What vs. who and which: Kind-denoting fillers and the complexity of *whether*-islands*

Theodora Alexopoulou and Frank Keller
University of Cambridge and University of Edinburgh

Abstract

We present results from three acceptability judgement experiments investigating the effect of discourse linking (d-linking) and animacy on *whether*-islands and interactions with resumption in Greek and English. Based on Anagnostopoulou's referentiality hierarchy, we test the acceptability of four types of *wh*-phrases, *what*, *what X*, *which X*, *which of X* in a range of configurations (simple questions and questions involving extractions out of (non-island) *that*-clauses and *whether*-islands). We further test interactions between animacy and d-linking in English. Our results show that d-linking improves *whether*-islands in both Greek and English. However, d-linking does not alter the overall interactions: *whether*-islands remain mostly less acceptable than *that*-clauses. While acceptability increases overall as predicted by the referentiality hypothesis, we obtain two unexpected contrasts: (i) a contrast between *which X* and *what* phrases and (ii) an independent effect of animacy; *who* is better than *what*, on a par with *which X* phrases. These contrasts affect the acceptability of *whether*-islands but not *that*-clauses.

We propose that what sets *what* phrases apart, is the contrast between kind-denoting (*what*) and ordinary individuals (*which,who*), which can be triggered by d-linking or animacy. This denotational hypothesis predicts that the distinction is only relevant for scopal islands like *whether*-islands. The denotational contrast affects the processing complexity of *whether*-islands. Kind-denoting wh-fillers have higher integration costs (in the sense of Gibson's complexity model). The denotation of the filler interacts with its complexity (e.g. *who* vs. *which X*); together, they may improve the acceptability of *whether*-islands; however, they cannot cancel their overall complexity, as they cannot cancel

*We would like to thank John Hawkins, Philip Hofmeister, Napoleon Katsos and the editors of this volume for comments and suggestions.The first author acknowledges support by *Education First*.

John Sprouse and Norbert Hornstein, eds., *Experimental Syntax and Island Effects,* 310–340. Cambridge: Cambridge University Press, 2013.

the main scope island (i.e. the question intercepting the filler-gap dependency); thus, manipulations of the filler cannot restore the acceptability of *whether*-islands. Finally, a crucial overall conclusion is that alleviation of integration costs of the filler (through d-liking/animacy) has a stronger effect on improving *whether*-islands compared to cancellation of locality costs (through resumption).

1 Introduction

One important challenge in the study of islands is the gradient nature of their acceptability and the still poorly understood interaction of different factors that have been shown to affect acceptability. In this study we focus on the role of d-linking in object wh-questions in two syntactic configurations, questions involving extractions out of (non-island) *that*-clauses and extractions out of *whether*-islands. We further investigate interactions between d-linking, animacy and resumption in these configurations in Greek and English. We first introduce the main phenomena and questions underlying the rationale of our experiments in section 2. In section 3 we present the experiments and in 4 we summarise the results and offer our theoretical analysis.

2 Background: d-linking, weak islands and resumption

2.1 D-linking and weak islands

Whether-islands are considered the par excellence weak or selective islands, in the sense that they selectively allow extractors to escape if they satisfy certain properties. What's the best way to characterise the properties of successful/escaping extractors has been a matter of intensive investigation in the syntactic and semantic literature over the last few decades. It was initially thought that the critical property is argumenthood as opposed to adjuncthood as illustrated in (1) (Huang 982a; Lasnik and Saito 1984; Chomsky 1986).

(1) a. Which problem did John ask how to phrase?
b. *How did John ask which problem to phrase?

(From Szabolcsi 2006, ex.82,83)

This generalisation was later substituted by referentiality or discourse linking (3) since non-referential arguments may not escape islands (2a) while referential/d-linked adjuncts can (2b) (see Szabolcsi 2006 and references therein).

(2) a. *What did John ask whether these pearls cost? (From Szabolcsi 2006, ex.87 attributed to Ross)
b. *How many books are you wondering whether to write next year?
c. How many books on the list are they wondering whether to publish next year? (From Szabolcsi 2006, ex.90,91)

With the emergence of a semantic scopal theory of weak islands (Szabolcsi and Zwarts 1993) the more pragmatic notion of discourse linking was replaced by the requirement that good extractors denote individuals. Thus, it is not d-linking *per se* that accounts for the contrast between (2b) and (2c). Rather, d-linking brings about the individuation of a domain so that *how many books* denotes an amount in (2b) but quantification over individuals in (2c). So the critical contrast is one between amounts/manners vs. individuals; d-linking triggers this contrast.

At the same time, a somewhat different notion of d-linking has been employed in the literature mostly drawing on the contrast between bare *wh*-phrases like *what* and *who* vs. *wh*-phrases introduced by *which* followed by an explicit lexical restriction.[1] The contrast has also been discussed for weak islands as (3a), claimed to be more acceptable than (3b) (as observed by Maling and Zaenen 1982, cited in Hofmeister and Sag 2010).

(3) a. Which article don't you remember who wrote?
b. What don't you know who wrote?

Hofmeister and Sag (2010) offer a systematic experimental investigation of the effect of d-linking on *whether*-islands by comparing acceptability judgements with reaction times in a self-paced reading task. Subjects read a declarative sentence like (4a) and then one comprehension question like the ones in (4b-d), varying the type of the filler, bare or which-X, and the type of embedded clause, *whether*-island (4b-c) and *that*-clause (4d).

(4) a. CONTEXT: Albert learned that the managers dismissed the employee with poor sales after the annual performance review.
b. BARE: Who did Albert learn whether they dismissed after the annual performance review?

[1]The contrast has been very dominant in theoretical discussions of multiple constituent questions going back to Karttunen (1977) and later Pesetsky (1987), (2000), Comorovski (989a),Cinque (1990),Rizzi (1990) among others. The interaction of such d-linking and superiority violations in multiple constituent questions has also been established experimentally (Featherston 2005, 2003, Meyer 2003; Hofmeister, Jaeger, Sag, Arnon, and Snider 2007).

c. WHICH: Which employee did Albert learn whether they dismissed after the annual performance review?

d. BASELINE: Who did Albert learn that they dismissed after the annual performance review?

(From Hofmeister and Sag 2010, ex.49)

They find that the WHICH condition is read faster than the BARE condition; crucially, the WHICH condition is read as fast as the BASELINE condition. In an acceptability task involving similar items[2] the WHICH condition was judged more acceptable than the BARE condition.

They take their results as evidence favouring a processing analysis of island constraints. Following Kluender (1992), (1998) and Kluender and Kutas (993b), they view islands as structures that strain the resources of the parser; the gradient acceptability effects are a reflection of the interaction between different factors and the resource limitations of the parser. Under Kluender's approach, discourse salience can be viewed as a way to increase the activation levels of the relevant discourse referent so that a d-linked filler may be more easily recovered when the gap is encountered than a non-d-linked one.[3] The underlying idea is closely linked to the notion of *forward accessibility* proposed by Ariel (1990, 1999), who argues that more discourse salient entities are more accessible at later stages of structural resolution. Hofmeister, Jaeger, Sag, Arnon, and Snider (2007) apply such ideas to the interaction of d-linking with multiple constituent questions.

However, Hofmeister and Sag (2010) go a step further and reject the pragmatic explanation as the critical factor explaining the processing advantage of d-linked fillers. Consider (4) above. Presumably, the set of employees is equally salient for the BARE as well as the WHICH condition in (4). But this contextual salience does not eliminate the advantage of the *which-X* phrase over bare *who*. They counterpropose that the critical property is the mere complexity of the filler. Structurally more complex fillers (by encoding richer semantic and syntactic information) establish highly activated antecedents and facilitate the resolution of the filler-gap dependency; in particular, complexity can facilitate the retrieval of the filler from memory by providing richer semantic and syntactic representations that can discriminate the target (Hofmeister 2011). The strongest evidence for their analysis comes from an additional experiment where questions involving extraction of "simple" adjuncts like (5a) were compared to questions with more

[2] In the acceptability task questions like (4c) were embedded as in *Only a few individuals repeated which employee Albert learned whether we dismissed after the annual performance evaluations.*

[3] On this, Kluender builds on Just and Carpenter (1992).

complex filler adjuncts like *for what period of time* as in (5b). Again, the sentences with the structurally complex fillers were read faster than the structures with the simpler fillers; and again, (5b) was read as fast as the baseline condition in (5c).[4] Such results indicate that it is complexity, independently of referentiality or discourse salience that leads to a facilitation of processing; *for what period of time* still denotes an amount rather than quantification over individual "periods of time", so that (5a) cannot be said to be "referential" in the way (2c) is.

(5) a. CONTEXT: Julie discerned that the survivor had managed to stay alive for eight days after the crash in harsh conditions.
b. SIMPLE: How long did Julie observe whether the passenger had survived in the unbelievably harsh conditions?
c. COMPLEX: For what period of time did Julie observe whether the passenger had survived in the unbelievably harsh conditions?
d. BASELINE: How long did Julie observe that the passenger had survived in the unbelievably harsh conditions?

2.2 Resumption, islands and embedding

Resumption has long been claimed to "save" island violations (Ross 1967; Kroch 1981) when a resumptive pronominal appears in place of an otherwise illicit gap as in (6). Example (6a) is a spontaneously produced example while (6b) type of examples were elicited experimentally by Ferreira and Swets (2005).

(6) a. We are afraid of *things* that we don't know what *they* are.
b. This is *the donkey* that I don't know where *it* lives.

(Ferreira and Swets 2005)

At the same time resumptive pronouns are more acceptable the more deeply embedded the gap is from the filler (Erteschik-Shir 1992; Tsimpli 1999; Dickey 1996; Alexopoulou and Keller 2007).

Despite the evidence from production, evidence from acceptability judgement experiments indicates that resumption fails to "save" island violations. In Alexopoulou and Keller (2007) we investigated the role of resumption in a range of syntactic configurations (non-islands, weak islands and strong islands) and for multiple levels of embedding (single, double and triple), in English, Greek and German. Resumption did not improve any of the structures or "save" any of the island vi-

[4]The critical contrasts in reaction times relate to specific segments of the string; it should be noted though that overall the effects in this experiment are weaker than the earlier experiment involving contrasts between bare and which *wh*-phrases.

olating ones. This picture was further confirmed by Xiang, Heestand, and Polinsky (2008) and Heestand, Xiang, and Polinsky (2011) who investigated the acceptability of resumption in a variety of islands in English and with varied experimental methodology (e.g. speeded judgments). Perhaps most characteristically, the very same subjects that produced examples like (6b) in the study of Ferreira and Swets (2005) rejected such sentences in the acceptability task.

Even though resumption fails to save islands, resumption does interact with islands and embedding; in particular, resumptives embedded in questions involving extractions from *that*-clauses as in (7b) and *whether*-islands as in (7c) were found significantly more acceptable than resumptives in simple questions like (7a) (Alexopoulou and Keller 2007). In Greek and German such resumptives were as acceptable as gaps. In other words, while resumption does not save islands, it does reverse the effect of embedding under a *that*-clause and the effect of *whether*-island observed in gap extractions. This is in line with online studies indicating a facilitating effect for embedded resumptives (Dickey 1996; Hofmeister and Norcliffe 2011).

(7) a. pion tha ton apolisume
who-ACC will him fire-1PL
Who will we fire?

b. Pion nomizi o Petros oti ishirizete
who-ACC think-3SG the-NOM Petros-NOM that claim-3SG
i Ana oti tha ton apolisume
the-NOM Ana that will him fire-1PL
Who does Petros think that Anna claims that we will fire?

c. Pion nomizi o Petros oti
who-ACC think-3SG the-NOM Petros-NOM that
anarotiete i Maria an tha ton apolisume
wonder-3SG the-NOM Ana if will him fire-1PL
Who does Petros think that Maria wonders whether we will fire?

2.3 Resumption and d-linking

D-linking has been argued to interact with both *intrusive* resumption, i.e. resumption used in place of an island violating gap (Sells 1984,1987) and with "grammatical" resumption, i.e. resumption that is freely available in certain structures, as for instance Greek Clitic Left Dislocation as discussed below.

Greek allows Clitic Left Dislocation (CLLD) as in (8a), where a typically referential topic appears preverbally and is linked to a clitic pronominal inside the clause (Anagnostopoulou 1994). Non d-linked *wh*-phrases are generally excluded from CLLD (8b); but acceptability is

claimed to improve with d-linking (8c) (Anagnostopoulou 1994; Cinque 1990; Rizzi 1997; Dobrovie-Sorin 1990).

(8) a. to Yani ton petihame sto
the-ACC Yanni-ACC him.CL.MSC.ACC met-1PL at-the
sinema tis proales
cinema the other-day
Yanis we met at the cinema the other day.

b. *pion ton petihate sto
who-ACC.MSC.SG him.CL.MSC.ACC met-2PL at-the
sinema tis proales
cinema the other-day?
Who did you meet (*him) at the cinema the other day?

c. ?pion fititi ton
who-ACC.MSC.SG student-ACC.MSC.SG him.CL.MSC.ACC
petihate sto sinema tis proales
met-2PL at-the cinema the other-day?
Which student did you meet (?him) at the cinema the other day?

Extending Pesetsky's notion of d-linking, Anagnostopoulou (1994) proposes that the acceptability of CLLD-ed *wh*-phrases increases according to the referentiality hierarchy below.

Anagnostopoulou's (1994) referential hierarchy Overt partitive wh-phrases (*which of your books*) $\prec$ Which-phrases (*which book*) $\prec$ What-phrases (*what book*) $\prec$ Bare wh-phrases (*who, what*)

The general intuition behind the hierarchy is clear, and, indeed Greek linguists agree that the acceptability of the pronominal in constituent questions increases as predicted by this hierarchy. However, it is not clear that the discriminating factor is always unambiguously referentiality. For instance, (9a) asks about kinds of books while (9b) about human individuals. According to the hierarchy above, (9a) is more referential than (9b). But it is not necessarily obvious why (9a), which restricts the possible answer to kinds of books, is more referential than (9b), which restricts the possible answers to people (possibly the set of men already shortlisted for the prize).

(9) a. What books make it to the Man Booker Prize shortlist each year?

b. Who will win the Man Booker Prize this year?

In other words, the contrast between bare wh-items like *who* and what-X phrases, if confirmed, is not necessarily one of referentiality;

the two types of *wh*-phrases have distinct denotations. We return to this point when we discuss our results.

At the same time there is evidence that d-linking interacts with intrusive resumption. In particular, Frazier and Clifton (2002) provide evidence from judgement experiments showing that *whether*-islands with resumptive pronouns as in (10), receive higher acceptability scores when the *wh*-phrase is d-linked (10a) than when it is not (10b).

(10) a. (*) Which students did the teacher wonder if they had gone to the library?
b. (*) Who did the teacher wonder if they had gone to the library?

(From Frazier and Clifton 2002, ex.15)

The results for intrusive pronominals as in (10), are on a par with results from intrasentential anaphora, where pronouns are shown to prefer d-linked antecedents over non d-linked ones (Frazier and Clifton 2002). However, since d-linking is claimed to improve *whether*-islands, it is not clear that (10a) is more acceptable than (10b) because of an interaction between d-linking and resumption. It could be an interaction between the *whether*-island and d-linking.

3 Experiments

3.1 General rationale and aims

The main aim of the studies presented here was to follow-up our previous experiments with an investigation of the effect of d-linking. As discussed earlier, resumption was shown to reverse the island effect but unable to restore the weak island violation to full acceptability. Our first aim is to measure the effect of d-linking in remedying weak island violations and establish whether, unlike resumption, d-linking may indeed restore violating structures to full acceptability. Further, whether a combined effect of d-linking and resumption may turn out to be cumulative and, perhaps, lead to restoring island violations to full acceptability.

A second aim is to understand interactions between d-linking and resumption. As discussed in section 2.3, d-linking has been argued to improve the acceptability of CLLD-ed *wh*-phrases in Greek. Experiment 1 below tests experimentally the validity of this claim and seeks to establish the magnitude of the d-linking effect on simple consituent resumptive Greek questions. The comparison of Greek with English in Experiment 2 seeks to clarify the nature of crosslinguistic variation. In Alexopoulou and Keller (2007) we found that resumption in simple, non-embedded questions was more acceptable in Greek than in English.

We explained this by assuming that resumptive wh-questions in Greek are cases of bad CLLD, i.e. cases where a quantificational antecedent heads a structure requiring a referential phrase. English constituent questions cannot be analysed as cases of CLLD and are, thus, of lower overall acceptability than their Greek counterparts and should not interact with d-linking.

On the other hand, intrusive resumptives may show sensitivity to d-linking in both English and Greek. Intrusive resumption has been shown to have properties of intrasentential anaphora (Sells 1984). Since intrasentential anaphora is sensitive to d-linking (Frazier and Clifton 2002), we expect intrusive resumption to also show sensitivity to d-linking.

Our final aim was to investigate the nature of d-linking *per se*. As discussed in section 2.1, there is still no consensus on whether the critical property underlying the effect of d-linking relates to a pragmatic distinction or to the structural complexity of the relevant phrases. In the experiments that follow we have adopted Anagnostopoulou's referentiality hierarchy. By indicating four distinct levels of d-linking, Anagnostopoulou's hierarchy allows us to investigate the gradient effect of d-linking across the hierarchy. It is worth comparing the predictions of the referentiality hypothesis with the complexity view. Hofmeister (2011) proposes an operational definition of complexity, according to which "For two descriptions x_1 and x_2 denoting a discourse entity e, if the semantic and syntactic feature-value pairs encoded by x_2 are a proper subset of the feature-value pairs encoded by x_1, then x_1 is more complex than x_2" (ibid,p.6). Under this definition a bare *wh*-phrase like *who* is less complex than a non-bare *which X* phrase on the assumption that the features encoded by the bare *wh*-phrase are a subset of the features encoded by a d-linked non-bare *wh*-phrase. However, the complexity hypothesis has nothing to say about the distinction between *which* and *what* phrases, since, as we will see shortly, these two phrases have distinct denotations, *which X* normally picking individuals while *what X* normally picking kinds.

The present studies aim to quantify the magnitude of the contrasts between adjacent points of the hierarchy, and therefore evaluate whether there is a gradual effect of referentiality/complexity and whether the semantic distinction between kind and non-kind denoting *wh*-phrases induces an separate effect. Preempting the discussion to come, we will see that this last contrast is the sharpest in our data, indeed supporting a *denotational* hypothesis to explain at least part of the d-linking/referentiality hierarchy.

In subsequent sections we present two acceptability judgement experiments investigating the interactions between d-linking and resumption in *whether*-islands in Greek and English. The English experiment is followed up by a second study investigating animacy and its inter-

action with d-linking.

3.2 *Wh*-Phrases in Greek

Experiment 1 investigated the acceptability of four different types of *wh*-phrases according to Anagnostopoulou's referentiality hierarchy; the *wh*-phrases are shown in (11).

(11) a. **bare wh** ti
what

b. **what X** ti tenia
what film.FEM.SG

c. **which X** pia tenia
who.FEM.SG film.FEM.SG

d. **which of X** pia apo tis
who.FEM.SG from the-FEM.PL.ACC
tenies
films.FEM.PL
which of the films

A few points on the morphosyntax and interpretation of the Greek *wh*-phrases: Greek *ti* (=*what*) does not inflect for gender, number or case, unlike *pios* (=*who*) which inflects for gender, number and case as indicated in the glosses in (11).[5] Greek does not distinguish between *who* and *which*; there is one interrogative pronoun, *pios* which inflects for gender, number and case like an adjective and may accompany an animate or inanimate noun.

Ti X usually refers to kinds and its felicity conditions are distinct from *pios X* which ranges over a set of (ordinary) individuals. Thus,

[5] *Ti* seems to have default singular neuter morphology; however, it does not agree on any of these features with accompanying nouns as indicated by the grammaticality of (i); in (ia) it can happily accompany singulars in all three genders while in (ib) the same nouns are in plural.

(i) a. ti tipos/ti tipo, ti yineka, ti
what type-MSC.SG.NOM/what type.MSC.SG.ACC, what woman.FEM.SG, what
aftokinito
car.NEUT.SG
what guy, what woman, what car

b. ti tipi/ti tipus, ti yinekes, ti
what guys.MSC.PL.NOM/what guys.MSC.PL.ACC, what women.FEM.PL, what
aftokinita
cars.NEUT.PL
what guys, what women, what cars

(12c), which gives a title as an answer to (12a) is an infelicitous answer for (12a) but a fine one for (13a). On the other hand, (12b), which gives a type of film, *comedy*, as an answer is a fine answer for (12a), but an infelicitous one for (13a).

(12) a. Q: ti tenia kerdhise to Oscar?
what film.FEM.SG won the Oscar?
What film won the Oscar?
b. A: komodia
comedy
A comedy.
c. A:% i omilia tu vasilia
the.NOM speech the-GEN king.GEN
A: % The King's speech.

(13) a. Q: pia tenia kerdhise to Oscar?
who-FEM.SG film.FEM.SG won the Oscar?
Which film won the Oscar?
b. A:% komodia
comedy
% A comedy.
c. A: i omilia tu vasilia
the.NOM speech the-GEN king.GEN
The King's speech.

Such data indicate a denotational contrast which appears orthogonal to the referentiality or complexity dimension. Note though that questions with *what* can elicit non-kind-denoting answers as in (14).

(14) a. Q: ti tha dhite sto sinema?
what will see-2PL at-the cinema?
What will you watch at the cinema?
b. A: tin omilia tu vasilia
the.ACC speech.FEM.SG the-GEN king-GEN
The king's speech.

3.3 Experiment 1: D-linking and resumption in *whether*-islands in Greek

This experiment tested the four different types of *wh*-phrase in (11) in questions involving extractions from two different types of embedded environments: a *that*-clause and a *whether*-island. Each type of question was presented in two variants, one with a gap and one with a resumptive pronominal. Simple questions, without embedding, were included as control cases. Sample stimuli are given in (15).

(15) **zero-embedding**

a. **Bare ti**: ti tha to/Ø dhun sto sinema?
what will it/Ø see-3PL at-the cinema?
What will they watch (it) at the cinema?

b. **Ti-X**: ti tenia tha ti/Ø dhun sto
what film.FEM.SG will her/Ø see-3PL at-the
sinema?
cinema?
What film will they watch (it) at the cinema?

c. **Pio-X**: pia tenia tha ti/Ø dhun
wh.FEM.SG film.FEM.SG will her/Ø see-3PL
sto sinema?
at-the cinema?
Which film will they watch at the cinema?

d. **Pio-of-X**: pia apo tis
wh.FEM.SG from the.FEM.PL.ACC
tenies tha ti/Ø dhun sto sinema?
films.FEM.PL will her/Ø see-3PL at-the cinema?
Which of the films will they watch at the cinema?

(16) **that-clause**

a. **Bare ti**: ti nomizi o Petros oti tha
what think.3SG the.NOM Petros.NOM that will
to/Ø dhun sto sinema?
it/Ø see-3PL at-the cinema?
What does Petros think they will watch (it) at the cinema?

b. **Ti-X**: ti tenia nomizi i Eleni
what film.FEM.SG think.3SG the.NOM Eleni.NOM
oti tha ti/Ø dhun sto sinema?
that will her/Ø see-3PL at-the cinema?
What film does Eleni think they will watch (it) at the cinema?

c. **Pio-X**: pia tenia nomizi o
wh.FEM.SG film.FEM.SG think.3SG the.NOM
Takis oti tha ti/Ø dhun sto sinema?
Takis.NOM that will her/Ø see-3PL at-the cinema?
Which film does Takis think they will watch at the cinema?

d. **Pio-of-X**: pia apo tis
wh.FEM.SG from the.FEM.PL.ACC
tenies nomizi i Sofia oti tha
films.FEM.PL think.3SG the.NOM Sophia.NOM that will

ti/Ø dhun sto sinema?
her/Ø see-3PL at-the cinema?
Which of the films does Sophia think they will watch at the cinema?

(17) **whether-clause**

a. **Bare ti**: ti anarotiete o Petros an tha
what wonder.3SG the.NOM Petros.NOM if will
to/Ø dhun sto sinema?
it/Ø see-3PL at-the cinema?
What does Petros wonder whether they will watch (it) at the cinema?

b. **Ti-X**: ti tenia anarotietai i
what film.FEM.SG wonder.3SG the.NOM
Eleni an tha ti/Ø dhun sto sinema?
Eleni.NOM if will her/Ø see-3PL at-the cinema?
What film does Eleni wonder they will watch (it) at the cinema?

c. **Pio-X**: pia tenia anarotiete o
wh.FEM.SG film.FEM.SG wonder.3SG the.NOM
Takis an tha ti/Ø dhun sto sinema?
Takis.NOM if will her/Ø see-3PL at-the cinema?
Which film does Takis wonder whether they will watch at the cinema?

d. **Pio-of-X**: pia apo tis
wh.FEM.SG from the.FEM.PL.ACC
tenies nomizi i Sofia oti tha
films.FEM.PL think.3SG the.NOM Sophia.NOM that will
ti/Ø dhun sto sinema?
her/Ø see-3PL at-the cinema?
Which of the films does Sophia wonder whether they will watch at the cinema?

3.3.1 Method

Twenty-nine participants were recruited over the Internet by postings to newsgroups and mailing lists. All participants were self-reported native speakers of Greek. Linguists and students of linguistics were excluded from the sample.

The experiment comprised two subdesigns. The first one tested *whether*-clauses and crossed the factors RESUMPTIVE (gap or pronoun), and WH-PHRASE (bare *ti*, *ti-X*, *pio-X*, *pio apo X*). The second subdesign tested *that*-clauses and compared them against an unembedded baseline condition. It crossed the factors EMBEDDING (no embedding, one level of embedding), resumption (gap, pronoun), and WH-PHRASE.

	what X	which X	which of X
what		(*)	**
what X			(**)
which X			

*$p < .05$; **$p < .01$; (): by participants only; []: by items only

Table 1: Result of Tukey test for the main effect of WH-PHRASE in Experiment 1 (*whether*-clauses)

Both subdesigns together had 24 cells. Eight lexicalisations were used for each cell, yielding a total of 192 stimuli. All stimuli involved inanimate *wh*-phrases. The stimulus set was divided into eight subsets of 24 stimuli by placing the items in a Latin square. A set of 24 fillers was used, covering the whole acceptability range.

The method used was magnitude estimation as proposed by Stevens (1975) for psychophysics and extended to linguistic stimuli by Bard, Robertson, and Sorace (1996) and Cowart (1997). Participants first saw a set of instructions that explained the concept of numerical magnitude estimation using line length. Participants were instructed to make length estimates relative to the first line they would see, the reference line. They were told to give the reference line an arbitrary number, and then assign a number to each following line so that it represented how long the line was in proportion to the reference line. Several example lines and corresponding numerical estimates were provided to illustrate the concept of proportionality. Then participants were told that linguistic acceptability could be judged in the same way as line length, i.e., by comparing the acceptability of a target sentence to that of a reference sentence. The task was illustrated by examples. The experiment was conducted over the web using WebExp (Keller, Gunasekharan, Mayo, and Corley 2009), an interactive software package for web-based psycholinguistic experimentation.

3.3.2 Results

The data were normalized by dividing each numeric judgment by the value that the subject had assigned to the reference sentence. This operation creates a common scale for all participants. Then the data were transformed by taking the decadic logarithm. This transformation ensures that the judgments are normally distributed and is standard practice for ME data (Bard, Robertson, and Sorace 1996). All analyses and figures are based on normalized, log-transformed judgments. Figure 1 graphs the mean judgments for both *whether*-clauses and *that*-clauses.

An ANOVA for subdesign 1 (*whether*-clauses) showed that sen-

	what X	which X	which of X
what		(**)	
what X		(**)	(**)
which X			

*$p < .05$; **$p < .01$; (): by participants only; []: by items only

Table 2: Result of Tukey test for the main effect of WH-PHRASE in Experiment 1 (*that*-clauses)

tences without resumptives were more acceptable than sentences with resumptives (main effect of RESUMPTION, significant by participants only, $F_1(1, 28) = 8.461$, $p = .007$; $F_2(1, 7) = 2.764$, $p = .140$). The type of *wh*-phrase in the sentence also had an effect on acceptability, with bare *it* being least acceptable and *pio apo ti* being most acceptable (main effect of WH-PHRASE, $F_1(3, 84) = 9.591$, $p < .001$; $F_2(3, 21) = 5.664$, $p = .005$). A post-hoc Tukey test was conducted to further investigate which of the *wh*-phrases differed significantly in acceptability (see Table 1). There was no significant interaction of RESUMPTION and WH-PHRASE.

An ANOVA for subdesign 2 (*that*-clauses and non-embedded baseline) confirmed that sentences without resumptives were more acceptable than sentences with resumptives (main effect of RESUMPTION, $F_1(1, 28) = 57.997$, $p < .001$; $F_2(1, 7) = 67.563$, $p < .001$). The type of *wh*-phrase also affected acceptability (main effect of WH-PHRASE, significant by participants only, $F_1(3, 84) = 8.210$, $p < .001$; $F_2(3, 21) = 2.797$, $p = .065$). Table 2 lists the results of a Tukey test to determine which *wh*-phrases differ significantly in acceptability. Whether sentences contained embedding or not did not significantly affect acceptability (no main effect of EMBEDDING).

There was a significant interaction of EMBEDDING and RESUMPTION ($F_1(1, 28) = 18.440$, $p < .001$; $F_2(1, 7) = 38.884$, $p < .001$): resumption leads to a larger reduction in acceptability in non-embedded sentences compared to the embedded sentences. This interaction is explored further by the Tukey test results in Table 3. Furthermore, we found that an interaction of RESUMPTION and WH-PHRASE (significant by participants only, $F_1(3, 84) = 5.859$, $p = .001$; $F_2(3, 21) = 1.514$, $p = .240$): the effect of resumption is larger for some *wh*-phrases than for other; again, a Tukey test was conducted to establish for which *wh*-phrases this holds (see Table 4). The other interactions failed to reach significance.

	no emb, res	emb, no res	emb, res
no emb, no res	**	[**]	**
no emb, res		**	**
emb, no res			**

*$p < .05$; **$p < .01$; (): by participants only; []: by items only

Table 3: Result of Tukey test for the interaction of EMBEDDING and RESUMPTION in Experiment 1 (*that*-clauses)

	no res, what X	no res, which X	no res, which of X	res, what	res, what X	res, which X	res, which of X
no res, what				**	**	(**)	(**)
no res, what X				*	*	(**)	(**)
no res, which X				**	**	(**)	(**)
no res, which of X				**	*	(**)	(**)
res, what						(**)	(**)
res, what X						(**)	(**)
res, which X							

*$p < .05$; **$p < .01$; (): by participants only; []: by items only

Table 4: Result of Tukey test for the interaction of RESUMPTION and WH-PHRASE in Experiment 1 (*that*-clauses)

3.3.3 Discussion

As predicted by the literature, d-linking improves the weak island violation and CLLD-ed *wh*-phrases. In these interactions, the validity of Anagnostopoulou's referentiality hierarchy is *prima facie* confirmed; the lowest point in this hierarchy, bare *ti*, is always worse than the highest point of the hierarchy, the explicit partitive. However, while acceptability increases in the direction of the hierarchy, the difference between intermediate points is not consistent. The sharpest contrast between two adjacent *wh*-phrases in the hierarchy is provided by cases of embedded pronominals in *that*-clauses: *pio-X* is better than *ti-X* and bare *ti* (18a). This is the only statistically significant difference (between adjacent points: *which of X* is also better than *ti (X)*.) The data therefore indicate a mild effect of referentiality but a stronger effect between *pio* and *ti* phrases, that is the contrast between individual denoting *wh*-phrases and kind-denoting *wh*-phrases.

(18) a. **That-clauses**
pio X $\prec$ **ti**, **ti X**

b. **Whether-clauses**
pio X ≺ ti

In *whether*-islands *pio-X* is significantly better than bare *ti*, but not *ti X*. Note that the pair *bare ti* and *ti-X* are of the same acceptability in both *that*-clauses and *whether*-islands. Why this weaker effect with *ti X*? As discussed earlier, *ti-X* phrases tend to denote kinds, but may also admit ordinary individuals. It is possible that, at least with some of our items, the main verb primed the set the *wh*-phrase ranges over yielding a more referential reading of bare *ti*, thus, weakening the contrast between *ti* and *ti X*, as well as the contrast between the *ti* conditions and *pio X*. For instance, in an example like (15b) the answer ranges over possible film titles even though the word *tenia(=film)* is not mentioned. This may also explain why the strongest contrast arises in the case of embedded pronominals: the verb arrives much later in structures like (17c); in the absence of information from the verb, the default kind-denoting interpretation of *ti* and *ti X* contrasts with the *pio X* in that the latter is a more acceptable pronoun antecedent.

Intrusive and non-intrusive resumptives improve as the *wh*-phrase becomes more referential on Anagnostopoulou's hierarchy. However, the acceptability of CLLD-ed *wh*-phrases remains overall low; in fact, CLLD-ed *wh*-phrases in simple questions are worse than the island violating structures. What improves CLLD-ed *wh*-phrases is embedding, either within a *that*-clause or a *whether*-island; interestingly, embedding improves resumption more than d-linking.

Finally d-linking improves *whether*-islands; it could be argued that d-linking restores the weak-island to full acceptability, since the most d-linked condition appears as acceptable as the non-island condition, i.e. extraction out of *that*-clause. However, there is no cumulative effect between d-linking and resumption. As in previous studies, resumptives are at best as acceptable as gaps; but note that this is so because the acceptability of gaps drops with embedding while the acceptability of resumptives improves with embedding; so while resumption reverses the island effect, it cannot cancel it.

An important overall conclusion is that, while d-linking does interact with the *whether*-island and with resumption, it does not change the overall interactions; weak island violations remain less acceptable that the non-island conditions, with the exception of the explicit partitive which approaches the acceptability of *that*-clauses.

3.4 Experiment 2: D-linking and resumption in *whether*-islands in English

This experiment replicated the Greek study for English. As discussed in section 2.1, d-linking has been argued to improve weak island viola-

tions (Maling and Zaenen 1982; Hofmeister and Sag 2010; Frazier and Clifton 2002) and also to interact with intrusive resumption (Frazier and Clifton 2002). Again based on Anagnostopoulou's referentiality hierarchy we tested the following four types of *wh*-phrases:

(19) a. **Bare what**: what
b. **What X**: what movie
c. **Which X**: which movie
d. **Which of X**: which of the movies

The first question is whether d-linking will be shown to improve *whether*-islands as in Greek. Second, whether and how d-linking interacts with resumption. As pointed out in section 2.2, English lacks any resumption of the type available in Greek CLLD, while effects related to intrusive resumption in English are of a smaller magnitude than Greek. We therefore don't expect any interaction between resumption and d-linking in simple questions but we expect intrusive resumption to interact with d-linking, though the overall effects should be weaker than Greek. Finally, by comparing the effect of d-linking in *whether*-islands with gaps and pronouns, it is possible to evaluate whether the results of Frazier and Clifton (2002) re examples like (10) are due to an interaction between d-linking and *whether*-islands or between d-linking and resumption.

The design was as in the Greek study; sample stimuli are given below.[6]

(20) **zero-embedding**
a. **Bare what**: What will they watch it/Øat the cinema?
b. **What X**: What movie will they watch it/Øat the cinema?
c. **Which X**: Which movie will they watch it/Øat the cinema?
d. **Which of X**: Which of the movies will they watch it/Øat the cinema?

(21) **that-clause**
a. **Bare what**: What does Mary think they will watch it/Øat the cinema?
b. **What X**: What movie does Sue think they will watch it/Øat the cinema?
c. **Which X**: Which movie does John think they will watch it/Øat the cinema?
d. **Which of X**: Which of the movies does Maggie think they will watch it/Øat the cinema?

[6]We thank Jim Blevins for going over the English stimuli to check their plausibility.

	what X	which X	which of X
what		(*)	
what X			
which X			

*$p < .05$; **$p < .01$; (): by participants only; []: by items only

Table 5: Result of Tukey test for the main effect of WH-PHRASE in Experiment 2 (*whether*-clauses)

(22) **whether-clause**

a. **Bare what**: What does Jean wonder whether they will watch it/Øat the cinema?

b. **What X**: What movie does Clare wonder whether they will watch it/Øat the cinema?

c. **Which X**: Which movie does Rachel wonder whether they will watch it/Øat the cinema?

d. **Which of X**: Which of the movies does Emily wonder whether they will watch it/Øat the cinema?

3.4.1 Method

Twenty-two participants were recruited over the Internet by postings to newsgroups and mailing lists. All participants were self-reported native speakers of English. Linguists and students of linguistics were excluded from the sample.

The design of the experiment mirrored that of Experiment 1, again comprising two subdesigns. The first one tested *whether*-clauses and crossed the factors RESUMPTIVE (gap or pronoun), and WH-PHRASE (bare *what*, *what X*, *which X*, *which of X*). The second subdesign tested *that*-clauses and compared them against an unembedded baseline condition. It crossed the factors EMBEDDING (no embedding, one level of embedding), resumption (gap, pronoun), and WH-PHRASE. The rest of the design and the experimental procedure were identical to those of Experiment 1.

3.4.2 Results

The data post-processed and analyzed as in Experiment 1. Figure 2 graphs the mean judgments for both *whether*-clauses and *that*-clauses.

An ANOVA for subdesign 1 (*whether*-clauses) demonstrated that sentences without resumptives were more acceptable than sentences with resumptives (main effect of RESUMPTION, $F_1(1, 21) = 9.712$, $p = .005$; $F_2(1, 7) = 19.053$, $p = .003$). We also found that *wh*-phrases varied in acceptability (main effect of WH-PHRASE, significant

by participants only, $F_1(3, 84) = 2.755$, $p = .050$; $F_2(3, 21) = 0.579$, $p = .635$). To establish which *wh*-phrases were significantly different from each other in acceptability, we conducted a post-hoc Tukey test, the results of which are given in Table 5. There was no significant interaction of RESUMPTION and WH-PHRASE.

An ANOVA for subdesign 2 (*that*-clauses and non-embedded baseline) confirmed that sentences without resumptives were more acceptable than sentences with resumptives (main effect of RESUMPTION, $F_1(1, 21) = 32.202$, $p < .001$; $F_2(1, 7) = 173.608$, $p < .001$). The other main effects and interactions failed to reach significance in this subdesign.

3.5 Discussion

The strongest effects in this study involve resumption and the *whether*-island. Both conditions induce a significant drop in acceptability; but the *whether*-islands are more acceptable than resumptive structures. D-linking improves *whether*-islands but does not improve resumption.

As in Greek, the referentiality hierarchy is confirmed in terms of the overall direction of the acceptability of the different types of *wh*-phrases. As in Greek *whether*-islands, the only significant difference is between *which X* and *what*, with *what X* being neither better than *what X* nor worse than *which X*. The crosslinguistic similarity is remarkable, suggesting a consistent contrast between *which* and *what* phrases. Interestingly, the morphological make-up of the *wh*-phrases in the two languages does not seem to have an effect. Greek *pio* is morphologically richer than English *which* and could be argued to be more complex. The crosslinguistic consistency of the contrast and the absence of a morphological effect provide further evidence for the interpretative nature of the contrast, i.e. the distinction between kind and individual denoting *wh*-phrases.[7]

Finally, d-linking improves *whether*-islands but does not interact with resumption. This suggests that the contrast in (10) found by Frazier and Clifton (2002) is due to the effect of d-linking on the *whether*-island rather than due to resumption.

[7] A reviewer suggests that the lower acceptability of the *what* condition in *whether*-islands could be due to a garden path as subjects may interpret *what* as the object of the matrix verb *wonder*. If this were so though, we should be getting a contrast between *what X* and *what* since the lexical restriction excludes the matrix object reading. The absence of a contrast between the two types of *what* phrases means that the lower acceptability of these phrases cannot be attributed to a garden path effect. Note also that this explanation could not carry over to the Greek data. As we saw, there is an effect of *wh*-phrase on simple resumptive questions (Fig. 1b) where *what* is unambiguously the matrix object and still receives the lowest acceptability score.

As in Greek, d-linking does not change the shape of the main interactions obtained in previous experiments. In particular, resumptives always remain less acceptable than gaps (with the exception of the explicit partitive in *whether*-islands where the overlap in acceptability is due to a drop in the acceptability of the gap). In addition, *whether*-islands always remain less acceptable than the non-island structures.

3.6 Experiment 3: D-linking, animacy, and resumption in *whether*-islands in English

This experiment investigated the potential effect of animacy. Thus, the following types of *wh*-phrases were tested, controlling for d-linking.

(23) a. **−Animate; −D-linked**: what
b. **+Animate; −D-linked**: who
c. **−Animate; +D-linked**: which book
d. **+Animate; +D-linked**: which colleague

We had two reasons to investigate animacy. The first had to do with the fact that, in contrast to Alexopoulou and Keller (2007), we did not obtain any interaction between resumption and embedding in Experiment 2; since the two studies differed in the animacy of the *wh*-phrase, we investigated the potential of animacy accounting for the different results.This possibility could be related to the properties of resumptive *it*, which as noted by Postal (1994), is sometimes excluded from environments where other resumptive elements may appear.

Secondly, we wanted to investigate whether animacy interacts with d-linking. Animacy has been shown to interact with the processing complexity of filler-gap dependencies (Traxler, Williams, Blozis, and Morris 2005) and in a descriptive sense, animacy makes a referent more salient. The question then is whether animacy can increase salience in the absence of d-linking, and induce effects similar to d-linking, i.e. improve the acceptability of *whether*-islands and intrusive resumption. If such an effect is elicited, whether it can lead to cumulative improvements in the case of +animate and +dlinked *wh*-phrases.

The rest of the design was similar to the previous two studies. In particular the design crossed the following factors: ISLAND (zero-embedding, *that*-clause, *whether*-island) X D-LINKING (bare, which X) X RESUMPTION (gap, pronoun) X ANIMACY (+/-animate). This yielded 24 cells; 8 lexicalisations were produced for the animate condition and 8 for the inanimate condition. As before experimental stimuli were interspersed with equal numbers of fillers.

Sample stimuli are given below:

(24) **−Animate; −D-linked**

a. **zero-embedding**: What will we watch it/Ø at the cinema?
b. **that-clause**: What does Alice think we will watch it/Ø at the cinema?
c. **whether-clause**: What does Clare wonder whether we will watch it/Ø at the cinema?

(25) **+Animate; −D-linked**

a. **zero-embedding**: Who will we fire him/Ø from the committee?
b. **that-clause**: Who does Ann claim we will fire him/Ø from the committee?
c. **whether-clause**: Who does Diane wonder whether we will fire him/Ø from the committee?

(26) **−Animate; +D-linked**

a. **zero-embedding**: Which movie will we watch it/Ø at the cinema?
b. **that-clause**: Which movie does Alice think we will watch it/Ø at the cinema?
c. **whether-clause**: Which movie does Clare wonder whether we will watch it/Ø at the cinema?

(27) **+Animate; +D-linked**

a. **zero-embedding**: Which colleague will we fire him/Ø from the committee?
b. **that-clause**: Which colleague does Ann claim we will fire him/Ø from the committee?
c. **whether-clause**: Which colleague does Diane wonder whether we will fire him/Ø from the committee?

3.6.1 Method

Thirty-seven participants from the same population as in Experiment 2 participated in this experiment.

The design of the experiment varied that of Experiment 2, again comprising two subdesigns. The first one tested *whether*-clauses and crossed the factors RESUMPTIVE (gap or pronoun), and D-LINKING (*who*/*what*, *which X*), and ANIMACY (animate, inanimate). The second subdesign tested *that*-clauses and compared them against an unembedded baseline condition. It crossed the factors EMBEDDING (no embedding, one level of embedding), resumption (gap, pronoun), and D-LINKING (*who*, *which X*), and ANIMACY (animate, inanimate). The experimental procedure was identical to that of Experiment 2.

	BARE what, inanim	which X, anim	which X, inanim
BARE who, anim	(*)		
BARE what, inanim		*	*
which X, anim			

*$p < .05$; **$p < .01$; (): by participants only; []: by items only

Table 6: Result of Tukey test for the interaction of D-LINKING and ANIMACY in Experiment 3 (*whether*-clauses)

	no emb, which X	emb, wh	emb, which X
no emb, wh			(*)
no emb, which X			
emb, wh			(*)

*$p < .05$; **$p < .01$; (): by participants only; []: by items only

Table 7: Result of Tukey test for the interaction of EMBEDDING and D-LINKING in Experiment 3 (*that*-clauses)

3.6.2 Results

The data post-processed and analyzed as in Experiment 2. Figure 3 graphs the mean judgments for both *whether*-clauses and *that*-clauses.

An ANOVA for subdesign 1 (*whether*-clauses) confirmed that sentences that do not contain resumptives are more acceptable than sentences with resumptives (main effect of RESUMPTION, significant by participants only, $F_1(1, 36) = 17.487$, $p < .001$; $F_2(1, 7) = 1.257$, $p = .299$). We also found that D-linked *wh*-phrases were more acceptable than non-D-linked ones (main effect of D-LINKING, $F_1(1, 36) = 16.518$, $p < .001$; $F_2(1, 7) = 7.898$, $p = .026$). Animacy affected acceptability

	wh, res	which X, no res	which X, res
wh, no res	**		**
wh, res		**	
which X, no res			**

*$p < .05$; **$p < .01$; (): by participants only; []: by items only

Table 8: Result of Tukey test for the interaction of D-LINKING and RESUMPTION in Experiment 3 (*that*-clauses)

only for non-D-linked *wh*-phrases (no main effect of ANIMACY, but interaction of D-LINKING and ANIMACY, significant by participants only, $F_1(1,36) = 5.515$, $p = .024$; $F_2(1,7) = 1.274$, $p = .296$)). A post-hoc Tukey test to investigate this interaction further revealed that non-D-linked inanimate *wh*-phrases were less acceptable than non-D-linked animate ones (see Table 6). No other interactions were significant.

An ANOVA for subdesign 2 (*that*-clauses and non-embedded baseline) confirmed once more that sentences without resumptives are more acceptable than sentences with resumptives (main effect of RESUMPTION, $F_1(1,36) = 46.931$, $p < .001$; $F_2(1,7) = 70.868$, $p < .001$). D-linking affected acceptability only for sentences with embedding (no main effects of EMBEDDING and D-LINKING, but an interaction these two factors, significant by participants only, $F_1(1,36) = 7.493$, $p = .010$; $F_2(1,7) = 3.520$, $p = .103$). A post-hoc Tukey test to investigate this interaction showed that non-D-linked embedded *wh*-phrases were less acceptable than non-D-linked embedded ones (see Table 7). We also found that the effect of resumption on acceptability was larger for sentences containing non-D-linked *wh*-phrases compared to ones containing D-linked *wh* (interaction of D-LINKING with RESUMPTION, significant by participants only, $F_1(1,36) = 6.962$, $p = .012$; $F_2(1,7) = 0.820$, $p = .395$). A post-hoc Tukey test was conducted to further investigate this interaction further (see Table 8). We found no effect of animacy on acceptability; also all remaining interactions were not significant.

3.7 Discussion

The most interesting finding is that animacy has an effect only in the non-d-linked condition; *what* (-DL,-Anim) is worse than *who* (-DL,+Anim), but *which colleague* is as acceptable as *which film*. In addition, *who*, while better than *what* is not worse than *which colleague* or *which book*. This is rather surprising in view of other studies in the literature where bare *who* received lower acceptability scores than *which X* phrases (see in particular Frazier and Clifton 2002 and Hofmeister and Sag 2010). Note that also in our data *who* is less acceptable than the *which X* conditions, but this difference did not reach significance.

What these results indicate is that, again, the sharpest contrast is the one between individual and kind denoting *wh*-phrases. Animacy is only indirectly relevant to this distinction in the case of bare *wh*-phrases. *Which X* phrases denote individuals irrespective of animacy and, thus, the animancy effect disappears in these conditions. Note that the denotational distinction cuts across bare and non-bare or d-linked *wh*-phrases as shown in (28), an unexpected outcome from the point of view of the referentiality hypothesis.

(28) **Exp. 3: whether-clauses**
which book, **which colleague**, **who** $\prec$ **what**

Unlike Experiment 2, there was an interaction between d-linking and resumption which indicates that intrusive resumption in English is sensitive to d-linking, in line with Frazier and Clifton (1989). But note that the interaction is much weaker than in Greek and does not result in improvement of resumptive structures due to d-linking; it just amounts to decreasing the difference between the pronoun and the gap.

Additionally, there was an interaction between embedding and d-linking; we only obtained this interaction in Exp. 3. It is possible that this interaction is triggered by the mysterious drop in acceptability in the non-embedded *which X* condition and it is not a real effect. We'll have no explanation to offer for this unexpected effects.

4 Analysis

4.1 Summary of results

The main facts to be accounted for:

1. General interaction between d-linking, resumption and islands
   (a) D-linking improves *whether*-islands; this is the strongest effect of d-linking in both English and Greek. It further improves resumption in Greek; (a weak interaction between resumption and d-linking was found in Exp. 3, for English).
   (b) Unlike d-linking, resumption does *not* improve the *whether*-island; however, it does reverse the island effect, so that embedded resumptives are better than non-embedded ones, at least in Greek.
   (c) Despite its effect, d-linking fails to change the shape of the main effects of *whether*-island and *resumption*; *whether*-islands remain mostly less acceptable than non-islands (*that*-clauses) and resumptive structures are at best as good as gap structures.
   (d) In Greek, resumption interacts with both types of embedded structures *whether*-islands and *that*-clauses.
   (e) Unlike resumption, d-linking interacts only with *whether*-islands but not with *that*-clauses.
   (f) In Greek the most d-linked *whether*-islands are of comparable acceptability with the non-island condition (*that*-clauses); in English neither d-linking nor animacy can restore *whether*-islands to full acceptability.
2. D-linking and animacy

(a) The results from the first two experiments partially confirm Anagnostopoulou's hypothesis in that acceptability decreases by and large as predicted. However, the increase in acceptability is not consistent throughout the hierarchy. We found a sharp contrast between *which X* and *what* phrases which is not predicted by the referentiality hypothesis. In addition, in Experiment 3, we found a distinction between the two bare phrases, *who* and *what* which, again, is not predicted.

(b) The results are crosslinguistically similar in that it is the *which X* condition which elicits significantly more acceptable sentences in both Greek and English, despite differences in the morphological make-up of the interrogative pronouns in the two languages.

(c) Finally, an animacy effect was obtained for bare *wh*-phrases, *who* and *what*, which did not extend to *wh*-phrases with lexical restrictions.

(d) Taken together, the results indicate that the contrast between kind and individual denoting *wh*-phrases is stronger than pragmatic or complexity differences between the different types of *wh*-phrases.

4.2 D-linking/animacy and the complexity of weak islands

We first account for the general effect of d-linking and animacy and their interaction with the *whether*-island and resumption. We return to a discussion of what the best characterisation of these effects is in section 4.3.

Filler-gap dependencies are generally more complex to process than their declarative counterparts; the complexity arises because a filler is carried across a number of intervening nodes separating it from its subcategorisor/gap (Frazier and Clifton 1989). The salience of the filler interacts with the complexity of such structures; more salient fillers remain highly activated in memory, thus, counterbalancing the strains imposed on the parser by longer dependencies (Just and Carpenter 1992; Kluender 1992; Kluender 1998; Hofmeister 007b). We follow this general intuition and apply it to the analysis of the *whether*-islands; we view the "islandhood" of these structures as the result of the interaction between the resource limitations of the parser and the structural complexity of *whether*-clauses, building on the insight of Kluender (1992) and more recent proposals by Hawkins (1999), (2004), Sprouse, Fukuda, and Kluender (2011) and Hofmeister and Sag (2010) among

others.[8] We implement these ideas by extending our earlier analysis for the interaction of islands and resumption in Alexopoulou and Keller (2007), building on the syntactic complexity model of Gibson (1998). A crucial conclusion is that linguistic devices decreasing integration costs (i.e. d-linking) have a stronger "saving" effect on islands than devices remedying locality costs (e.g. resumption).

According to Gibson's model, the linguistic complexity involved in filler-gap dependencies is de-composed into two components. First, it's the memory cost of the syntactic prediction associated with the filler, that is, the cost of carrying a filler down the dependency until the predicted verb/subcategorisor is encountered. This cost is locality or "distance"-based, calculated on the basis of intervening discourse referents between the filler (or the prediction associated with the filler) and the subcategorisor/gap. The second cost is that of the linguistic integration of new input into structure; in the case of filler-gap dependencies, the integration of the filler with the subcategorisor. The linguistic integration cost has a locality/ "distance"-based component, amounting to a kind of "backward" memory cost calculated on the basis of intervening referents processed between the point of integration and the point the prediction was made.[9] In Alexopoulou and Keller (2007) we argued that locality/distance-based costs should be calculated in terms of intervening syntactic/phrasal heads instead of discourse referents; this assumption is crucial in order to account for the contrast between *that*-clauses and *whether*-islands as we will see below. The linguistic integration cost has a second component dependent on the complexity of the integration *per se*, i.e., dependent on the type of element being integrated; for instance, new discourse referents (e.g. indefinite NPs) are assumed to involve a higher integration cost than old/established discourse referents.

Gibson's complexity metric provides a mix of locality (forward and backward) and integration costs, which allow an account of the interactions between *whether*-islands, d-linking and resumption. In Alexopoulou and Keller (2007) we argued that the acceptability of the *whether*-islands and their interaction with resumption sets them apart from strong islands and makes them pattern with *that*-clauses with the main difference between the two a quantitative one, amounting to the magnitude of the obtained interactions. We argued that the similarities favor a processing explanation for the two, but, their differences

[8] The findings of Sprouse, Wagers, and Phillips (2012) are a potential challenge to analyses of islands appealing to the limitations of the memory resources of the parser as they find no correlation between memory capacity and acceptability of island violations.

[9] To be precise, Gibson states that the distance-based integration cost is the number of new discourse referents that have been processed "since h_1 was last highly activated"; for current purposes we assume that "high" activation corresponds to the point the filler was introduced.

also indicate a structural contrast. In both structures, the integration of the complementiser is associated with a higher integration cost, as the (prediction associated with the) filler crosses a clause boundary (Frazier and Clifton 1989). In syntactic terms, the higher integration cost is linked to the intermediate trace at the specifier of C, i.e. cyclicity. However, unlike *that*, *whether* is a scope element, in the sense of Szabolcsi and Zwarts (1993), introducing a new scope domain, a new question, before the matrix *wh*-phrase is integrated into the structure. Thus, the linguistic integration cost of *whether* is much higher than that of *that*, leading to a decrease in acceptability. Under this view, there is no syntactic island constraint *per se*. However, there is a structural difference between *that*-clauses and *whether*-islands, namely that *whether*-islands but not *that*-clauses involve an additional scopal element within the filler-gap dependency; this structural contrast is the source of additional processing complexity—due to increased linguistic integration costs. Crucially, the effect of this structural contrast on processing persists, despite the improving effect of d-linking.

Let us now turn to d-linking. Hofmeister and Sag (2010) provide evidence that d-linked fillers like *which employee* are read slower than non d-linked fillers like *who*, but this initial cost is offset by the fact that segments following the subcategorising verb are read faster for d-linked fillers rather than non-d-linked ones, a fact which indicates that d-linking eases the resolution of the filler-gap dependency. In Gibson's model, this can be linked to the integration costs associated with the integration of the filler with the gap/subcategorising verb; let us, then, stipulate that d-linked phrases have lower integration costs (we return in the next section to why this might be so). This assumption can capture the gradient effects. As integration costs decrease with d-linking, the complexity of *whether*-islands decreases and their acceptability improves. However, the integration costs arising from the scope interaction, associated with *whether* itself, cannot be cancelled by d-linking; thus, the effect of the *whether*-island on acceptability is present even for d-linked *wh*-phrases.

Turning to resumption, we have argued that resumption cancels the "backward" locality costs associated with the integration of the filler. In particular, resumption leads the parser to drop the syntactic route of solving the filler-gap dependency and initiates an anaphoric resolution of the dependency, by searching for the filler as a discourse antecedent in intrasentential anaphora (Alexopoulou and Keller 2007; Erteschik-Shir 1992; Hawkins 1999; Dickey 1996). In this respect, resumption dispenses with the "backward" memory costs of the integration of the filler and, in particular, costs associated with the filler prediction at clause boundaries, i.e. the cyclicity of the syntactic resolution of the

dependency.[10] Since resumption initiates an anaphoric resolution of the dependency, it is not surprising that it interacts with d-linking, if better established discourse referents make better antecedents. However, resumption may not cancel the "forward" memory costs already incurred when the parser hits the pronoun. Thus, while resumption reverses the "island" effect by cancelling the "backward" locality costs, it comes too late to "save" the island.

One emerging conclusion is that, ultimately, integration costs matter more than locality costs.[11] The main culprit for the increased processing complexity of *that*-clauses is the cost of carrying the filler across a clause boundary, i.e. the higher integration cost of intermediate C. Further, the contrast between (non-island) *that*-clauses and *whether*-islands is again linked to higher integration costs for *whether*, which involves the integration of a new scope domain. Finally, d-linking significantly improves *whether*-islands while resumption fails to save them. So, again, it is the decrease of integration costs (d-linking) that is more effective than alleviation of locality costs (resumption).

Before we turn to the question of why or how d-linking and animacy can reduce integration costs, let us consider briefly an alternative view of the processing complexity of *whether*-islands and their interaction with d-linking offered in Hofmeister and Sag (2010); they also share the intuition that *whether*-islands are semantically more complex than *that*-clauses since questions are generally analysed as involving sets of propositions and *whether* questions require the evaluation of the positive and negative alternatives. But they don't relate this semantic complexity to structural complexity specific to *whether*-islands. The crucial difference between their analysis and ours is that our analysis assumes a processing (integration) cost specifically linked to the interaction of two scope domains. While it is true that questions are generally semantically more complex, it is also true that it is in particular questions intercepting a filler-gap dependency that give rise to the *whether*-island effect. *Whether*-questions as complements of verbs in declarative sentences are much more acceptable than *whether*-questions intercepting filler-gap dependencies (Alexopoulou and Keller 2007). In other words, the *whether*-island arises precisely because of the interac-

[10] A reviewer wonders why online studies show no re-analysis effects with resumptive structures, if the parser drops the "gap-resolution" for the "resumptive" one; instead, Dickey (1996) and Hofmeister and Norcliffe (2011) find that resumptive pronouns bring a facilitating effect. Unlike other cases of reanalysis, the two competing analyses in our case, the gap and the resumptive one, are variants of essentially the same dependency/filler-gap resolution, so, rather than reanalysis, this looks like a case of local adjustment on the realisation of the dependency. Further, the absence of reanalysis effects may indicate that the parser's prediction is rather underspecified between a resolution with a subcategorisor or with a pronominal.

[11] Thanks to J.Hawkins for this point.

tion of two scope domains; and it is precisely the absence of such scope interaction in *that*-clauses that induces only mild processing difficulty in comparison with the *whether*-island.

4.3 D-linking and animacy

In this section we address two final questions: (i) what is the best characterisation of the obtained d-linking and animacy effects?(ii) why should d-linking decrease integration costs of the filler at the retrieval point and why this interaction appears relevant only for *whether*-islands but not *that*-clauses.

Together, the experimental results can be summarised by (29). We have excluded *what X* since, with the exception of Greek *that*-clauses, it was not significantly different from either *what* or *which X*.

(29) **which of X, which X, who $\prec$ what**

In terms of tendencies, we found *which of X* to be better than *which X* which, in turn, was better than *who*. This picture is consistent with the referentiality hypothesis as well as the complexity view which predicts that structurally and semantically richer representations facilitate the retrieval of the filler and, therefore, can improve the acceptability of *whether*-islands (Hofmeister and Sag 2010; Hofmeister 2011).

However, there are two further aspects of the results that need explanation. The first one is why *what* phrases are set apart from all others. The second is why the type of *wh*-phrase affects the acceptability of *whether*-islands but not *that*-clauses. Both facts can be understood by a "denotational hypothesis", namely, that what matters in *whether*-islands is that the filler/extracted phrase denotes an individual, since individuals (as opposed to amounts for instance) can participate in the semantic operations relevant to *whether*-islands (Szabolcsi and Zwarts 1993). Kind denoting *wh*-phrases are not ordinary individuals, rather they become individuals after the application of semantic shifting operations (Chierchia 1998). D-linking and animacy interact with the distinction between kinds and ordinary individuals in as much as they may implicitly encode the contrast (e.g. in bare *who* vs. bare *what*) or through establishing sets of discourse referents (though a lexical restriction in d-linked phrases). In other words, the relevance of d-linking and animacy here is very similar with the role of d-linking in bringing about the individuation of amounts as proposed by Szabolcsi and Zwarts (1993) (see discussion in the introduction). None of this matters for *that*-clauses since there is no scope island involved and the denotation of the extractor is not crucial.

In view of the denotational hypothesis, we can then assume that D-

linked/animate wh-phrases have lower integration costs because they help define individuals and therefore facilitate the semantic operations relevant to *whether*-islands, while *what* phrases have higher integration costs due to their denoting less typical individuals. Over and above the distinction between individuals and kinds, the complexity of the wh-filler, may independently affect the overall complexity of the structure by facilitating retrieval of the filler. The precise interaction between complexity and denotation can only be determined on further data.

5 Conclusion

We have investigated experimentally the role of d-linking in *whether*-islands in English and Greek. Our results confirm that the type of *wh*-phrase has an effect on the acceptability of *whether*-islands as well as resumption. We obtained a broad confirmation for Anagnostopoulou's referentiality hierarchy. However, we also obtained two unexpected effects: (i) a main contrast between *what X* and *what* and (ii) an animacy effect for bare *wh*-phrases; *who* was better than *what*. In these results *what* is set apart from all other *wh*-phrases. We propose that the critical contrast is between kind-denoting and ordinary individuals, a contrast relevant for the semantic operations involved in the interpretation of *whether*-islands (Szabolcsi and Zwarts 1993). This is why, the distinction between kind-denoting and ordinary individuals only affects *whether*-islands, but not on *that*-clauses; the denotation of the extractor is not crucial in *that*-clauses.

We embed this analysis in our analysis of the complexity of *whether*-islands. We view the "islandhood" of *whether*-islands as the result of increased complexity due to the interaction of two scope domains. We assume that ordinary-individual-denoting *wh*-phrases have lower integration costs than kind-denoting *wh*-phrases. The complexity of the wh-phrase (e.g. *who* vs. *which X*) may additionally decrease or increase integration costs, since structurally richer descriptions are shown to facilitate retrieval of fillers. Together, the effect of complexity and the denotation of the filler, may affect acceptability; however, they cannot cancel the overall complexity of the *whether*-island, which results from higher integration costs associated with the introduction of a new question intercepting the filler-gap dependency. In other words, while d-linking/animacy can improve *whether*-islands, it cannot restore them to full acceptability: *whether*-islands are mostly less acceptable than *that*-clauses.

Compared to resumption, d-linking/animacy has a stronger "rescuing" effect, in that it does lead to improvement in the acceptability of *whether*-islands. By contrast, resumption, fails to "save" or improve these violations. Our theoretical interpretation of this con-

trast entails that amelioration of integration costs, which is linked to d-linking/animacy, is more effective than (partial) cancelling of locality/distance-based costs, which is linked to resumption.

References

Alexopoulou, T. and F. Keller (2007). Locality, cyclicity and resumption: at the interface between the gramamr and the human sentence processor. *Language 83*(1), 110–160.

Anagnostopoulou, E. (1994). *Clitic Dependencies in Modern Greek*. Ph. D. thesis, University of Salzburg.

Ariel, M. (1990). *Acessing Noun Phrase Antecedents*. London: Routledge.

Ariel, M. (1999). Cognitive universals and linguistic conventions: the case of resumptive pronouns. *Studies in Language 23*(2), 217–269.

Bard, E. G., D. Robertson, and A. Sorace (1996). Magnitude estimation of linguistic acceptability. *Language 72*, 32–68.

Chierchia, G. (1998). Reference to kinds across languages. *Natural Language Semantics 6*, 339–405.

Chomsky, N. (1986). *Barriers*. Cambridge, MA: MIT Press.

Cinque, G. (1990). *Types of A-bar Dependencies*, Volume 17. MIT Press. Linguistic Inquiry Monographs.

Comorovski, I. (1989a). *Discourse and the syntax of multiple constituent questions*. Ph. D. thesis, Cornell University.

Cowart, W. (1997). *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. Thousand Oaks, CA: Sage Publications.

Dickey, M. W. (1996). Constraints on the sentence processor and the distribution of resumptive pronouns. In M. W. Dickey and S. Tunstall (Eds.), *Linguistics in the Laboratory*, Number 19 in Occasional Papers in Linguistics, pp. 157–192. Amherst, MA: University of Massachusetts.

Dobrovie-Sorin, C. (1990). Clitic doubling, *wh*-movement and quantification in Romanian. *Linguistic Inquiry 21*(3), 351–397.

Erteschik-Shir, N. (1992). Resumptive pronouns in islands. In H. Goodluck and M. Rochemont (Eds.), *Island Constraints: Theory, Acquisition and Processing*, Volume 15 of *Studies in theoretical psycholinguistics*, pp. 89–108. Kluwer Academic Publishers.

Featherston, S. (2003). The relation between judgement and frequency data in syntactic well-formedeness: the decathlon model. In *Proceedings of AMLAP 2003*.

Featherston, S. (2005). Magnitude estimation and what it can do for your syntax: Some wh-constraints in German. *Lingua 115*(11), 1525–1550.

Ferreira, F. and B. Swets (2005). The production and comprehension of resumptive pronouns in relative clause "island" contexts. In A. Cutler (Ed.), *Twenty-first Century Psycholinguistics: Four Cornerstones*, pp. 263–278. Mahway, NJ: Lawrence Erlbaum Associates.

Frazier, L. and C. Clifton (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes 4*(2), 93–126.

Frazier, L. and C. Clifton (2002). Processing "d-linked" phrases. *Journal of Psycholinguistic Research 31*(6), 633–659.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition 68*, 1–76.

Hawkins, J. A. (1999). Processing complexity and filler-gap dependencies across grammars. *Language 75*, 244–285.

Hawkins, J. A. (2004). *Efficiency and Complexity in Grammars*. Oxford University Press.

Heestand, D., M. Xiang, and M. Polinsky (2011). Resumption still does not rescue islands. *Linguistic Inquiry 42*, 138–152.

Hofmeister, P. (2007b). *Representational Complexity and memory retrieval in language comprehension*. Ph. D. thesis, Stanford University, Stanford CA.

Hofmeister, P. (2011). Representational complexity and memory retrieval in language comprehension. *Language and Cognitive Processes 26*(3), 376–405.

Hofmeister, P., T. F. Jaeger, I. A. Sag, I. Arnon, and N. Snider (2007). Locality and accessibility in wh-questions. In S. Featherston and W. Sternefeld (Eds.), *Roots: Linguistics in Search of its Evidential Base*. Berlin: Mouton de Gruyter.

Hofmeister, P. and E. Norcliffe (2011). The effect of resumptive pronouns on sentence comprehension. CUNY 2011, poster.

Hofmeister, P. and I. A. Sag (2010). Cognitive constraints and island effects. *Language 86*(2), 366–415.

Huang, C. (1982a). *Logical relations in Chinese and the theory of grammar*. Ph. D. thesis, MIT.

Just, M. and P. Carpenter (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review 99*, 122–149.

Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics & Philosophy 1*, 3–44.

Keller, F., S. Gunasekharan, N. Mayo, and M. Corley (2009). Timing accuracy of web experiments: A case study using the WebExp

software package. *Behavior Research Methods 41*(1), 1–12.

Kluender, R. (1992). Deriving island constraints from principles of predication. In H. Goodluck and M. Rochemont (Eds.), *Island Constraints: Theory, Acquisition and Processing*, Volume 15 of *Studies in theoretical psycholinguistics*, pp. 223–58. Kluwer Academic Publishers.

Kluender, R. (1998). On the distinction between strong and weak islands: A processing perspective. In P. W. Culicover and L. McNally (Eds.), *The limits of syntax*, pp. 241–269. New York: Academic Press.

Kluender, R. and M. Kutas (1993b). Subjacency as a processing phenomenon. *Language and Cognitive Processes 8*, 573–633.

Kroch, A. (1981). On the role of resumptive pronouns in amnestying island constraint violations. In *Proceedings of the Meeting of the Chicago Linguistics Society*, Chicago, pp. 125–135. Chicago Linguistics Society.

Lasnik, H. and M. Saito (1984). On the nature of proper government. *Linguistic Inquiry 15*, 235–289.

Maling, J. and A. Zaenen (1982). A phrase structure account of Scandinavian extraction phenomena. In P. Jacobson and G. Pullum (Eds.), *The nature of syntactic representation*, pp. 229–282. Dordrecht:Reidel.

Meyer, R. (2003). Superiority effects in Russian, Polish and Czech: Comparative evidence from studies on linguistic acceptability. In *Proceedings of the 12th Conference on Formal Approaches to Slavic Linguistics*, Ottawa.

Pesetsky, D. (1987). Wh-in situ: Movement and unselective binding. In E. Reuland and A. ter Meulen (Eds.), *The representation of (in)definiteness*. Cambridge, MA: MIT Press.

Pesetsky, D. (2000). *Phrasal Movement and its kin*. Cambridge, MA: MIT Press.

Postal, P. M. (1994). Contrasting extraction types. *Journal of Linguistics 30*, 159–186.

Rizzi, L. (1990). *Relativized Minimality*. Cambridge, Massachusetts: MIT Press.

Rizzi, L. (1997). The Fine Structure of the Left Periphery. In L. Haegeman (Ed.), *Elements of Grammar: Handbook of Generative Syntax*, pp. 281—337. Dordrecht: Kluwer.

Ross, J. R. (1967). *Constraints on Variables in Syntax*. Ph. D. thesis, MIT. Published as (Ross 1986).

Ross, J. R. (1986). *Infinite Syntax!* Norwood, NJ: Ablex.

Sells, P. (1984). *Syntax and Semantics of Resumptive Pronouns*. Ph. D. thesis, university of Massachusetts at Amherst.

Sells, P. (1987). Binding resumptive pronouns. *Linguistics and Philosophy 10*, 261–298.

Sprouse, J., H. O. S. Fukuda, and R. Kluender (2011). Reverse island effects and the backward search for a licensor in multiple wh-questions. *Syntax 14*(2), 179–203.

Sprouse, J., M. Wagers, and C. Phillips (2012). A test of the relation between working memory capacity and syntactic island effects. *Language 88*(1), 82–123.

Stevens, S. S. (1975). *Psychophysics: Introduction to its Perceptual, Neural, and Social Prospects*. New York: John Wiley.

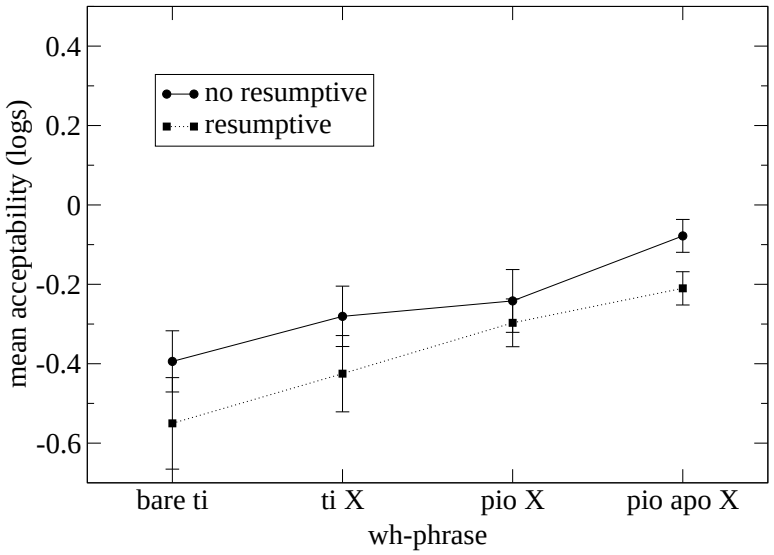
Szabolcsi, A. (2006). Strong vs. Weak Islands. In M. Everaet, H. van Riemsdijk, R. Goedemans, and B. Hollebrandse (Eds.), *The Blackwell Companion to Syntax*, Volume IV. Blackwell Publishing.

Szabolcsi, A. and F. Zwarts (1993). Weak islands and an algebraic semantics for scope taking. *Natural Language Semantics 1*, 235–284.

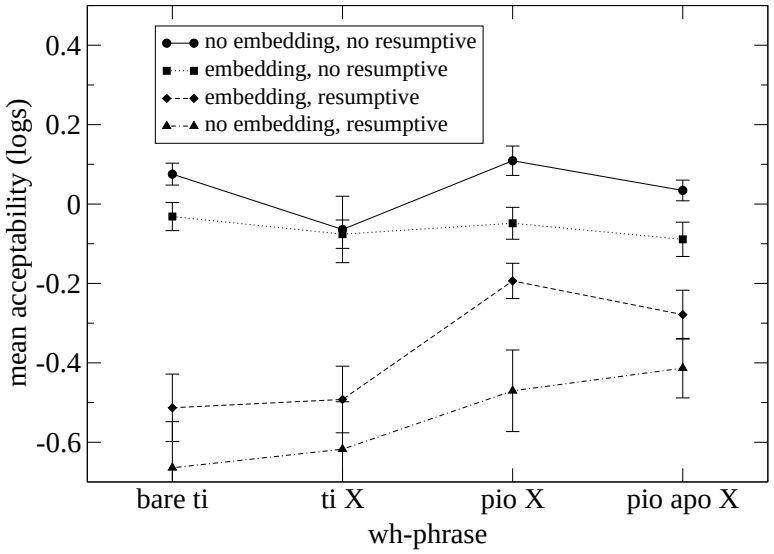
Traxler, M. J., R. S. Williams, S. A. Blozis, and R. K. Morris (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language 53*, 204–224.

Tsimpli, I.-M. (1999). Null operators, clitics and identification: a comparison between Greek and English. In A. Alexiadou, G. Horrocks, and M. Stavrou (Eds.), *Studies in Greek syntax*, Volume 43 of *Studies in Natural Language and Linguistic Theory*, pp. 241–262. Kluwer Academic Publishers.

Xiang, M., D. Heestand, and M. Polinsky (2008). Resumptive pronouns: (non)-rescuing function? In *CUNY 2008*.

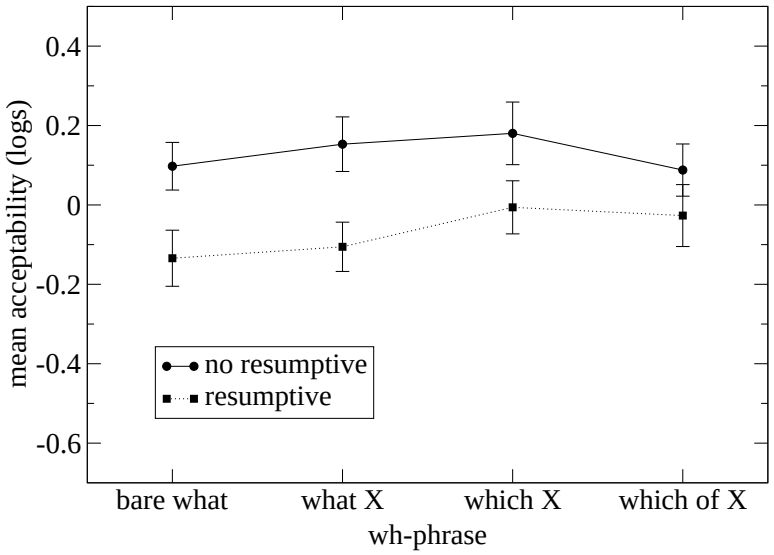


(a) Extraction from *whether*-clauses

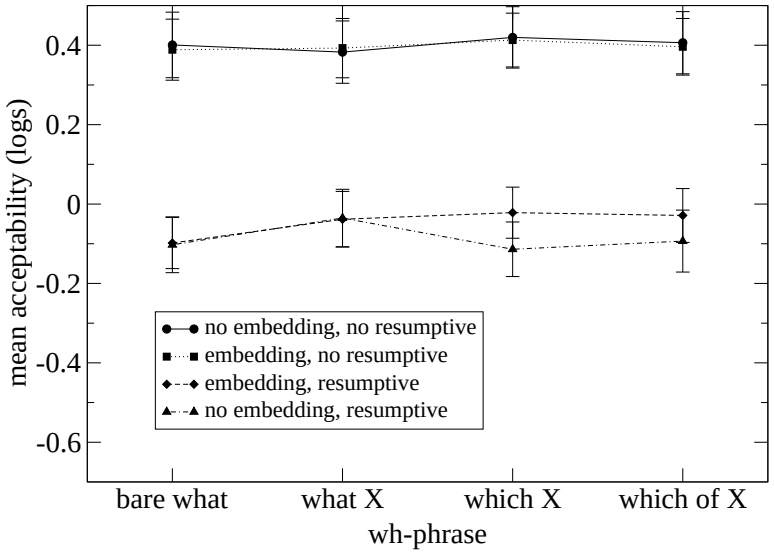


(b) Extraction *that*-clauses, with unembedded baseline

Figure 1: Effect of embedding, resumption, and *wh*-phrase on subject extraction in Greek in Experiment 1

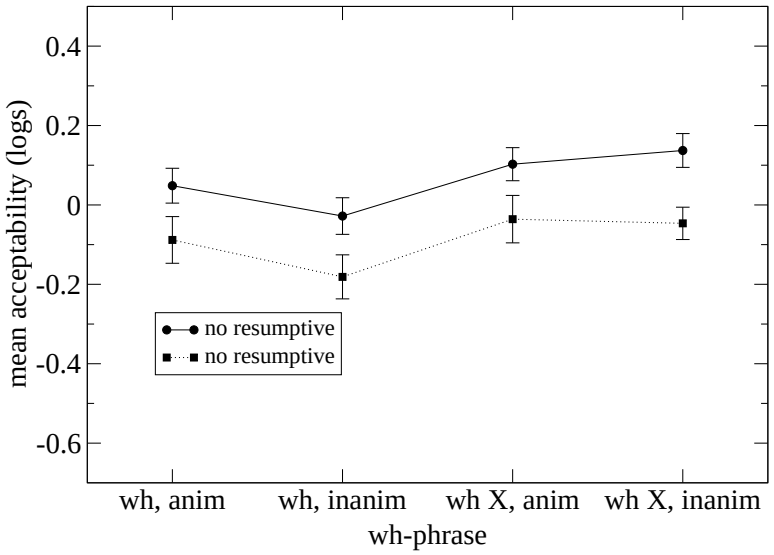


(a) Extraction from *whether*-clauses

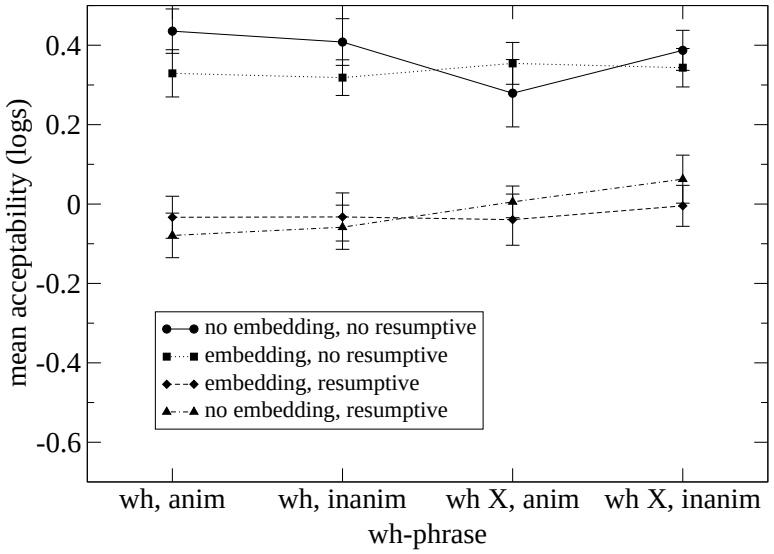


(b) Extraction *that*-clauses, with unembedded baseline

Figure 2: Effect of embedding, resumption, and *wh*-phrase on subject extraction in English in Experiment 2



(a) Extraction from *whether*-clauses



(b) Extraction *that*-clauses, with unembedded baseline

Figure 3: Effect of embedding, resumption, D-linking, and animacy on subject extraction in English in Experiment 3