

QUD effects on epistemic containment principle: An experimental study¹

Sunwoo JEONG — *Department of Linguistics, Stanford University*

Abstract. The Epistemic Containment Principle (ECP) requires that epistemic modals take wider scope than strong quantifiers such as *every* or *most* (von Fintel and Iatridou, 2003). Although fairly robust in its realization, a few systemic classes of counterexamples to the ECP have been noted. Based on these, previous work has argued for two claims: subjective modals obey the ECP, whereas objective ones don't (Tancredi, 2007; Anand and Hacquard, 2008); and *every* respects the ECP, whereas *each* violates it (Tancredi, 2007). This paper argues that explicit Questions Under Discussion (QUDs; Roberts, 1996; Ginzburg, 1996) also systematically influence the ECP: scopal orderings that provide relevant answers to the given QUDs are preferred, and this tendency can override the ECP. To support this claim, the paper presents an experimental study. The results corroborate the existence of systematic QUD effects on the ECP, and support the view that the ECP is derived from a confluence of various pragmatic and lexical biases.

Keywords: Epistemic Containment Principle (ECP), epistemic modals, Question Under Discussion (QUD), quantifiers, scopal ambiguity, experimental semantics.

1. Introduction

The Epistemic Containment Principle (henceforth ECP) is a widely known descriptive and theoretical claim according to which epistemic modals must take wider scope than strong quantifiers such as *every* or *most* (von Fintel and Iatridou, 2003). The ECP can capture, for instance, why a sentence like (1) sounds infelicitous. As the scopal ordering that would have yielded a felicitous meaning (1b) is in effect ruled out by the ECP, the only possible interpretation that is left is (1a), which results in an unlikely meaning ((1a) is tenable only if multiple people can collectively constitute 'the murderer').

- (1) #*Every student might be the murderer.* (von Fintel and Iatridou, 2003)
a. MIGHT \succ EVERY: #It is possible that every student is the murderer.
b. #EVERY \succ MIGHT: For every student x, it is possible that she is the murderer.

Although fairly robust in its realization, a few systemic classes of counterexamples to the ECP have been noted. They involve the distinction between subjective vs. objective epistemic modals, and differences in quantifier types.

- (2) a. Objective vs. subjective (doxastic)
Objectively speaking, every student might be the murderer. (Tancredi, 2007)
b. Quantifier type
Each student might be the murderer. (Tancredi, 2007)

¹ I would like to thank Chris Potts, Cleo Condoravdi, Judith Degen, Dan Lassiter, Simon Todd, the anonymous reviewers of the abstract, and audiences at *Sinn und Bedeutung 22* for many insightful comments on this project and/or the paper. This research was funded in part by: NSF BCS-1456077.

For instance, epistemic modals that quantify over objective knowledge states have been shown to allow quantifiers to scope over them, as indicated by the felicity of (2a). Based on this, Tancredi (2007) concluded that only subjective epistemic modals (which he calls ‘doxastic’ and assumes to be the default interpretations) that quantify over the speaker’s subjective belief states observe the ECP (cf. Lyons, 1977; Anand and Hacquard, 2008). In addition, quantifiers such as *each* have been shown to be able to violate the ECP, as indicated by the felicity of (2b).

These counterexamples raise questions about the nature and the status of the ECP. First, is the ECP a hard-wired structural constraint, as von Stechow and Iatridou (2003) had initially conceptualized,² or is the intuition behind it better characterized as a gradient tendency, as Anand and Hacquard (2008) suggest? Secondly, if the ECP can be reduced to a kind of gradient preference, how robust is this preference, and what factors come into play in shaping this preference? Answering these questions will go some way towards explaining why exactly the ECP arises in the first place, and how it connects with the more general tendency for epistemic modals to scope wide.

The aim of this paper is to engage with these questions by pursuing two specific empirically tractable goals. The first is to measure the robustness of the intuition behind the ECP via quantitative methods. The second is to introduce a new contextual factor that also seems to systematically influence the realization of the ECP, namely, Questions Under Discussion (henceforth QUDs; Roberts, 1996). The paper argues that listeners tend to prefer scopal orderings that provide relevant answers to the given QUDs, and that this preference can override the ECP.

To achieve these goals, an experimental study is presented. The results of the study corroborate the significant effects of QUDs on the ECP, while also demonstrating that violations of the ECP can occur even for subjective modals and for the quantifier *every*. Based on these data, the paper propounds the view that the ECP arises from a combination of various pragmatic and lexical biases. The resulting account is shown to have broader implications for thinking about the scopal preferences of epistemic modals, as well as how context comes into play in shaping these preferences.

2. QUD effects on the ECP: Probing the intuition

Suppose that the same sentence from (1), repeated in (3b), was uttered in answer to an explicit question in (3a).

- (3) a. *Which of the four students is the murderer?*
 b. *Every student might be the murderer.*

The sentence sounds distinctly better in (3) than in (1), although the quantifier *every* and the subjective modal interpretation have remained constant. This relative felicity seems to stem from the fact that the ECP-violating EVERY \succ MIGHT interpretation provides a directly relevant answer to the explicit QUD in (3).

² As a structural constraint, the ECP is characterized by von Stechow and Iatridou (2003) as follows: At LF, a quantifier cannot bind its trace across an epistemic modal.

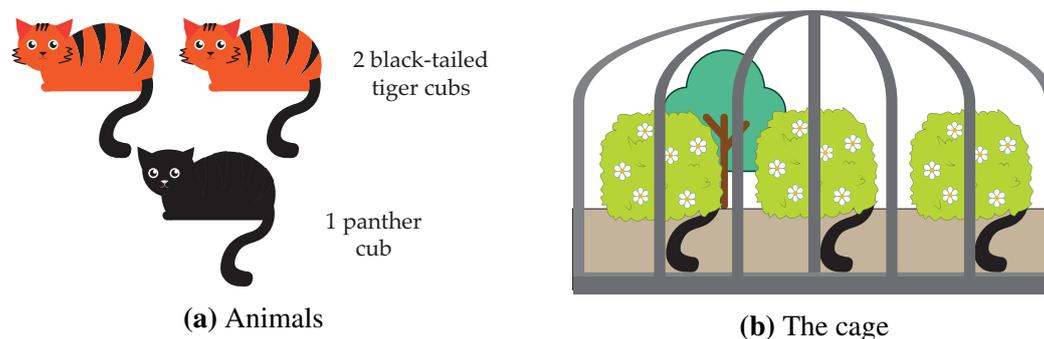


Figure 1: Ron is at a zoo. Ron knows for certain that the cage is currently housing exactly two tiger cubs and exactly one panther cub. Ron also knows for certain that the animals look as in (a). Ron peers through the cage, which looks as in (b).

QUDs (Roberts, 1996; cf. Ginzburg, 1996) signal what is at-issue, and provide a systematic way of capturing the information structure in the discourse. They have been argued to be at the heart of a variety of linguistic phenomena such as focus and focus-sensitive expressions (Beaver and Clark, 2008), projection behavior (Simons et al., 2010), and the meaning contribution of diverse discourse particles. This work indicates that the at-issueness status as prescribed by QUDs has far reaching repercussions for a wide range of linguistic expressions and the way they are interpreted. Given this, it seems reasonable to expect that QUDs would also impact scopal disambiguation processes. In particular, scopal orderings that provide relevant answers to the QUDs are likely to be preferred over ones that don't (cf. Gualmini et al., 2008).

By linking this broader QUD intuition with the ECP phenomenon, we may entertain the following hypothesis: Suppose that the ECP can be recast as a kind of default but defeasible scopal disambiguation strategy or preference which is derived from more primitive lexical and pragmatic biases. Since it operates as a mere preference rather than a structural constraint, we expect it to be overridden by independent scopal preferences arising from QUDs when the two are in conflict, if the preferences stemming from QUDs are stronger. The QUD-based scopal preferences are likely to be stronger when the QUDs are explicitly spelled out.

Let us henceforth call this the QUD hypothesis and probe our intuitions about it with the aid of a paradigm that doesn't involve felicity judgments. (The same paradigm is implemented in a larger scale in the experimental study presented in sec. 4.) For instance, consider the context outlined in Fig. 1. In this context, suppose that Ron utters the modal statement given in (4):

- (4) **“Every bush might have a tiger.”**
- a. MIGHT > EVERY: FALSE
It is possible that every bush has a tiger.
 - b. EVERY > MIGHT: TRUE
For every bush x, it is possible that x has a tiger.

Given the context in Fig. 1, the presence vs. absence of the ECP can be probed by examining whether Ron's answer is interpreted as true or false. Given what Ron knows and believes, Ron's statement in (4) is interpreted as true only under the ECP violating scopal ordering (4b), but

false under the ECP observing one (4a).³ This holds regardless of which QUD (4) is addressing. Since Ron knows that there are only 2 tigers in total, it is not possible that all three bushes have tigers (4a). In other words, in *none* of Ron's epistemically and doxastically accessible worlds is it the case that every bush has a tiger. In contrast, since Ron also knows that tigers and panthers have identical looking tails, for each bush *x*, it is equally possible that *x* has a tiger (4b). In other words, Ron's epistemically and doxastically accessible worlds include worlds in which any given bush *x* has a tiger. In sum, if Ron's statement in (4) is evaluated as true, it indicates that the ECP was violated, and if it is evaluated as false, it indicates that the ECP was observed.⁴

Building on this basic premise, let us now introduce explicit QUDs. Suppose that Ron produced his model statement in (4) as an answer to one of the two explicit questions in (5) and (6). The question was raised by his friend Luna who arrived late at the scene (and is thus more ignorant than Ron about the number and the shape of tigers involved).

- | | |
|---|---|
| (5) HOW-QUD:
<i>“How many of the three bushes
 have a tiger?”</i> | (6) WHICH-QUD:
<i>“Which of the three bushes
 has a tiger?”</i> |
|---|---|

The emerging intuition is that our true vs. false judgment of Ron's statement in (4), which maps onto the violation vs. observation of the ECP, respectively, depends crucially on the type of explicit QUD that (4) is addressing. In answer to the question in (5), henceforth the HOW-QUD, Ron's statement is more likely to be evaluated as false, suggesting that the QUD is nudging us towards the ECP-observing scopal interpretation. In answer to the question in (6) however, henceforth the WHICH-QUD, Ron's statement is more likely to be evaluated as true, suggesting that the QUD is nudging us towards the ECP-violating scopal interpretation. Intuitively, the reason for this seems to be as follows: the ECP-violating interpretation (4b) provides a relevant answer to the WHICH-QUD but not to the HOW-QUD, whereas the ECP-observing interpretation (4a) provides a relevant answer to the HOW-QUD but likely not to the WHICH-QUD (cf. see sec. 3). In sum, ECP-violating interpretations seem to become more accessible when they can provide relevant answers to the explicit QUDs.

If the intuition outlined so far is on the right track, it suggests that the QUD bias does indeed override the ECP, which in turn suggests that the ECP is at best a defeasible bias rather than a categorical constraint. The hypothesized QUD effect also seems to crosscut other factors that have been known to influence the ECP. The sentence in (4) includes the quantifier *every* but still seems to allow for the ECP-violation depending on the QUD. Likewise, given the context of ignorance (Fig. 1b) and in the absence of explicit adverbials like ‘objectively speaking’,

³ One may wonder if providing true/false judgments to subjective/doxastic modal statements is an unintuitive task. However, people do seem to be able to make truth value judgments about subjective epistemic modal statements, although the body of knowledge/beliefs against which such a modal statement is evaluated may shift (von Fintel and Gillies, 2011). Since we, as readers, are led to share identical beliefs as those of Ron's (w.r.t. the tigers and the bushes) by virtue of Fig. 1, the proposed equivalence between true vs. false judgments and ECP violation vs. observation would hold regardless of this potential shift.

⁴ This equivalence holds under the assumption that Ron's knowledge and belief states described in Fig. 1 are fully taken into account when generating the relevant modal base for the statement. See sec. 4.3.2 for more discussion about the validity of this assumption.

the sentence in (4) most likely elicits the default subjective/doxastic modal interpretation but nevertheless seems to allow for the ECP-violation depending on the QUD.

The experimental study to come (sec. 4) aims to provide quantitative support for the QUD hypothesis described so far by implementing a paradigm like Fig. 1 and (4)–(6). Before presenting the experiment however, let us first get a better handle on why the proposed relevance relations hold between the WHICH-QUD and the ECP-violating interpretation (4b) on the one hand, and the HOW-QUD and the ECP-observing interpretation (4a) on the other.

3. Defining relevance for modalized statements

The QUD hypothesis in sec. 2 was formulated on the basis of the assumption that EVERY \succ MIGHT propositions like (4b) provide relevant answers to certain types of questions like (6), while MIGHT \succ EVERY propositions like (4a) provide relevant answers to other types of questions like (5). These judgments seem to be corroborated by native speaker intuitions, but we may want to formulate them in a more precise manner as they will figure as core background assumptions in the experiment.

There are several ways of gauging whether a given proposition counts as a relevant answer to a question. Since it is not the main goal of the paper to argue for a specific theory of relevance, we will consider three possible ways of defining ‘relevance’ and establish that the hypothesized relevance relations are predicted by all of them, although to different degrees.

A proposition is standardly analyzed as being a relevant answer to a question Q if it identifies or rules out a member of the question denotation Q (cf. Groenendijk and Stokhof, 1984; Roberts, 1996; Dayal, 2016). However, hedged/modal answers (e.g., *might p*, *I think that p*, etc.) can also count as relevant answers to simple/non-modal questions (e.g., *whether p?*), although modal propositions do not themselves identify/rule out any member of such Q s. In these cases, what matters seems to be the prejacent: if the prejacent of the modal statements identify/rule out a member of Q , they count as relevant answers. We will therefore posit the following core premise across all three accounts: In the case of a modal statement, the proposition with which we evaluate its relevance to a given question is its prejacent p (see Beaver and Clark (2008) and Kaufmann (2016) for similar views).

Following this assumption, let us zoom in on the prejacent of the ECP-observing (4a) and ECP-violating (4b) when evaluating their relevance to different QUDs. They are presented again in (7) and (8) in a more detailed form; B stands for the modal base and the underlined parts pick out the prejacent. The prejacent of (7) is straightforward; it is the proposition: *every bush has a tiger*. Determining the prejacent of (8) requires more flexibility: when a modal takes a narrow scope as in (8), the propositional argument of the modal contains a free variable x , resulting in: *x has a tiger*. In this case, we define its prejacent as follows: any member of the set of propositions that results when the free variable x of the propositional argument is assigned to a member of the restrictor of *every*. To give a unified account, we also posit that the prejacent of (7) is the sole member of an analogous prejacent set in (7b), which is a singleton set.

- (7) *might* \succ *every*
- a. $\text{MIGHT}(B)(\text{every bush}(\lambda x(\text{a tiger}(\lambda y(x \text{ has } y))))))$
 - b. prejacent set: $\{ \llbracket \text{every bush has a tiger} \rrbracket^g \}$
- (8) *every* \succ *might*
- a. $\text{every bush}(\lambda x(\text{MIGHT}(B)(\text{a tiger}(\lambda y(x \text{ has } y))))))$
 - b. prejacent set: $\{ \llbracket \text{bush}(x) \wedge \text{has a tiger}(x) \rrbracket^{g[x \rightarrow b]} : b \in D_e \}$

With these assumptions in place, one way of defining relevance is as in (9b).⁵

- (9) a. **Q**: a set of propositions that are possible answers to Q (cf. Hamblin 1971)
- b. **Relevance** (ver. 1): A proposition p is a relevant answer to a question Q iff: $p \in Q$

Assuming that a given question Q denotes a Hamblin set Q , i.e., a set of contextually constrained possible answers to Q as in (9a), the HOW-QUD in (5): *How many of the 3 bushes have a tiger?* would have the denotation in (10a), and the WHICH-QUD in (6): *Which of the 3 bushes has a tiger?* would have the denotation in (10a), where $\llbracket \text{bush} \rrbracket = \{b_1, b_2, b_3\}$.

- (10) a. $\llbracket \text{HOW-QUD} \rrbracket = \{ (\llbracket \text{no bush has a tiger} \rrbracket), \llbracket \text{one bush has a tiger} \rrbracket, \llbracket \text{two bushes have a tiger} \rrbracket, \llbracket \text{three bushes have a tiger} \rrbracket \}$
- b. $\llbracket \text{WHICH-QUD} \rrbracket = \{ \llbracket \text{has a tiger} \rrbracket(b_1), \llbracket \text{has a tiger} \rrbracket(b_2), \llbracket \text{has a tiger} \rrbracket(b_3), \llbracket \text{has a tiger} \rrbracket(b_1 + b_2), \llbracket \text{has a tiger} \rrbracket(b_2 + b_3), \llbracket \text{has a tiger} \rrbracket(b_1 + b_3), \llbracket \text{has a tiger} \rrbracket(b_1 + b_2 + b_3) \}$

From (9b), it follows that the ECP-observing (7) is a relevant answer to both the HOW-QUD in (5) and the WHICH-QUD in (6): The prejacent p of (7) is a member of the denotations of both QUDs as the proposition $\llbracket \text{every bush has a tiger} \rrbracket$ is contextually equivalent to $\llbracket \text{three bushes have a tiger} \rrbracket$ and $\llbracket \text{has a tiger} \rrbracket(b_1 + b_2 + b_3)$. In contrast, the ECP-violating (8) is only a relevant answer to the WHICH-QUD but *not* a relevant answer to the HOW-QUD. This is because all of its possible prejacent from (8b), e.g., $\llbracket \text{has a tiger} \rrbracket(b_2)$, are members of $\llbracket \text{WHICH-QUD} \rrbracket$, whereas none of them are members of $\llbracket \text{HOW-QUD} \rrbracket$.

We remain agnostic about whether the denotation of the HOW-QUD should include $\llbracket \text{no bush has a tiger} \rrbracket$; the same prediction comes out irrespective of this choice. We also take the conservative approach of assuming that the partitive wh-phrase ‘Which of x’ is associated with the domain of not just atomic individuals but also plural ones; if only atomic individuals are allowed, as is standardly assumed to be the case for ‘Which x’ (Dayal, 2016), then the denotation of WHICH-QUD would be $\{ \llbracket \text{has a tiger} \rrbracket(b_1), \llbracket \text{has a tiger} \rrbracket(b_2), \llbracket \text{has a tiger} \rrbracket(b_3) \}$ and the current account would predict an even stronger asymmetry: that the ECP-violating (8) is *only* relevant to the WHICH-QUD and the ECP-observing (7) is *only* relevant to the HOW-QUD.

Relevance can also be defined in terms of partitions introduced by a given question (Groenendijk and Stokhof, 1984). A partition of Q can be derived from a Hamblin set Q (Fox, 2017),

⁵ Note that (9b) does not impose that the relevant answer be a true answer. The notion of ‘relevance’ that we are after is only concerned with whether the proposition is directly germane to a given question.

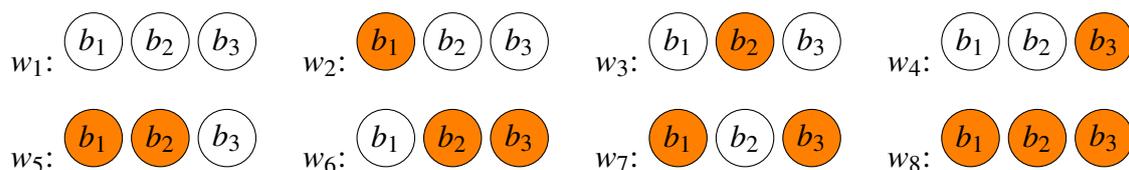


Figure 2: Possible worlds in which there is a total of 3 bushes in the domain. Orange-colored circles represent bushes with a tiger; empty circles represent bushes without a tiger.

as shown in (11a). Unlike propositions in Hamblin sets that can pick out overlapping worlds, partitions cut up the worlds into equivalence classes. Each cell in the partition(Q) correspond to an exhaustive answer. Oftentimes, answers to questions are non-exhaustive in form but understood exhaustively. Given this, we may posit an exhaustivity operator as in (11b) which (roughly) strengthens p to *only* p (F^+ stands for focus alternatives), and define relevance as in (11c). Simply put, p is a relevant answer to Q if upon strengthening to $\text{EXH}(p)$, it identifies a cell in $\text{partition}(Q)$.

- (11) a. **Partition(Q):** a set of equivalence classes under the relation:
 $w \sim w'$ iff $\forall p \in Q[p(w) = p(w')]$ (Groenendijk and Stokhof, 1984; Fox, 2017)
- b. $\llbracket \text{EXH } \phi \rrbracket = \llbracket \phi \rrbracket \wedge (\bigwedge_{P \in \llbracket \phi \rrbracket_{F^+}} \neg P)$ (Klinedinst and Rothschild, 2011: 11)
- c. **Relevance** (ver. 2): p is a relevant answer to Q iff: $\text{EXH}(p) \in \text{Partition}(Q)$

From (11c), it follows that the ECP-observing (7) is a relevant answer to both the HOW-QUD in (5) and the WHICH-QUD in (6), whereas the ECP-violating (8) is only a relevant answer to the WHICH-QUD, but not the HOW-QUD. To see more concretely why this holds, let us imagine the possible worlds in Fig. 2 as constrained by the context. Given these possible worlds, the partitions introduced by the $\llbracket \text{HOW-QUD} \rrbracket$ and the $\llbracket \text{WHICH-QUD} \rrbracket$ are as in (12a) and (12b).

- (12) a. $\text{Partition}(\llbracket \text{HOW-QUD} \rrbracket) = \{\{w_1\}, \{w_2, w_3, w_4\}, \{w_5, w_6, w_7\}, \{w_8\}\}$
- b. $\text{Partition}(\llbracket \text{WHICH-QUD} \rrbracket) = \{\{w_1\}, \{w_2\}, \{w_3\}, \{w_4\}, \{w_5\}, \{w_6\}, \{w_7\}, \{w_8\}\}$

Given (12a) and (12b), the exhaustified prejacent of the ECP-observing (7), presented in (13a), is a member of both $\text{Partition}(\llbracket \text{HOW-QUD} \rrbracket)$ and $\text{Partition}(\llbracket \text{WHICH-QUD} \rrbracket)$. In contrast, any exhaustified prejacent of the ECP-violating (8), one of which is exemplified in (13b), is a member of $\text{Partition}(\llbracket \text{WHICH-QUD} \rrbracket)$ but crucially not a member of $\text{Partition}(\llbracket \text{HOW-QUD} \rrbracket)$. This combined with (11c) predicts the same kind of asymmetry captured by the previous account.

- (13) a. $\text{EXH}(\llbracket \text{every bush has a tiger} \rrbracket) = \{w_8\}$ (cf. $\llbracket \text{every bush has a tiger} \rrbracket = \{w_8\}$)
- b. $\text{EXH}(\llbracket \text{has a tiger} \rrbracket(b_1)) = \{w_2\}$ (cf. $\llbracket \text{has a tiger} \rrbracket(b_1) = \{w_2, w_5, w_7, w_8\}$)

The two accounts outlined so far already generate the asymmetric relevance relations we need. However, the asymmetry predicted by both seems to be weaker than expected, as the ECP-observing (7) is predicted to provide relevant answers to both QUDs. The prejacent of (7) actually seem like an odd answer to the WHICH-QUD (6), but this intuition is not captured by the accounts (unless, as noted above, we posit a domain of atomic individuals for (6)). Intuitively, (7) sounds odd in response to (6), because the questioner of (6) seems to assume

some kind of non-maximality: she thinks that a unique bush or at most two bushes have tigers. The prejacent of (7) goes against this assumption, while the prejacent of (8) satisfies it. Such an intuition can be incorporated into the third account in (14b), which posits $\text{Partition}(Q, cs_Q)$. This partition retains only the cells in $\text{Partition}(Q)$ that are consistent with the questioner's assumptions, as in (14a). This account predicts a stronger asymmetry, where the ECP-observing (7) only provides a relevant answer to the HOW-QUD (this is because unlike $\text{Partition}(\llbracket \text{WHICH-QUD} \rrbracket)$, $\text{Partition}(\llbracket \text{WHICH-QUD} \rrbracket, cs_Q)$ no longer contains $\{w_8\}$), and the ECP-violating (8) only provides a relevant answer to the WHICH-QUD.

- (14) a. **Partition**(Q, cs_Q): $\{P \cap cs_Q : P \in \text{Partition}(Q)\} \setminus \{\emptyset\}$ where cs_Q stands for the context set (set of possible worlds) consistent with the questioner's assumptions
 b. **Relevance** (ver. 3): p is a relevant answer to Q iff: $\text{EXH}(p) \in \text{Partition}(Q, cs_Q)$

The discussion in this section suggests that the relevance intuitions we began with are warranted and can be spelled out in different ways. Having established the needed relevance relations, the next section presents the main experimental study.

4. Experiment

The experiment presented in this section tests the QUD hypothesis outlined in sec. 2. The paradigm it adopts is largely identical to the one already presented in Fig. 1 and (5)–(4) in sec. 2. In the experiment, participants familiarized themselves with a series of situations by reading the prompts and the associated visual stimuli. Against varying contextual backdrops, the main speakers in the target trials uttered sentences of the form: *every X might have a Y*. The sentence was either uttered out of the blue (no clear surrounding dialogue) or in response to an explicit QUD. The situation and the visual stimuli were set up in such a way that the sentence uttered by the main speaker would be interpreted as true only under the ECP-violating scopal ordering (EVERY \succ MIGHT) but false under the ECP-observing scopal ordering (MIGHT \succ EVERY). The main task of the participants was to judge whether the sentences spoken by the main speakers were true or false. In sum, as adumbrated in sec. 2, the core assumption that underlies this experimental design is that we can track participants' ECP violating interpretations in an intuitive way by examining their True/False responses.

4.1. Methods

4.1.1. Participants

600 native speakers of American English were recruited as participants from Amazon Mechanical Turk. They were paid \$0.50 to participate.

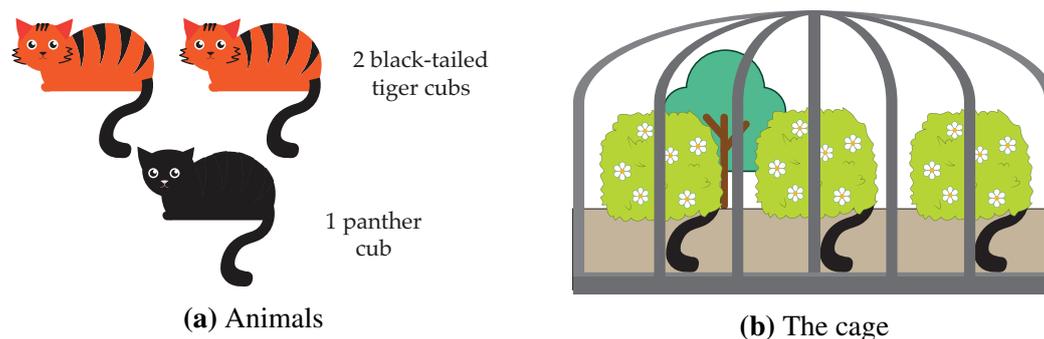


Figure 3: Sample visual stimuli for the three target conditions: HOW, WHICH, and NOQUD
Context prompt: Ron is at a zoo. Ron knows for certain that the cage is currently housing exactly two tiger cubs and exactly one panther cub. Ron also knows for certain that the animals look as in (a). Ron peers through the cage, which looks as in (b).

4.1.2. Materials

The visual stimuli and the prompts for each trial had different configurations depending on whether it was a target condition or a baseline condition. There were three target conditions and three baseline conditions.

The three target conditions, HOW, WHICH, and NOQUD, were associated with an identical range of visual stimuli and context prompts, but differed in the presence vs. absence of an explicit QUD and the type of QUD. The visual stimuli shared the basic paradigm exemplified in Fig. 1. All stimuli established situations in which the main speakers accounted for objects (tiger cubs, cherry toppings, butterfly cocoons, etc.) distributed across containers (bushes, icecream sundaes, beakers, etc.).⁶ For ease of reference, an example of the visual stimuli and a condensed version of the prompt that we already saw in Fig. 1 are reproduced in Fig. 3.

To ensure that a full correspondence is established between ECP violations and True responses on the one hand, and ECP observations and False responses on the other, the speakers' epistemic states were made clear via the visual stimuli and the prompt. In each situation, the speakers were shown to definitively know the cardinality of the items (1 or 2 items distributed across 3 containers) as well as the fact that there is visual ambiguity in identifying them. This information served to constrain the epistemic/doxastic modal bases (Kratzer, 1981) in intended ways, so that as long as the participants incorporated them in their interpretations, the associated modal statements would be evaluated as true under the ECP violating ordering (EVERY \succ MIGHT) but false under the ECP observing one (MIGHT \succ EVERY).

Along with these prompts and visual stimuli, the three target conditions introduced the QUD manipulations summarized in (15). The WHICH condition introduced WHICH-QUDs in the format of *Which of X has a Y?*, such as: *Which of the three bushes has a tiger?* The HOW condition introduced HOW-QUDs in the format of *How many of X have a Y?*, such as: *How many of the three bushes have a tiger?*, and the NOQUD condition did not have any explicit

⁶ The link to the actual experiment is provided in the appendix.

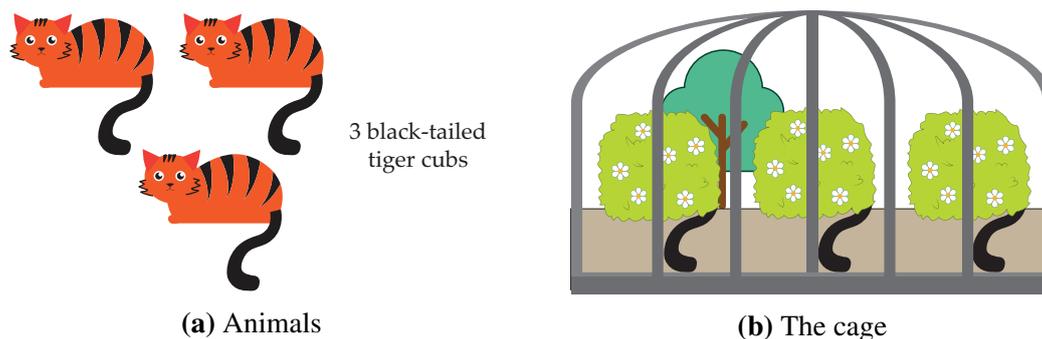


Figure 4: Sample visual stimuli for the TRUE baseline condition

Context prompt: Ron is at a zoo. Ron knows that the cage is currently housing exactly three tiger cubs. Ron also knows that the tigers look as in (a). Ron peers through the cage, which looks as in (b).

QUDs. If an explicit QUD was present, it immediately preceded the target modal sentence.

(15) Three target conditions

a. WHICH condition

A friend arrives and asks the main speaker: *“Which of X has a Y?”*

The main speaker replies: *“Every X might have a Y.”*

b. HOW condition

A friend arrives and asks the main speaker: *“How many of X have a Y?”*

The main speaker replies: *“Every X might have a Y.”*

c. NOQUD condition:

(no explicit QUD; no mention of another friend)

The main speaker says: *“Every X might have a Y.”*

In addition to these three target conditions, the experiment also included three baseline conditions: TRUE, FALSE, and NO-M. As we will see later, these provided analytically useful points of comparison and ensured that the experimental design worked in the intended way. First, in the TRUE condition, visual stimuli and context prompts were designed to generate modal bases that would render the associated modal statements True under *both* the ECP-observing scopal ordering *and* the ECP-violating scopal ordering. The modal statements that were presented were identical to those from the target conditions: *Every X might have a Y*. The statements were uttered by the speaker in the absence of any explicit QUD. A sample prompt and visual stimuli using the same items from Fig. 3 are given in Fig. 4. In this context, Ron’s utterance: *Every bush might have a tiger*. would be evaluated as True regardless of the choice in scopal ordering.

Second, in the FALSE condition, visual stimuli and context prompts were designed to generate modal bases that would render the associated modal statements False under *both* the ECP-observing scopal ordering *and* the ECP-violating scopal ordering. Again, the modal statements that were presented were identical to those from the target conditions: *Every X might have a Y*, and the statements were uttered by the speaker in the absence of any explicit QUD. A sample prompt and visual stimuli are given in Fig. 5. In this context, Ron’s utterance: *Every bush might have a tiger*. would be evaluated as False regardless of the choice in scopal ordering.

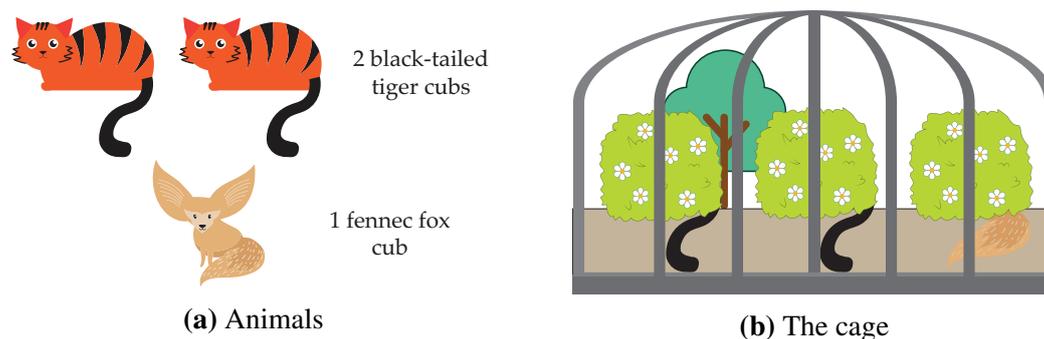


Figure 5: Sample visual stimuli for the FALSE baseline condition

Context prompt: Ron is at a zoo. Ron knows that the cage is currently housing exactly two tiger cubs and exactly one fennec fox cub. Ron also knows that the animals look as in (a). Ron peers through the cage, which looks as in (b).

In particular, the last bush in the cage in Fig. 5b clearly has a fox tail and is incompatible with hiding a tiger; it thus renders the associated modal statement False even under the ECP-violating scopal ordering.

Finally, in the NO-M condition, a range of visual stimuli and prompts that were identical to the ones in target conditions (e.g., Fig. 3) was presented. However, the sentences uttered by the main speakers were non-modal statements in the form of: *Every X has a Y*, such as: *Every bush has a tiger*. Given the context (in particular, the cardinality information such as Fig. 3a), these non-modal statements (which are equivalent to the prejacent of the modals in the ECP-observing MIGHT \succ EVERY interpretations) would be evaluated as False.

4.1.3. Procedure

The experiment had five trials. One of the trials was a filler trial, and the remaining four were target and baseline trials. Each participant saw all three target conditions (WHICH, HOW, NO-QUD) and one of the three baseline conditions (TRUE, FALSE, NO-M). Each of the five trials was associated with five distinct scenarios and visual stimuli, such that no participant saw the same type of scenario/item across conditions/trials.

The filler trials involved cases where the questioner asks identification questions such as *Which pot has a dessert lotus shoot?*, and the main speaker responds with unambiguously true or false answers (given the visual information) involving neither modals nor quantifiers; e.g., *The pot in the middle has a dessert lotus shoot*. Responses to fillers were later checked to confirm that participants paid attention to the experimental tasks.

In each trial, participants answered the questions summarized in (16) after familiarizing themselves with the set-up and the target utterance. The main task was the True/False judgments in (16a), but there was also a gradient rating task (16b) as well as an optional free response question (16c). The experiment lasted an average of 8 minutes.

(16) **Questions in each trial**

- a. Q1: Is what [the speaker] said True or False? (forced choice)
- b. Q2: How confident are you about your response to Q1? (ratings from 0–100)
- c. (Optional) Any comments?

4.2. Predictions

Given the relevance relation argued for in sec. 3, the following predictions emerge for the target conditions: (i) The WHICH condition will elicit significantly more True responses, i.e., more ECP violations, than the HOW condition. (ii) The NOQUD condition will pattern in between the WHICH condition and the HOW condition, as participants may reconstruct a range of different QUDs. The predictions for the baseline conditions are straightforward: the TRUE condition is expected to elicit predominantly True responses; the FALSE and the NO-M condition are expected to elicit predominantly False responses.

4.3. Results

Participants' True/False responses (in %) depending on the 6 conditions are plotted in Fig. 6, along with a summary of the conditions (Table 1). The vertical axis represents the 6 conditions, and the horizontal axis represents percent values. True responses are coded in green, and False responses in red. Error bars represent 95% confidence intervals.

The data were analyzed using a mixed effects logistic regression model with by-participant random intercepts, predicting True/False responses (dependent variable) from the 6 conditions (independent variable). By-situation random intercepts were initially posited as well, but were later dropped as they did not capture any significant variance. The model was fitted using the *lmerTest* package (Kuznetsova et al., 2016) in *R* (R Core Team, 2015). A summary of the fixed effects can be found in the link in the appendix.

4.3.1. Target conditions: QUD effects on the ECP

As predicted, the WHICH condition elicited significantly more True responses, which translates into more ECP violations, than the HOW condition ($\beta = -0.41$, $SE = 0.13$, $z = -3.11$, $p < 0.01$). Fig. 6 captures this: the green bar (True responses) in the WHICH condition is distinctly higher than the one in the HOW condition. In comparison, there was no significant difference between the NOQUD condition and the HOW condition, and between the NOQUD condition and the WHICH condition. While these differences were not significant, Fig. 6 demonstrates that the bars of the NOQUD condition fall somewhere in-between those of the WHICH condition and the HOW condition, as expected.

conditions	QUD and sentence stimuli
NO-M	(No explicit QUD) <i>Every X has a Y.</i>
FALSE	(Fig. 5; no explicit QUD) <i>Every X might have a Y.</i>
HOW	<i>How many of X have a Y?</i> <i>Every X might have a Y.</i>
NOQUD	(No explicit QUD) <i>Every X might have a Y.</i>
WHICH	<i>Which of X has a Y?</i> <i>Every X might have a Y.</i>
TRUE	(Fig. 4; no explicit QUD) <i>Every X might have a Y.</i>

Table 1: Summary of conditions

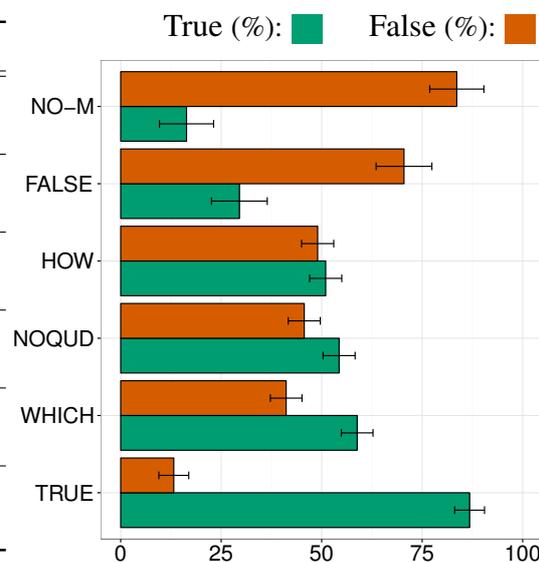


Figure 6: Percent True/False

In sum, the results support the main hypothesis that QUDs significantly affect the ECP: Scopal orderings that provide relevant answers to the QUDs are favored, and this tendency can override the ECP preference (as in the case of the WHICH condition).

More globally, all three target conditions including the HOW condition elicited non-negligible proportion of True responses, indicating ECP-violations. Fig. 6 demonstrates that the proportions of True responses for all three target conditions are above 50%. This is unexpected if the ECP operated as a categorical constraint or even as a strong bias. If this was the case, then we would expect the ECP to be near-categorically observed as long as certain preconditions are satisfied (e.g., in the HOW condition where the QUD bias aligns with the ECP bias and the statements involve subjective interpretations and the quantifier *every*), predicting near-categorical False responses for such conditions. To probe the implications of the surprisingly robust number of True responses, let us conduct a more detailed examination of the baseline conditions and participants’ free responses.

4.3.2. No-M vs. False condition: unrealistic modal bases

The availability of True responses across all three target conditions suggests that the ECP may be violated more easily than previously assumed. In order for us to arrive at this conclusion however, we first need to rule out alternative explanations. In particular, is it possible that the core experimental assumption, namely, that True responses are fully equivalent to violations of the ECP, was not always met? This assumption was grounded on the fact that the experimental trials always provided explicit information about the speakers’ epistemic states. As long as this information was included in creating the modal bases, the ECP-observing orderings could not possibly yield True evaluations; only ECP-violating orderings would allow them.

Nevertheless, participants might occasionally have granted more leeway in constructing the modal bases than is strictly allowed from the visual information. If this is the case, then some of the True responses in the target conditions may correspond not to genuine ECP violations, but rather to cases where participants posited unrealistic modal bases. For example, given a scenario like Fig. 3, certain participants might have included in the modal base worlds in which the ECP-observing prejacent ‘every bush has a tiger’ is true, despite the fact that the visual information and the speaker’s epistemic/doxastic state clearly ruled this out.⁷

Is there a way of probing whether such exceptional interpretations occurred, and if so, how often? Comparing the two baseline conditions: FALSE and NO-M might provide a window into this issue. The FALSE condition, as expected, elicited primarily false responses (more than 70%). However, it also allowed for significantly more true responses than the NO-M condition ($\beta = -1.00$, $SE = 0.35$, $z = 2.83$, $p < 0.01$). This is not expected if we assume that participants always fully included the contextual/visual information when construing the modal bases of the speakers: in the FALSE condition such as Fig. 5, the information in Fig. 5b rules out the ECP-violating interpretation being true, while the information in Fig. 5a (shared by the target conditions) rules out the ECP-observing interpretation being true.⁸ Nevertheless, the FALSE condition occasionally allowed True responses, suggesting that participants may not always have taken this information fully into account. Based on this, we may conclude that analogous cases of laxer construal of modal bases existed, albeit to a small degree, in the three target conditions as well.

4.3.3. Target vs. False condition: no ECP?

However, the significant number of True responses in the three target conditions cannot all be reduced to exceptional cases where participants posited laxer, unrealistic modal bases. If this were the case, we would at least expect these conditions to pattern with the FALSE condition. In other words, the rate at which such laxer construals occurred (and manifested as True responses despite observing the ECP) would be at best equivalent to the FALSE condition, and most likely lower.⁹ As it turns out however, not only the WHICH condition, but also the NOQUD and the HOW condition elicited significantly more True responses than the FALSE condition (e.g., comparing the HOW condition with the FALSE condition: $\beta = 1.17$, $SE = 0.22$, $z = 5.23$, $p < 0.001$). This suggests that significant parts of the True responses in all three target conditions are indeed reflections of genuine ECP violations.

This state of affairs is corroborated further by participants’ free responses. A few comments that unambiguously confirm the availability of the ECP violating scopal interpretations (*every*

⁷ Since the target modal sentences did not have any explicit adverbials like ‘based solely on *X*’ (cf. Portner, 2007), it seems highly unlikely that this happened frequently.

⁸ The partial visual information given in Fig. 5b introduces weaker, more defeasible information than the information about cardinality given in Fig. 5a. So it is also possible that many of the True responses in the FALSE condition reflect participants’ uncertainty about the information given in Fig. 5b, rather than indicate that participants reconstructed unrealistic modal bases.

⁹ This is because as mentioned in footnote 8, the information about cardinality such as Fig. 3a is a much stronger piece of information and is thus harder to ignore than the partial visual information such as Fig. 5b.

\succ *might*) are presented in (17).¹⁰ All 3 target conditions elicited some amount of free responses in the vein of (17), confirming the general availability of ECP violations across conditions.

- (17)
- a. “I think she means each statue has the potential to have a blue sapphire not that all of them will.”
 - b. “It is tricky to know if George means there are 3/3 moss butterflies or (if he means) each beaker (could) possibly contain one.”
 - c. “this is true because you don’t know which one it is. Obviously one of them doesn’t have a maraschino cherry but it’s possible that the one you choose will.”
 - d. “Every bush *MIGHT* have one until you find the two and then the last bush would not have one.”

In sum, the results for the three target conditions suggest that the ECP violating interpretations are more available than previously assumed, and confirm the main hypothesis that the QUD-based scopal biases boost this availability further.

4.3.4. Target vs. True condition: evidence for the ECP

Faced with this rather pervasive availability of ECP-violating scopal orderings, one may begin to wonder if our intuitions about the ECP were perhaps illusory. However, the results of the experiment also suggest that some kind of bias that works towards creating the intuition behind the ECP does exist. If the ECP-violating scopal orderings were as available as the ECP-observing ones, then we would expect the three target conditions to pattern more like the TRUE condition (modulo the effects of QUDs), as participants would have had full access to the scopal ordering that renders the statement true. Since people tend to prefer interpretations that render the statement true when there is ambiguity (Gualmini et al., 2008; cf. Meyer and Sauerland, 2009), we would expect participants to predominantly choose the ECP-violating ordering if there weren’t any bias against it. However, all three conditions also elicited significantly fewer True responses than the TRUE condition (e.g., comparing the WHICH condition with the TRUE condition: $\beta = -1.92$, $SE = 0.21$, $z = -9.07$, $p < 0.001$); and the proportion of True responses for the three target conditions fell somewhere in-between the FALSE condition and the TRUE condition, differing significantly from either of the baselines. This suggests that there exists some kind of gradient preference towards the ECP-observing scopal ordering that is activated to a different degree depending on various factors (one of which, as we saw, is the QUD).

4.3.5. Variability of the intuition behind the ECP

To some extent, scopal preferences varied depending on the language user. Comments like (17) indicate that certain participants had no trouble resorting to ECP-violating scopal interpretations. In comparison, comments like (18) indicate that other participants were more strongly

¹⁰ The full list of free responses is provided in the .csv file that can be accessed via the link in the appendix.

biased against them.¹¹

- (18) a. “*Interesting. This depends on how you interpret ‘every pot might...’ I would lean towards (this meaning) that all 3 pots would have a desert lotus shoot and not just that each pot might potentially have a desert lotus root.*”
 b. “*The more I think about it I guess every bush MIGHT have a tiger cub but it just is not the correct way to say this.*”

Participants’ certainty ratings (i.e., their answers to Q2) also provide indirect evidence for the existence of this variability.¹² If participants shared essentially the same kind of scopal preference, and if the contrast between True vs. False responses (and also between the comments in (17) and (18)) are just manifestations of their stochastic decisions when faced with ambiguity, we would expect their certainty ratings for the three target conditions to be significantly lower than those for the three baseline conditions. However, no significant difference in certainty ratings emerged across the 6 conditions when a mixed effects regression model was fitted (with certainty ratings as the dependent variable and conditions as the independent variable; and with by-participant random intercepts). This suggests that proponents of (17) vs. (18) were certain about their respective intuitions, which varied significantly from each other.

5. Discussion

The current experimental study provides evidence for the view that the ECP is at best a de-feasible/gradient bias whose manifestation is subject to a variety of factors (cf. Anand and Hacquard, 2008), one of which is the QUD.

Does this ‘ECP bias’ have an independent presence, or can it be reduced to a combination of more primitive lexical and pragmatic biases? The paper suggests that the latter option is more likely, and that the ECP bias primarily reflects lexical biases associated with various quantifiers and/or modals. It is widely known that epistemic modals like *might* tend to take wide scopes. It is also known that different types of quantifiers are susceptible to the ECP to different degrees: *each* seems to more easily allow ECP violations than *every* (Tancredi, 2007);¹³ and *every* in turn seems to more easily allow ECP violations than *all*.¹⁴ This state of affairs can be straightforwardly captured if one posits that quantifiers come pre-equipped with different scopal preferences. More specifically, they prefer or allow wide scopes in the following order: *each* > *every* > *all*. Since *might* prefers to take wide scope and *every* doesn’t, something like the ECP bias is predicted to emerge. Such a view is indirectly supported by

¹¹ In addition to user-dependent variation, other factors such as intonation also seem to influence the activation of ECP, as indicated by comments such as “*It depends on if she emphasizes the word ‘every’ or ‘might’.*”

¹² Since each participant only did a single trial for each condition, we cannot know if a given participant would have provided consistent responses across multiple trials of the same condition. If this turned out to be the case, and if distinct groups of answer patterns were to emerge, we would have more direct evidence.

¹³ This intuition was confirmed via a separate experimental study that mainly tested the effects of evidential adverbials on the ECP. Since the study is not directly relevant to the main argument of the current paper, it is not presented here. However, the results as well as the experiment itself can be found in the link in the Appendix.

¹⁴ The present account would predict that ECP-violations would be even less available for the quantifier *all* (compared to *every*), although not entirely inaccessible if other contextual cues like QUDs strongly point towards ECP-violating interpretations.

participants' comments like (19). These highlight the role of lexical alternatives in shaping how the ECP bias manifests itself across different types of quantifiers.

- (19) a. *The correct phrase should be 'each pot might have a desert lotus shoot'. The word 'every' implies that all pots inclusively together have desert lotus shoots.*
 b. *The leaves on both kinds of plants have the same appearance. So it's possible that any of the pots could have a lotus shoot. Though saying 'every' isn't as clear as saying 'any'.*

For instance, (19a) suggests that *each* is a better choice than *every* for conveying $\forall \succ \diamond$, most likely because *each* is associated with a stronger bias towards taking a wide scope than *every*; it is thus a less ambiguous option for signaling the intended meaning. The availability of *each* in conveying the ECP-violating proposition seems to give rise to additional pragmatic inferences that further strengthen the ECP bias associated with *every*: Listeners may reason that if the speaker had intended to convey the ECP-violating proposition, she likely would have used *each*. Since she opted for *every* instead, she most likely intends to convey the ECP-observing proposition (unless QUDs or other contextual information indicate otherwise).

As a final note, we may want to discuss what to make of the apparent discrepancy between the current experimental results (which highlight the defeasibility of the ECP) and the strong ECP intuitions initially reported in von Stechow and Iatridou (2003). There are various possible explanations for this. First, the implicit QUDs evoked by von Stechow and Iatridou (2003)'s examples might have been more along the line of HOW-QUDs than WHICH-QUDs. Second, providing explicit questions (which were absent in von Stechow and Iatridou (2003), although the contextual settings they introduced were fairly rich otherwise) along with the modal statements might have been the real game-changer and thus greatly boosted the availability of ECP violating interpretations in the experiment. Finally, the participants in the current experiment might have become more attuned to the latent ECP-violating scopal ordering through the course of their exposure to statements involving quantifiers and modals across multiple trials. The paper remains agnostic as to what could have been the most important source of the discrepancy. The main take-away seems to be that the ECP-violating interpretations cannot be entirely ruled out, although in many cases, diverse factors will conspire against them.

6. Conclusion

This paper presented an experimental study of the effects of explicit QUDs on the ECP. Based on the experimental results, it argued that the ECP can be recast as a gradient scopal preference which arises from a confluence of more primitive lexical and pragmatic biases.

References

- Anand, P. and V. Hacquard (2008). Epistemics with attitude. In *Semantics and Linguistic Theory*, Volume 18, pp. 37–54.
 Beaver, D. and B. Clark (2008). *Sense and sensitivity: How focus determines meaning*. Wiley–Blackwell.

- Dayal, V. (2016). *Questions*. Oxford: Oxford University Press.
- Fox, D. (2017). Exhaustivity and the presupposition of questions - comments on dayal (1996). In *Talk given at Sinn und Bedeutung 26*.
- Ginzburg, J. (1996). Dynamics and the semantics of dialogue. In J. Seligman and D. Westerstahl (Eds.), *Language, Logic, and Computation*, Volume 1, pp. 221–237. Stanford, CA: CSLI Lecture Notes. CSLI.
- Groenendijk, J. and M. Stokhof (1984). *Studies on the Semantics of Questions and the Pragmatics of Answers*. Ph. D. thesis, University of Amsterdam.
- Gualmini, A., S. Hulsey, V. Hacquard, and D. Fox (2008). The question–answer requirement for scope assignment. *Natural language semantics* 16(3), 205–237.
- Hamblin, C. L. (1971). Mathematical models of dialogue. *Theoria* 37(2), 130–155.
- Kaufmann, M. (2016). Fine-tuning natural language imperatives. *Journal of Logic and Computation*.
- Klinedinst, N. and D. Rothschild (2011). Exhaustivity in questions with non-factives. *Semantics and Pragmatics* 4(2), 1–23.
- Kratzer, A. (1981). The notional category of modality. *Words, worlds, and contexts*, 38–74.
- Kuznetsova, A., P. B. Brockhoff, and R. H. B. Christensen (2016). Imertest package (version 2.0.30).
- Lyons, J. (1977). *Semantics*, Volume 2. Cambridge: Cambridge University Press.
- Meyer, M.-C. and U. Sauerland (2009). A pragmatic constraint on ambiguity detection. *Natural Language & Linguistic Theory* 27(1), 139–150.
- Portner, P. (2007). Imperatives and modals. *Natural Language Semantics* 15(4), 351–383.
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.
- Roberts, C. (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics. *The Ohio State University (OSU) Working Papers in Linguistics*, 91–136.
- Simons, M., J. Tonhauser, D. Beaver, and C. Roberts (2010). What projects and why. In *Semantics and linguistic theory*, Volume 20, pp. 309–327.
- Tancredi, C. (2007). A multi-modal theory of I-semantics. Manuscript, University of Tokyo.
- von Stechow, K. and A. S. Gillies (2011). 'might' made right. In A. Egan and B. Weatherson (Eds.), *Epistemic Modality*, pp. 108–130. New York: Oxford University Press.
- von Stechow, K. and S. Iatridou (2003). Epistemic containment. *Linguistic Inquiry* 34(2), 173–198.

7. Appendix

Links to all the experiments, data, and statistical models can be found at:
<https://github.com/sunwooj/ecp/>