Exhaustivity in Attentional Pragmatics: Still Quantity, but of different stuff

Matthijs Westera
University of Amsterdam / Universitat Pompeu Fabra

Manuscript, September 2017

Abstract. This paper presents Attentional Pragmatics, a theory of exhaustivity that is based on the idea that exhaustivity arises when relevant propositions aren't mentioned, which is subtly but essentially different from the standard recipe. The idea is implemented in terms of attention-drawing, and exhaustivity is explained as deriving, not from the traditional maxim of information quantity, but from its attentional counterpart. This is demonstrated to avoid or solve a number of well-known (potential) challenges for the standard recipe, namely Hurford disjunctions, the epistemic step, and the exhaustivity effects of hints. Certain partial correspondences between Attentional Pragmatics and existing approaches will be formally proven. This paper is not concerned with embedded exhaustivity (other than those in matrix-level Hurford disjunctions), leaving these for other pragmatic or semantic mechanisms to explain, some of which will be listed.

Keywords: exhaustivity, attention, Quantity, Hurford disjunctions, epistemic step, hints.

1 Introduction

In this paper we present and motivate a new pragmatic theory of some (but not all) exhaustivity effects, and compare it in detail to the standard pragmatic recipe based on Grice's (1975) maxim of Quantity (e.g., Horn 1972; Gazdar 1979; Schulz and Van Rooij 2006; Spector 2007; Geurts 2011). The new theory, which we call *Attentional Pragmatics*, solves three actual or potential problems for the standard pragmatic recipe, each brought forward by proponents of the grammatical approach. Altogether, with Attentional Pragmatics we aim to establish as plausible a reasonably intuitive but hitherto overlooked pragmatic alternative to the standard recipe.[1]

Attentional Pragmatics is based on the following simple assumption:

Assumption 1. Exhaustivity arises when relevant propositions are not mentioned.

To illustrate, consider first B's response in (1), from which we may infer that B intends it as an exhaustive answer to the question (at least if it is pronounced with falling intonation):

(1) A: Who (of John, Mary and Bill) was at the party?
B: John was there. *(Implied: Mary and Bill weren't, according to B)*

[1]I am very grateful to Jeroen Groenendijk and Floris Roelofsen for valuable criticism on all iterations of the work presented here, as well as to three anonymous reviewers for Sinn und Bedeutung. Financial support from the Netherlands Organization for Scientific Research (NWO) is gratefully acknowledged.

An intuitive explanation for this implication could be the following: if B had considered it possible that Mary was at the party as well, she would have mentioned it, for instance by uttering (2) instead:

(2) B: John was there, or both John and Mary.

This may sound a bit explicit and artificial, but more natural examples can be constructed that we assume exhibit the same structure as (1)/(2), e.g., "three"/"three or four" and "some"/"some or all". In each case the first response may be interpreted exhaustively in a way that the second may not. This suggests that assumption 1 is indeed on the right track. Attentional Pragmatics turns this simple assumption into a precise theory of exhaustivity, spelling out "mentioning" in terms of *drawing attention* to things, governed by a new set of conversational maxims. Perhaps surprisingly, the intuitive generalization in assumption 1 has not received much attention. This is because a different intuitive generalization has been getting in the way: that exhaustivity would arise when relevant propositions are not *asserted*. The latter constitutes the core of the standard pragmatic recipe for exhaustivity in the literature, and it is different from assumption 1 because one can mention a proposition without asserting it. Consequences of this difference will be discussed in detail in this paper.

As we will show, Attentional Pragmatics solves the following (actual or potential) problems for the standard recipe:

I. **the granularity problem:** the Gricean maxim of Quantity and hence the standard recipe as a whole cannot (easily) distinguish between examples like (1) and (2) (e.g., Gazdar 1979; Schulz and Van Rooij 2006; Chierchia et al. 2012; Katzir and Singh 2013);

II. **the epistemic step:** Gricean Quantity implications fall short of exhaustivity, the opinionatedness assumption invoked by the standard recipe to bridge the gap is ad hoc (e.g., Chierchia et al. 2012; cf. Soames 1982; Sauerland 2004), and exhaustivity seems to occur also without it; recent experimental results seemingly in favor of the opinionatedness assumption have been misinterpreted;

III. **the exemption problem:** exhaustivity seems to occur on utterances that are arguably exempt from Gricean Quantity, namely the hints of a quizmaster (Fox 2014).

In the second half of this paper these problems will be discussed in detail. The granularity problem (I) is frequently acknowledged to be a genuine problem for the standard recipe, the epistemic step (II) and the exemption problem (III) less so. We aim to make plausible that all three are genuine problems for the standard recipe, and that Attentional Pragmatics appears more successful in this regard. However, the main aim of this paper is not empirical but theoretical: we present a new theory that, given the pre-theoretical plausibility of assumption 1, is worth developing in its own right, and worth comparing to existing approaches in detail. Where our comparison involves uncertain empirical data, these cases will be highlighted as important questions for future experimental work.

The above (potential) problems have been brought forward as reasons for abandoning the standard recipe especially in favor of the *grammatical approach* (e.g., Fox 2007; Chierchia et al. 2009, 2012; Katzir and Singh 2013; Mayr and Romoli 2016). Since we will show that a pragmatic solution to these problems exists after all, this paper neutralizes part of the motivation for the grammatical approach – though only as far as un-embedded exhaustivity goes. The pragmatics/grammar debate concerning exhaustivity has revolved to a great extent around cases where exhaustivity effects appear to be embedded under various grammatical operators. Such effects are difficult or impossible to account for within the standard pragmatic recipe, i.e., through the Gricean maxim of Quantity. Although Attentional Pragmatics will account for some supposedly embedded cases, it is not intended (and cannot) account for all of them. This means that adopting Attentional Pragmatics commits one to explaining these effects by other means, which may in principle be either semantic/grammatical – e.g., modal operators (Uegaki, 2015); metalinguistic negation (Horn, 1985; Geurts, 1998); or covert exhaustivity operators (Chierchia et al., 2012) – or more pragmatic

in some sense – e.g., typicality inferences (Van Tiel, 2014) or implic-i-ture (Bach 1994; 'explicature' in Recanati 2004). Indeed, it is not a priori unlikely that language and cognitive processes could give rise to exhaustivity effects in more ways than one. It is in part an empirical issue whether the various exhaustivity effects should be treated as a unified phenomenon, and a theoretical issue whether they can be (e.g., Russell 2006; Geurts 2011; Simons 2011). To our understanding no answers to these questions have commanded general agreement thus far. While it is important to remain aware of the full range of exhaustivity(-like) effects, it is also crucial to keep investigating to what extent certain subclasses of exhaustivity effects may receive an adequate explanation. For this reason and for reasons of space, we will not compare Attentional Pragmatics to the grammatical approach (except occasionally, in passing); such a comparison must be left for another occasion.

Outline A conceptual and formal framework and the theory of Attentional Pragmatics are presented in section 2. Section 3 applies it to several examples, leading up to a general characterization of exhaustivity as predicted by the theory. In section 4 the theory is compared to the standard recipe, with particular attention to problems I.-III. given above. Section 5 concludes with a discussion and suggestions for future work.

2 Attentional Pragmatics

For a clear exposition of the theory, it is necessary to briefly outline a conceptual and formal pragmatic framework – essentially a streamlined and terminologically slightly safer version of much existing work in pragmatics. A conceptual framework will be outlined in section 2.1 and its formalization in section 2.2. Subsequently, sections 2.3 and 2.4 define the theory of Attentional Pragmatics, which consists of a set of maxims governing information (the *I-maxims*) and a set of maxims governing attention (the *A-maxims*). Finally, section 2.5 uses the maxims to retroactively motivate certain assumptions made in section 2.2 about the speaker intentions in the relevant examples.

2.1 Conceptual framework

A core assumption, shared with much work on pragmatics, is the following:

Assumption 2. Speakers have certain communicative intentions (e.g., to share a piece of information) – let us call the objects of such intentions (e.g., a piece of information that the speaker intends to share) "intents".

Intents are to be distinguished from the (semantic) *contents* of a sentence. The "intent"/"content" terminology is adopted from Bach and Harnish 1979, where it generalized Grice's (1967) distinction between utterer's meaning and sentence meaning. This paper will say only very little about semantic contents, i.e., sentence meaning; we will just presuppose that an adequate semantics can be devised that assigns to the relevant expressions contents which, constrained by the maxim of Manner, can clearly communicate the assumed intents. Consistently maintaining a terminological distinction between contents and intents will help to avoid certain misunderstandings about what Attentional Pragmatics does and doesn't commit us to, as we will see.

For a communicative intention to be rational it must serve a certain goal, e.g., that a certain piece of information becomes common ground. For a pragmatic account of exhaustivity (whether the standard recipe or Attentional Pragmatics) it is essential that speakers typically pursue several such goals at once:

Assumption 3. Speakers group their goals based on features like subject matter, importance, efficiency and transparency. Utterances are aimed not at individual goals but at such groupings. Let us call these groups "themes", and formalize them as sets of propositions, each representing a piece of information that ought to be made common ground.

For instance, in (1), "John was there.", it seems impossible to explain the exhaustivity implication that Mary wasn't there unless we assume that establishing Mary's presence was a conversational goal alongside establishing John's presence – say, that both goals were part of the "party presence" theme introduced by speaker A. In calling these organizations of goals "themes" we follow, among others, Balogh (2009). Themes are more commonly called "questions under discussion" (e.g., Roberts 2012), but this gives rise to a potentially harmful confusion between questions as (groupings of) discourse goals (here called "themes"), questions as utterances, questions as meanings of interrogative sentences and questions as meanings of embedded interrogative-like constructions.

Attentional Pragmatics is intended to implement the basic intuition captured in assumption 1: that exhaustivity arises when relevant possibilities are not mentioned. To that end, it assumes, following Ciardelli et al. 2009, that interlocutors intentionally draw each others' attention to things, i.e., that drawing attention is not just side-effect of sharing information but a communicative intention in its own right. Thus, on top of the common assumptions that utterances have themes and intents, we also need a particular type of intent:

Assumption 4. Besides *informational intents* (at least for assertions), utterances also have *attentional intents*. An attentional intent will be formalized as a set of propositions, representing those pieces of information to which a speaker intends to draw the audience's attention.

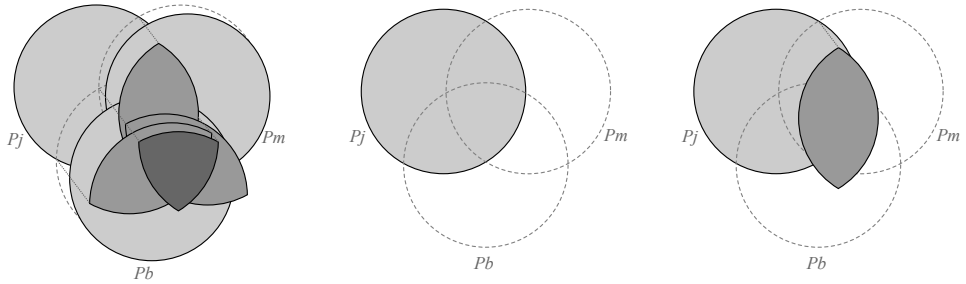
An utterance may of course draw attention to many things, propositions and otherwise, including (but not limited to) everything explicitly mentioned in the utterance – but not everything will be part of what the speaker *intended* to draw attention to, i.e., part of the attentional intent.[2]

Being the objects of a type of communicative intention, attentional intents have to be governed by a set of rationality constraints, or conversational maxims, paralleling those which Grice (1975) proposed govern informational intents. Thus, alongside the informational maxims (or I-maxims) of Quality, Relation and Quantity, there will be a set of attentional maxims (A-maxims) – and we will see that exhaustivity derives from the latter. All these maxims will be defined further below.

2.2 Formalism

For the sake of explicitness and precision, the assumed intents and themes for the relevant examples, as well as a definition of the conversational maxims, will be specified formally using *Intensional Logic* (Montague, 1973), albeit with doxastic rather than alethic modality and with some additional notation conventions. We refer to the exposition in Gamut 1991 (vol.2) for the basic formalism. In a nutshell, Intensional Logic is a form of type theory with convenient shorthands for quantification over, abstraction over and application to possible worlds. The operators $\Box$ and $\Diamond$ signify universal and existential quantification over worlds, in our case worlds in the speaker's doxastic state, i.e., the modalities of believing (or taking oneself to know) and considering possible. The operators $^\wedge$ and $^\vee$ signify abstraction over and application to worlds, which in this paper will be used almost exclusively to switch between propositions and their truth values. That is, $^\wedge\varphi$ can generally be read as "(the proposition) that φ", and $^\vee p$ as "the proposition p is true here". To illustrate, the

[2] Assumption 4, that utterances have attentional *intents*, does not necessarily commit one to the assumption of a corresponding dimension of attentional semantic *content* – although we are not against such an assumption. We return to this in section 2.5.



(a) (b) (c)

Figure 1: The theme and attentional intents of (3a), (3b) and (3c).

formula $\Box(Pj \wedge \mathcal{T}_0(\hat{}Pj))$ might express that the speaker takes herself to know that John was at the party and that the proposition that John was at the party is an element of the main theme, i.e., that establishing this proposition is a primary goal of the utterance.

As is common, expressions of certain relevant types will be distinguished typographically. Besides using lowercase for individual constants/variables ($a, b, c, \ldots$ of type e) and uppercase for predicates ($A, B, C, \ldots$ of type $\langle e, t\rangle$), we will use lowercase calligraphic for propositions ($\mathcal{a}, \mathcal{b}, \mathcal{c}, \ldots$ of type $\langle s, t\rangle$) and uppercase calligraphic for sets of propositions such as themes and attentional intents ($\mathcal{A}, \mathcal{B}, \mathcal{C}, \ldots$ of type $\langle\langle s, t\rangle, t\rangle$). Furthermore, as a notational convention, for all unary, first-order predicate constants P, we may write, e.g., $Pjmb$ to mean $Pj \wedge Pm \wedge Pb$. In addition, it will occasionally be convenient to conceive of functions of type $\langle a, t\rangle$ as sets of things of type a, and to use the usual set-theoretical operations and relations within the object language (these can be defined in Intensional Logic as mere notational shorthands; Zimmermann 1989). Lastly, for any set-type expression X we will write $X^{\cap}$ to mean the closure of X under intersection.

To illustrate, and also to get acquainted with the formalism, let us consider examples (1) and (2) with which this paper started, repeated below as (3) with the themes and intents we assume.

(3) a. A: Who (of John, Mary and Bill) was at the party?

$\mathcal{T}_0 = \{\hat{}Pj, \hat{}Pm, \hat{}Pb\}^{\cap} \qquad \mathcal{A}_0 = \{\hat{}Pj, \hat{}Pm, \hat{}Pb\}^{\cap}$

b. B: John was there.

$\mathcal{T}_0 = \{\hat{}Pj, \hat{}Pm, \hat{}Pb\}^{\cap} \qquad p_0 = \hat{}Pj \qquad \mathcal{A}_0 = \{\hat{}Pj\}$

c. B: John was there, or John and Mary.

$\mathcal{T}_0 = \{\hat{}Pj, \hat{}Pm, \hat{}Pb\}^{\cap} \qquad p_0 = \hat{}Pj \qquad \mathcal{A}_0 = \{\hat{}Pj, \hat{}Pjm\}$

For each utterance, $\mathcal{T}_0$ denotes the main theme (a set that is, in this case, closed under intersection), p_0 denotes the main informational intent (what is asserted) and $\mathcal{A}_0$ denotes the main attentional intent (the set of propositions to which the speaker primarily intends to draw attention). The assumed attentional intents of (3a), (3b) and (3c) are depicted, from left to right, in figure 1. This type of pictorial representation will be relied upon again further below. In it, each attentional intent is depicted as a Venn diagram on the set of all possible worlds, based on the three atomic propositions of John's,

Mary's, and Bill's presence (the circles). Overlapping propositions in the attentional intent – the gray regions – are pulled apart in a third dimension, towards the reader as it were, for clearer presentation.

The assumed themes and intents align with some related proposals. For instance, the attentional intent of the question corresponds to Hamblin's (1973) semantics of interrogatives (and the assumed theme will follow from this through the maxims to be defined below); the attentional intents for the assertions correspond to assumptions about disjunction in, e.g., Aloni 2001; Schulz and Van Rooij 2006; Alonso-Ovalle 2008; Ciardelli et al. 2009. We will motivate the assumed attentional intents in more detail in section 2.5; in the current section we will only briefly reflect on the assumed theme. Crucially, the theme of (3) is asymmetrical, i.e., it contains only people's presences, not their absences. This is necessary to avoid the *symmetry problem*, i.e., the problem that the exclusion (through an exhaustivity implication) of both a proposition and its negation would entail a contradiction (e.g., Kroch 1972; Groenendijk and Stokhof 1984; Katzir 2007; Block 2008; Chierchia et al. 2012). Although the aforementioned authors assume that themes (or relevance) are necessarily symmetrical, this view finds an adversary in what Horn (2001) calls the *Asymmetry Thesis*, which states that our interests in conversation tend to be asymmetrical. For instance, according to Horn (1978) and Leech (1981) we tend to be much more interested in what there is than in what there isn't.[3] Since space does not permit a discussion of the symmetry problem in the current paper, the reader who (unlike us) takes relevance to be necessarily symmetrical can perhaps conceive of the themes in our approach not as relevance plain and simple, but as relevance *filtered by* some asymmetrical mechanism, e.g., lexical scales or considerations of brevity or complexity (for discussion, see, e.g., Matsumoto 1995; Russell 2006; Katzir 2007).

Lastly, in order to prove interesting things about the relevant utterances, and about exhaustivity more generally, expressions of Intensional Logic will be interpreted on a subclass of models, the *admissible models*, defined so as to interpret certain constants in a certain way, comparable to Montague's meaning postulates. In particular, admissible models fix the interpretation of constants representing the conversational maxims, e.g., the constant I-Relation will be interpreted by admissible models basically as the Gricean maxim of Relation. Admissible models must also validate the KD45 belief axioms, plus *intent introspection* and *theme introspection*, which ensure that the speaker knows the interpretations of constants denoting intents and themes of the utterance (e.g., p_0 for the main informational intent) – we can leave their definitions implicit.[4] Thus:

Definition 1 (Admissible model)**.** A model $\mathbf{M}$ (or $\langle \mathbf{M}, w_0 \rangle$) is an *admissible model* iff:

1. $\mathbf{M}$ validates (makes true in all its worlds) the KD45 belief axioms;
2. $\mathbf{M}$ validates intent and theme introspection; and
3. $\mathbf{M}$ validates the definitions of the maxims, to be given below.

And it is an admissible model *for a given example* if, and only if, in addition:

4. w_0 validates all formal statements given in the example (in a gray box); and
5. W is sufficiently large, namely, every contingent first-order (hence non-modal) formula, that can be constructed from only constants used in the example, variables, connectives and quantifiers, is true in some $w \in W$.

Admissible models enable us to formalize the relevant parts of a given example and prove potentially

[3] Horn proposes that negative information tends not to be primarily relevant, but only secondarily, after the earlier consideration of its positive counterpart. We think that the mere secondary relevance of negative information could explain why exhaustivity is (in the relevant examples) not asserted but merely implicated.

[4] To adopt intent introspection and theme introspection is not to say that speakers always know what they mean and why, or that Attentional Pragmatics would yield no predictions when they don't – it merely reflects our assumption that this is the case in the examples of interest in the current paper.

interesting things about it, e.g., that an utterance complies with the maxim of I(nformational)-Relation, which is the case if in all admissible models for the example the constant I-Relation returns true in the actual world, when applied to the relevant intents and themes, and given the beliefs and goals of the speaker. We will see more concrete examples soon.

2.3 Maxims governing informational intents

Attentional Pragmatics contains maxims governing informational intents and maxims governing attentional intents, although only the latter are (somewhat) novel. The *I(nformation)-maxims* closely resemble Grice's (1989, ch.2) maxims (except Manner, which will not play an explicit role in this paper), and I-Quantity in particular is defined as for instance in Harnish 1976:

Definition 2.

1. $\text{I-Quality}(p) = \Box^{\vee} p$
   "Intend to share only information you take to be true."
2. $\text{I-Relation}(p, \mathcal{T}) = \mathcal{T}(p)$
   "Intend to share only information that is thematic."
3. $\text{I-Quantity}(p, \mathcal{T}) = \forall q \left(\begin{pmatrix} \text{I-Quality}(q) \wedge \\ \text{I-Relation}(q, \mathcal{T}) \end{pmatrix} \rightarrow (p \subseteq q) \right)$
   "Intend to share all thematic information you take to be true."

Let compliance with *all* I-Maxims be defined as follows:

Definition 3.
$$\text{I-Maxims}(p, \mathcal{T}) = \begin{pmatrix} \text{I-Quality}(p) \wedge \\ \text{I-Relation}(p, \mathcal{T}) \wedge \\ \text{I-Quantity}(p, \mathcal{T}) \end{pmatrix}$$

On the basis of these definitions we can prove some general results. For instance, a speaker will always know whether a given informational intent complies with the I-maxims or not:

Fact 1. For all admissible models $\mathbf{M}$ and any constants p_i, $\mathcal{T}_j$:

$$\mathbf{M} \models \Box\, \text{I-Quality}(p_i) \leftrightarrow \text{I-Quality}(p_i)$$
$$\mathbf{M} \models \Box\, \text{I-Relation}(p_i, \mathcal{T}_j) \leftrightarrow \text{I-Relation}(p_i, \mathcal{T}_j)$$
$$\mathbf{M} \models \Box\, \text{I-Quantity}(p_i, \mathcal{T}_j) \leftrightarrow \text{I-Quantity}(p_i, \mathcal{T}_j)$$

Moreover, if there exists a compliant intent then it is the only one:

Fact 2. For all admissible models $\mathbf{M}$ and constant $\mathcal{T}_i$:

$$\mathbf{M} \models \forall p \forall q \left(\begin{pmatrix} \Box\, \text{I-Maxims}(p, \mathcal{T}_i) \wedge \\ \Box\, \text{I-Maxims}(q, \mathcal{T}_i) \end{pmatrix} \rightarrow (p = q) \right)$$

And this holds also without the modal boxes ($\Box$).

This is because, if two intents comply with all I-maxims, and specifically with I-Relation and I-Quality, then compliance with I-Quantity demands that each intent is contained in the other,

hence that the two intents are equivalent. Fact 2 shows that the maxims are to some extent deterministic, i.e., that given a certain theme and a speaker's beliefs, if there is a way to comply, we can predict which intent the speaker will be using.[5] Besides these very general results, we can also compute concrete predictions for a given example, e.g.:

Fact 3. For all admissible models $\langle \mathbf{M}, w_0 \rangle$ for example (3b):

$$\mathbf{M}, w_0 \models \Box\,\text{I-}\textsc{Quantity}(p_0, \mathcal{T}_0) \rightarrow (\neg\Box Pm \wedge \neg\Box Pb)$$

Note that this I-Quantity implication is a mere absence of belief ($\neg\Box Pm$), which on its own does not entail the presence of a belief that Mary wasn't there ($\Box\neg Pm$), i.e., the predicted implication falls short of an exhaustivity implication. This difference between I-Quantity and exhaustivity is the epistemic step (Sauerland, 2004), to be discussed in section 4.3.

Two features of the current I-maxims require a brief clarification. First, I-Quantity is so demanding that so-called "mention-some" contexts can be represented only by assuming a restriction of the theme (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 theme. But as Schulz and Van Rooij (2006) note such an explanation is required regardless of how one seeks to represent the two possible contexts, e.g., pragmatically or by assuming a semantic ambiguity, and in the rest of this paper it can be set aside. Second, since I-Relation does not allow for negations, disjunctions or intersections of thematic propositions, nor indirect, partial or merely probable answers – unless of course these just happen to be thematic in their own right – such discourse moves will have to be accounted for by other means. For instance, we can still explain why an indirect answer can be an appropriate discourse move by assuming (and independently motivating) that although its primary intent does not directly resolve the theme and as such does not comply with I-Relation, it has a secondary intent – a conversational implicature – that does. Other strictly non-compliant maneuvers must be dealt with by assuming theme shifts, constrained by a suitable theory of which themes a rational speaker may pursue, i.e., a "theme pragmatics". This division of labor between "theme pragmatics" and "intent pragmatics" – or between choosing goals and selecting the means to achieve them – is primarily theory-internal and therefore a mere matter of convenience. Its convenience cannot be demonstrated, however, without being more explicit about theme pragmatics, e.g., about rhetorical relations (Asher and Lascarides, 2003) or discourse strategies (Roberts, 2012), and about the ways in which certain relations between intents and themes are prosodically marked (Roberts, 2012; Gunlogson, 2008; Westera, 2013; Malamud and Stephenson, 2015). For present purposes these issues can be safely set aside.

2.4 Maxims governing attentional intents

The *A(ttention)-maxims* follow the same general recipe as the I-maxims, except for the addition of the maxim of A-Parsimony, which we will motivate in some detail below:

[5] Moreover, because for each non-contradictory intent there exists an admissible model in which it complies with the I-maxims, fact 2 implies that for all non-contradictory intents there exists a model in which it is the unique compliant intent. This shows that the I-maxims distinguish all type-theoretically (or set-theoretically) distinct informational intents, hence that the mathematical objects by means of which we model informational intents are no richer than necessary for the given characterization of their rational usage. Something similar will hold for the A-maxims, further below. These results are comparable in status to the *antisymmetry* of an order like entailment or meaning inclusion; see Roelofsen 2013a for a linguistic perspective.

Definition 4.

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$(\mathcal{A}, \mathcal{T}) = \forall a(\mathcal{A}(a) \rightarrow \mathcal{T}(a))$
   "Intend to draw attention only to thematic propositions."
3. A-Parsimony$(\mathcal{A}, \mathcal{T}) = \forall a \left(\begin{array}{c} (\mathcal{A}(a) \wedge \text{A-Quality}(\{a\})) \rightarrow \\ \Diamond \left({}^{\vee} a \wedge \forall b \left(\left(\begin{array}{c} b \subset a \wedge \\ \text{A-Relation}(\{b\}, \mathcal{T}) \end{array} \right) \rightarrow \neg^{\vee} b \right) \right) \end{array} \right)$
   "Intend to draw attention to a proposition only if, if you consider it possible, you consider it possible independently of any more specific thematic proposition(s)."
4. A-Quantity$(\mathcal{A}, \mathcal{T}) = \forall a \left(\left(\begin{array}{c} \text{A-Quality}(\{a\}) \wedge \\ \text{A-Relation}(\{a\}, \mathcal{T}) \wedge \\ \text{A-Parsimony}(\{a\}, \mathcal{T}) \end{array} \right) \rightarrow \mathcal{A}(a) \right)$
   "Intend to draw attention to all thematic propositions you consider independently possible."

As on the informational side, let compliance with *all* A-Maxims be defined as follows:

Definition 5.

$$\text{A-Maxims}(\mathcal{A}, \mathcal{T}) = \left(\begin{array}{c} \text{A-Quality}(\mathcal{A}) \wedge \\ \text{A-Relation}(\mathcal{A}, \mathcal{T}) \wedge \\ \text{A-Parsimony}(\mathcal{A}, \mathcal{T}) \wedge \\ \text{A-Quantity}(\mathcal{A}, \mathcal{T}) \end{array} \right)$$

The A-maxims are intended as independently plausible, reasonable constraints, just as is commonly assumed for the I-maxims. Indeed, it will be generally beneficial for an intelligent being to keep track of all (A-Quantity) pertinent goals (A-Relation) that can still be achieved (A-Quality) – say, pieces of information, surrounding a certain subject matter, that ought to be made common ground and that are potentially true – and we will motivate A-Parsimony shortly. Moreover, similar constraints have been assumed before. A maxim like A-Quality is assumed by Roelofsen, who calls it "Attentive Sincerity" (2013b; building on Ciardelli et al. 2009); it corresponds also to "Genuineness" in Zimmermann 2000 (p.270; he explores a semantic and a pragmatic account), and to the "Viability" constraint in Biezma and Rawlins 2012 (p.46, formulated for questions). Biezma and Rawlins also 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 the same basic idea is found in Grice 1989: that disjunctions would serve to specify possibilities "that relate in the same way to a given topic".

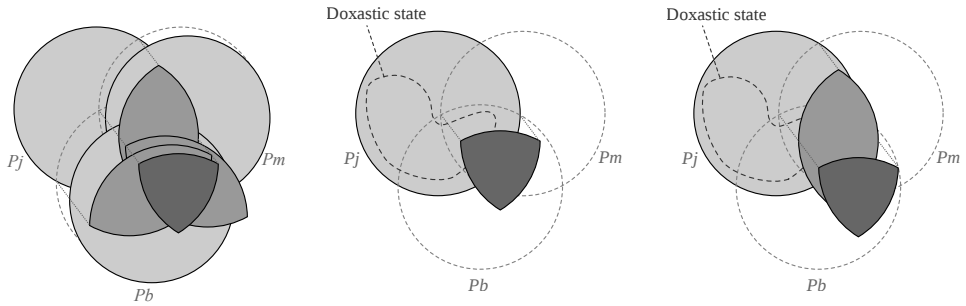
The maxim of A-Parsimony can be understood and motivated as follows. Consider a speaker B who believes that if John and Mary were at the party, then so was Bill ($\Box(Pjm \rightarrow Pb)$). Now, consider the following two utterances made by this speaker:

(4) *(B believes that if John and Mary were at the party, then so was Bill.)*

a. B: John was at the party, or John, Mary and Bill.

$\Box(Pjm \rightarrow Pb)$
$\mathcal{T}_0 = \{{}^{\wedge}Pj, {}^{\wedge}Pm, {}^{\wedge}Pb\}^{\cap}$ $\quad \mathcal{A}_0 = \{{}^{\wedge}Pj, {}^{\wedge}Pjmb\}$ $\quad p_0 = {}^{\wedge}Pj$

b. (?) B: John was there, or John and Mary, or John, Mary and Bill.



(a) (b) (c)

Figure 2: The theme and attentional intents of (4), with the speaker's doxastic state.

$\Box(Pjm \to Pb)$
$\mathcal{T}_0 = \{\hat{}Pj, \hat{}Pm, \hat{}Pb\}^{\cap}$ $\quad \mathcal{A}_0 = \{\hat{}Pj, \hat{}Pjm, \hat{}Pjmb\}$ $\quad p_0 = \hat{}Pj$

We take it that, given the assumed speaker beliefs, the utterance in (4b) is intuitively quite strange: why, given B's beliefs, did B include the middle disjunct – or, in terms of attentional intents, why did B intend to draw attention to John and Mary's joint presence ($\hat{}Pjm$)? The maxim of A-Parsimony predicts precisely this strangeness: drawing attention to John and Mary's joint presence is not parsimonious, i.e., superfluous, because B does not consider it possible independently of the presence of all three of them ($\hat{}Pjmb$). Indeed, it can be formally proven for (4b) that the attentional intent (if it complies with A-Quality) must violate A-Parsimony. In contrast, for (4a) there does exist an admissible model in which all the maxims are complied with. To visualize the foregoing, the theme and both attentional intents of (4) are depicted in figure 2. Relative to the theme in figure 2a, and given the indicated doxastic state, the attentional intent in figure 2b complies with A-Parsimony, while the attentional intent in figure 2c violates it. This is because the latter contains a proposition ($\hat{}Pjm$) that the speaker does not consider possible independently of the more specific proposition ($\hat{}Pjmb$) to which attention is already drawn. A-Parsimony rules out such attentional redundancy.[6] Note that A-Parsimony is similar to A-Quality in that both require that the speaker must consider certain propositions possible. Indeed, without the condition A-Quality($\{a\}$) in the definition of A-Parsimony, compliance with A-Parsimony would have entailed compliance with A-Quality. But we separate the two maxims because they are conceptually quite distinct: A-Quality expresses a qualitative requirement, whereas A-Parsimony expresses a quantitative preference for drawing attention to more specific propositions – it just happens to also involve a qualitative criterion.

The maxim of A-Parsimony, with its "considering propositions independently possible", may intuitive feel as if it is enforcing *local exhaustification* of the disjuncts, as assumed in the grammatical approach. This is not a coincidence, given that A-Parsimony will be used to solve a puzzle that the grammatical approach aims to solve with local exhaustification. Nevertheless, A-Parsimony is a global pragmatic criterion, in the sense that (i) it operates on speaker intentions, not on the individual disjuncts that can be used for communicating such speaker intentions, and (ii) it can be

[6] No analogous maxim of "I-Parsimony" was defined among the I-maxims. In principle, some sort of parsimony requirement on informational intents is conceivable, e.g., "don't intend to provide information that is already common ground" or "don't intend to provide the same information twice"; but the first of these can be adequately captured by a constraint on the goals that one may rationally pursue (i.e., don't pursue goals that are already achieved), and the second is more adequately captured by the maxim of Manner, left implicit in this paper (and note that an analogous non-redundancy requirement should hold also for attentional intents, and *isn't* captured by A-Parsimony).

understood as an independently reasonable constraint: if one has two goals (say, two propositions that should be made common ground, or, non-linguistically, two bricks that must be stacked in a certain way) and if there is no way to achieve the one without the other, then it is a waste of energy and working memory to keep paying attention to both of these goals individually. Moreover, while A-Parsimony achieves something not unlike local exhaustification (see section 4.2), the maxim itself is quite different from the constraint invoked by the grammatical approach to this end, namely Hurford's Constraint: whereas Hurford's Constraint has been motivated in terms of informational redundancy (e.g., Katzir and Singh 2013), A-Parsimony cares only about attentional redundancy.[7] For reasons of space, however, we must leave a detailed comparison to the grammatical approach for another occasion, as announced in the introduction.

Lastly, just as for the I-maxims, some general facts can be proven that yield some more insight into the A-maxims. For instance, just as on the informational side, a speaker will always know whether the A-maxims are complied with (thanks to our assumption, in admissible models, that the speaker knows the exact nature of the theme):

Fact 4. For all admissible models $\mathbf{M}$ and constants $\mathcal{A}_i$, $\mathcal{T}_j$:

$$\mathbf{M} \models \Box\,\text{A-}\textsc{Quality}(\mathcal{A}_i) \leftrightarrow \text{A-}\textsc{Quality}(\mathcal{A}_i)$$
$$\mathbf{M} \models \Box\,\text{A-}\textsc{Relation}(\mathcal{A}_i, \mathcal{T}_j) \leftrightarrow \text{A-}\textsc{Relation}(\mathcal{A}_i, \mathcal{T}_j)$$
$$\mathbf{M} \models \Box\,\text{A-}\textsc{Quantity}(\mathcal{A}_i, \mathcal{T}_j) \leftrightarrow \text{A-}\textsc{Quantity}(\mathcal{A}_i, \mathcal{T}_j)$$
$$\mathbf{M} \models \Box\,\text{A-}\textsc{Parsimony}(\mathcal{A}_i, \mathcal{T}_j) \leftrightarrow \text{A-}\textsc{Parsimony}(\mathcal{A}_i, \mathcal{T}_j)$$

Moreover, if there is a compliant attentional intent then it is the only one, showing that the A-maxims are deterministic in the same sense as the I-maxims:

Fact 5. For all admissible models $\mathbf{M}$, and constants $\mathcal{T}_i$:

$$\mathbf{M} \models \forall\mathcal{A}\forall\mathcal{B}\left(\left(\begin{array}{c}\Box\,\text{A-}\textsc{Maxims}(\mathcal{A}, \mathcal{T}_i) \wedge \\ \Box\,\text{A-}\textsc{Maxims}(\mathcal{B}, \mathcal{T}_i)\end{array}\right) \rightarrow (\mathcal{A} = \mathcal{B})\right)$$

And this holds also without the modal boxes.

This is easily seen from fact that all A-maxims except A-Quantity operate pointwisely: A-Quantity essentially collects all the singleton sets that each individually comply with A-Quality, A-Relation and A-Parsimony.

2.5 Motivating the assumed attentional intents

Before applying the maxims to the relevant examples and deriving exhaustivity – as we will do in the next section – let us consider why the attentional intents of these examples must be as assumed. We already mentioned that it aligns with some of the related notions in the literature. However, we can in addition motivate the assumed attentional intents – to a large extent – in terms of the assumed informational intents, namely by deriving certain constraints from the I-maxims and the A-maxims. For instance, relative to themes of the sort that we have assumed, the informational intent must always be the union of the attentional intent:

[7] Ciardelli and Roelofsen (2016) have recently noted that the redundancy-based motivation for Hurford's Constraint falls short if one's notion of meaning (or speaker intention) is as fine-grained as, e.g., attentional intents in the current approach: disjuncts that do not make an informational contribution may well make an attentional contribution, after all. Hence, Hurford's Constraint is less plausible from the perpective of Attentional Pragmatics.

Fact 6. For all admissible models $\mathbf{M}$ such that $\mathcal{T}_i$ denotes a theme that is closed under intersection:

$$\mathbf{M}, w_0 \models \forall p \forall \mathcal{A} \left(\left(\begin{array}{c} \text{I-MAXIMS}(p, \mathcal{T}_i) \wedge \\ \text{A-MAXIMS}(\mathcal{A}, \mathcal{T}_i) \end{array} \right) \rightarrow \left(p = \bigcup \mathcal{A} \right) \right)$$

A proof is included in the appendix. This result means that when addressing a theme that is closed under intersection, the attentional intent must be a *cover* of the informational intent, composed of thematic propositions. Moreover, given A-Parsimony, it must be a *strongest* cover, in the sense that it must not contain any proposition that can itself be covered by a set of more specific thematic propositions. But it need not be a *minimal* cover, i.e., it may contain propositions that are not strictly necessary to cover the intent, as for "John, or John and Mary". All of this is in line with the assumed intents, but it doesn't quite suffice to determine these intents with certainty. A more complete characterization of attentional intent can be derived, at least for assertions in *disjunctive normal form* (i.e., a disjunction of (conjunctions of) simple (positive or negative) propositions), if we assume that all uttered material must contribute to the clear communication of either the informational or the attentional intent, which is what a maxim of Manner/Conciseness might require. By restricting our attention to these types of utterances, the following can be derived:

Fact 7. Any utterance of a sentence φ in disjunctive normal form, that has only an informational and an attentional intent, will have as an attentional intent:

$$\{\text{info}(\psi) \mid \psi \text{ is a disjunct of } \varphi\}$$

provided (i) all (negations of) atomic sentences that occur in φ are thematic, (ii) the theme is closed under union and intersection, and (iii) the utterance complies with the maxims.

The result relies on fact 6 for the inclusion in the attentional intent of the weakest disjuncts (like $^\wedge Pj$ in (3c)), which are necessary for the attentional intent to cover the informational intent, and on Manner/Conciseness for the inclusion of non-weakest disjuncts (like $^\wedge Pjm$ in (3c)), which would otherwise not serve a purpose. The attentional intents assumed thus far follow the pattern of fact 7, given the assumed theme. Hence, given the way in which the maxims are defined, and given the assumed theme and the informational intents, the assumed attentional intents couldn't have been any different.[8]

The above considerations do not directly apply to questions: facts 6 and 7 constrain the attentional intent in terms of an informational intent, but questions arguably lack the latter (see also section 4.4). This means that the attentional intent of A's "wh"-question in the relevant examples, unlike the attentional intents of B's assertions, is a substantive assumption in its own right. One might try to motivate the assumed intent by grounding it in a particular question semantics from the literature (e.g., Hamblin 1973). Fortunately, since the sole purpose of A's initiatives in this paper is to introduce a particular theme for speaker B to address, the topic of "wh"-question semantics can be set aside. After all, in principle we could have used an explicit disjunctive question instead, like (5a), or even a corresponding assertion (5b):

(5) A: a. Was John at the party, or Mary, or Bill, or John and Mary, or John and Bill, or Mary and Bill, or all three of them?

[8]To clarify: this is not a reformulation of the determinacy established in fact 5: there the uniqueness of a compliant attentional intent was established given the theme and the speaker's beliefs. For an audience (or linguist) who needs to understand the attentional intent of an utterance, the belief state of the speaker is typically one of the unknowns, hence fact 5 is useless there. In contrast, fact 7 does help an audience (or linguist) to determine the attentional intents of the relevant utterances, namely on the basis of the themes and the informational intents.

b. John was at the party, or Mary, or Bill, or John and Mary, or John and Bill, or Mary and Bill, or all three of them.

B: John was there, or all three of them.

B's response is still reasonably natural, and seems to give rise to the same relevant exhaustivity implications as in (4) – but now fact 7 enables us to identify the attentional intent at least of (5b), and, via A-Relation, the theme it sets up for speaker B. This is all we need to know for applying Attentional Pragmatics to B's subsequent response, hence we can safely (and will) leave a motivation for the assumed attentional intent of the original "wh"-question for another occasion.[9]

Fact 7 aligns with semantic proposition-set notions in the literature that superficially resemble attentional intents, e.g., sets of Hamblin alternatives for disjunction (Alonso-Ovalle, 2008) or "attentive semantics" (Ciardelli et al., 2009). In particular, according to common definitions of these notions, these sets will likewise contain, for utterances in disjunctive normal form, a proposition for each disjunct (though unlike (basic) *inquisitive* semantics; see Ciardelli and Roelofsen 2016 for discussion). This alignment means that the way in which attention is rationally used in conversation – supposing our A-maxims are correct – can perhaps help explain why these proposition set notions behave the way they do under conjunction and disjunction. For instance, the A-maxims can explain why a disjunction but not a conjunction would introduce alternatives, despite the (we think) intuitive fact that conjunctions draw attention to their conjuncts as much as disjunctions draw attention to their disjuncts: they differ only in whether the speaker could have rationally *intended* to draw attention to the conjuncts/disjuncts, i.e., whether they are elements of the attentional intent, or whether the attention drawn to them is a mere side-effect.[10]

3 Deriving exhaustivity

We will first apply the maxims to some concrete examples, before presenting a general result and a derivative exhaustivity operator. The operator will be useful later, when comparing Attentional Pragmatics to the standard pragmatic recipe based on I-Quantity and to the grammatical approach.

3.1 Some examples

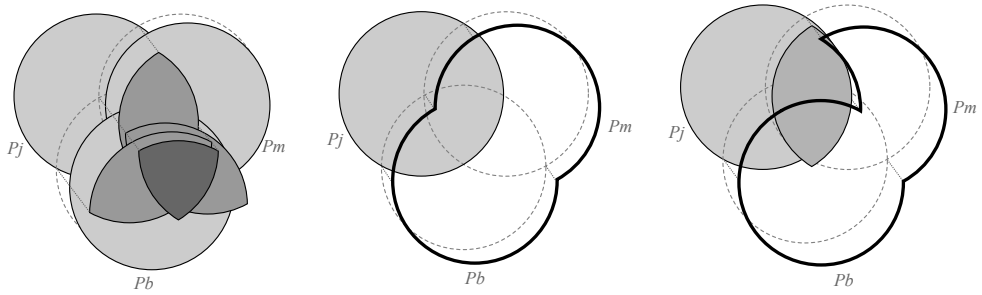
Consider once again the contrast with which this paper started, with the assumed themes and intents as given in (3), repeated here:

(3) a. A: Who (of John, Mary and Bill) was at the party?

$\mathcal{T}_0 = \{^\wedge Pj, ^\wedge Pm, ^\wedge Pb\}^\cap \qquad \mathcal{A}_0 = \{^\wedge Pj, ^\wedge Pm, ^\wedge Pb\}^\cap$

[9]The fact that the "wh"-question in (4), the disjunctive question in (5a) and the disjunctive assertion in (5a) all give rise to the same relevant exhaustivity effects on B's response can also be used to motivate the attentional intent we assumed for the "wh"-question, perhaps favoring a particular question semantics.

[10]More generally, the foregoing entails that, for the simple types of utterances with which this paper is concerned, we can explain how attentional intents are reliably communicated without having to assume an "attentional" semantic dimension like Hamblin alternatives. After all, it suffices to assume – more minimally and innocently – that utterances simply inevitably draw attention to the meanings of any and every (proper and improper) constituent of the sentence (though not necessarily in equal degrees), regardless of how these constituents fit together syntactically. That of these various semantic objects attention was *intentionally* drawn only to the meanings of the disjuncts, i.e., that only these are part of the attentional intent, can be determined on the basis of the informational intent, the theme and the maxims, through fact 7. Now, semantic proposition sets appear to be necessary or at least very convenient elsewhere in semantics/pragmatics, so we are not opposed to using them; but it is good to know that for a relevant class of utterances Attentional Pragmatics *explains* such notions rather than having to assume them.



(a) (b) (c)

Figure 3: The theme, attentional intents and and exhaustivity implications of (3b) and (3c).

b. B: John was there.

$\mathcal{T}_0 = \{^\wedge Pj, ^\wedge Pm, ^\wedge Pb\}^\cap \qquad p_0 = ^\wedge Pj \qquad \mathcal{A}_0 = \{^\wedge Pj\}$

c. B: John was there, or John and Mary.

$\mathcal{T}_0 = \{^\wedge Pj, ^\wedge Pm, ^\wedge Pb\}^\cap \qquad p_0 = ^\wedge Pj \qquad \mathcal{A}_0 = \{^\wedge Pj, ^\wedge Pjm\}$

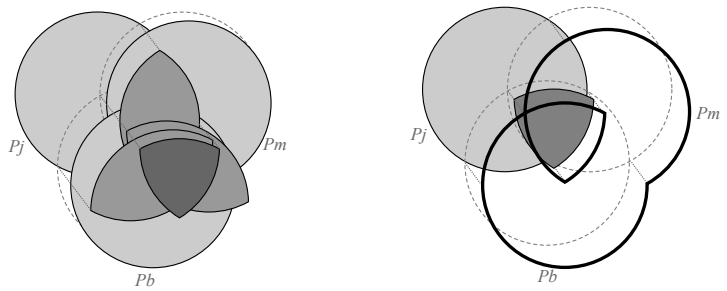
The exhaustivity implication we wish to derive for (3b) is that B believes that Mary and Bill are not at the party ($\Box\neg Pm$, $\Box\neg Pb$). For (3c) we want to derive the same for Bill ($\Box\neg Pb$), but not for Mary – with regard to Mary we may want to derive merely that the speaker does not consider Mary's presence possible independently of John's ($\Box(Pm \to Pj)$). The desired implications are depicted schematically in figure 3, which is identical to figure 1 given earlier except for the bold outlines, which contain those worlds that the exhaustivity implications would exclude from the speaker's doxastic state. Formally, exhaustivity follows from the assumption that B takes her attentional intent to comply with A-Quantity ($\Box$ A-Quantity$(\mathcal{A}_0, \mathcal{T}_0)$). That is, for (3b), as depicted in figure 3(b), we get:[11]

Fact 8. For all admissible models $\langle \mathbf{M}, w_0 \rangle$ for (3b):

$$\mathbf{M}, w_0 \models \Box \text{ A-Quantity}(\mathcal{A}_0, \mathcal{T}_0) \to (\Box\neg Pb \wedge \Box\neg Pm)$$

And for (3c), as depicted in figure 3(c), we get:

[11] No proofs will be given for these facts; but the appendix does contain a proof for the more general fact 10 below.



(a) (b)

Figure 4: The theme, attentional intent and exhaustivity implication of example (6).

Fact 9. For all admissible models $\langle \mathbf{M}, w_0 \rangle$ for (3c):

$$\mathbf{M}, w_0 \models \Box \text{ A-QUANTITY}(\mathcal{A}_0, \mathcal{T}_0) \rightarrow (\Box\neg Pb \wedge \Box(Pm \rightarrow Pjm))$$

And there exists such a model where:

$$\mathbf{M}, w_0 \not\models \Box \text{ A-QUANTITY}(\mathcal{A}_0, \mathcal{T}_0) \rightarrow \Box\neg Pm$$

Instead, for all admissible models for (3c) we have:

$$\mathbf{M}, w_0 \models \Box \text{ A-QUALITY}(\mathcal{A}_0) \rightarrow \neg\Box\neg Pm$$

Here the implication that Mary's presence is not possible independently of John's ($\Box(Pm \rightarrow Pjm)$) derives from A-Quantity's conditioning on A-Parsimony: B didn't draw attention to Mary's presence, so she must not consider it possible independently of John and Mary's joint presence. But note that, in this case, the same can be inferred from I-Quality: believing that John was at the party ($\Box Pj$) entails believing that he is if Mary is ($\Box(Pm \rightarrow Pjm)$). A-Parsimony is necessary, however, to explain why no attention is drawn to Mary's presence ($^\wedge Pm$) despite this being thematic and possible: it isn't possible independently of John and Mary's joint presence ($^\wedge Pjm$). A-Parsimony is also necessary for the exhaustivity implication of the variant in (4a) given earlier, repeated here in (6) (though without the particular speaker beliefs):

(6) B: John was at the party, or John, Mary and Bill.

$\mathcal{T}_0 = \{^\wedge Pj, ^\wedge Pm, ^\wedge Pb\}^\cap \qquad p_0 = ^\wedge Pj \qquad \mathcal{A}_0 = \{^\wedge Pj, ^\wedge Pjmb\}$

Figures 4(a) and (b) depict the assumed theme and the attentional intent, in which the bold outline again contains those worlds excluded by the exhaustivity implication from B's doxastic state. The implication is that, according to B, Mary or Bill can be at the party only if everyone is. As before, the implication that if Mary wa there then so were John and Bill follows from A-Quantity's conditioning on A-Parsimony; but this time, unlike in (3c), it doesn't already follow from I-Quality (since $\Box^\vee Pj$ does not entail $\Box(Pm \rightarrow Pjmb)$). Hence, A-Parsimony really makes a difference.

3.2 General characterization

To derive a more general result, let us restrict our attention to cases where the theme has the property of *chain completeness*, which means that for every chain of ever more specific, thematic propositions, their infinitary intersection is also thematic. The following can be proven:

Fact 10. For all admissible models $\mathbf{M}$, for arbitrary constants $\mathcal{A}_i$ and $\mathcal{T}_j$, where the theme denoted by $\mathcal{T}_j$ is chain-complete:

$$\mathbf{M} \models \Box\,\text{A-Quantity}(\mathcal{A}_i, \mathcal{T}_j) \rightarrow \forall a \left(\begin{array}{l} (\mathcal{T}_j(a) \wedge \neg\mathcal{A}_i(a)) \rightarrow \\ \quad \Box(\neg^{\vee} a \vee \exists b(\mathcal{A}_i(b) \wedge (b \subset a) \wedge {}^{\vee} b)) \end{array} \right)$$

And if in $\mathbf{M}$ the speaker's beliefs are accurate (factivity), then:

$$\mathbf{M} \models \Box\,\text{A-Quantity}(\mathcal{A}_i, \mathcal{T}_j) \rightarrow \forall a \left(\begin{array}{l} (\mathcal{T}_j(a) \wedge \neg\mathcal{A}_i(a)) \rightarrow \\ \quad (\neg^{\vee} a \vee \exists b(\mathcal{A}_i(b) \wedge (b \subset a) \wedge {}^{\vee} b)) \end{array} \right)$$

In words: compliance with A-Quantity implies that, for every proposition that is thematic and to which no attention is intended to be drawn, the speaker must think that it does not obtain or that, if it does obtain, a more specific proposition obtains to which attention is intended to be drawn. The second result in fact 10 is a strengthening (or simplification) for models in which the speaker's beliefs are accurate: every thematic proposition to which no attention is drawn is either false, or only true together with a more specific proposition to which attention is drawn. A proof is given in the appendix, and there the role of the chain-completeness restriction is also made clear.[12]

We can define a notational shorthand for the second result of fact 10, namely by taking its intension ($^{\wedge}$), while making sure through abstraction and application (a trick used by Zimmermann 1989) that the intent and theme constants are interpreted in the initial world of evaluation. We capture this in an exhaustivity operator Exh_{AP} (A(ttentional)P(ragmatics)):[13]

Definition 6. For $\mathcal{A}$ and $\mathcal{T}$ any constant or variable of type $\langle\langle s,t\rangle,t\rangle$, let the following notational shorthand be defined:

$$\text{Exh}_{\text{AP}}(\mathcal{A},\mathcal{T}) \stackrel{\text{def}}{=} \lambda\mathcal{T}'\left(\lambda\mathcal{A}'\ {}^{\wedge}\forall a\left(\begin{array}{l}(\mathcal{T}'(a)\wedge\neg\mathcal{A}'(a))\rightarrow\\ \quad(\neg^{\vee}a\vee\exists b(\mathcal{A}'(b)\wedge(b\subset a)\wedge{}^{\vee}b))\end{array}\right)(\mathcal{A})\right)(\mathcal{T})$$

Note that this exhaustivity operator is not a substantive assumption of the theory, but a mere notational shorthand for the exhaustivity implications that are predicted by the theory anyway, at least in admissible models, given factivity and compliance with A-Quantity. The operator can also be formulated in the metalanguage, in a more set-theoretical fashion:[14]

[12] The restriction to chain-complete themes is only a presentational choice, that allows a simpler formulation of the main result. It is not indicative of, say, some sort of defect in the maxims. We do not think that, relative to a theme that is not chain-complete, a speaker could rationally behave differently from what the current maxims predict, namely, to not draw attention to any particular proposition in the chain. If anything, a rational speaker may want to consider switching to a chain-complete theme instead.

[13] Admissible models must be understood, from now on, to validate also this definition, but we leave the required update of definition 1 implicit.

[14] Parameters $\mathbf{M}, w, g$ for the interpretation function $[\![.]\!]$, omitted for readability, are the same throughout.

(a) (b)

Figure 5: The theme and attentional intent (with exhaustivity implication) of (7B).

Fact 11. For arbitrary constants or variables $\mathcal{A}$ and $\mathcal{T}$, relative to an admissible model:

$$[\![\textsc{Exh}_{\textsc{AP}}(\mathcal{A},\mathcal{T})]\!] = \bigcap_{\substack{a\in[\![\mathcal{T}]\!]\\ a\notin[\![\mathcal{A}]\!]}} (\overline{a} \cup \bigcup_{\substack{b\in[\![\mathcal{A}]\!]\\ b\subset a}} b)$$

This obtains fairly directly, and we omit a formal proof: the universal quantifier in definition 6 corresponds here to generalized intersection; negation to complementation, disjunction to union and existential quantification to generalized union. As a first illustration, notice that the complements of the bold outlines in figures 3 and 4 given earlier correspond precisely to the sets of worlds characterized by the exhaustivity operator, when applied to the relevant themes and intents.

For a more interesting illustration of the operator, let us consider a more elaborate example:

(7) A: How many kids does John have?
B: John has one, three, or five kids.

$\mathcal{T}_0 = \{^\wedge K0, ^\wedge K1, ^\wedge K2, ^\wedge K3, ^\wedge K4, \ldots\} \qquad \mathcal{A}_0 = \{^\wedge K1, ^\wedge K3, ^\wedge K5\}$
$p_0 = ^\wedge K1 \quad$ (equivalent to $^\wedge(K1 \vee K3 \vee K5)$)

Here, $^\wedge Kn$ denotes the proposition that John has n kids. I assume an "at least"-interpretation of numerals, e.g., that if John has 3 kids, he therefore also has 2 kids, i.e., $K3 \rightarrow K2$.[15] The theme and the attentional intent of (7B) are depicted in figure 5, the nesting a consequence of the "at least"-interpretation. The striped regions together contain the worlds that are excluded by the operator (like the bold outlines before), which can be computed as follows. For readability I will now write bare numerals n as a shorthand for $^\wedge Kn$, i.e., the proposition that John has at least n kids. The following equivalences then hold in any admissible model for (7B):

$$\begin{aligned}
\textsc{Exh}_{\textsc{AP}}(\mathcal{A}_0,\mathcal{T}_0) &= (\overline{0}\cup 1\cup 3\cup 5) \quad \cap \quad (\overline{2}\cup 3\cup 5) \quad \cap \quad (\overline{4}\cup 5) \quad \cap \overline{6} \cap \overline{7} \cap \ldots \\
&= \quad 1 \quad \cap \quad (\overline{2}\cup 3) \quad \cap \quad (\overline{4}\cup 5) \quad \cap \overline{6} \\
&= (1\cap\overline{2}\cap\overline{4}\cap\overline{6}) \cup \ldots \cup (1\cap 3\cap\overline{4}\cap\overline{6}) \cup \ldots \cup (1\cap 3\cap 5\cap\overline{6}) \\
&= \quad (1\cap\overline{2}) \quad \cup \quad (3\cap\overline{4}) \quad \cup \quad (5\cap\overline{6})
\end{aligned}$$

The step from the second to the third line involves distributing unions over intersections, and eliding ("...") from the result all intersections that amount to the empty set. The last line says that John has exactly one, exactly three, or exactly five kids. Formally:

Fact 12. For all admissible models $\langle \mathbf{M}, w_0\rangle$ for (7B):

$$\mathbf{M}, w_0 \models \textsc{Exh}(\mathcal{A}_0,\mathcal{T}_0) = (^\wedge K1 \cap \overline{^\wedge K2}) \cup (^\wedge K3 \cap \overline{^\wedge K4}) \cup (^\wedge K5 \cap \overline{^\wedge K6})$$

[15] This treatment of numerals is not uncontroversial (for recent discussion see Kennedy, 2013), but the essential structure of the example could in principle be replicated using a "who"-question and disjunctions of conjunctions, as in (6) further above, avoiding numerals at the cost of longer and more artificial utterances.

The set of worlds thus computed corresponds to the gray, non-striped region in figure 5b.

A word of caution: the exhaustivity operator is not a substantive component of the theory, but a mere abbreviation of the predictions of the pragmatic account. Hence, before we apply the operator to an example we must ensure that the pragmatic assumptions on which the account relies actually obtain, or, we will say more formally, that the model is *operable*:

Definition 7 (Operable model)**.** An admissible model $\langle \mathbf{M}, w_0 \rangle$ is *operable* if and only if the speaker's beliefs are accurate (factivity), in w_0 the relevant intents comply with the maxims relative to the relevant themes, and the set of thematic propositions in w_0 is chain-complete.

This restriction must be kept in mind when applying the operator to particular examples, but also when comparing the operator to existing operators in the literature: such a comparison will be meaningful only in operable models. As we will see further below, with this restriction in mind, the current operator is in relevant respects conservative with regard to existing operators in the literature. We will discuss the *minimal worlds* operator $\textsc{Exh}_{\text{mw}}$, which has been shown to abbreviate the traditional pragmatic approach based on I-Quantity (Spector 2007; Schulz and Van Rooij 2006) and a more sophisticated *dynamic* operator $\textsc{Exh}_{\text{dyn}}$ from Schulz and Van Rooij 2006, which basically applies an operator like $\textsc{Exh}_{\text{mw}}$ separately to each disjunct of an utterance.

4 Comparison to the standard recipe

4.1 The standard recipe in a nutshell

Where Attentional Pragmatics derives exhaustivity from the maxim of A-Quantity, the standard pragmatic recipe is based on the maxim of I-Quantity: that rational speakers intend to share all thematic (or relevant) information they take to be true. To illustrate, consider again (3b), "John was there.". From I-Quantity it follows that, since B did not intend to convey that Mary was at the party, she must not have had the belief that she was. Indeed, fact 3 given earlier stated that any admissible model for (3b) verifies the following:

$$\Box\, \text{I-}\textsc{Quantity}(p_0, \mathcal{T}_0) \rightarrow (\neg\Box Pm \wedge \neg\Box Pb)$$

The I-Quantity implication is almost like exhaustivity, but not quite: from the absence of a positive belief ($\neg\Box Pm$) we may not in general conclude the presence of a negative belief ($\Box\neg Pm$) – 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 assume, normally or in the relevant situations, 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 speaker B in (1) takes herself to know whether or not Mary was at the party, the I-Quantity implication can be strengthened, yielding exhaustivity:

$$\neg\Box Pm \wedge \overbrace{(\Box Pm \vee \Box\neg Pm)}^{\text{opinionatedness assumption}} \models \Box\neg Pm$$

I-Quantity plus opinionatedness is the essence of the standard recipe, which has been adopted by many (e.g., Horn 1972; Gazdar 1979; Schulz and Van Rooij 2006; Spector 2007; Geurts 2011).

Just like we derived the operator $\textsc{Exh}_{\text{AP}}$ from the maxim of A-Quantity, Schulz and Van Rooij (2006); Spector (2007) have shown that the outcome of the standard recipe can be captured in an exhaustivity operator. Let us denote this so-called *minimal worlds* operator by $\textsc{Exh}_{\text{mw}}$, and temporarily add it to the formal language, with the following semantics (as before, the omitted parameters of [.] are the same everywhere):

Definition 8. For arbitrary constants/variables p and $\mathcal{T}$, let:

$$[\textsc{Exh}_{\text{mw}}(p, \mathcal{T})] \stackrel{\text{def}}{=} \{w \in [p] \mid \text{there is no } w' \in [p] \text{ such that: } \{a \mid a \in [\mathcal{T}], w' \in a\} \subset \{a \mid a \in [\mathcal{T}], w \in a\}\}$$

That is, the proposition denoted by p must be true in the relevant worlds w, together with a set of other thematic propositions that is *minimal* compared to the sets of true thematic propositions in other worlds w' in which the proposition denoted by p is true. As it turns out, the minimal worlds operator aligns with the current operator if the attentional intent is a singleton set containing the informational intent:

Fact 13. Take any utterance with intents denoted by p_i and $\mathcal{A}_j$ such that $\mathcal{A}_j = \{p_i\}$ is true, and theme denoted by $\mathcal{T}_k$. For any admissible, operable model $\langle \mathbf{M}, w_0 \rangle$ for such an utterance:

$$\mathbf{M}, w_0 \models \textsc{Exh}_{\text{mw}}(p_i, \mathcal{T}_k) = p_i \cap \textsc{Exh}_{\text{AP}}(\mathcal{A}_j, \mathcal{T}_k)$$

A proof is given in the appendix.[16] In effect, the above correspondence shows that the operators align when attention isn't there to make a difference (as in (3b), but unlike (3c)). Even in such cases, though, fact 13 does not imply that the standard recipe and Attentional Pragmatics generate the same empirical predictions, because the operators derive from different pragmatic assumptions: compliance with A-Quantity for $\textsc{Exh}_{\text{AP}}$ but compliance with I-Quantity and an opinionatedness assumption for $\textsc{Exh}_{\text{mw}}$. As a consequence the predictions of the two accounts will vary with the availability of the assumptions on which they rely, as we will see.

In what follows we will discuss several cases where the predictions of the two accounts indeed diverge. In particular, as announced in the introduction to this paper, we will discuss three (potential) problems for the standard recipe which, if they are indeed problems, are solved by Attentional Pragmatics:

I. **the granularity problem:** I-Quantity and hence the standard recipe as a whole cannot (easily) distinguish between examples like (1) and (2);

II. **the epistemic step:** the opinionatedness assumption invoked by the standard recipe to bridge the gap between I-Quantity and exhaustivity is ad hoc, and exhaustivity seems to occur also without it;

III. **the exemption problem:** exhaustivity seems to occur on utterances that are arguably exempt from I-Quantity, namely the hints of a quizmaster.

We aim to make plausible, in what follows, that all of these are indeed genuine problems for the standard recipe, and will explain how Attentional Pragmatics solves them. As we will see, to the extent that the empirical facts are clear they seem to favor Attentional Pragmatics. In cases where the empirical jury is to our awareness still out, the following will serve primarily to clarify the main ways in which Attentional Pragmatics differs from the standard recipe, leaving an empirical verdict for another occasion – the primary aim of this paper, recall, is theoretical.

[16] See Spector 2016 for a close correspondence, in turn, between the operator $\textsc{Exh}_{\text{mw}}$ and other existing operators from the literature, i.e., from Krifka 1993 and Fox 2007.

4.2 The granularity problem

The standard recipe cannot (easily) distinguish between examples like (3b) and (3c), with "or both" – a consequence of the fact that I-Quantity and its derivative operator $\textsc{Exh}_{\text{mw}}$ operate only on informational intents, combined with the assumption that (3b) and (3c) have the same primary informational intent.[17] The only room to maneuver, for the standard recipe, lies in the opinionatedness assumption: it must be made sensitive somehow to the disjuncts of the utterance, such that the speaker is not considered opinionated about disjuncts like "or both". To our awareness two approaches exist that try something along these lines: Gazdar's (1979) approach based on "clausal implicatures", and Schulz and Van Rooij's (2006) approach based on dynamic semantics. We will briefly discuss each approach.[18]

Gazdar's (1979) "clausal implicatures" Gazdar tries to curb the opinionatedness assumption by assuming that utterances have "clausal implicatures" to the effect that a speaker should be uncertain about any embedded clause of an uttered sentence, e.g., the disjuncts of a disjunction.[19] The speaker would then be assumed to be only as opinionated as these clausal implicatures allow. Gazdar seeks to derive these clausal implicatures from I-Quantity, combined with the assumption that both the embedded clause and its negation are thematic (or relevant): the speaker's uncertainty about the truth of the embedded clause then follows from the fact that the speaker asserted neither the embedded clause nor its negation. Thus, for (3c), "John, or both John and Mary", Gazdar would assume that both John and Mary's joint presence and its negation are thematic, in order to derive an ignorance implication from I-Quantity that then blocks the exhaustivity implication that Mary wasn't there. Now, we do not think that embedded clauses and their negations are always thematic – e.g., it is conceivable that one would utter "John believes that φ" when one doesn't care about the truth of φ, or when one already knows that φ is true (or false) – but for the disjuncts of a disjunction Gazdar's assumption may not seem unreasonable.

Nevertheless, the assumption that both the disjuncts and their negations are thematic is problematic: lest it be ad hoc, one would have to assume for (3c) that not only John and Mary's joint presence and its negation are thematic, but also Mary's presence and its negation, John's presence and its negation, Bill's presence and its negation, and so on, i.e., that the theme is closed under negation. But by assuming closure under negation more generally one runs into the well-known symmetry problem, which we briefly discussed in section 2.2: in (3c) closure under negation would block the exhaustivity implication that Bill isn't there, which is undesirable, and an analogous assumption for (3b) would wrongly block the implication that Mary isn't there. In sum, the assumption about themes that is required for deriving clausal implicatures from I-Quantity in order to block *some* exhaustivity implications, will, when generalized, block *all* of them.

Instead of trying to derive opinionatedness-blocking implications from I-Quantity, as Gazdar (1979) does, one could try to derive them more directly by assuming that one should consider all disjuncts possible. The speaker can then be assumed to be only as opinionated as these possibility implications allow, and this combined with I-Quantity would deliver the right sort of exhaustivity. We are unaware of existing accounts of exhaustivity along these lines, but a question that such an account would raise is why a speaker would have to consider all disjuncts possible. In Attentional

[17] We are unaware of proposals that the primary informational intents of (3b) and (3c) would differ in a way that could explain the difference in exhaustivity. (The primary intents would differ according to the grammatical approach, because it incorporates the exhaustivity into the primary intents – but as such the difference between the intents could not be used to explain the difference in exhaustivity, i.e., to explain itself.)

[18] An altogether different type of solution to the granularity problem has been proposed within the grammatical approach, a discussion of which, as we announced, must be left for another occasion. But see our remarks on A-Parsimony and Hurford's Constraint in section 2.4.

[19] Gazdar used "implicature" to mean "implication", and we will simply adopt his composite "clausal implicature", but not without stressing that the distinction between implication and implicature is very important (Bach, 2006): implications are often not part of what is meant, and the truth of implicatures is often not implied.

Pragmatics this is explained by the maxim of A-Quality combined with the prediction that, in the relevant examples, the disjuncts are what the speaker intends to draw attention to (fact 7). But if attention-drawing is a communicative intention, i.e., if we assume that utterances have attentional intents, then we need a set of maxims that determines which attentional intents are rational – A-Quality on its own is insufficient (cf. the determinacy established in fact 5). The attempt to derive clausal implicatures from A-Quality may therefore ultimately lead to something like Attentional Pragmatics, in which, as it turns out, I-Quantity and the opinionatedness assumption are no longer necessary for exhaustivity: the attentional maxims are sufficiently restrictive on their own. Further below we will see that severing the ties between exhaustivity and I-Quantity and the opinionatedness assumption in this way appears to be adequate.

Schulz and Van Rooij's (2006) dynamic operator Schulz and Van Rooij (2006), after pragmatically motivating the minimal worlds operator $\textsc{Exh}_{\text{mw}}$ given above, reformulate their operator in dynamic semantics so as to render it sensitive to the discourse referents introduced by an utterance. Combined with the assumption that a disjunction introduces its disjuncts as discourse referents, this new operator $\textsc{Exh}_{\text{dyn}}$ overcomes the granularity problem. It is insightful to formally compare $\textsc{Exh}_{\text{dyn}}$ to our operator $\textsc{Exh}_{\text{AP}}$. Their operator $\textsc{Exh}_{\text{dyn}}$ is similar to the minimal worlds operator $\textsc{Exh}_{\text{mw}}$, the difference being that $\textsc{Exh}_{\text{dyn}}$ does not minimize the set of true thematic propositions among *all* worlds in the informational intent, but only within certain subsets. We will bypass the details of how they determine these subsets – in a nutshell, they compare only world-assignment pairs that share the same assignment (discourse referents). At least for disjunctive utterances that introduce a discourse referent for each disjunct, this amounts simply to comparing only worlds within some proposition expressed by a disjunct, i.e., within some proposition in the attentional intent. This is how we (re)define their operator:

Definition 9. For arbitrary constants/variables $\mathcal{A}$, and $\mathcal{T}$, let:

$$[\![\textsc{Exh}_{\text{dyn}}(\mathcal{A},\mathcal{T})]\!] = \{w \mid \text{for some } a \in [\![\mathcal{A}]\!]\text{: } w \in a \text{ and there is no } w' \in a \text{ s.t. } \{b \mid b \in [\![\mathcal{T}]\!], w' \in b\} \subset \{b \mid b \in [\![\mathcal{T}]\!], w \in b\}\}$$

Whether this definition corresponds exactly to theirs depends on the degree to which the current attentional intents align with what they consider to be discourse referents. For our simple examples the two notions seem to align. More generally, however, to our understanding discourse referents are the things that are available for subsequent anaphoric reference, which could be anything to which the utterance draws attention. An attentional intent, in contrast, will be more restricted: it contains only propositions, and only those to which attention was intentionally drawn (namely, normally, those which are thematic and independently possible). It is an important task to map the relations between the different attention-related and proposition-set-like notions in the literature, but it is also a substantial challenge, and one which is (fortunately) not essential for the main points of this paper.

The operator $\textsc{Exh}_{\text{dyn}}$ can distinguish between examples like (3b) and (3c), as well as account for cases like (6). Indeed, our operators align quite generally:

Fact 14. Take any utterance with intents denoted by p_i and $\mathcal{A}_j$ and a theme denoted by $\mathcal{T}_k$, where the theme is closed under intersection. For any admissible, operable model $\langle \mathbf{M}, w_0 \rangle$ for such an utterance:

$$\mathbf{M}, w_0 \models \textsc{Exh}_{\text{dyn}}(\mathcal{A}_j, \mathcal{T}_k) = (p_i \cap \textsc{Exh}_{\text{AP}}(\mathcal{A}_j, \mathcal{T}_k))$$

A proof is given in the appendix. The correspondence highlights that the contribution of this paper

is not the exhaustivity operator in itself, as a purely formal construct, but its derivation from Attentional Pragmatics. In contrast, Schulz and Van Rooij offer only a partial explanation for their operator – they do not motivate its sensitivity to discourse referents, which is precisely what gives it an edge over EXH_{mw}. Moreover, the part which they do derive from pragmatic considerations (basically EXH_{mw}) relies on I-Quantity plus a "maximal opinionatedness" assumption (namely, the assumption that the speaker is as opinionated as is compatible with her primary informational intent). This means that, despite the formal correspondence between our operators, our accounts yield different predictions in cases where I-Quantity does not apply or where no assumption of maximal opinionatedness is present, as we will see in subsequent subsections.[20]

The dynamic operator EXH_{dyn}, by virtue of its definition, effectively applies a minimal worlds operator EXH_{mw} to each individual disjunct. This is basically what the *grammatical approach* seeks to do by applying covert grammatical exhaustivity operators to the disjuncts (Chierchia et al., 2009). As such, fact 14 also entails a formal correspondence, albeit limited to the relevant examples, between Attentional Pragmatics and the grammatical approach. Again, whether this formal correspondence also entails a descriptive/empirical correspondence depends on when the grammatical approach predicts these covert exhaustivity operators to be present, a matter that we cannot discuss here for reasons of space. In any case, an important difference is that, whereas the grammatical approach predicts that individual disjuncts are interpreted exhaustively, i.e., that there are local exhaustivity effects, the same cannot be said for Attentional Pragmatics. The difference between the two approaches can be paraphrased as follows (the example appeared earlier as (6)):

(8) B: John, or John, Mary and Bill.

a. **Grammatical approach:**
John *and no one else* was there, or John, Mary and Bill. $(Pj \wedge \neg Pm \wedge \neg Pb) \vee Pjmb$

b. **Attentional Pragmatics:**
John was there, or John, Mary and Bill – *and if Mary or Bill was then everyone was.*
$(Pj \vee Pjmb) \wedge ((Pm \vee Pb) \rightarrow Pjmb)$

Although the two paraphrases (and formulae) are classically, informationally equivalent, only (8a) really involves the exhaustive interpretation of an individual disjunct. What this highlights is that a globalist pragmatic theory can explain the appearance, at the level of overall truth conditions of the assertion plus the implicature, of a local exhaustivity effect, at least in the relevant examples. Of course a prerequisite for achieving this is that one's globalist pragmatic theory be sensitive to a dimension of speaker meaning that, like attentional intents, reflects the syntactic structure of the uttered sentence more closely than informational intents do (cf. Simons 2011). In the case of Attentional Pragmatics, the maxim that is primarily responsible for local-esque exhaustivity effects is A-Parsimony, which we motivated in section 2.4 as a global, independently reasonable pragmatic constraint. There we also mentioned that A-Parsimony is substantially different from the constraint by which the grammatical approach seeks to enforce local exhaustification in the above examples, namely Hurford's Constraint.

Summing up, Gazdar's (1979) attempt to solve the granularity problem in terms of clausal implicatures derived from I-Quantity is unsatisfactory – and an attempt to derive them instead from A-Quality could potentially lead to Attentional Pragmatics. Schulz and Van Rooij's (2006) approach is formally equivalent (in the relevant circumstances) to the exhaustivity operator derived

[20] A somewhat similar but likewise not pragmatically derived operator is defined in Alonso-Ovalle 2008. In a nutshell, the operator simply negates propositions that are "innocently excludable", a notion adopted from Fox 2007 but made sensitive to the set of disjuncts of an utterance. Unlike EXH_{AP} and EXH_{dyn}, Alonso-Ovalle's operator doesn't work for examples like (6) and (7), because it will never exclude only *part* of a proposition, which is what is needed there.

from Attentional Pragmatics, and as such it solves the granularity problem, but it is incomplete as an explanation.

4.3 The epistemic step: exhaustivity without an opinionatedness assumption

Recall that Attentional Pragmatics predicts exhaustivity without relying on an opinionatedness assumption: it suffices to assume that the speaker complies with A-Quantity, i.e., intends to draw attention to all thematic propositions that are possible independently of something stronger to which attention is drawn. Thus, the epistemic step between I-Quantity and exhaustivity (Sauerland, 2004) does not exist between A-Quantity and exhaustivity. This may be an attractive feature of the current approach: it seems questionable that it would be rational to go about the world assuming – without evidence – that our interlocutors are opinionated. Indeed, the standard recipe'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).

However, the opinionatedness assumption is embraced by many (e.g., Soames 1982; Horn 1984; Matsumoto 1995; Green 1996; Russell 2006; Schulz and Van Rooij 2006; Geurts 2011). For instance, Geurts (p.30) writes that connecting "weak" implications like $\neg\Box Pm$ to "strong" implications like $\Box\neg Pm$ via the opinionatedness assumption is "one of the main virtues" of the approach. This appreciation of the epistemic step is due largely to the fact that exhaustivity implications seem to disappear if the speaker denies her own opinionatedness, as Soames (1982) noted:

(9) A: Who was at the party?
B: I'm not sure about the others, but John was there.

Similar examples can be found throughout the literature. Breheny et al. (2013) present experimental results that seem to support the reliance of exhaustivity implications on an opinionatedness assumption also in less explicit cases. They presented participants with utterances that could measurably trigger or not trigger an exhaustivity implication, while manipulating the participant's knowledge about the speaker. When the participants knew that the speaker had only partial knowledge about the theme, the exhaustivity implications were significantly weaker than when no such knowledge was present. Goodman and Stuhlmüller (2013) present similar results.

However, we think that these experimental results as well as Soames's original observation have been misinterpreted. After all, in the relevant examples 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 *un*opinionated is to be unable to give an exhaustive answer, it is unsurprising that assuming *un*opinionatedness has some effect on the perceived exhaustivity implications, regardless of whether their presence would have depended on an opinionatedness assumption to begin with.[21] A better test case for the standard recipe'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 (a similar example is considered in Westera 2014):

[21] We would explain the effect of an *un*opinionatedness assumption on exhaustivity, as in (9), as follows: because B's answer cannot be exhaustive (given her uncertainty about the others), it cannot comply with the maxims relative to the full prior theme of who was at the party, but must be interpreted either as violating a maxim – in which case one might expect linguistic cues of a maxim violation, like perhaps rising or fall-rising intonation – or as addressing an implicitly more restricted theme, i.e., "*...of the people I do know about*, John was there". Relative to such a restricted theme, an accordingly weaker exhaustivity implication would be predicted, one which does not contradict the speaker's self-proclaimed *un*opinionatedness. A further exploration of this type of analysis is unnecessary for present purposes, and it would require more precise theories of maxim violations and theme pragmatics than we can offer in the present paper.

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

In this conversation, A explicitly refrains from assuming B's opinionatedness, though crucially without assuming her *un*opinionatedness 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). This was only a first exploration, not a proper empirical study, as our aim in this paper is not to defend any substantial empirical claims. However, as we have pointed out, purported evidence to the contrary, i.e., in favor of the opinionatedness assumption, has in fact been misinterpreted, so perhaps our tentative claim with regard to (10) should not be dismissed too easily. But what matters is that (10) would be a case where the predictions of the standard recipe and Attentional Pragmatics diverge, and that, as such, a proper empirical assessment of cases like (10) could be very valuable.

A possible objection to our interpretation of (10) 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 "complete" (e.g., Pierrehumbert and Hirschberg 1990), a notion that is made more precise by Westera (2013) in terms of compliance with the Gricean maxims, and by Malamud and Stephenson (2015) in terms of meta-pragmatic issues remaining unresolved. 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). In contrast, we are unaware of any precise, predictive theory of intonational meaning that would associate falling intonation with opinionatedness. A challenge for such a hypothetical theory would be that it would have to be the right *degree* of opinionatedness, or one might wrongly predict the same exhaustivity implication for (3b) and (3c), with "...or both", which implies opinionatedness in one respect and ignorance in another (Schulz and Van Rooij, 2006). 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 that is not ad hoc: the (partial) opinionatedness of a speaker follows from the requirement that speakers draw attention to all thematic propositions they consider independently 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.

4.4 The exemption problem: exhaustivity without I-Quantity

As Fox (2014) notes, a quizmaster who is giving a hint would not normally be expected to share all the relevant information in their possession, i.e., hints seem exempt from I-Quantity. Nevertheless, we agree with Fox that exhaustivity can still be present (depending on intonation, as usual):

(11) Quizmaster: (Of these three boxes over here,) there is money in box A or in box C.
(implied: according to the quizmaster not in both, and not in B)

Fox notes with regard to an analogous example that while exhaustivity seems present, there is no genuine I-Quantity implication from which it could derive, i.e., one would not normally be justified in inferring that the quizmaster lacks the belief that the money is in box A, or the belief that the money is in box C. Hence, there is no way for the standard recipe, which relies on I-Quantity, to predict exhaustivity for (11). Although we agree with Fox's judgment, to our awareness no proper empirical evaluation of this type of example has been conducted. Pending a proper empirical evaluation, then, the main point of the following will be that the predictions of th standard recipe and Attentional Pragmatics differ with regard to cases like (11). However, for ease of speaking, let us suppose in what follows that our judgments regarding (11) are representative.

One could try to account for (11) within Attentional Pragmatics by supposing that, while quizmasters 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 usual, through A-Quantity. However, while we find it quite intuitive that a quizmaster's hints would be exempt from I-Quantity but not from A-Quantity, and this seems like an easy explanation, the explanation would not be entirely satisfactory. For one, the intuition that a quizmaster shouldn't violate A-Quantity may well be a mere consequence of the explanandum, i.e., that hints imply exhaustivity, rather than the starting point of an explanation. Moreover, even if the assumption that hints are exempt from I-Quantity but not from A-Quantity could be part of an explanation, the assumed difference between these maxims would itself be in need of explanation, lest it be ad hoc. A significant challenge in this regard is that it doesn't seem to instantiate a broader pattern of I-maxims vs. A-maxims, because the maxims of I/A-*Quality* would have to behave in the opposite way: while (11) seems exempt from A-Quality – the quizmaster needn't actually consider it possible that there is money in box A – it doesn't seem to be exempt from I-Quality, i.e., there has to money in box A or box C.

A more satisfying explanation involves pretense.[22] We find the following assumption plausible:

Assumption 5. Quizmasters may pretend to know *less* about the quiz answers than what they actually know, but not to know *more* about the quiz answers than what they actually know (i.e., information that is (potentially) false), lest they be guilty of misleading.

It follows from this assumption 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 and A-Quality both imply not-knowing (namely $\neg\Box$ and $\Diamond$ ($= \neg\Box\neg$), respectively), which can therefore be pretended, whereas I-Quality and A-Quantity both imply knowing (namely $\Box$ and $\neg\Diamond$ ($= \Box\neg$), respectively), which must therefore be genuine. In this way assumption 5 explains the (non-)exemption pattern of the different maxims in a quiz situation, which avoids having to make any ad-hoc assumptions in this regard. If this is on the right track, then Attentional Pragmatics indeed predicts exhaustivity for (11), unlike the standard recipe.

The foregoing line of explanation, based on assumption 5, can work for any theory that delivers exhaustivity directly rather than via not-knowing, including for instance the grammatical approach. After all, assumption 5 entails that any implied knowing must be genuine, and this includes exhaustivity regardless of whether it is delivered by A-Quantity or by a grammatical operator plus I-Quality. Hence, we agree with Fox's arguments that cases like (11) – if our judgments are representative – favor the grammatical approach over the standard recipe. But what we have shown is that there is a pragmatic alternative that, because it delivers exhaustivity directly, can

[22] Pretense is considered also by Fox (2014), but only as a hypothetical and ultimately inadequate attempt to save the standard recipe: if compliance with I-Quantity is pretended, there simply is no way to derive *genuine* exhaustivity from it.

solve this exemption puzzle as well: Attentional Pragmatics.

5 Discussion and outlook

This paper presented Attentional Pragmatics, a theory based on the basic idea that exhaustivity arises when relevant propositions are not mentioned (assumption 1). This basic idea was implemented in terms of the notion of attentional intent and a set of maxims that govern it. It is different from the standard pragmatic recipe, which is based on the more common idea that exhaustivity would arise when relevant propositions are not *asserted*. This difference, we have shown, results in different predictions with regard to a range of (potential) challenges for the standard recipe: the granularity problem, the epistemic step, and the exemption problem (hints). In a nutshell, in cases where the two approaches yield different predictions, Attentional Pragmatics is, as far as the empirical facts are clear, descriptively more accurate. Altogether, it seems that exhaustivity is probably a matter of attention, not information, i.e., it derives from A-Quantity, not I-Quantity. If true, this would mandate a thorough revision of the extensive experimental and theoretical literature on the topic.

As we announced in the introduction, we did not compare Attentional Pragmatics to the grammatical approach, except occasionally, in passing. The availability of covert exhaustivity operators in the grammatical approach potentially enables one to bypass I-Quantity and the opinionatedness assumption, which would solve the exemption problem (Fox, 2014) and the epistemic step (Chierchia et al. 2012). It would also potentially solve the granularity problem, namely by applying operators to each individual disjunct and letting the resulting symmetry of the disjuncts block subsequent exhaustification (Chierchia et al., 2009). The main challenge for the grammatical approach is to turn "potentially" in each of the foregoing phrases into "actually", which requires, foremost, an explanation of when and where these covert exhaustivity operators are present (Russell, 2006; Chierchia et al., 2012; Geurts, 2013). Recent proposals have explored the potential of Hurford's Constraint in this regard (Chierchia et al., 2009; Katzir and Singh, 2013; Mayr and Romoli, 2016). As we mentioned when clarifying A-Parsimony, motivations for Hurford's Constraint based on informational redundancy fall short if Attentional Pragmatics is adopted, because disjuncts that are informationally redundant may well make an attentional contribution. However, *attentional* redundancy does play a role in our account: the maxim of A-Parsimony. For this reason we believe that a detailed comparison of the two approaches is important not only for a better understanding of exhaustivity, but also for a better understanding of the role of redundancy in semantics/pragmatics. We leave must this for another occasion.

Attentional Pragmatics is not intended to (and cannot) account for embedded exhaustivity effects (recall that it treats the embedded exhaustivity effects of Hurford disjunctions as only apparent). As such, the scope of Attentional Pragmatics (like the standard recipe) is more limited in principle than the possible scope of the grammatical approach. But as we said in the introduction to this paper, language and cognition may give rise to exhaustivity effects in more ways than one – some possible sources are modal operators (Uegaki, 2015), metalinguistic negation (Horn, 1985; Geurts, 1998), typicality inferences (Van Tiel, 2014), impliciture (Bach 1994) – and indeed grammatical exhaustivity operators (Chierchia et al., 2012), the reality of which would not in principle be incompatible with Attentional Pragmatics. Until we have a better, more integrated understanding of these various potential sources of exhaustivity, it is important to keep investigating to what extent certain subclasses of exhaustivity effects may receive an adequate explanation, such as un-embedded exhaustivity effects in Attentional Pragmatics. Accordingly, the empirical partiality of any given theory of exhaustivity, which seems to us inevitable at this stage, does not necessarily constitute an argument against the theory – though ultimately it may.

Lastly, as a potential future application of Attentional Pragmatics we wish to highlight *questions*.

Questions arguably lack a primary informational intent for the I-maxims to apply to – e.g., questions lack assertive force (Frege, 1918; Krifka, 2011); or their informational content is a tautology (Ciardelli et al., 2013). Nevertheless, questions have been noted to imply exhaustivity (e.g., Bartels 1999; Roelofsen and Van Gool 2010; Biezma and Rawlins 2012; though depending on intonation). We are unaware of any pragmatic explanation for the exhaustivity effects of questions. However, Attentional Pragmatics may apply as usual: although questions do not primarily serve to provide information, they may well serve to draw attention to things, and as such are subject to the A-maxims just like assertions – see Biezma and Rawlins 2012 for an account of questions in line with this attentional view. Thus, Attentional Pragmatics may offer a uniform account of the exhaustivity effects of questions and assertions.[23] We leave a detailed exploration of this suggestion, and a comparison to the extensive literature on the exhaustivity effects of questions, for another occasion.

Appendix: proofs of facts 6, 10, 13, and 14

Proof of fact 6

Take an arbitrary admissible model $\mathbf{M}$ in which $\mathcal{T}_i$ denotes a theme closed under intersection. Take an arbitrary assignment function g and suppose that it assigns to p and $\mathcal{A}$ intents that comply with the maxims:

$$\mathbf{M}, g \models \text{I-MAXIMS}(p, \mathcal{T}_i) \wedge \text{A-MAXIMS}(\mathcal{A}, \mathcal{T}_i)$$

We will show that $p = \bigcup \mathcal{A}$ is true, by showing that neither $p \supset \bigcup \mathcal{A}$ nor $p \not\supseteq \bigcup \mathcal{A}$ is true.

Suppose (to obtain a contradiction) that $p \supset \bigcup \mathcal{A}$ is true. This means that no attention is drawn to the proposition assigned to p, i.e., $p \notin \mathcal{A}$. Since the proposition assigned to p is thematic and considered true (I-Relation, I-Quality), and hence thematic and considered possible, the reason (given compliance with A-Quantity) why no attention is drawn to it must be A-Parsimony, i.e., there must be a more specific thematic proposition independently of which p is not believed to be possible. That is, there exists an assignment function g' that differs from g only in the assignment to, say, a, such that:

$$\mathbf{M}, g' \models a \in \mathcal{T}_i \wedge a \subset p \wedge \Box(^\vee p \rightarrow {}^\vee a)$$

But since the proposition denoted by p itself is believed to be true (I-Quality), so must the denotation of a. And since the latter is both thematic and believed to be true, compliance of the proposition denoted by p with I-Quantity demands that $p \subseteq a$ is true, which is not the case. Contradiction.

Alternatively, suppose (again to obtain a contradiction) that $p \not\supseteq \bigcup \mathcal{A}$ is true. This means that attention is drawn to some proposition that is a superset of or logically independent of (properly overlaps) the proposition denoted by p. That is, there exists an assignment function g' that differs from g at most in, say, a, such that:

$$\mathbf{M}, g' \models a \in \mathcal{A} \wedge \neg(a \subseteq p)$$

Now, we know that the denotations of p and a are thematic (I-Relation and A-Relation, respectively), hence so is their intersection, i.e., $\mathcal{T}_i(p \cap a)$ is true. Moreover, since $\Box{}^\vee p$ is true (I-Quality), so

[23] It is important to note that a uniform account of the exhaustivity *implication* is compatible in principle with the common claim that the exhaustivity implications of assertions would serve the communication of an indirect intent, i.e., a Gricean conversational implicature, unlike the exhaustivity implications of questions, which, in contrast, would be presupposed (Bartels 1999; Aloni and Égré 2010; Biezma and Rawlins 2012). That is, the question of whether a given pragmatic implication serves the communication of an intent (or, "is at-issue") is separate from the question of how the implication is derived to begin with (Bach, 2006). It is only the latter question that Attentional Pragmatics would answer.

is $\Box({}^{\vee}a \rightarrow {}^{\vee}(p \cap a))$, which means that the proposition denoted by a is not considered possible independently of its intersection with p. It follows that, by A-Parsimony, no attention ought to have been drawn to the proposition denoted by a. Contradiction.

In sum, neither $p \supset \bigcup \mathcal{A}$ nor $p \not\supseteq \bigcup \mathcal{A}$ is true. It follows that $p = \bigcup \mathcal{A}$ is true.

Proof of fact 10

Take an arbitrary admissible model $\mathbf{M}$ in which the denotation of $\mathcal{T}_j$ is chain-completene. Take an arbitrary world w in this model. Suppose that the speaker takes A-Quantity to be complied with, i.e., $\mathbf{M}, w \models \Box$ A-Quantity$(\mathcal{A}_i, \mathcal{T}_j)$. Given intent introspection and theme introspection, this means that A-Quantity is actually complied with in w:

$$\mathbf{M}, w \models \forall a \left(\left(\begin{array}{c} \text{A-Quality}(\{a\}) \wedge \\ \text{A-Relation}(\{a\}, \mathcal{T}_j) \wedge \\ \text{A-Parsimony}(\{a\}, \mathcal{T}_j) \end{array} \right) \rightarrow \mathcal{A}_i(a) \right) \tag{1}$$

Take an arbitrary function g that assigns to a a thematic proposition, i.e., suppose that:

$$\mathbf{M}, w, g \models \mathcal{T}_j(a) \tag{2}$$

Suppose, furthermore, that no attention is drawn to a in w:

$$\mathbf{M}, w, g \models \neg \mathcal{A}_i(a) \tag{3}$$

Since $\mathcal{A}_i(a)$ is false in w, the antecedent in (1) cannot be true either, hence at least one of its conjuncts must be false. A-Relation cannot be blamed, because the proposition denoted by a is thematic in w (from supposition (2)), so it must be either A-Quality or A-Parsimony. Let us explore the latter.

Suppose that the singleton intent denoted by $\{a\}$ does not comply with A-Parsimony in w, i.e., $\mathbf{M}, w, g \models \neg$ A-Parsimony$(\{a\}, \mathcal{T}_0)$. This amounts to:

$$\mathbf{M}, w, g \models \Diamond {}^{\vee}a \wedge \Box({}^{\vee}a \rightarrow \exists b\, (b \subset a \wedge \mathcal{T}_0(b) \wedge {}^{\vee}b)) \tag{4}$$

It follows that there exists a world w' that is belief-accessible from w, such that the proposition assigned to a is true in w', and, by the second conjunct, some stronger proposition can be assigned to b that is true and thematic in w'. This means that in the original world w, the proposition assigned to b must be considered possible and, by theme introspection, thematic. Hence, we have:

$$\mathbf{M}, w, g \models \exists b(b \subset a \wedge \mathcal{T}_0(b) \wedge \Diamond {}^{\vee}b)$$

Since this stronger proposition b is thematic and possible, A-Quantity (which is complied with according to (1)) requires that it be an element of the attentional intent denoted by $\mathcal{A}_i$ in w, unless A-Parsimony prevents this, i.e., unless there is an even stronger thematic and possible proposition (say, c), independently of which b in turn is not considered possible. And so on, potentially *ad infinitum*.

To curb this potential infinitude, we assumed that the set of thematic propositions is *chain-complete*, i.e., that for every chain of increasingly specific thematic propositions $a_0, a_1, \ldots$ (i.e., such that every $a_{i+1} \subset a_i$), their infinitary intersection $\bigcap\{a_0, a_1, \ldots\}$ is also thematic. This guarantees that there exists a maximally specific thematic and possible proposition, and according to A-Quantity that must be an element of the attentional intent denoted by $\mathcal{A}_i$. This means that we can strengthen supposition (4) by adding the conjunct $\mathcal{A}_i(b)$, which after dropping the conjunct $\mathcal{T}_0(b)$ yields the following:

$$\mathbf{M}, w, g \models \Diamond {}^{\vee}a \wedge \Box({}^{\vee}a \rightarrow \exists b\, (b \subset a \wedge \mathcal{A}_i(b) \wedge {}^{\vee}b))$$

This was derived, recall, from the supposition that the singleton intent denoted by $\{a\}$ does not comply with A-Parsimony in w, i.e., that A-Parsimony is the reason why the proposition assigned to a is not an element of the attentional intent denoted by $\mathcal{A}_i$. The other possible reason was A-Quality, i.e., $\neg\Diamond^{\vee}a$. Hence, we can conclude the disjunction of these two reasons:

$$\mathbf{M}, w, g \models \neg\Diamond^{\vee}a \vee (\Diamond^{\vee}a \wedge \Box({}^{\vee}a \rightarrow \exists b\,(b \subset a \wedge \mathcal{A}_i(b) \wedge {}^{\vee}b)))$$

which implies: $\mathbf{M}, w, g \models \Box(\neg{}^{\vee}a \vee \exists b\,(b \subset a \wedge \mathcal{A}_i(b) \wedge {}^{\vee}b))$

By retracting suppositions (2) and (3) we obtain:

$$\mathbf{M}, w \models \forall a \left(\begin{array}{l} (\mathcal{T}_j(a) \wedge \neg\mathcal{A}_i(a)) \rightarrow \\ \qquad \Box(\neg{}^{\vee}a \vee \exists b(\mathcal{A}_i(b) \wedge b \subset a \wedge {}^{\vee}b)) \end{array} \right)$$

And by retracting supposition (1), i.e., that A-Quantity is (believed to be) complied with, we obtain the first result in fact 10. The second result directly derives from this through factivity.

Proof of fact 13

Let us first interpret the right-hand side of the equivalence, relying on fact 11, omitting the parameters $\mathbf{M}, w_0, g$ for the sake of readability:

$$\begin{aligned} [\![p_i \cap \textsc{Exh}_{\textsc{AP}}(\{p_i\}, \mathcal{T}_k)]\!] &= [\![p_i]\!] \cap [\![\textsc{Exh}_{\textsc{AP}}(\{[\![p_i]\!]\}, [\![\mathcal{T}_k]\!])]\!] \\ &= [\![p_i]\!] \cap \bigcap_{\substack{a \in [\![\mathcal{T}_k]\!] \\ a \notin \{[\![p_i]\!]\}}} (\overline{a} \cup \bigcup_{\substack{b \in \{[\![p_i]\!]\} \\ b \subset a}} b) \\ &= [\![p_i]\!] \cap \bigcap_{\substack{a \in [\![\mathcal{T}_k]\!] \\ a \neq [\![p_i]\!]}} (\overline{a} \cup \bigcup_{\substack{b = [\![p_i]\!] \\ b \subset a}} b) \end{aligned}$$

Continuing from the previous page:

$$\begin{aligned} [\![p_i]\!] \cap \bigcap_{\substack{a \in [\![\mathcal{T}_k]\!] \\ a \neq [\![p_i]\!]}} (\overline{a} \cup \bigcup_{\substack{b = [\![p_i]\!] \\ b \subset a}} b) &= [\![p_i]\!] \cap (\bigcap_{\substack{a \in [\![\mathcal{T}_k]\!] \\ a \supset [\![p_i]\!]}} (\overline{a} \cup \bigcup_{\substack{b = [\![p_i]\!] \\ b \subset a}} b)) \cap (\bigcap_{\substack{a \in [\![\mathcal{T}_k]\!] \\ a \not\supseteq [\![p_i]\!]}} (\overline{a} \cup \bigcup_{\substack{b = [\![p_i]\!] \\ b \subset a}} b)) \\ &= [\![p_i]\!] \cap (\bigcap_{\substack{a \in [\![\mathcal{T}_k]\!] \\ a \supset [\![p_i]\!]}} (\overline{a} \cup [\![p_i]\!])) \cap (\bigcap_{\substack{a \in [\![\mathcal{T}_k]\!] \\ a \not\supseteq [\![p_i]\!]}} (\overline{a} \cup \top)) \\ &= [\![p_i]\!] \cap (\bigcap_{\substack{a \in [\![\mathcal{T}_k]\!] \\ a \supset [\![p_i]\!]}} [\![p_i]\!]) \cap (\bigcap_{\substack{a \in [\![\mathcal{T}_k]\!] \\ a \not\supseteq [\![p_i]\!]}} \overline{a}) \\ &= [\![p_i]\!] \cap \bigcap_{\substack{a \in [\![\mathcal{T}_k]\!] \\ a \not\supseteq [\![p_i]\!]}} \overline{a} \end{aligned}$$

That is, when the attentional intent is the singleton set denoted by $\{p_i\}$, the current exhaustivity operator simply excludes every thematic proposition that is not entailed by the proposition denoted by p_i. What results is a set of worlds in which the proposition denoted by p_i is true and no (other) thematic propositions are true except those that are entailed by the proposition denoted by p_i.

Depending on the theme, the set of worlds thus characterized may be empty (namely if there is a set of propositions that are not entailed by p_i but that together cover (a superset of) p_i – excluding each of those from p_i will yield the empty set; see also section **??**). But the set is necessarily non-empty in any admissible, operable utterance model for the relevant type of utterance, i.e.:

$$[\![p_i \cap \textsc{Exh}_{\textsc{AP}}(\mathcal{A}_j, \mathcal{T}_k)]\!]_{\mathbf{M}, w_0, g} \neq \emptyset$$

This holds because both $[\![p_i]\!]$ and $[\![\textsc{Exh}_{\textsc{AP}}(\mathcal{A}_j, \mathcal{T}_k)]\!]$ must contain at least the actual world w_0: this follows, respectively, from compliance with I-Quality and factivity, and from compliance with A-Quantity, thematic competence and factivity.

Since the set is non-empty, the worlds it contains – i.e., in which the proposition denoted by p_i is true and no other thematic propositions except those that are entailed by the proposition denoted by p_i – will be the ones where the set of true thematic propositions is *minimized* as far as the truth of the proposition denoted by p_i allows. And these are exactly the worlds that $\textsc{Exh}_{\text{mw}}$ selects, according to definition 8.

Proof of fact 14

We prove the equivalence of the two operators by proving inclusion right-to-left and then left-to-right. First right-to-left: in an arbitrary admissible, operable model $\langle \mathbf{M}, w_0 \rangle$ of the specified type of utterance, take a world $w \in [\![p_i \cap \textsc{Exh}_{\textsc{AP}}(\mathcal{A}_j, \mathcal{T}_k)]\!]_{\mathbf{M},w_0,g}$. Given that $p_i = \bigcup \mathcal{A}_j$ is true in w_0 (fact 6), there must be some $a \in [\![\mathcal{A}_j]\!]_{\mathbf{M},w_0,g}$ such that $w \in a$. Moreover, given the chain-completeness restriction on themes in operable models, and given compliance with A-Relation, there must be a most specific (strongest, smallest) proposition a of that sort. From our exhaustivity operator it follows that every thematic proposition to which no attention is drawn is either false in w, or entailed by this most specific a. Hence, w makes the proposition a true and anything entailed by it, but no other thematic propositions. Within a, then, there is no $w' \in a$ where the set of true thematic propositions is smaller than in w. Hence (by definition) $w \in [\![\textsc{Exh}_{\text{dyn}}(\mathcal{A}_j, \mathcal{T}_k)]\!]_{\mathbf{M},w_0,g}$.

Conversely, take an arbitrary world $w \in [\![\textsc{Exh}_{\text{dyn}}(\mathcal{A}_j, \mathcal{T}_k)]\!]_{\mathbf{M},w_0,g}$. According to the definition of $\textsc{Exh}_{\text{dyn}}$, there must be some $a \in [\![\mathcal{A}_j]\!]_{\mathbf{M},w_0,g}$ such that $w \in a$ and w makes a minimal number of thematic propositions true, compared to other $w' \in a$. Given the chain-completeness restriction and compliance with A-Relation, there must be a most specific (strongest, smallest) a of that sort. Within this most specific a, any minimal set of true thematic propositions will contain a and anything entailed by it, but nothing else. (This is because, if a minimal set of true thematic propositions had contained another thematic proposition a', then the intersection $a \cap a'$ would have been thematic as well (by assumption of closure under intersection), and a would not have been possible independently of these more specific intersections, contrary to A-Parsimony, and would not have been included in the attentional intent.) Hence, this world w is contained in a, to which attention is drawn, but in no more specific thematic proposition. By definition, our operator contains all such worlds. Moreover, given that $p_i = \bigcup \mathcal{A}_j$ is true in w_0, we have that $w \in [\![p_i]\!]_{\mathbf{M},w_0,g}$, and hence $w \in [\![p_i \cap \textsc{Exh}_{\textsc{AP}}(\mathcal{A}_j, \mathcal{T}_k)]\!]_{\mathbf{M},w_0,g}$.

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. (2008). Innocent exclusion in an alternative semantics. *Natural Language Semantics 16*, 115–128.

Asher, N. and A. Lascarides (2003). *Logics of Conversation*. Cambridge University Press.

Bach, K. (1994). Conversational impliciture. *Mind and Language 9*, 124–62.

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.

Balogh, K. (2009). *Theme with variations: a context-based analysis of focus.* Ph. D. thesis, University of Amsterdam.

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.

Block, E. (2008). Is the symmetry problem really a problem? Unpublished manuscript; retrieved from `http://web.eecs.umich.edu/~rthomaso/lpw08/block.pdf`.

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.

Chierchia, G., D. Fox, and B. Spector (2009). Hurfords 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 Heusinger (Eds.), *Semantics: An International Handbook of Natural Language Meaning*, Volume 2, pp. 2297–2332. Mouton de Gruyter.

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 (2016). Hurford's constraint, the semantics of disjunction, and the nature of alternatives. Ms. February 2016; retrieved from `http://semanticsarchive.net/Archive/TViYjY0O/hurford.pdf`.

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.

Frege, G. (1918). Der gedanke: Logische untersuchungen, erster teil. *Beitrage zur Philosophie des Deutschen Idealismus 1*, 58–77.

Gamut, L. T. F. (1991). *Language, Logic and Meaning (vols. 1 and 2).* Chicago University Press.

Gazdar, G. (1979). *Pragmatics: Implicature, Presupposition, and Logical Form.* New York: Academic Press.

Geurts, B. (1998). The mechanisms of denial. *Language*, 274–307.

Geurts, B. (2011). *Quantity Implicatures.* Cambridge University Press.

Geurts, B. (2013). A plea for covert operations. In M. Aloni, M. Franke, and F. Roelofsen (Eds.), *The dynamic, inquisitive, and visionary life of φ, $?\varphi$, and $\Diamond\varphi$: a festschrift for Jeroen Groenendijk, Martin Stokhof, and Frank Veltman*, pp. 105–107. ILLC Publications.

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. (1967). Utterer's meaning, sentence-meaning and word-meaning. In *Studies in the Way of Words*, pp. 117–137. 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 M. Stokhof (1984). *Studies on the Semantics of Questions and the Pragmatics of Answers.* Ph. D. thesis, University of Amsterdam.

Gunlogson, C. (2008). A question of commitment. *Belgian Journal of Linguistics 22*, 101–136.

Hamblin, C. L. (1973). Questions in Montague English. *Foundations of Language 10*, 41–53.

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. (1978). Lexical incorporation, implicature, and the least effort hypothesis. In D. Farkas, W. M. Jacobsen, and K. W. Todrys (Eds.), *CLS: Papers from the Parasession on the Lexicon*, pp. 196–209. Chicago Linguistic Society.

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. (1985). Metalinguistic negation and pragmatic ambiguity. *Language 61*(1), 121–174.

Horn, L. R. (2001). *A Natural History of Negation*. David Hume Series on Philosophy and Cognitive Science Reissues. Standford: CSLI Publications. First published in 1989.

Katzir, R. (2007). Structurally-defined alternatives. *Linguistics and Philosophy 30*(6), 669–690.

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*.

Kennedy, C. (2013). A scalar semantics for scalar readings of number words. In I. Caponigro and C. Cecchetto (Eds.), *From grammar to meaning: The spontaneous logicality of language*, Cambridge Studies in Linguistics, pp. 172–200. Cambridge University Press.

Krifka, M. (1993). Focus and presupposition in dynamic interpretation. *Journal of Semantics 10*(4), 269–300.

Krifka, M. (2011). Questions and question-acts. Presented at the 18th Amsterdam Colloquium, December 2011, available via `amor.cms.hu-berlin.de/~h2816i3x/talks`.

Kroch, A. (1972). Lexical and inferred meanings for some time adverbs. *Quarterly Progress Reports of the Research Laboratory of Electronics 104*, 260–267.

Leech, G. (1981). Pragmatics and conversational rhetoric. In H. Parret, M. Sbisa, and J. Verschueren (Eds.), *Possibilities and limitations of pragmatics*, pp. 413–442. Amsterdam: John Benjamins.

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).

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.

Recanati, F. (2004). *Literal meaning*. Cambridge University Press.

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. (2013a). Algebraic foundations for the semantic treatment of inquisitive content. *Synthese 190*(1), 79–102.

Roelofsen, F. (2013b). 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.

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.

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.

Spector, B. (2016). Comparing exhaustivity operators. *Semantics and Pragmatics 9*.

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.

Westera, M. (2013). 'Attention, I'm violating a maxim!' A unifying account of the final rise. In R. Fernández and A. Isard (Eds.), *Proceedings of the Seventeenth Workshop on the Semantics and Pragmatics of Dialogue (SemDial 17)*.

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.

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.