expressions in Romanian target a higher position in the left periphery than bare ones and that, more generally, the order of *wh* phrases in the left periphery reflects a feature-based cartography of positions.

In section 2, we followed Rizzi 2011 and claimed that the additional feature on lexically restricted *wh* phrases is [+Top], thereby expressing the connection between the D-linking interpretation (requiring answers to be drawn from a set that counts as old information for speaker and hearer) and left-peripheral topics. Now, bare *wh* expressions can also be D-linked in appropriate discourse conditions (Pesetsky 1987) and are thus expected to be potential bearers of [+Top]. Yet the Romanian data is clear on this point: an object *wh* can only precede a subject *wh* if it has the form "which *x*," that is, if it is an expression containing a *wh* word 'which' and a lexical restriction *x* ('book' in (10c)). Bare but D-linked objects cannot precede *wh* subjects in Romanian. Thus, it seems that [+Top] does not make the right cut and that the relevant feature is instead one that encodes the lexical restriction. This is further supported by experiment 3 in Villata, Rizzi & Franck 2016, which shows that lexical restriction has an effect on *wh*-island acceptability independently of D-linking. For this reason, we provisionally adopt from Friedmann, Belletti & Rizzi 2009 the label [+N(ominal)] for the extra feature borne by "which *x*."

The *wh* field in the left periphery of Romanian (and, presumably, of other multiple-*wh*-fronting languages in which the order of *wh* elements is basically rigid, e.g., Bulgarian) distinguishes a probing head with the feature [+Wh] and a higher head

bearing the probing features [+Wh, +N]. The lexically restricted *wh* expressions in Romanian target this higher head.

From the earliest work on LF movement, an important working hypothesis has been that overt and covert syntactic operations are uniform, differing only in their morphological and phonological realization. It has also been assumed that an operation that takes place overtly in one language can take place covertly in another, the choice being governed by a parameter. Thus, the overt patterns of movement in Bulgarian and Romanian are expected to reproduce themselves covertly in languages that lack multiple overt *wh* movement. We therefore assume that in English and French multiple questions, lexically restricted *wh* expressions may also target a higher head than bare *wh* elements.

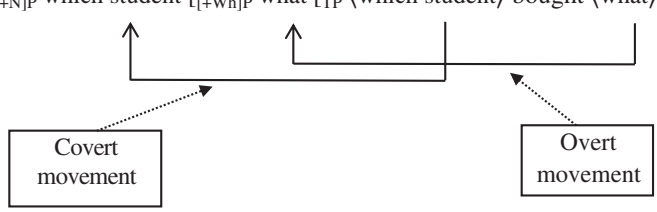
The *wh* field in languages like Bulgarian and Romanian displays a rich and fairly rigid cartography of positions (Krapova & Cinque 2008, Soare 2009): for example, a bare subject *wh* precedes a bare object *wh*, which in turn precedes an adverbial *wh*. In these languages, the various positions are also sensitive to the [±human] distinction (and, in Bulgarian, to the presence or absence of a resumptive clitic associated with the fronted *wh* expression). There is no evidence for such fine-grained cartographic structure in English or French. They appear to syntactically distinguish only between lexically restricted and bare *wh* elements. Given the coarser granularity of the *wh* field in these languages, it is plausible that the probe here is featurally less precise than it is in Bulgarian and Romanian. We suggest that the pure *wh* head in these languages may probe both bare and lexically restricted goals. Only the latter, however, have access to the higher, featurally more specific position. Lexically restricted *wh* expressions (or their relevant features) thus have the option of targeting a higher head but may also target the lower [+Wh] head or tuck in below it, if its specifier is filled by a previously (overtly) moved *wh*.

Consider, in light of this discussion, the derivation of inverse inclusion (5a) and inclusion (5b), starting with the former. In (11), *what* is overtly moved to the specifier of the "pure" [+Wh] head. *Which student* implements the option of moving covertly to the higher [+Wh, +N] head, yielding intersecting chains at LF. The sentence—which appears to instantiate a Superiority violation—is actually predicted to be grammatical by the Krapova & Cinque 2008 interpretation of Relativized Minimality.

(11) Inverse inclusion

Mary wonders what which student bought.

LF: . . . $[_{[+Wh,+N]P}$ which student $[_{[+Wh]P}$ what $[_{TP}$ ⟨which student⟩ bought ⟨what⟩]]]

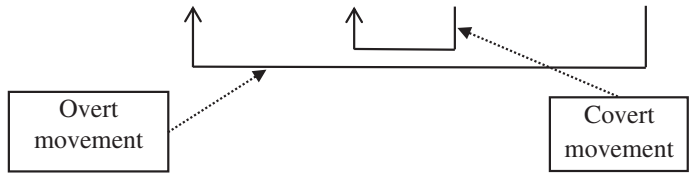


In the LF of the inclusion configuration, either *which book* is the specifier of the [+Wh, +N] head and *who* the specifier of the [+Wh] head, as in (12a), or *which book* is the specifier of the [+Wh] head and *who* tucks in below it, as in (12b) (perhaps incorporating into the head of the [+Wh] projection, as suggested above).[7] Both derivations give rise to nesting chains, predicted to be ungrammatical.

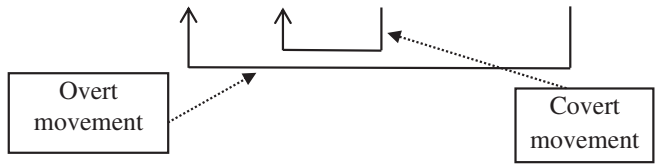
(12) Inclusion

I wonder which book who bought.

a. LF_1: . . . [$_{[+Wh,+N]P}$ which book [$_{[+Wh]P}$ who [$_{TP}$ ⟨who⟩ bought ⟨which book⟩]]]



b. LF_2: . . . [$_{[+Wh]P}$ which book who [$_{TP}$ ⟨who⟩ bought ⟨which book⟩]]]



The experimental data accord with these predictions: The inverse-inclusion condition is systematically judged more acceptable than the inclusion condition. (At the end of this section, we return to the question of why inverse inclusion is nevertheless not judged at ceiling level.)[8]

The analysis presented in the preceding paragraphs is clearly inspired by the treatment of multiple-*wh* questions and D-linking in Pesetsky 1987, 2000. For Pesetsky, as for us, multiple-*wh* questions require additional landing sites in the left periphery (multiple specifiers of C in his 2000 approach; two (or more) featurally distinct heads—with a single specifier each—in ours). Our approaches differ, however, in that for Pesetsky, multiple questions containing D-linked phrases are associated with a complementizer that takes a single specifier, as opposed to one that

[7] With Richards 2001, it must be assumed that tucking in constrains movement and is not a representational constraint. An expression P can only tuck in below an expression Q if Q has already moved. Moving P to a low position and then moving Q to a higher position is not a case of tucking in. In our proposal, Q can only move to a higher position than P if it is probed by a higher head, as in the case of lexically restricted *wh*.

[8] The literature reports that sentences such as (12) are grammatical in Bulgarian and Romanian (cf. (10c)). Given the fact that in these languages, lexically restricted *wh* expressions *must* target a higher probe than bare *wh* expressions, it could be argued that the two chains do not interact and do not enter into the evaluation of chain intervention. Krapova & Cinque construe this idea in terms of distinct "spaces" in the left periphery and argue that intervention is only calculated internally to each space.

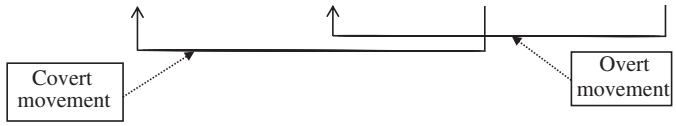
takes multiple specifiers, while for us, the presence of lexically restricted phrases induces more, not fewer positional options: single specifiers of several featurally distinct heads. (On the need to distinguish D-linking from lexical restriction, see the discussion following (10).)[9]

In our system, complex identity exploits the availability of two left-peripheral positions for lexically restricted *wh* phrases. Consider (13), in which the object *wh* is overtly moved to Spec,[+Wh]P and the subject covertly moved above it, to Spec, [+Wh, +N]P.

(13) Complex identity

Which book did which student read?

LF_1: . . . $[_{[+Wh,+N]P}$ which student $[_{[+Wh]P}$ which book $[_{TP}$ ⟨which student⟩ read ⟨which book⟩]]]



One analysis can thus account not only for the subtle differences in acceptability between inclusion and inverse inclusion but also for the fact that complex identity is judged more acceptable than bare identity. What our syntactic approach does not capture is the observation that complex identity receives a higher rating than inverse inclusion in the French experiment (see figure 1, section 3.5) and in Hofmeister et al. 2007, 2013's results on English. We leave this matter open for future research, noting that Hofmeister et al. 2007, 2013 as well as Villata, Rizzi & Franck 2016 and Villata 2017 all attribute the accrued acceptability of complex identity to an extragrammatical factor, unrelated to chain formation.

Let us now see how the proposed analysis carries over to multiple-*wh* questions that do not violate Superiority:

(14) a. Who bought what?
b. Who bought which book?
c. Which student bought what?

For Relativized Minimality to be respected, the multiple questions in (14) must involve intersecting chains. To bring that about, covert movement of the in-situ *wh* must target a position below the overtly moved *wh* subject. In particular, the

[9] The two approaches also differ in their empirical coverage. Pesetsky compares multiple questions with bare *wh* phrases to multiple questions in which all the *wh* elements are lexically restricted. His approach is designed to deal with the contrast between our bare-identity and complex-identity configurations. Pesetsky does not discuss inclusion and inverse inclusion. Pesetsky 1987 argues that D-linked *wh* phrases can be interpreted in situ and that simplex *wh* phrases must move. Pesetsky 2000 refines this view and argues that, in Superiority-violating configurations, only the in-situ strategy is available. One can envisage an extension of Pesetsky's analysis to the contrasts between the inverse-inclusion and inclusion configurations that our research has highlighted. Such an extension would require explaining which one of the two complementizers (the single-specifier one or the multiple-specifier one) is required in inclusion and inverse inclusion and whether a bare *wh* in the inverse-inclusion configuration can satisfy Attract Closest.

lexically restricted object in (14b) must implement the option of *not* moving to the specifier of the higher [+Wh, +N] probe. If it did, the (*who*, ⟨*who*⟩) chain would end up nested within the (*which book*, ⟨*which book*⟩) chain, in violation of Relativized Minimality.

At LF, then, the sentences in (14) are derived as in (15). In all three, an intersecting configuration of chains is obtained.

(15) a. Who what ⟨who⟩ bought ⟨what⟩?
b. Who which book ⟨who⟩ bought ⟨which book⟩?
c. Which student what ⟨which student⟩ bought ⟨what⟩?

In (15a), covert movement of *what* (or of its features) tucks in below *who*, which moves overtly to the specifier of the [+Wh] head. In (15b), likewise, covert movement of *which book* (or of its features) tucks in below *who*. In (15c), *which student* moves overtly to the specifier of the [+Wh, +N] head, and covert movement of *what* targets the specifier of the lower [+Wh] head; alternatively, *which student* overtly moves to the specifier of the [+Wh] head and covert movement of *what* (or of its features) tucks in below it.

Note, now, that the LF in (15c) is identical to that of inverse inclusion in (11). But then, why do inverse-inclusion sentences receive a score of less than 4 on a seven-point Likert scale, when they contain intersecting chains just like the non-Superiority-violating (14c)? A closer look at the chain links crossed over in the derivations of (11) and (14c)/(15c) is needed to characterize the difference between the two sentence types. In the inverse-inclusion case (11), the overt movement chain crosses over *which student* in its argument position, whereas in the chain formed covertly, the intervening chain link is the head of the (*what*, ⟨*what*⟩) chain, an A′ position. In (14c)/(15c), as in the other sentences in (14)/(15), neither overt nor covert movement crosses over a chain head in an A′ position, only over the tails of the *wh* chains, in A positions. The less-than-ceiling acceptability score of (11), as compared with the perfectly grammatical non-Superiority-violation case of (14c)/(15c), suggests that we capitalize on this difference.

Our suggestion is that crossing over a chain link in an operator position is costlier than crossing over a link in an argument position and that it is this added cost that underlies the less-than-perfect status of inverse inclusion (see also Chomsky 2001: (16)). No detectable cost is incurred by crossing the link in the A position (the tail of the chain). The highest link of an A′ chain, its head, is the link that is interpreted as the operator and that bears the criterial (in the sense of Rizzi 2006) feature. The lower link, the copy, is interpreted as the variable.

According to Krapova & Cinque 2008's interpretation of Relativized Minimality, only full chains are computed as interveners. We suggest that link intervention exists as well but is milder. To detect it, it is helpful to probe the data experimentally and evaluate the statistical significance of the results. Our experiment demonstrates that the degradation of inverse inclusion is, indeed, less robust than that of inclusion, which involves nested chains, but real nonetheless.

5. On the Difference between Weak Islands and Superiority Configurations

Recall the basic puzzle: the pattern of acceptability of inclusion and inverse inclusion in *wh*-island violations is reversed in Superiority violations. While inclusion fares better than inverse inclusion in the former, it is rated as less acceptable in the latter. We accounted for the relatively degraded status of inclusion in Superiority violations by appealing to Krapova & Cinque 2008's proposal that only nested chains give rise to (robust) Relativized Minimality violations. Inclusion configurations involve nested chains, while the chains in inverse-inclusion configurations intersect.

The *wh*-island violations tested in Villata, Rizzi & Franck 2016 all involve nested chains in which a direct object is extracted out of a *wh* island formed by the embedded subject, as in (16). Krapova & Cinque's interpretation of Relativized Minimality as a condition sensitive to chain geometry consequently predicts *wh*-island violations to be *uniformly* excluded by Relativized Minimality. This is contrary to the empirical findings.

(16) wh_1 do you wonder wh_2 $\langle wh_2 \rangle$ bought $\langle wh_1 \rangle$?

There is thus a tension between Krapova & Cinque's statement of Relativized Minimality as a condition on chain intervention—relevant for the computation of multiple-*wh* questions—and featural Relativized Minimality, which is sensitive to the feature content of interveners but is apparently oblivious to whether the interveners are full chains or segments of chains.

This tension brings to light some fundamental differences between intervention effects in multiple-*wh* questions (Superiority) and extraction from *wh* islands. One relevant difference between the two is that the *wh* expressions in a multiple question are interpreted in the same left periphery and have overlapping or identical scopes. Each *wh* expression in the island configuration is, by contrast, interpreted in a different left periphery, in a different clause.

We suggest that Relativized Minimality is implemented differently in the two cases: chain intervention in the Krapova & Cinque 2008 sense is only relevant to the computation of chains within the same clause and is hence oblivious to the fact that (16) involves chain nesting.

Unlike chain intervention, featural Relativized Minimality is relevant both in multiple-*wh* questions and in *wh*-island configurations, although its impact is different in the two cases. In *wh*-island contexts, where chain geometry is not relevant, the set-theoretic implementation of featural Relativized Minimality developed in Friedmann, Belletti & Rizzi 2009 and elaborated on in Villata, Rizzi & Franck 2016 accounts for the observed differences in the acceptability of the inclusion and inverse-inclusion configurations. [+N] fundamentally plays a set-theoretic role in featural Relativized Minimality: {[+Wh]} is a subset of {[+Wh], [+N]}, and Relativized Minimality is sensitive to the subset–superset relations among feature bundles.

In Superiority violations, the features on the *wh* expressions designate different landing sites for movement. We have argued, in particular, that [+Wh, +N] enables a *wh* expression to target a higher position in the left-peripheral *wh* field. It is

reasonable to think that the different feature-sensitive landing sites for *wh* elements are also present in the island configuration, such that an island-forming [+Wh, +N] expression like *which student* in the inverse-inclusion configuration exemplified in (4), **What do you wonder which student bought?*, may target a higher position than a bare *wh* word. However, in this case the choice of landing site is beside the point, since the *wh* word moved out of the island will cross over the island-forming *wh* whether it sits in Spec,[+Wh]P or in the higher Spec,[+Wh, +N]P.

It is, of course, tempting to try to unify the set-theoretic calculus with the structural/geometric computation. The temptation is surely enhanced by the correspondence between cartographic structure and the featural makeup of the *wh* in the Bulgarian and Romanian left peripheries, briefly discussed in section 4. But we leave this project for future work.[10]

6. Summary and Conclusion

In an experiment manipulating lexical restriction in sentences containing Superiority violations in French, we found that inverse-inclusion configurations were more acceptable than inclusion configurations, replicating English evidence reported by Hofmeister et al. 2007, 2013. This pattern of acceptability is the opposite of the pattern found for *wh* islands. Building on Villata, Rizzi & Franck 2016, we argued that the difference in acceptability judgments between these two structures is rooted in the fact that the derivation of multiple-*wh* questions involves covert movement of the in-situ *wh*. We developed an explicit implementation of covert movement, such that bare *wh* elements covertly move to a position below the overtly moved *wh* while lexically restricted *wh* elements have the added option of moving to a higher position, due to their particular feature content.[11] We agreed with Rizzi 2017 in arguing that Krapova & Cinque 2008's interpretation of Relativized Minimality accounts for the degraded status of the inclusion configuration in Superiority violations. The higher, but not fully acceptable status of inverse inclusion in Superiority violations led us to develop the idea that intervention by the head of a chain also enters into the computation of acceptability. We summarize our proposal as follows.

[10] Another open question is whether featural Relativized Minimality impacts link intervention in non-Superiority-violating multiple questions. One might expect some gradation in the acceptability of the sentences in (14), under the assumption that the chain-creating algorithm is sensitive to the [+N] feature, even when it applies covertly and crosses over the tail of a chain in an argumental position. It would not be surprising if speakers noted subtle differences between the bare–bare configuration in (14a), the bare–lexically restricted one in (14b), and the lexically restricted–bare configuration in (14c)—but this would need to be established experimentally.

[11] Hofmeister et al. 2007 also attempts to account for the higher acceptability of inverse inclusion than of inclusion in Superiority violations. The claim is that "intervener accessibility [which is higher for lexically restricted interveners than for bare ones] impacts the processing of *wh*-dependencies as much as, or even more than filler accessibility" (p. 200). *Filler* refers to the extracted *wh* element that has to be integrated with a dedicated gap position during the processing of *wh* dependencies. But if intervener accessibility affects the processing of *wh* dependencies more than filler accessibility, the same pattern of acceptability should be found in *wh* islands and Superiority violations, contrary to the facts. This illustrates the point that an account that fails to consider the fine-grained structural differences between Superiority violations and *wh* islands cannot explain the reverse pattern of acceptability between inverse inclusion and inclusion attested by the experimental results.

(17) a.	Non-Superiority-violating	Intervention by the foot of the chain	Full acceptability (but see footnote 9)
b.	Inverse inclusion	Intervention by the head of the chain	Mild degradation
c.	Inclusion	Intervention by the full chain	Severe degradation

Full-chain intervention draws a line between (17a) and (17b) on the one hand and (17c) on the other. With nested chains, as in the inclusion configuration, degradation is severe. The improvement perceived in the inverse-inclusion configuration is due to the fact that the two *wh* chains are not nested but rather intersect; however, the formation of the chain rooted in subject position crosses over the head of the chain rooted in the object, which does reduce acceptability somewhat. In non-Superiority-violating configurations, on the other hand, the two chains intersect and the intervening link is only the foot of the chain; thus, no degradation is perceived.

Our theoretical conclusion is that Relativized Minimality should be thought of as a family of constraints on what can intervene in nonlocal dependencies in syntax. In some contexts, dependencies are sensitive to the feature content of interveners. In other contexts, the mere nesting of a full chain inside another renders such a dependency ungrammatical, and the preference for inverse inclusion over inclusion is due to a subtle sensitivity to whether an intervening chain link is an operator or a variable.

References

Atkinson, E., A. Aaron, K. Rawlins & A. Omaki. 2015. Similarity of *wh*-phrases and acceptability variation in *wh*-islands. *Frontiers in Psychology* 6: article 2048.

Bates, D., M. Maechler, B. Bolker & S. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1):1–48.

Chomsky, N. 1973. Conditions on transformations. In *A festschrift for Morris Halle*, ed. S. Anderson & P. Kiparsky, 232–286. New York: Holt, Reinhart & Winston.

Chomsky, N. 1981. *Lectures on government and binding*. Dordrecht, the Netherlands: Foris Publications.

Chomsky, N. 2001. Derivation by phase. In *Ken Hale: A life in language*, ed. M. Kenstowicz, 1–50. Cambridge, MA: MIT Press.

Cinque, G. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. Oxford: Oxford University Press.

Fox, D. 2003. On logical form. In *Minimalist syntax*, ed. R. Hendrick, 82–123. Oxford: Blackwell Publishing.

Friedmann, N., A. Belletti & L. Rizzi. 2009. Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua* 119:67–88.

Gelman, A. & J. Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Grohmann, K. 2006. Top issues in questions: Topics—topicalization—topicalizability. In *Wh-movement: Moving on*, ed. L. Cheng & N. Corver, 249–288. Cambridge, MA: MIT Press.

Hofmeister, P., F. Jaeger, I. Sag, I. Arnon & N. Snider. 2007. Locality and accessibility in *wh*-questions. In *Roots: Linguistics in search of its evidential base*, ed. S. Featherston & W. Sternefeld, 185–206. Berlin: Mouton de Gruyter.

Hofmeister, P., F. Jaeger, I. Sag, I. Arnon & N. Snider. 2013. The source ambiguity problem: Distinguishing effects of grammar and processing on acceptability judgments. *Language and Cognitive Processes* 28:48–87.

Hornstein, N. 1995. *Logical Form: From GB to Minimalism*. Cambridge, MA: Blackwell Publishers.
Huang, C.-T. J. 1982. Logical relations in Chinese and the theory of grammar. Doctoral thesis, Massachusetts Institute of Technology, Cambridge.
Huang, C.-T. J. 1995. Logical Form. In *Government and Binding Theory and the Minimalist Program*, ed. G. Webelhuth, 127–173. Oxford: Blackwell Publishers.
Johnson, D. & J. C. Creech. 1983. Ordinal measures in multiple indicator models: A simulation study of categorization error. *American Sociological Review* 48:398–407.
Krapova, I. & G. Cinque. 2008. On the order of *wh*-phrases in Bulgarian multiple *wh*-fronting. In *Formal description of Slavic languages: The fifth conference, Leipzig 2003*, ed. G. Zybatow, L. Szucsich, U. Junghanns & R. Meyer, 318–336. Frankfurt, Germany: Peter Lang Publishing.
Kuno, S. & J. Robinson. 1972. Multiple wh questions. *Linguistic Inquiry* 3:463–487.
Mirdamadi, F. 2018. Intervention effects in non-local dependencies: Evidence from Persian. Doctoral thesis, University of Geneva, Geneva.
Mirdamadi, F., U. Shlonsky & J. Franck. 2016. The effect of the Persian object marker *ra* and of lexical restriction on the acceptability of extraction across wh islands. *Rivista di Grammatica Generativa* 38:239–246.
Pesetsky, D. 1982. Paths and categories. Doctoral thesis, Massachusetts Institute of Technology, Cambridge.
Pesetsky, D. 1987. *Wh*-in-situ: Movement and unselective binding. In *The representation of (in) definiteness*, ed. E. Reuland & A. ter Meulen, 98–129. Cambridge, MA: MIT Press.
Pesetsky, D. 2000. *Phrasal movement and its kin*. Cambridge: MA: MIT Press.
R Development Core Team. 2015. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
Richards, N. 2001. *Movement in language: Interactions and architectures*. Oxford: Oxford University Press.
Rizzi, L. 1990. *Relativized Minimality*. Cambridge, MA: MIT Press.
Rizzi, L. 2004. Locality and the left periphery. In *Structures and beyond*, ed. A. Belletti, 223–251. Oxford: Oxford University Press.
Rizzi, L. 2006. On the form of chains: Criterial positions and ECP effects. In Wh-*movement*: *Moving on*, ed. L. Cheng & N. Corver, 97–134. Cambridge, MA: MIT Press.
Rizzi, L. 2011. Minimality. In *The Oxford handbook of linguistic minimalism*, ed. C. Boeckx, 220–238. Oxford: Oxford University Press.
Rizzi, L. 2017. Comparing extractions from wh-islands and superiority effects. In *Festschrift für Martin Prinzhorn*, ed. C. Mayr & E. Williams, 253–261, *Wiener Linguistische Gazette* 82.
Roberts, I. G. 2010. *Agreement and head movement: Clitics, incorporation, and defective goals*. Cambridge, MA: MIT Press.
Ross, J. 1967. Constraints on variables in syntax. Doctoral thesis, Massachusetts Institute of Technology, Cambridge.
Rudin, C. 1988. On multiple questions and multiple wh front. *Natural Language and Linguistic Theory* 6. 4:445–501.
Schneider, W., A. Eschman & A. Zuccolotto. 2012. *E-Prime user's guide*. Pittsburgh: Psychology Software Tools.
Soare, G. 2009. The syntax–information structure interface: A comparative view from Romanian. Doctoral thesis, University of Geneva, Geneva.
Starke, M. 2001. Move dissolves into merge: A theory of locality. Doctoral thesis, University of Geneva, Geneva.
Villata, S. 2017. Intervention effects in sentence processing. Doctoral thesis, University of Geneva, Geneva.
Villata, S., L. Rizzi & J. Franck. 2016. Intervention effects and Relativized Minimality: New experimental evidence from graded judgments. *Lingua* 179:76–96.

Ur Shlonsky
Département de Linguistique
Faculté des Lettres
Université de Genève
5, rue De-Candolle
1211 Genève 4
Switzerland

ur.shlonsky@unige.ch

Sandra Villata
Department of Linguistics and Department of Psychological Sciences
University of Connecticut
365 Fairfield Way, Oak Hall, unit 1145
Storrs, CT 06269
USA

sandra.villata@uconn.edu

Julie Franck
Faculté de Psychologie et des Sciences de l'Éducation
Université de Genéve
Uni-Pignon
42, boulevard du Pont d'Arve
1205 Genève
Switzerland

julie.franck@unige.ch

Appendix

Here we report on a 2 × 2 linear mixed-effects model using lexical restriction of the extracted *wh* (Wh_1) and lexical restriction of the intervening *wh* (Wh_2) as fixed factors, for the sake of comparison with the data analyses in Hofmeister et al. 2007, 2013. Other than the different fixed-factor structure, the analyses reported here follow the same logic as those described in section 3.4.

Results from the French study revealed three things: a main effect of Wh_1, with the conditions involving a lexically restricted extractee being rated higher than those with a bare extractee ($M = 3.2$ vs. $M = 3.6$; $F(1, 41.6) = 11.324$, $p = .0016$); a main effect of Wh_2, with the conditions involving a lexically restricted intervener being rated higher than those with a bare intervener ($M = 2.88$ vs. $M = 3.91$; $F(1, 28.86) = 127.284$, $p < .001$); and a significant interaction ($F(1, 1{,}032.63) = 30.608$, $p < .001$). Subsequent models exploring the interaction revealed a significant effect of Wh_1 when Wh_2 is lexically restricted ($\beta = 0.2933$, $SE = 0.035$, $t = 8.317$, $p < .001$), but no effect when Wh_2 is bare ($\beta = 0.29$, $SE = 0.035$, $t = 8.317$, $p < .001$). Thus, complex identity was rated higher than inverse inclusion, but no difference was attested between bare identity and inclusion.