

Attentional Pragmatics: a new pragmatic approach to exhaustivity

Matthijs Westera

Submitted for review, January 2020

Word count: 12117

Abstract. Where the neo-Gricean approach to exhaustivity is based on the idea that exhaustivity arises when relevant propositions are not *asserted*, this paper explores a new pragmatic approach based on the idea that exhaustivity arises when relevant propositions are not *mentioned*, or more precisely, when the speaker did not intend to draw attention to them. This seemingly subtle shift from information to attention results in different predictions on a range of challenges for the neo-Gricean approach, of which this paper discusses three: exhaustivity on the hints of a quizmaster, exhaustivity on questions, and exhaustivity without an opinionatedness assumption. The resulting theory, Attentional Pragmatics, is compared in detail to both the neo-Gricean approach and the grammatical approach to exhaustivity, although it is inevitable that substantial parts of the empirical domain are either treated elsewhere or left for another occasion. A number of interesting new challenges are identified for the grammatical approach.

Keywords: pragmatics, exhaustivity, disjunction, attention, alternatives, Quantity, questions, opinionatedness, hints.

1 Introduction

Exhaustivity is a widespread phenomenon whereby, roughly, hearing one thing can lead one to infer the negation of something stronger, e.g., from a disjunction the negation of a conjunction:¹

- (1) Alf was at the party, or Beth. (*Implied: not both*)

The utterance implies that Alf and Beth weren't both there, even though under a classical, logical treatment of "or" and "and" this doesn't follow, and in natural language too it is possible to add "...and maybe both" without contradiction. This phenomenon is variably known as exhaustivity, scalar or quantity inference, implicature or implication.

¹Disjunctive examples in this paper are intended to carry focus accents on each disjunct and falling intonation at the end.

For many years the only approach to exhaustivity implicature was the neo-Gricean one, which formalizes the following intuitive explanation: if the speaker had believed the stronger information that both Alph and Beth attended, then they would have asserted that – since the speaker didn't, they must believe it's false. However, the last two decades have arguably witnessed its demise, and popularity in a pragmatic approach to exhaustivity more generally has faded. The engine behind this is the grammatical approach to exhaustivity, which, motivated by certain challenges for the neo-Gricean approach, has sought to bypass it altogether and treat exhaustivity, instead, as the result of a grammatical operation. This paper argues that the neo-Gricean approach's demise (and the grammatical approach's success) should not entail that we give up exploring pragmatic routes to exhaustivity more generally, and develops one such route in detail.

The new route, called *Attentional Pragmatics* because it puts attention center stage, solves a number of challenges for the neo-Gricean approach. The core idea is quite simple, and but a minor deviation from the intuition underlying the neo-Gricean approach:

- **Neo-Gricean approach:** Exhaustivity arises when relevant propositions aren't *asserted*.
- **Attentional Pragmatics:** Exhaustivity arises when relevant propositions aren't *mentioned*.

The crucial but perhaps superficially subtle difference is that that one can mention an alternative without asserting it. For instance, in uttering a disjunction one mentions the disjuncts without asserting either one; and in asking a question one can mention various propositions without asserting anything whatsoever. Of course the relevant notion of 'mentioning' will have to be more precisely characterized, which is done in Attentional Pragmatics in terms of the notion of *attention*. The heart of Attentional Pragmatics is a set of maxims requiring, roughly, that a reasonable speaker should – in addition to trying to assert all (and only) relevant propositions they consider true, as required by the Gricean maxims – also aim to draw attention to all (and only) relevant propositions they consider possible.

The current paper presents and motivates Attentional Pragmatics, and compares it to both the neo-Gricean approach and the grammatical approach with regard to three empirical puzzles: the appearance of exhaustivity on the hints of a quizmaster (Fox 2014), on questions (e.g., Biezma and Rawlins 2012), and in contexts where the speaker is not assumed to be opinionated (Westera 2014, cf. Sauerland 2004). Elsewhere (Westera 2017a, 2020a[under review]) I highlight how the shift from the neo-Gricean approach to Attentional Pragmatics also enables a better pragmatic treatment of Hurford disjunctions, i.e., disjunctions where one disjunct entails the other, and there I offer a detailed comparison also between Attentional Pragmatics and the grammatical approach in this regard. This topic had to be left out from the present paper for reasons of scope.

Section 2 situates this work in the literature by providing an overview of Gricean pragmatics, existing approaches to exhaustivity (the neo-Gricean approach and the grammatical approach), and existing work on attention in semantics/pragmatics. Section 3 defines and motivates the substance of Attentional

Pragmatics, namely, a set of maxims that govern what a reasonable speaker should and shouldn't draw attention to, and shows how this predicts exhaustivity. Section 4 compares the resulting account of exhaustivity to the existing approaches with regard to the three puzzles mentioned above. Section 5 concludes.

2 Background and related work

2.1 Gricean pragmatics

It is important to distinguish the sum of all information provided by an utterance from the information which the speaker *intended* to provide (i.e., what the speaker *meant*), and that in turn from the information which the utterance expressed by virtue of its semantics alone. I will refer to these notions as *informational potential*, *informational intent*, and *informational content*, where the “intent”/“content” division, adopted from Bach and Harnish 1979, is a snappy rewording of Grice's (1967b) “utterer's meaning”/“sentence meaning”. This terminology has the advantage of generalizing to attention in the current paper (i.e., attentional potential, intent and content), and consistently wielding it will help clarify what Attentional Pragmatics does and doesn't commit us to.

The core of Gricean pragmatic theory is a set of conversational maxims that govern which informational intents are rational/cooperative for a given utterance, given the speaker's beliefs about what is true and relevant. These maxims require, in a nutshell, that informational intents be true, relevant, sufficiently informative and clearly conveyed. (In practice one often finds the maxims being applied to the semantic content instead of the intent – a consequence in part of fact that the content/intent distinction isn't always explicitly maintained; see Bach 2001; Westera and Boleda 2019 for discussion.) I will refer to these maxims as the I(nformation)-maxims, anticipating the definition of a set of A(ttention)-maxims in section 3 which form the spine of Attentional Pragmatics. I will make both sets of maxims explicit to allow a precise comparison between the neo-Gricean approach and Attentional Pragmatics.

For the I-maxims I adopt the exact definition given in Westera 2017b, which in turn closely follows Grice's (1975) original proposal and subsequent implementations. In particular, the maxim of (I-)Quantity, which is the starting point of the neo-Gricean approach, will be equivalent in the relevant details to the version in Harnish 1976 (for a discussion of the details of the present definition and alternatives I refer to Westera 2017b). Following Carlson 1982; Ginzburg 1996; Roberts 1996 among others, compliance with the maxims is defined relative to a question under discussion (QUD). This embodies the assumption that the set of all pieces of information that are broadly speaking worth sharing in a given discourse is subdivided into topically and strategically organized subsets – typically paraphrasable as questions – where any given informational intent needs to be relevant to only one such subset, i.e., be an answer to one QUD. Formally, the I-maxims are defined as functions taking an informational intent p (proposition) and a question under discussion Q (set of propositions) and returning a truth value representing whether the maxim is complied with or

not:²

The I-maxims:

1. I-QUALITY(p) = $\Box\vee p$
“Intend to share only propositions you take to be true.”
2. I-RELATION(p, \mathcal{Q}) = $\mathcal{Q}(p)$
“Intend to share only propositions in the QUD.”
3. I-QUANTITY(p, \mathcal{Q}) = $\forall q \left(\left(\begin{array}{l} \text{I-QUALITY}(q) \wedge \\ \text{I-RELATION}(q, \mathcal{Q}) \end{array} \right) \rightarrow (p \subseteq q) \right)$
“Intend to share all propositions in the QUD you take to be true.”

I refer to Westera 2017b for an overview and justification of various formal and empirical consequences of this exact definition, e.g., that a speaker will always know whether a given informational intent complies with the I-maxims or not; that if there exists a compliant intent then it is the only one; and that some but not all maxims can clash, i.e., not be jointly satisfiable, which yields predictions concerning the way in which speakers indicate clashes.

As mentioned, the above definition of the I-maxims follows existing proposals. But I have found that two of its features in particular often evoke questions from a critical audience, and as such require a brief clarification. First, I-Quantity (as in, e.g., Harnish 1976) is so demanding that so-called “mention-some” contexts can be represented only by assuming a restriction of the QUD (and a domain restriction of the question that introduces it), say, to those propositions that are the most relevant for practical purposes, or to those propositions that come to the answerer’s mind first. As an account of mention-some contexts this would be incomplete, of course, without an explanation of how an addressee may figure out the intended restriction of the QUD. But as Schulz and Van Rooij (2006) note such an explanation is required regardless of how one seeks to represent the mention-some/mention-all distinction, e.g., pragmatically or by assuming a semantic ambiguity, hence in the rest of this paper it can be set aside.

Second, since I-Relation does not explicitly make negations, disjunctions or intersections of answers to the QUD relevant, nor indirect, partial or merely probable answers, these would either have to be assumed to be proper elements of the QUD itself (as I will assume for intersections below), or be accounted for by other means. For instance, one could try to explain why an indirect answer can be an appropriate discourse move by assuming that although its primary intent does not directly resolve the QUD and as such does not comply with I-Relation, the utterance has a secondary intent – a conversational implicature – that does. Alternatively one could invoke (independently motivated) QUD shifts, such as a shift from the question of what the world is like to a question of what your epistemic state is like – the latter seems to be needed to explain why “I’m not sure” is a felicitous response to basically any question. For reflection

²The definition uses Montague’s (1973) Intensional Logic, but with doxastic modality. Since the formulas in this paper will be fairly simple (and always paraphrased by English) there is no need to separately introduce the formalism here. Just a quick reminder will do: $\vee p$ can be read as “proposition p is true”, $\wedge \varphi$ as “the proposition that φ ”, $\Box \varphi$ as “the speaker believes that φ ” and $\Diamond \varphi$ as “the speaker considers it possible that φ ”.

on these possible choices and their consequences I must refer again to Westera 2017b.

2.2 Existing approaches to exhaustivity

2.2.1 The neo-Gricean approach: I-Quantity plus opinionatedness

Exhaustivity has been approached within pragmatics in what has come to be known as the *neo-Gricean approach*. To illustrate, consider again the example from the introduction:

- (1) Alf was at the party, or Beth. (*Implied: not both*)

The neo-Gricean approach is based on the following intuition: if the speaker had believed the stronger information that both Alph and Beth attended, then according to the maxim of I-Quantity they would have asserted that – since the speaker didn’t, they must believe it is false.

To see how this works formally, we need to specify the primary informational intent p for (1) and its QUD \mathcal{Q} .³

$$\mathcal{Q} = \{\wedge Pa, \wedge Pb, \wedge Pab\} \quad p = \wedge(Pa \vee Pb)$$

The QUD contains the propositions denoted by the disjuncts and their intersection (i.e., Alf and Beth’s joint presence, $\wedge Pab$). The inclusion of the disjuncts ought to be derivable with the help of a theory of focus intonation – the accents on the disjuncts are crucial for getting a reading compatible with this type of QUD (e.g., Roelofsen and Van Gool 2010; Meertens 2019) – but for reasons of scope I will simply presuppose it in this paper. The inclusion in the QUD of their intersection aligns with the common assumption that QUDs are by default closed under intersection (e.g., Schulz and Van Rooij 2006; Spector 2007), an assumption warranted by the fact that (i) the intersection of two propositions p and q in a given QUD plausibly has whichever features p and q share by virtue of which they were grouped in the same QUD (e.g., subject matter, strategic use), and (ii) if information p is worth making common ground, and likewise for information q , then their intersection is certainly worth making common ground as well. Nevertheless, closure under intersection is only a default assumption, permissive of exceptions. For instance, if the person who introduces a new QUD already knows that the intersection is false, there is no reason for them to include it in the QUD. In section 4.2 we will see an empirical consequence of this possibility.

With the assumed QUD and informational intent we can prove that the following holds, in a certain relevant subclass of models (formally: those models which validate the definitions of the maxims, along with a number of principles such as

³For convenience let Pab denote $Pa \wedge Pb$. It will also (sometimes) be convenient to conceive of functions to truth values as (their canonical) sets, and I will use set-theoretical notation with its usual meaning where this improves legibility (this can be defined within Intensional Logic itself; Zimmermann 1989).

the introspection axioms for belief and that the speaker knows what the QUD is; for formal proofs see Westera 2017b):

$$\text{I-QUANTITY}(p, Q) \rightarrow \neg \Box Pab$$

That is, if the utterance complies with I-Quantity, the speaker must not have the belief that both Alf and Beth were at the party. Note that this I-Quantity implication is almost exhaustivity, but not quite: from the absence of a positive belief ($\neg \Box Pab$) one may not in general conclude the presence of a negative belief ($\Box \neg Pab$) – the speaker could simply be ignorant, after all. That the I-Quantity implication falls short of exhaustivity was already pointed out by Soames (1982: p.534), in a discussion of Gazdar’s (1979) account. More recently, Sauerland (2004) called the gap between I-Quantity and exhaustivity, i.e., between not-believing and believing-that-not, the *epistemic step*. To explain how participants in a conversation take the epistemic step, Soames proposed that they normally assume each other’s *opinionatedness* (Horn (2001) notes that this proposal can be found already in Mill 1867). Indeed, if we assume for whatever reason that the speaker in (1) takes herself to know whether or not both were at the party, the I-Quantity implication can be strengthened, yielding exhaustivity:

$$\neg \Box Pab \wedge \overbrace{(\Box Pab \vee \Box \neg Pab)}^{\text{opinionatedness assumption}} \rightarrow \Box \neg Pab$$

I-Quantity plus opinionatedness is the essence of the neo-Gricean approach (e.g., Horn 1972; Gazdar 1979; Schulz and Van Rooij 2006; Spector 2007; Geurts 2011).⁴

The neo-Gricean approach has been adopted by many, but the last decade or two have witnessed its apparent demise. The main engine behind this has been the literature proposing what has come to be known as the *grammatical approach* to exhaustivity implicature (e.g., Fox 2007; Chierchia et al. 2009, 2012; Katzir and Singh 2013; Mayr and Romoli 2016). The grammatical approach assumes that exhaustivity is generated directly by covert grammatical operators – I will summarize it further below. Sticking with the neo-Gricean approach for now, the most central argument raised within the grammatical approach against it is that it would be unable to account for cases where exhaustivity effects seems to exist in a grammatically embedded position. For instance, consider embedding example (1) in a belief context (though with the same intonation as before):

- (2) Gemma believes that Alf or Beth was at the party.
(Implied: Gemma believes that not both were there)

The utterance has a reading where the exhaustivity effect “not both” is itself embedded under belief. This would be problematic for the pragmatic approach, which operates on informational intents – communicative intentions – as these

⁴The neo-Gricean approach applies not just to disjunctions but to any situation in which the QUD contains a stronger proposition than what is asserted, including so-called scalar cases such as “some” implying “not all” and “warm” implying “not hot”, and the same will be true of Attentional Pragmatics. This paper is focused on disjunction for the sake of exposition, and also because one of the puzzles in section 4 concerns questions, and non-disjunctive propositional questions with falling intonation are pragmatically marked (e.g., they have an “assertive” flavor; Bartels 1999).

are (by and large) had by entire utterances and not by their parts (but cf. Simons 2011). For reasons of scope I must background the argument from embedded exhaustivity in the current paper; Attentional Pragmatics will be compared to the neo-Gricean and the grammatical approach only on a set of independent challenges (section 4). However, given its centrality in the literature, I owe the reader a justification for this choice (as well as references to places where the argument is more properly addressed). I turn to this next, before discussing the grammatical approach itself in a bit more detail.

2.2.2 Backgrounding embedded exhaustivity (in this paper)

I think that the centrality of the argument from embedded cases in the pragmatics/grammar debate (i) is not entirely warranted, (ii) has stalled the pragmatics/grammar debate and (iii) has, by putting pragmatics on the defensive, caused some of the grammatical approach's own assumptions to be insufficiently scrutinized from a pragmatic perspective. I will briefly justify these three claims.

Concerning (i), the reason why embedded exhaustivity is not a knock-down argument against the pragmatic approach is that there are various other pathways, besides the grammatical approach, that can lead to seemingly embedded exhaustivity implicatures (Geurts, 2011), e.g., different types of pragmatic implicature, pragmatic implicature (Bach, 1994), typicality inference (e.g., Van Tiel 2014), lexical semantics, and interactions of these mechanisms. In fact, different modal verbs give rise to different kinds of exhaustivity effects, suggesting that the lexical semantics of these predicates has at least some role to play (e.g., Uegaki 2015), either instead of or in interplay with grammar and/or pragmatics. Many of these alternative pathways are perfectly compatible with maintaining a pragmatic approach for *unembedded* exhaustivity, and it is an open and in part empirical issue whether the various exhaustivity effects can and should be treated as a unified phenomenon at all (e.g., Russell 2006; Geurts 2011; Simons 2011).

As for (ii), the notion that embedded exhaustivity would present a knock-down argument against the pragmatic approach has caused interest in the latter to fade. This is unfortunate, because without developing the pragmatic approach further we will not gain a proper understanding of the explanatory power we might lose by replacing pragmatics by grammar – most scholars seem to agree that a pragmatic approach should be favored *if it would work*. Another important reason to keep exploring the pragmatic approach is that developments on the pragmatic side may carry over to the grammatical approach. Although the current proposal is motivated pragmatically, there is nothing intrinsically ‘pragmatic’ about the core idea itself, i.e., that exhaustivity arises not when relevant propositions aren't asserted, but when no attention is drawn to them. In principle a grammatical exhaustivity operator could be defined as being sensitive not just to the informational content of its complement, but also to its attentional content. Indeed, Alonso-Ovalle (2008) defines an operator that is sensitive to the Alternative Semantic value of its complement (e.g., the disjuncts of a disjunction), but its potential has not been explored much within the grammatical approach.⁵

⁵It could impact for instance the way in which the grammatical approach handles disjunctions (e.g., Hurford disjunctions), as well as the complex interactions between exhaustivity and

A connection between grammatical and pragmatic exhaustification is also conceivable at a more explanatory level, at least under a (partly) functionalist and/or diachronic perspective on language. If a grammatical exhaustivity operator is to be understood as a silent morpheme, we must explain how it entered the lexicon (across languages), and one possible explanation is that it arose as a grammaticalization of a pragmatic inference. In that case, understanding what the original pragmatic inference may have been like is crucial to understanding the resulting operator (even if since then it may of course have diverged from it further). This should lend some additional value to the present undertaking, even for those already sold on the grammatical approach.

Lastly, (iii), the focus on embedded exhaustivity has put the pragmatic approach on the defensive – but there are (of course) things about the grammatical approach that warrant pragmatic scrutiny, as the next subsection will cover. I think that the pragmatic approach’s fading popularity is due in part to these aspects, discussed next, not having received the attention from pragmatics they deserve.

2.2.3 The grammatical approach

The grammatical approach assumes covert operators in the grammatical structure of a sentence that directly render their complement exhaustive relative to some (grammatically determined/constrained) set of alternatives (e.g., Fox 2007; Chierchia et al. 2009, 2012; Katzir and Singh 2013; Mayr and Romoli 2016). For instance, for disjunctions such as (1) the grammatical approach assumes an exhaustivity operator on each disjunct:

- (3) EXH(Alf was at the party) or EXH(Beth was at the party).

The result of interpreting this structure is roughly paraphrasable as “Only Alf was at the party or only Beth”, which indeed has the desired “not both” implication (provided the disjuncts are each other’s alternatives, to be excluded by EXH). Exhaustivity operators and grammatically constrained alternatives have turned out to afford the grammatical approach the flexibility to tackle an impressive range of data, including of course embedded cases of exhaustivity but also superficially unrelated phenomena such as free choice (Fox 2006). But for present purposes I will concentrate on three ingredients that I think raise issues from a pragmatic perspective, and which seem not to have received the attention they deserve. Note that my goal here is not to prove that these are somehow unsurmountable problems, but merely to suggest that the lack (to my awareness) of sufficient justification/explanation for these aspects provides some further reason to keep exploring alternative (pragmatic) accounts of exhaustivity alongside it, i.e., some additional motivation for the present paper.

First, by making exhaustivity part of the main informational content of an expression, it becomes part of the main informational intent, i.e., what is asserted, as opposed to an indirect, secondary intent, i.e., what is conversationally implicated in the sense of Grice 1967a. This marks a major deviation from

interrogativity.

prior characterizations of exhaustivity, but it seems to have slipped under the radar, perhaps in part because it has been obfuscated by the continued use of the term “implicature” in the grammatical approach (cf. Geurts 2011).⁶ Major as this deviation may be, it has not to my awareness been empirically assessed or otherwise defended other than by showing that making exhaustivity part of the informational content allows one to model *the overall truth-conditions* (disregarding the assertion/implicature distinction) of various unembedded and embedded exhaustivity effects with a single operator. This is a great result, but one should wonder whether the distinction between main and secondary intent (implicature) can in the end somehow be recovered and, if not, what exactly we would lose.

Second, in order to regulate the presence/absence of (covert) exhaustivity operators the grammatical approach relies on the Strongest Meaning Hypothesis, that in case of ambiguity one should opt for the strongest possible interpretation – hence insert an exhaustivity operator whenever this strengthens the overall meaning (e.g., Chierchia et al. 2012; Mayr and Romoli 2016; based on Dalrymple et al. 1998). From a pragmatic perspective this assumption is suspicious in two ways. First, delivering information *on top of* the primary assertion is precisely what a pragmatic approach would do; conversational implicature by definition is a secondary speech act, one which comes on top of the first, hence which cannot weaken the overall information provided. The fact that the grammatical approach needs to essentially stipulate as basic something so intrinsic to the pragmatic approach should raise an eyebrow. Second, the Strongest Meaning Hypothesis is not rational from a purely probabilistic perspective – the strongest interpretation will often not be the most probable one unless abiding by this rule is stipulated to somehow be a convention of the language. Probabilistically, one should opt not for the strongest interpretation but for the interpretation that is the most likely, i.e., the one which most likely corresponds to what the speaker had in mind (combining what is uttered with priors about what is likely true, knowledge about what is likely relevant, and so on). Empirically, too, it isn’t obvious that the Strongest Meaning Hypothesis is on the right track. It was originally formulated as a model of the interpretation of (plural, reciprocal) noun phrases (Dalrymple et al. 1998), but in that domain it has recently been shown to yield the wrong predictions: Poortman (2017) argues on empirical grounds for an alternative, the Maximum Typicality Hypothesis, which happens to align better with a probabilistic perspective, too.

Third and last, another principle that regulates when exhaustivity operators ought to be present is Hurford’s Constraint, which states that disjunctions are infelicitous if one disjunct entails the other – accordingly, insert an exhaustivity operator if this can serve to break such an entailment (Chierchia et al. 2009, after Hurford 1974). For instance, in the disjunction “Alf, or Beth, or both”, the third disjunct entails the other two, violating Hurford’s Constraint, unless the other two disjuncts are interpreted exhaustively (as “only Alf” and “only Beth”, respectively). Initially a mere descriptive generalization, numerous authors have since sought to motivate Hurford’s Constraint in terms of redundancy (e.g., Singh 2008; Katzir and Singh 2013; Mayr and Romoli 2016): if one disjunct

⁶The pragmatic approach itself carries part of the blame, for having watered down the notion of implicature by using it essentially interchangeably with “implication” and “inference” (as criticized in Bach 2006).

entails another, then it does not contribute to the overall informational content (reflected by the logical absorption laws: $p \vee (p \wedge q) \equiv p$), hence it would be redundant and therefore the utterance as a whole infelicitous.

Plausible as this may seem, an important consequence of adopting this motivation for Hurford’s Constraint is pointed out by Ciardelli and Roelofsen (2017): the appeal to redundancy is warranted only within a semantic/pragmatic framework that is sufficiently coarse-grained for the entailing disjunct to indeed be redundant (i.e., a semantics that satisfies the absorption laws). This is true for instance of classical (information-only) semantics, as well as Inquisitive Semantics, which models only the minimally informative answers to a question (Ciardelli et al., 2013). But it is not the case for systems that have been used in pragmatics, such as Alternative Semantics for disjunction (Alonso-Ovalle 2006; or equivalently Unrestricted Inquisitive Semantics of Ciardelli 2009), and in Dynamic Semantics where the disjuncts introduce discourse referents (Schulz and Van Rooij, 2006). Conceiving of Hurford’s Constraint in terms of redundancy is incompatible also with the attention-drawing use of disjunction on which the account of exhaustivity in the present paper will depend: disjuncts such as “or both” can crucially make an attentional contribution even when informationally redundant. Again, this is not intended as an argument against the grammatical approach; but it shows that adopting Hurford’s Constraint (at least in the absence of an alternative, non-redundancy-based explanation) comes at a cost, in particular from a pragmatic perspective, and this should encourage us to keep exploring avenues that can do without it. For a detailed formal application of Attentional Pragmatics to (unembedded) Hurford disjunctions see Westera 2017a; for more discussion, an application to embedded cases and a detailed comparison to the grammatical approach I must refer to Westera 2020a[under review].

The foregoing serves as additional motivation for the present aim to continue developing the pragmatic side of the literature on exhaustivity, which this paper will do in the shape of Attentional Pragmatics. I will return to both the neo-Gricean and the grammatical approach in section 4, where these are compared to Attentional Pragmatics on several empirical puzzles. However, these puzzles will be mostly unrelated to the challenges posed by embedded exhaustivity, as well as to the three aspects of the grammatical approach discussed above.

2.3 Previous work on attention in semantics/pragmatics

The importance of attention in semantics/pragmatics is quite generally acknowledged. For instance, the interpretation of pronouns is guided in part by considering which individuals are attentionally salient, i.e., which are available discourse referents. More relevantly to the present undertaking, the common idea that expressions such as disjunction introduce their disjuncts as alternatives (e.g., Alonso-Ovalle 2006; Ciardelli et al. 2009) can also be given an attentional understanding: they draw attention to their disjuncts – e.g., Aloni (2001) models this precisely by assuming that disjunction introduces its disjuncts as discourse referents.

Central to the current paper is the idea that attention drawing not merely happens and affects interpretation, but happens *intentionally* – i.e., it is a

type of communicative intention in its own right, just like sharing information. Parallelling the informational notions, let the *attentional potential* of an utterance be the set of all things to which it draws attention, let an *attentional intent* be a subset of that to which the speaker intended to draw attention (in the sense of a communicative intention), and let the *attentional content* be the set of things to which an utterance draws attention by virtue of its semantics alone. In these terms, that utterances have attentional contents and potentials is uncontroversial. It is the assumption that utterances have attentional intents that sets Attentional Pragmatics apart from most work; though see Ciardelli et al. (2009), who likewise put attention on a par with information in this regard. What we gain from the assumption that utterances have attentional intents is that it gives us a new anchor for pragmatic explanation. Attentional intents being communicative intentions, they should be governed by a set of constraints in their own right, a set of attentional maxims, and these will form the core of Attentional Pragmatics.

On the informational side it is common to derive the primary informational intent directly from a semantic notion of informational content that requires only a minimum amount of contextualization (e.g., fixing referents of pronouns, quantifier domain restrictions). Similarly, on the attentional side one could invoke a separate dimension of the semantics to deliver the attentional intents required by Attentional Pragmatics. A suitable semantic backbone in this regard would be Alternative Semantics for disjunction (Alonso-Ovalle, 2006) or, equivalently, Unrestricted Inquisitive Semantics (Ciardelli et al. 2009), as the attentional intents assumed in this paper (and beyond) correspond to the semantic objects these systems assign to the relevant sentences. However, we may not need a designated attentional semantics to begin with: it can be formally shown that in many cases (utterances in disjunctive normal form being a particularly relevant subclass) the attentional intent of an utterance can be reliably identified by reasoning about the QUD, the informational intent and the maxims, without assuming a special attentional dimension of the semantics. For formal proofs of this finding, as well as an assessment of the suitability of Alternative Semantics and other options as a semantic backbone for Attentional Pragmatics, I must refer to Westera 2017b, chapter 6.

Some work on Attentional Pragmatics has appeared or is currently under review, although this is the first manuscript in journal format. My PhD thesis (Westera, 2017b) provides the most comprehensive overview. The predictions of Attentional Pragmatics at a purely formal level are described more succinctly in Westera 2017a, where its core prediction is captured in an exhaustivity operator that is conservative in relevant respects with regard to previous operators in the literature. Meertens 2019 relies on Attentional Pragmatics to account for the contrast between polar and alternative questions. In Westera 2020b it is used to explain a difference between the “not both” implication of disjunctive assertions and disjunctive questions, a puzzle to which I will briefly return in section 4.2. Lastly, in Westera 2020a[under review] a detailed comparison is offered of pragmatic approaches including Attentional Pragmatics to the grammatical approach with regard to unembedded and embedded Hurford disjunctions.

3 Attentional Pragmatics

3.1 Definition of the A(ttention)-Maxims

The A-maxims follow the same general recipe as the I-maxims introduced above, relative to a QUD \mathcal{Q} , this time for an attentional intent \mathcal{A} :

The A-maxims:

1. A-QUALITY(\mathcal{A}) = $\forall a(\mathcal{A}(a) \rightarrow \diamond^{\vee} a)$
 “Intend to draw attention only to propositions that you consider possible.”
2. A-RELATION(\mathcal{A}, \mathcal{Q}) = $\forall a(\mathcal{A}(a) \rightarrow \mathcal{Q}(a))$
 “Intend to draw attention only to relevant propositions.”
3. A-QUANTITY(\mathcal{A}, \mathcal{Q}) = $\forall a \left(\left(\begin{array}{c} \text{A-QUALITY}(\{a\}) \wedge \\ \text{A-RELATION}(\{a\}, \mathcal{Q}) \end{array} \right) \rightarrow \mathcal{A}(a) \right)$
 “Intend to draw attention to all relevant propositions you consider independently possible.”

Once the assumption is granted that attention-drawing is a type of communicative intention, I trust that these maxims are intuitively plausible, e.g., it will be generally beneficial for a rational interlocutor to keep track of all (A-Quantity) pertinent discourse goals (A-Relation) that can still be achieved (A-Quality). Moreover, very similar constraints have been assumed in the prior literature. A maxim like A-Quality is assumed by Roelofsen (2013; building on Ciardelli et al. 2009), who calls it “Attentive Sincerity”; it corresponds also to “Genuineness” in Zimmermann 2000 (p.270), and to the “Viability” constraint in Biezma and Rawlins 2012 (p.46); a similar intuition underlies Gazdar’s (1979) “clausal implicatures”. Biezma and Rawlins moreover assume a constraint like A-Relation, though incorporated as a presupposition in their question semantics; Simons (2001) assumes a comparable “relatedness condition” on disjunctions, and in fact the same idea is found already in Grice 1989, who states that disjunction serves to specify possibilities “that relate in the same way to a given topic”.

The above presentation of Attentional Pragmatics is slightly simplified compared to Westera (2017b, a). The full-fledged A-maxims include a requirement that every proposition in the attentional intent should be considered not only possible (A-Quality), but possible independently of any stronger proposition in the attentional intent.⁷ This is motivated as a type of economy requirement: drawing attention to a proposition a is redundant if attention is already drawn to a more specific proposition b and it is known that a cannot be true without b being true. To illustrate, if you know that either Alph is at the party or all three of Alph, Beth and Gemma, but never only two of them, then uttering (4a) is fine while uttering (4b) is bad:

⁷In Westera 2017b, a this is achieved by inserting a separate maxim, but the same can be achieved more succinctly by refining A-Quality:

$$\text{A-QUALITY}(\mathcal{A}) = \forall a(\mathcal{A}(a) \rightarrow \diamond(\vee a \wedge \forall b((b \subset a \wedge \mathcal{A}(b)) \rightarrow \neg^{\vee} b)))$$

- (4) a. Alph is at the party, or Alph, Beth and Gemma.
 b. ?? Alph is at the party, or Alph and Beth, or Alph, Beth and Gemma.

Example (4b) is bad because the second disjunct is not considered possible independently of the third. The full-fledged definition matters for Hurford disjunctions such as these, including in particular those like (4a) with “distant entailing disjuncts” (Fox and Spector 2018), but for present purposes we can use the simpler definition (see Westera 2017a; Westera 2020a[under review] for discussion).

3.2 Deriving exhaustivity

For the aims of this paper it is sufficient to consider how exhaustivity is predicted for the simple disjunction from the introduction, repeated here:

- (1) Alf was at the party, or Beth. (*Implied: not both*)

To apply Attentional Pragmatics we need to specify the primary attentional intent \mathcal{A} , in addition to the informational intent p and QUD \mathcal{Q} given earlier:

$$\mathcal{Q} = \{\wedge Pa, \wedge Pb, \wedge Pab\} \quad p = \wedge(Pa \vee Pb) \quad \mathcal{A} = \{\wedge Pa, \wedge Pb\}$$

That is, I assume that the speaker intends to draw attention to the propositions expressed by the two disjuncts. This attentional intent corresponds to previous assumptions about disjunction in, e.g., Aloni 2001; Schulz and Van Rooij 2006; Alonso-Ovalle 2006; Ciardelli et al. 2009 (see my comments on identifying attentional intents in section 2.3).

With the assumed QUD and attentional intent for (1) the following can be proven:

$$\text{A-QUANTITY}(\mathcal{A}, \mathcal{Q}) \rightarrow \Box \neg Pab$$

Concise proof: that no attention is drawn to Alf and Beth’s joint presence ($\wedge Pab \notin \mathcal{A}$) implies, by contraposition of the conditional of A-Quantity, that it must either not be relevant ($\neg \text{A-RELATION}(\{\wedge Pab\}, \mathcal{Q})$) or not be considered possible ($\neg \text{A-QUALITY}(\{\wedge Pab\})$), and since we know it is in fact relevant, the explanation can only be the latter. That is, assuming compliance with the maxim of A-Quantity entails that the speaker must believe that not both Alf and Beth were at the party, which is of course the exhaustivity implicature we set out to explain.

Attentional Pragmatics furthermore predicts that the way to prevent an exhaustivity implicature is by including a third disjunct “or both” in order to draw attention to Alf and Beth’s joint presence:

- (5) Alf was at the party, or Beth, or both.

By drawing attention to Alf and Beth’s joint presence (i.e., ensuring that $\wedge Pab \in \mathcal{A}$), this utterance satisfies the consequent of A-Quantity and even implies, through A-Quality, that the speaker considers their joint presence possible,

which is the polar opposite of exhaustivity ($\diamond Pab \equiv \neg \square \neg Pab$). Altogether, the theory does justice to the motivating intuition given in the introduction: that exhaustivity arises when a relevant proposition isn't mentioned. For a more general specification of the predictions of Attentional Pragmatics I refer to Westera 2017a, where a formal exhaustivity operator is defined as a shorthand for what the maxims achieve, and compared to existing operators in the literature.⁸ In circumstances where the I-maxims and A-maxims are complied with the operator is formally conservative with respect to existing operators, in particular the operator in Schulz and Van Rooij 2006 (which for descriptive coverage was stipulated to take the individual disjuncts into account).

4 Comparison with regard to three puzzles

4.1 Exhaustivity on the hints of a quizmaster

An interesting puzzle for the neo-Gricean approach is presented in Fox 2014. When a quizmaster is giving a hint to a contestant, the quizmaster is not expected to share all the information they have – hints are partial answers by definition. Another way of framing this is that hints are exempt from the maxim of I-Quantity; not all true relevant information needs to be shared. And yet, surprisingly for the neo-Gricean approach which relies on it, exhaustivity can still be present:⁹

- (6) Quizmaster: (Of these three boxes over here,) there is money in box A or in box C. (*implied: not in both, and not in B*)

Exhaustivity is present here: if the quizmaster had known that there was money in box B, or in both A and C, they could be accused of misleading the contestant. And yet no genuine I-Quantity implication is present here to the effect that, e.g., the quizmaster necessarily lacks the belief that the money is in box A – in fact the quizmaster may well know exactly that that's where the money is. Accordingly, as Fox notes, there is no way for the neo-Gricean approach, which relies crucially on I-Quantity, to predict exhaustivity for (6). Moreover, the reasonable assumption that the quizmaster is at least *pretending* to be complying with I-Quantity cannot help in this case, as Fox notes: if the I-Quantity implication is pretended, there is no way for an implication that is derived from it to not also be merely pretended – but the exhaustivity is genuine.

Since Attentional Pragmatics derives exhaustivity from A-Quantity instead of I-Quantity, one could try to account for (6) by supposing that, while quizmasters

⁸For the sake of completeness, the operator in Westera 2017a is defined as follows:

$$\text{EXH}(\mathcal{A}, \mathcal{Q}) = \bigcap_{\substack{a \in \mathcal{Q} \\ a \notin \mathcal{A}}} (\bar{a} \cup \bigcup_{\substack{b \in \mathcal{A} \\ b \subset a}} b)$$

⁹Intonation is again crucial – example (6) ought to be read, as before, with falling intonation and focus on the two disjuncts. I want to stress this here because an alternative intonation is particularly natural for hints, and partial answers more generally: fall-rise (Constant, 2012; Wagner, 2012; Westera, 2019). Fall-rise being a signal for anti-exhaustivity, reading the example that way would appear to contradict Fox's argument that exhaustivity is present, so this intonation must be avoided.

are not supposed to provide all the information in their possession, perhaps they do have to mention all possible choices for a quiz participant, i.e., comply with A-Quantity, lest they be found guilty of misleading. If true, this would deliver exhaustivity as before, through A-Quantity. However, while I find it quite intuitive that a quizmaster’s hints would be exempt from I-Quantity but not from A-Quantity, it isn’t entirely satisfactory without an explanation for why the different exemption status of the different maxims.¹⁰ Puzzlingly, it doesn’t seem to instantiate a broader pattern of I-maxims vs. A-maxims, because the maxims of I/A-*Quality* seem to behave in the opposite way: while (6) is exempt from A-Quantity – the quizmaster needn’t actually consider it possible that there is money in box A – it isn’t exempt from I-Quantity, i.e., there has to be money in box A or box C.

As an explanation for this exemption/non-exemption pattern, and for what is going on in (6), I propose the following:

- Quizmasters may pretend to know *less* about the quiz answers than they actually do, but not to know *more* about the quiz answers than they actually do (i.e., information that may be false).

This is definitely plausible; one could imagine reading this as part of the job instructions for a quizmaster. Crucially, it follows that pragmatic implications of not-knowing ($\neg\Box$) need not be genuine, whereas pragmatic implications of knowing (\Box) must be genuine, and this accounts precisely for the apparent exemption status of the different maxims:

- I-Quantity: implies knowing (\Box), hence hints are subject to it.
- I-Quantity: implies not-knowing ($\neg\Box$), hence hints are exempt.
- A-Quantity: implies not-knowing (\Diamond , which is equivalent to $\neg\Box\neg$), hence hints are exempt.
- A-Quantity: implies knowing ($\neg\Diamond$, which is equivalent to $\Box\neg$), hence hints are subject to it.

Given Attentional Pragmatics, the prediction that hints are exempt from I-Quantity whilst subject to A-Quantity solves precisely the puzzle posed by (6), where exhaustivity is present without a genuine I-Quantity implication.

In fact, the foregoing line of explanation works for any theory that delivers the knowing-that-not implication of exhaustivity directly, rather than via a not-knowing implication as in the neo-Gricean approach. This is true of Attentional Pragmatics, but also of the grammatical approach: the crucial part of the explanation is the assumption that a quizmaster’s implied knowledge about the quiz answers must be genuine, and this includes exhaustivity regardless of whether it is delivered by A-Quantity or by a grammatical operator plus I-Quantity. This corroborates Fox’s arguments that cases like (6) favor the grammatical approach over the neo-Gricean approach. But it refines them by showing that a pragmatic alternative is available too, namely, Attentional Pragmatics.

¹⁰Moreover, the intuition that a quizmaster isn’t exempt from A-Quantity isn’t really to be trusted: it may well be a mere consequence of the explanandum, i.e., that hints imply exhaustivity, rather than an explanation for it.

4.2 Exhaustivity in questions

Where a quizmaster’s hints are a somewhat exotic example of an utterance that is exempt from I-Quantity, questions are a more ordinary case. Questions are exempt from I-Quantity, or all the I-maxims for that matter, because they do not have a primary informational intent to which these maxims could apply, or in other words they lack assertive force (e.g., Frege 1918; Krifka 2011). Nevertheless, questions have been noted to imply exhaustivity (e.g., Bartels 1999; Roelofsen and Van Gool 2010; Biezma and Rawlins 2012):¹¹

(7) Was Alf at the party, or Beth? *(implied: not both)*

Questions’ exemption from I-Quantity makes their implying exhaustivity problematic for the neo-Gricean approach. By contrast, Attentional Pragmatics again offers a straightforward starting point for an explanation: although questions do not serve to provide information, they arguably do serve to draw attention to things, and as such are expected to be subject to the A-maxims just like assertions. Indeed, by drawing attention to relevant propositions one would effectively be indicating (because they are relevant) the desire that they be made common ground, which is conceivably the primary purpose of (information-seeking) questions – see Biezma and Rawlins 2012 for an account of questions in line with this view. Thus, since (7) is subject to A-Quantity just like the assertion in (1) with which this paper started, it is expected to imply exhaustivity in exactly the same way: since no attention is drawn to Alf and Beth’s joint presence, the speaker must not consider it possible.

Simple as this may be, it isn’t quite right. A precondition for the foregoing explanation is that Alf and Beth’s joint presence is relevant, which earlier I motivated, following previous work, by assuming that QUDs are closed under intersection. But here we need to be careful. Closure of QUDs under intersection is a default assumption, assumed to be true of QUDs in the absence of reasons to believe otherwise. As such, it is a safe assumption when it concerns a QUD that was not explicitly introduced, e.g., the QUD of an assertion in isolation, but it may not be safe when it concerns a QUD that is made explicit. The latter is the case in (7): the speaker who utters the interrogative disjunction is the one introducing the QUD, and they would not include a given intersection in their QUD if they already knew it to be false. It follows from this that the intersection cannot be relevant, hence that the exhaustivity of (7) must be explained otherwise.

Let me go through this explanation with a bit more precision. Suppose counterfactually (to obtain a contradiction) that Alf and Beth’s joint presence is relevant in (7), i.e., is an element of the QUD. In that case, since no attention was drawn to it, compliance with A-Quantity would imply that the speaker must consider it false – this would be exhaustivity in the usual manner. But if the speaker already knew the intersection was false, then – since they are the one introducing the QUD – why would they have included it? From Roberts 2012 I adopt the assumption that QUDs should not contain propositions that the

¹¹As before the disjunction needs to be read with focus on the disjuncts and a final fall, i.e., the type of intonation triggering an ‘alternative question’ interpretation.

person introducing them already knows are false – or more generally, one should not set goals that one already knows are unachievable (found for instance in Bratman 1987). Therefore, contrary to the initial supposition, Alf and Beth’s joint presence cannot be relevant in (7), and exhaustivity must be explained slightly differently: the speaker must not consider Alf and Beth’s joint presence possible because otherwise they would have included it in the QUD (and drawn attention to it).

The foregoing suggests that whereas exhaustivity in the case of assertions is *the exclusion of relevant propositions*, exhaustivity in the case of questions is *the exclusion of propositions that would have been relevant had they been deemed possible*. Both explanations rely crucially on A-Quantity and the fact that no attention was drawn to Alf and Beth’s joint presence, but they diverge from there, based on the fact that closure under intersection applies differently when the speaker is the one introducing the QUD. The idea that exhaustivity in questions is primarily about unmentioned propositions not being relevant (rather than, as for assertions, unmentioned relevant propositions not being true) is independently motivated in Biezma and Rawlins 2012, and the resulting overall explanation for the ‘not both’ implication is similar to the one sketched in Groenendijk and Roelofsen 2009, namely, that if the speaker had considered “both” relevant they would have listed it as a possible answer.

But if the end result is the same for assertions and questions, i.e., a “not both” implication, then what difference does it make? For reasons of scope I will summarize just one empirical consequence, and refer to Westera 2020b for a more detailed argumentation. It is often assumed that in the case of assertions the “not both” implication, and exhaustivity more generally, is not just an implication but a conversational implicature, i.e., an indirectly communicated informational intent. By contrast, the same implication in the case of questions is commonly assumed to be a presupposition (Bartels 1999; Aloni and Égré 2010; Biezma and Rawlins 2012), i.e., it is implied by the utterance to be true but not part of what the speaker meant, i.e., not an informational intent. The explanation presented in Westera 2020b is, in a nutshell, as follows. In the case of assertions, “both” is relevant, i.e., establishing it is a conversational goal, hence establishing its negation is a secondary conversational goal – it is important to prune unachievable goals and keep the conversation focused on those which are achievable.¹² Assuming that secondary goals are grouped in a secondary QUD, this is exactly the type of QUD that the “not both” implicature would serve to address (since implicatures are informational intents in their own right, they too must be relevant to a QUD), which explains why it is reasonable to have this informational intent in the case of assertions. By contrast, in the case of questions “both” is not relevant to begin with, hence there is no secondary QUD containing “not both”, hence no QUD that could be addressed by means of the “not both” implication, which therefore does not serve the communication of an intent.

Notwithstanding these interesting differences in the way exhaustivity is explained for questions compared to assertions, the main point of this subsection is that

¹²This view on the conversational purpose of the “not both” implication aligns with Horn 1989: the typical reason for conveying negative information is the prior consideration of its positive counterpart – we tend not to be interested in negative information in its own right.

questions are subject to A-Quantity but not I-Quantity, hence deriving exhaustivity from A-Quantity instead of I-Quantity lets Attentional Pragmatics extend to questions better than the neo-Gricean approach. What about the grammatical approach? Questions have only recently begun to draw some attention there, e.g., Ciardelli and Roelofsen (2017) suggest that Hurford’s Constraint, one of the core principles in the grammatical approach, holds for both questions and assertions – though they do not seek to account for the exhaustivity implications of questions. The grammatical approach treats exhaustivity as a semantic entailment, and as such directly predicts “not both” to be part of what is meant in the case of assertions, i.e., as part of what is asserted (the primary informational intent), and its truth is implied accordingly by the maxim of I-Quality. But in the case of questions nothing is asserted, so the maxim of I-Quality doesn’t have anything to apply to (or alternatively, in some accounts interrogativity turns the informational content/intent into a tautology), and accordingly any exhaustivity generated below the interrogativity does not end up being implied to be true. A more technical way of putting this is that entailments don’t “project” out of interrogativity, so neither does exhaustivity as conceived in the grammatical approach. It is technically possible of course to define a special-purpose exhaustivity operator for interrogatives, one which makes exhaustivity escape interrogativity as a presupposition (Biezma and Rawlins (2012) define such an operator, though independently of the grammatical approach to exhaustivity), but such special treatment isn’t desirable; it fails to be explanatory unless independent empirical motivation and ideally explanation is given for this different treatment.

Summing up, Attentional Pragmatics accounts for exhaustivity in a way that generalizes from assertions to questions, while also predicting some differences, in particular that the “not both” implication serves to communicate an intent in the case of assertions but not questions. Meanwhile, exhaustivity in questions poses a challenge for the neo-Gricean approach, and it isn’t clear at present how the grammatical approach would handle it, either.

4.3 Exhaustivity without an opinionatedness assumption?

Recall that the neo-Gricean approach relies on an opinionatedness assumption (e.g., for (1), the assumption that $\Box Pab \vee \Box \neg Pab$), because the I-Quantity implication on its own ($\neg \Box Pab$) falls short of exhaustivity ($\Box \neg Pab$) – this is the epistemic step (Sauerland, 2004). By contrast, Attentional Pragmatics predicts exhaustivity without relying on an opinionatedness assumption; the A-Quantity implication is strong enough as it is. The reason for this is that A-Quantity is more demanding than I-Quantity: it requires that attention be drawn not just to propositions that are believed to be true, but to any propositions that are considered even merely possible. One might wonder whether this is an advantage or disadvantage: what is the explanatory value of an opinionatedness assumption to begin with, and can exhaustivity occur without it?

The explanatory value of the opinionatedness assumption is unclear at best. To my awareness it has not been argued to be an independently reasonable assumption, and the neo-Gricean approach’s reliance on an opinionatedness assumption has been criticized by some for being ad hoc (Groenendijk and Stokhof 1984; Chierchia et al. 2012; Westera 2014). In fact, it doesn’t seem

rational to go about the world assuming, without evidence, that our interlocutors are necessarily opinionated about whatever the QUD happens to be. A more rational strategy, at least from a probabilistic standpoint, would be to assume opinionatedness only where this seems (almost) certainly true, and to simply not assume it whenever it is anything less than almost certainly true. Nevertheless, the opinionatedness assumption has been embraced by many (e.g., Soames 1982; Horn 1984; Matsumoto 1995; Green 1996; Russell 2006; Schulz and Van Rooij 2006; Geurts 2011).

This appreciation of the epistemic step is due largely to the fact that exhaustivity implications seem to disappear if the speaker denies their own opinionatedness, as Soames (1982) noted:

- (8) A: Who was at the party?
B: I'm not sure about Alf, but Beth was there.

B's hedge, denying their opinionatedness about Alf, causes the usual exhaustivity implication "not Alf" to be absent. Similarly explicit examples can be found throughout the literature. Breheny et al. (2013) consider less explicit cases, where the speaker's (un)opinionatedness is set up by means of a description of the context, and present experimental results using these cases that seem to support the reliance of exhaustivity implications on an opinionatedness assumption. They presented participants with utterances that could measurably trigger or not trigger an exhaustivity implication, while contextually manipulating how much knowledge the speaker has about the QUD. When the participants knew that the speaker had only partial knowledge about the QUD, the exhaustivity implications were significantly weaker than when no such knowledge was present. Goodman and Stuhlmüller (2013) present comparable results.

However, there is reason to believe that these experimental results as well as Soames's original observation have been misinterpreted. After all, in the relevant examples (such as (8), but also the more contextualized cases of Breheny et al. 2013) not only is the supposed opinionatedness assumption *denied*; it is in fact *replaced by an un-opinionatedness assumption*, or even knowledge to that effect. And since to be *unopinionated* is to be unable to give an exhaustive answer, it is unsurprising that knowing in advance that the speaker is *unopinionated* has some effect on the perceived exhaustivity implications, regardless of whether their presence would have depended on an opinionatedness assumption to begin with. After all, when we know that the speaker is not opinionated about Alf, we know in advance that we must not interpret the answer "Beth" as excluding Alf's presence, and we will try to find a way of interpreting the utterance that is compatible with this uncertainty. For instance, we may interpret the main part of B's response in (8) as addressing an implicitly more restricted QUD: "but *of the people I do know about* Beth was there". As long as there is a way to accommodate the speaker's unopinionatedness about part of the QUD, there will be a reading of the utterance that lacks exhaustivity – and this is true even if exhaustivity itself does not depend on an opinionatedness assumption.

A better test case for the neo-Gricean approach's reliance on an opinionatedness assumption would be to deny not the speaker's opinionatedness, but merely the

opinionatedness *assumption*, and see if exhaustivity still appears. The following example is such a test case (I considered a similar example in Westera 2014):

- (9) A: I *may* be asking the wrong person – you probably don’t know this –
but do you have any idea who was at the picnic?
B: John was there, Bob, Mary, and Sue. (*with falling intonation*)
A: So you do know! Great, thanks. Only four though, that’s a bit
disappointing...

Here, A explicitly refrains from assuming B’s opinionatedness, though crucially (unlike (8)) without assuming their *unopinionatedness* instead, and in the end A still takes B’s answer to be exhaustive. When presenting an analogous example (but without A’s final response) to nine informants, in a simple questionnaire with auditory stimulus, eight agreed that B’s response implies exhaustivity (with strength 5 out of 5). A proper experimental investigation lies outside the present scope, but this tentative result at least beats the purported but misinterpreted evidence in favor of an opinionatedness assumption. Combined with its being essentially ad hoc this speaks in favor of an approach that does not rely on it, such as Attentional Pragmatics or the grammatical approach (see below).

A possible objection to my interpretation of (9) is that B’s opinionatedness may be conveyed by means of her falling intonation contour, a suggestion found in Hara 2005; Schulz and Van Rooij 2006 – and indeed falling intonation tends to sound more authoritative and hence opinionated than a final rising pitch (e.g., Ohala, 1983). Now, while it is true that intonation plays a role in exhaustivity and that falling intonation sounds opinionated, this does not tell us much about the direction of the inferential/explanatory arrow between opinionatedness and exhaustivity – exhaustivity implies opinionatedness, after all. In fact, a long strand of work on intonational meaning treats rises/falls as indicating not opinionatedness but whether the utterance is pragmatically a “complete” contribution to the discourse (e.g., Pierrehumbert and Hirschberg 1990), a notion that is made more precise for instance by Malamud and Stephenson (2015) in terms of meta-pragmatic issues remaining unresolved, and in Westera 2017b, 2018 in terms of a failure to fully comply with the Gricean maxims. While permitting an important role for intonation in conveying opinionatedness, this strand of accounts of intonational meaning still points to something like the maxims as the real source of exhaustivity (and *thereby* opinionatedness). By contrast, there is no general theory of intonational meaning, to my awareness, that would directly associate falling intonation with opinionatedness; it remains an ad hoc assumption with the sole purpose of bridging the gap between I-Quantity and exhaustivity.

As Schulz and Van Rooij (2006) note, a challenge for any account based on opinionatedness (whether an opinionatedness assumption or through intonation) is that it needs to deliver *the right amount* of opinionatedness. For instance, for the plain disjunction in (1), the speaker is opinionated about Alf and Beth’s joint presence but not about Alf’s individual presence, or Beth’s for that matter, otherwise the speaker could have and should have asserted the relevant disjunct(s). Indeed, regardless of whether opinionatedness is a contextual assumption, conveyed intonationally and/or implied by compliance with the maxims, it must

somehow be sensitive to the things mentioned in the utterance, say, the things to which attention is drawn. This calls for an explanation, and Attentional Pragmatics provides one: the right amount of opinionatedness of a speaker follows from the requirement that speakers draw attention to all relevant propositions they consider possible (A-Quantity), a constraint that is not a mere reformulation of opinionatedness, but which can be understood as a trait of general rational behavior: it is important to keep track of the propositions that are to be made (and can still be made) common ground, regardless of whether one is opinionated or not.

I will end this section by considering the grammatical approach, on which the foregoing may seem to have little or no bearing. After all, the grammatical approach bypasses I-Quantity and opinionatedness altogether: exhaustivity (hence the right amount of opinionatedness) would be generated by a grammatical operator and implied to be true (as part of the main informational intent) simply through the maxim of I-Quality. Nevertheless, the grammatical approach relies on a very similar assumption for predicting when an exhaustivity operator ought to be present to begin with, namely the Strongest Meaning Hypothesis: an operator is present whenever this strengthens the overall meaning of the utterance. Schulz and Van Rooij (2006) frame the opinionatedness assumption in a way that is strikingly similar: that one should assume that the speaker is as informed as possible about the QUD – we could call this a “Strongest Beliefs Hypothesis” to emphasize the parallelism with the Strongest Meaning Hypothesis. In fact, the Strongest Meaning Hypothesis can be derived from the “Strongest Beliefs Hypothesis” (i.e., opinionatedness) together with I-Quantity: since I-Quantity requires that whichever information a speaker has about the QUD must be part of what the speaker means, assuming the strongest beliefs about the QUD entails assuming the strongest meaning. Moreover, my argument in section 2 why the Strongest Meaning Hypothesis does not seem plausible is similar to the reason why the opinionatedness assumption is ad hoc. In light of the foregoing it seems desirable to try to come up with a grammatical approach that does not rely on it – just to see what is possible – but I must leave exploring this direction for another occasion.

5 Conclusion

This paper presented Attentional Pragmatics, a pragmatic theory based on the idea that exhaustivity arises when relevant propositions are not mentioned, or more precisely, when the speaker did not intend to draw attention to them. By contrast, the neo-Gricean approach is based on the more common idea that exhaustivity would arise when relevant propositions are not *asserted*. This superficially subtle difference results in very different predictions on a range of challenges for the neo-Gricean approach, of which I could discuss three in the scope of this paper: exhaustivity on the hints of a quizmaster, exhaustivity on questions, and exhaustivity without an opinionatedness assumption. The differences suggest that, if exhaustivity is a pragmatic phenomenon, it is a matter of attention rather than of information, i.e., it derives from A-Quantity, not I-Quantity.

At the outset I motivated why we need to continue developing pragmatic approaches to exhaustivity despite the challenges posed within the grammatical approach, foremost embedded exhaustivity (but see Westera 2020a[under review]). I also emphasized that the core idea in this paper is not intrinsically pragmatic – one could imagine a grammatical approach to exhaustivity that is sensitive to the attentional component of utterances. Given the persistence of considerations of informational strength also in the grammatical approach, the findings in this paper invite a re-evaluation of the extensive experimental and theoretical literature on exhaustivity, both pragmatic and grammatical.

References

- Aloni, M. (2001). *Quantification under Conceptual Covers*. Ph. D. thesis, University of Amsterdam.
- Aloni, M. and P. Égré (2010). Alternative questions and knowledge attributions. *The Philosophical Quarterly* 60(238), 1–27.
- Alonso-Ovalle, L. (2006). *Disjunction in Alternative Semantics*. Ph. D. thesis, University of Massachusetts, Amherst.
- Alonso-Ovalle, L. (2008). Innocent exclusion in an alternative semantics. *Natural Language Semantics* 16, 115–128.
- Bach, K. (1994). Conversational implicature. *Mind and Language* 9, 124–62.
- Bach, K. (2001). You don’t say. *Synthese* 128, 15–44.
- Bach, K. (2006). The top 10 misconceptions about implicature. In *Drawing the boundaries of meaning: Neo-Gricean studies in pragmatics and semantics in honor of Laurence R. Horn*, pp. 21–30. John Benjamins Publishing Company.
- Bach, K. and R. Harnish (1979). *Linguistic communication and speech acts*. Cambridge, Massachusetts and London: MIT Press.
- Bartels, C. (1999). *The intonation of English statements and questions: a compositional interpretation*. Routledge.
- Biezma, M. and K. Rawlins (2012). Responding to alternative and polar questions. *Linguistics and Philosophy* 35(5), 361–406.
- Bratman, M. (1987). *Intentions, plans, and practical reason*. Cambridge, Mass.: Harvard University Press.
- Breheny, R., H. Ferguson, and N. Katsos (2013). Taking the epistemic step: Towards a model of on-line access to conversational implicatures. *Cognition* 126, 423–440.
- Carlson, L. (1982). Dialogue games: an approach to discourse analysis. *Synthese Language Library* 17.
- Chierchia, G., D. Fox, and B. Spector (2009). Hurford’s constraint and the theory of scalar implicatures. *Presuppositions and implicatures* 60, 47–62.
- Chierchia, G., D. Fox, and B. Spector (2012). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In C. Maienborn, P. Portner, and K. von Stechow (Eds.), *Semantics: An International Handbook of Natural Language Meaning*, Volume 2, pp. 2297–2332. Mouton de Gruyter.
- Ciardelli, I. (2009). Inquisitive semantics and intermediate logics. Master Thesis, University of Amsterdam, available at <http://dare.uva.nl/document/443433>.

- Ciardelli, I., J. Groenendijk, and F. Roelofsen (2009). Attention! *Might* in inquisitive semantics. In S. Ito and E. Cormany (Eds.), *Semantics and Linguistic Theory (SALT) 19*.
- Ciardelli, I., J. Groenendijk, and F. Roelofsen (2013). Inquisitive semantics: a new notion of meaning. *Language and Linguistics Compass* 7(9), 459–476.
- Ciardelli, I. and F. Roelofsen (2017). Hurford’s constraint, the semantics of disjunction, and the nature of alternatives.
- Constant, N. (2012). English rise-fall-rise: a study in the semantics and pragmatics of intonation. *Linguistics and Philosophy* 35(5), 407–442.
- Dalrymple, M., M. Kanazawa, Y. Kim, S. Mchombo, and S. Peters (1998). Reciprocal expressions and the concept of reciprocity. *Linguistics and Philosophy* 21(2), 159–210.
- Fox (2006). Free choice and the theory of scalar implicatures. From his 2007 book.
- Fox, D. (2007). Free choice disjunction and the theory of scalar implicatures. In U. Sauerland and P. Stateva (Eds.), *Presupposition and implicature in compositional semantics*, pp. 71–120. Basingstoke: Palgrave Macmillan.
- Fox, D. (2014). Cancelling the maxim of quantity: Another challenge for a Gricean theory of scalar implicatures. *Semantics and Pragmatics* 7(5), 1–20.
- Fox, D. and B. Spector (2018). Economy and embedded exhaustification. *Natural Language Semantics* 26(1), 1–50.
- Frege, G. (1918). Der gedanke: Logische untersuchungen, erster teil. *Beitrage zur Philosophie des Deutschen Idealismus* 1, 58–77.
- Gazdar, G. (1979). *Pragmatics: Implicature, Presupposition, and Logical Form*. New York: Academic Press.
- Geurts, B. (2011). *Quantity Implicatures*. Cambridge University Press.
- Ginzburg, J. (1996). Dynamics and the semantics of dialogue. In J. Seligman and D. Westerståhl (Eds.), *Language, Logic, and Computation*, Volume 1.
- Goodman, N. D. and A. Stuhlmüller (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in cognitive science* 5(1), 173–184.
- Green, G. M. (1996). *Pragmatics and natural language understanding*. Psychology Press.
- Grice, H. (1967a). Logic and conversation. In *Studies in the Way of Words*, pp. 22–40. Harvard University Press.
- Grice, H. (1967b). Utterer’s meaning and intentions. In *Studies in the Way of Words*, pp. 86–116. Harvard University Press.
- Grice, H. (1989). *Studies in the Way of Words*. Harvard University Press.
- Grice, H. P. (1975). Logic and conversation. In P. Cole and J. Morgan (Eds.), *Syntax and Semantics*, Volume 3, pp. 41–58.
- Groenendijk, J. and F. Roelofsen (2009). Inquisitive semantics and pragmatics. In J. M. Larrazabal and L. Zubeldia (Eds.), *Meaning, Content, and Argument: Proceedings of the ILCLI International Workshop on Semantics, Pragmatics, and Rhetoric*.
- Groenendijk, J. and M. Stokhof (1984). *Studies on the Semantics of Questions and the Pragmatics of Answers*. Ph. D. thesis, University of Amsterdam.
- Hara, Y. (2005). Contrastives and gricean principles. In P. Dekker and M. Franke (Eds.), *Proceedings of the Fifteenth Amsterdam Colloquium*, pp. 101–106. University of Amsterdam.
- Harnish, R. (1976). Logical form and implicature. In T. G. Bever, J. J. Katz,

- and D. T. Langendoen (Eds.), *An Integrated Theory of Linguistic Ability*. New York: Crowel.
- Horn, L. R. (1972). *On the Semantic Properties of Logical Operators in English*. Ph. D. thesis, University of California Los Angeles.
- Horn, L. R. (1984). Towards a new taxonomy of pragmatic inference: Q-based and R-based implicatures. In D. Schiffrin (Ed.), *Meaning, Form, and Use in Context*, pp. 11–42. Georgetown University Press.
- Horn, L. R. (1989). *A Natural History of Negation*. Chicago: University of Chicago Press.
- Horn, L. R. (2001). *A Natural History of Negation*. David Hume Series on Philosophy and Cognitive Science Reissues. Stanford: CSLI Publications. First published in 1989.
- Hurford, J. R. (1974). Exclusive or inclusive disjunction. *Foundations of Language* 11(3), 409–411.
- Katzir, R. and R. Singh (2013). Hurford disjunctions: embedded exhaustification and structural economy. In U. Etxeberria, A. Fălăuş, A. Irurtzun, and B. Leferman (Eds.), *Proceedings of Sinn und Bedeutung* 18.
- Krifka, M. (2011). Questions and question-acts. Presented at the 18th Amsterdam Colloquium, December 2011, available via amor.cms.hu-berlin.de/~h2816i3x/talks.
- Malamud, S. A. and T. Stephenson (2015). Three ways to avoid commitments: Declarative force modifiers in the conversational scoreboard. *Journal of Semantics* 32(2), 275–311.
- Matsumoto, Y. (1995). The conversational conditions on horn scales. *Linguistics and Philosophy* 18, 21–60.
- Mayr, C. and J. Romoli (2016). A puzzle for theories of redundancy: Exhaustification, incrementality, and the notion of local context. *Semantics and Pragmatics* 9(7).
- Meertens, E. (2019). How prosody disambiguates between alternative and polar questions. In *Proceedings of the 22nd Amsterdam Colloquium*.
- Mill, J. S. (1867). *An Examination of Sir William Hamilton's Philosophy*. Retrieved from <http://oll.libertyfund.org/>.
- Montague, R. (1973). The proper treatment of quantification in ordinary English. *Approaches to Natural Language* 49, 221–242.
- Ohala, J. (1983). Cross-language use of pitch: an ethological view. *Phonetica* 40.
- Pierrehumbert, J. B. and J. Hirschberg (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, and M. E. Pollack (Eds.), *Intensions in communication*, pp. 271–311. Cambridge, MA: MIT Press.
- Poortman, E. (2017). *Concepts and Plural Predication: The Effects of Conceptual Knowledge on the Interpretation of Reciprocal and Conjunctive Plural Constructions*. Ph. D. thesis, Utrecht University.
- Roberts, C. (1996). Information structure in discourse. In J. Yoon and A. Kathol (Eds.), *OSU Working Papers in Linguistics*, Volume 49, pp. 91–136. Ohio State University.
- Roberts, C. (2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5, 6:1–96. Publication of a 1996 manuscript.
- Roelofsen, F. (2013). A bare bone attentive semantics for *might*. In M. Aloni, M. Franke, and F. Roelofsen (Eds.), *The dynamic, inquisitive, and visionary*

- life of φ , $?\varphi$, and $\diamond\phi$: a festschrift for Jeroen Groenendijk, Martin Stokhof, and Frank Veltman, pp. 190–215. ILLC publications.
- Roelofsen, F. and S. van Gool (2010). Disjunctive questions, intonation, and highlighting. In M. Aloni, H. Bastiaanse, T. de Jager, and K. Schulz (Eds.), *Logic, Language, and Meaning: Selected Papers from the Seventeenth Amsterdam Colloquium*, pp. 384–394. Springer.
- Russell, B. (2006). Against grammatical computation of scalar implicatures. *Journal of Semantics* 23, 361–382.
- Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy* 27(3), 367–391.
- Schulz, K. and R. van Rooij (2006). Pragmatic meaning and non-monotonic reasoning: The case of exhaustive interpretation. *Linguistics and Philosophy* 29(2), 205–250.
- Simons, M. (2001). Disjunction and alternativeness. *Linguistics and Philosophy* 24.
- Simons, M. (2011). Dynamic pragmatics, or why we shouldn’t be afraid of embedded implicatures. In N. Ashton, A. Chereches, and D. Lutz (Eds.), *Semantics and Linguistic Theory (SALT) 21*, pp. 609–633.
- Singh, R. (2008). On the interpretation of disjunction: Asymmetric, incremental, and eager for inconsistency. *Linguistics and Philosophy* 31(2), 245–260.
- Soames, S. (1982). How presuppositions are inherited: a solution to the projection problem. *Linguistic Inquiry* 13(3), 483–545.
- Spector, B. (2007). Scalar implicatures: Exhaustivity and Gricean reasoning. In M. Aloni, A. Butler, and P. Dekker (Eds.), *Questions in Dynamic Semantics*, pp. 225–250. Elsevier.
- Van Tiel, B. (2014). *Quantity Matters: Implicatures, typicality and truth*. Ph. D. thesis, Radboud Universiteit Nijmegen.
- Uegaki, W. (2015). *Interpreting questions under attitudes*. Ph. D. thesis, Massachusetts Institute of Technology.
- Wagner, M. (2012). Contrastive topics decomposed. *Semantics and Pragmatics* 5, 1–54.
- Westera, M. (2014). Exhaustivity through the Maxim of Relation. In Y. I. Nakano, K. Satoh, and D. Bekki (Eds.), *Selected Papers from the Tenth Logic and Engineering of Natural Language Semantics (LENLS 10)*, New Frontiers in Artificial Intelligence, pp. 193–205. Springer.
- Westera, M. (2017a). An attention-based explanation for some exhaustivity operators. In *Proceedings of Sinn und Bedeutung*. University of Edinburgh.
- Westera, M. (2017b). *Exhaustivity and intonation: a unified theory*. Ph. D. thesis, ILLC, University of Amsterdam.
- Westera, M. (2018). Rising declaratives of the quality-suspending kind. *Glossa: a journal of general linguistics* 3(1).
- Westera, M. (2019). Rise-fall-rise as a marker of secondary QUDs. In D. Gutzmann and K. Turgay (Eds.), *Secondary content: The semantics and pragmatics of side issues*. Leiden: Brill.
- Westera, M. (2020a). Hurford disjunctions: an in-depth comparison of the grammatical and the pragmatic approach. Under review; manuscript available from <https://semanticsarchive.net/Archive/jk5YzYy0/>.
- Westera, M. (2020b). Implying or implicating ‘not both’ in declaratives and interrogatives. To appear in *Proceedings of Sinn und Bedeutung* 24.
- Westera, M. and G. Boleda (2019). Don’t blame distributional semantics if it

- can't do entailment. In *Proceedings of the 13th International Conference on Computational Semantics (IWCS)*, pp. 120–133.
- Zimmermann, T. E. (1989). Intensional logic and two-sorted type theory. *Journal of Symbolic Logic* 54(1), 65–77.
- Zimmermann, T. E. (2000). Free choice disjunction and epistemic possibility. *Natural Language Semantics* 8, 255–290.