Modularity and iterated rationality models of scalar implicatures*

Danny Fox and Roni Katzir

February 25, 2019

Abstract

In the linguistics literature, the derivation of scalar implicatures has often been handled in a relatively modular way, using computations that are sensitive to logical relations among alternatives such as entailment but are blind to other notions such as the probabilities that participants in a conversation might associate with these alternatives (or with related propositions). In recent years, a family of models that we refer to as *iterated rationality models* (IRMs) have offered an interestingly different perspective on such alternative-sensitive processes in terms of multiple iterations of probabilistic reasoning. Our paper attempts to identify and disentangle some of the central assumptions that are bundled within IRMs in the literature in order to better understand the predictions and maneuver space of such models and to facilitate their comparison with more traditional approaches. We focus in particular on an argument for probabilities made in Franke 2011 and some of the intricacies it involves. We conclude that the argument is not only compatible with a modular perspective on the computation of SIs, but in fact is crucially dependent on it.

1 Introduction

Work on Scalar Implicatures (SIs) has grown significantly over the past fifteen years, leading to new empirical generalizations and a rich body of competing theoretical perspectives. Under one family of proposals, SIs result from pragmatic strengthening of weak logical representations (the pragmatic approach, as outlined in Grice 1975 and developed further in Horn 1972 and much later work). Under another family of proposals, SIs are logical entailments of enriched grammatical representations (the grammatical approach; see Chierchia 2004, Fox 2007 and Chierchia et al. 2012), often using a covert operator, notated as *Exh*, that is akin to 'only' in its semantics and that derives SIs in much the same way that 'only' derives its entailments. An obvious conceptual argument in favor of the pragmatic approach is that it does not require the formal

*Acknowledgments: To be added.

mechanisms of grammatical enrichment.[1] On the other hand, it has been argued that the grammatical approach allows for a more principled theory of pragmatic reasoning (see Fox 2007, 2014). This conceptual impasse motivates the search for empirical considerations that might distinguish the two approaches.

1.1 Modularity and probability

One relevant consideration pertains to modularity. If the pragmatic approach is correct, one would expect the computation of SIs to be sensitive, at least in principle, to any type of information that is available to participants in a conversation. If the grammatical approach is correct, the computation could be highly modular and, in particular, informationally encapsulated. Hirschberg (1985/1991) has famously argued that *ad hoc* scales can serve as the basis for the computation of SIs. Her arguments can be taken to support the claim that the notion of entailment relevant for the computation of SIs is that of entailment given contextual information (contextual entailment) rather than the modular notion of logical entailment predicted by proposals that follow the grammatical approach. Specifically, "John sent the letter" does not logically entail "John put the letter in an envelope" but might contextually entail it, in which case the two sentences would belong to an *ad hoc* scale. This scale, in turn, is appealed to in Hirschberg's account of certain implicatures ("John put the letter in an envelope" could lead to the implicature that he did not send the letter).

While Hirschberg's examples seem to call for access to general contextual information in the computation of SIs, Fox and Hackl (2006) and Magri (2009, 2011) have presented evidence that goes in the opposite direction and appears to require a highly modular computation. More recently, Magri (2017) has argued for a resolution of this empirical conflict based on the grammatical approach. Specifically, he has shown that minor modifications of Hirschberg's cases behave differently and suggested a reanalysis of her basic cases using only logical scales. So, as things stand, it seems to us that the empirical facts pertaining to modularity favor the grammatical approach.

A reason to revisit the question of modularity in SIs comes from the relatively recent development of models within the pragmatic approach in which probabilities play a crucial role in the computation of SIs. If arguments can be presented for the crucial use of probabilities in the computation of SIs, this would seem, at least at first sight, to present important evidence against the type of modularity assumed under the grammatical approach, motivating, in particular, the search for alternative theories of the phenomena presented by Fox and Hackl and by Magri. The goal of the current paper is, thus, to extract arguments from current literature in favor of introducing probabilistic considerations into the computation of SIs and to see to what extent they favor a non-modular perspective.

Our tentative conclusion will be that, to the extent that such arguments exist, they do not favor non-modular theories. In fact, the opposite is true: at least within the scope

[1] The mechanisms of grammatical enrichment under the grammatical approach (at least under the *Exh*-based version of the approach) are rather simple and needed for a theory of focus sensitivity outside the domain of SIs (see Fox and Katzir 2011 and Katzir 2014). Hence the conceptual argument against the grammatical approach is much weaker than the very cogent argument against lexical ambiguity found in, e.g., Grice 1989 and Horn 1989.

of our examination, the arguments can survive only if a highly modular perspective on probabilities is taken. Moreover, in evaluating the arguments themselves, we will see that they rely on arbitrary decisions and analyses that are in no way superior to what can be achieved in the relatively simple modular theories proposed within the grammatical approach.

1.2 External and internal probabilities

Finding arguments that the computation of SIs is sensitive to probabilities is not going to be simple. The distribution of SIs is, of course, expected to be affected by a variety of contextual variables, among other things by the belief states of participants in a conversation, and in particular by their assessments of the probabilities of various states of affairs. But such interactions can easily be accounted for outside the system that computes SIs. Consider, for example, the potential inference from "Some of the students did their homework" to "It is not the case that many of the students did their homework". Presumably, the likelier the speaker is to know and care about the truth of "Many of the students did their homework", the likelier the inference is to be made. See Fox 2007, Goodman and Stuhlmüller 2013, and Chemla and Singh 2014a,b for discussion, including experimental evidence supporting this assessment.

However, this kind of correlation is unhelpful in evaluating whether probabilities play a role in the computation of SIs.[2] For example, under the grammatical view this probabilistic correlation can be derived through the probability that the potential alternative "Many of the students did their homework" will be used, i.e. will end up being a member of the restrictor argument of Exh (and perhaps even the probability that a parse with Exh will be used in the first place), while the computation of SIs itself remains modular and probability-free. (This is analogous to the observation that, while considerations of likelihood can affect how structural ambiguity is resolved in examples such as "Kim saw the student with the telescope", this does not constitute an argument for incorporating probabilities into the syntax.) Supporting a role for probabilities in the computation of SIs, then, is more complicated than noting that the presence of SIs correlates with the likelihood of various speaker/hearer beliefs.[3]

[2] Here we disagree with Goodman and Stuhlmüller (2013) who write: "This interaction between language understanding and general knowledge is not predicted by strongly modular theories that place scalar implicature within a semantics module (Chierchia et al. 2012). We show further that the interaction of knowledge and implicature is fine grained: The details of a speaker's belief distribution affect the details of an implicature." (p. 174)

[3] Somewhat separately from the present point, one of Goodman and Stuhlmüller (2013)'s results raises an interesting challenge for the grammatical approach, as noted by Irene Heim (in a class taught at MIT, Fall 2013). This challenge concerns an asymmetry between SIs and 'only', which, on the grammatical view, are almost identical. To see this asymmetry, consider a speaker who has seen the contents of exactly two out of three envelopes and says "One of the three envelopes contains a check". The inference drawn from this utterance is typically that at least one of the three envelopes contains a check and that at most two of them do (the $1 \leq n \leq 2$ inference). With 'only', on the other hand, the corresponding sentence ("Only one of the three envelopes contains a check") entails that exactly one of the three envelopes contains a check. This discrepancy, however, is resolved if, as argued by Meyer (2013, 2014, 2015) and further defended by Fox (2016) and Buccola and Haida (2018), the grammar contains a covert assertion operator, K, and if that operator can only attach close to the root and therefore always outscopes 'only' but can be outscoped by Exh. If that is the case, the sentence "One of the envelopes contains a check" can have a strengthened meaning paraphraseable as I only know/assert that one of the envelopes contains a check ($Exh \gg K$),

To better understand the question of sensitivity to probabilities, it is useful to try to associate with every theory of SIs a function, f, that takes as input a proposition (the basic meaning of a sentence), and various other arguments and returns a strengthened meaning. To say that a theory assumes sensitivity to probabilities is to say that one of the arguments of the f associated with this theory is a probability distribution.[4] To say that the computation is sensitive to probabilities is to say that this holds for the f associated with the correct theory.

The most obvious argument that the computation of SIs is sensitive to probabilities would consist in a demonstration that there is no way to state the function f without a probability distribution as argument – there is simply not enough information. Arguments of this form are not uncommon in semantics, including in the domain of SIs. For example, there is a well-known argument that one cannot write a function that simply takes a proposition as argument and returns a strengthened meaning. This argument is based on the so-called *symmetry problem* and is commonly taken to show that the computation of SIs is sensitive to formal alternatives. We are not aware of a parallel argument for probabilities. (As just mentioned, the general observation that SIs correlate with the likelihood of the speaker's opinionatedness does not constitute such an argument.) Still, many current models assume that probabilities are crucial. Our goal in writing this paper was to focus on places where a role for probabilities within the computation of SIs is not simply assumed but actually argued for.

The most promising argument we are familiar with for the use of probabilities in the computation of SIs is made by Franke (2009, 2011). Franke argues for a unified account of a potentially disparate set of phenomena within a probabilistic framework (see van Rooij 2010 for further discussion). An obvious question is whether this can be turned into an argument of the form described above (that is, that f requires a probability distribution as argument). We will argue that it cannot. In fact, we will show that for Franke's proposal to derive correct predictions, the relevant probability distributions must be uniform, which in turns means that f has no probabilistic argument (see fn. 4). In other words, while Franke's proposal uses what we may refer to as *internal* probabilities (that is, probabilities within the mechanism of f), it does not rely on *external* probabilities (which are probabilistic inputs to f).

A theory that makes use of internal probabilities but not of external probabilities does not argue for non-modularity. On the contrary, it further bolsters arguments for modularity. The question, of course, is which modular theory is correct. While we will not be able to address this question here, we will attempt to present the neces-

while the variant with overt only can only be read as "I know/assert that only one of the envelopes contains a check" ($K \gg Exh$). The $Exh \gg K$ parse contextually entails the $1 \leq n \leq 2$ inference: from the fact that the speaker saw the contents of two of the envelopes, it follows that if n were greater than 2, the speaker would have known that n is at least 2, contrary to a logical inference of $Exh \gg K$. So while the challenge is interesting, it is separate from our point regarding probabilities.

[4] By 'the f associated with T' we mean the simplest f that one can associate with T (when there is a unique simplest f). It is, of course, possible to add arguments to f that would not affect its output (arguments with respect to which f is constant). If such gratuitous arguments exist, f can be simplified by their elimination and will thus not be associated with any theory. Our observation that Franke's theory needs to assume a uniform probability distribution is, hence, a demonstration that if f were to take a probability distribution as argument, it would not be affected by this argument, hence that the (simplest) f associated with Franke's theory does not take a probability distribution as argument. Of course, what we say would not be affected if it turns out that there is more than one simplest f – simply replace the definite 'the' with the indefinite 'a'.

sary ingredients of Franke's theory in the hope that this can contribute towards theory comparison.

Our goals in this paper are rather limited. Specifically, we will attempt to outline Franke's argument for the use of probabilities in the computation of SIs doing our best to divorce it from the type of framework that Franke assumes. Along the way, we will attempt to identify some of the challenges that this kind of argument will have to face, and to explain why, if correct, it is an argument against external probabilities and hence an argument for modularity. But even these limited goals will require a rather long journey most of which will not be concerned with the question of modularity to which we will return at the very end.

1.3 Iterated rationality models and paper outline

Franke 2009, 2011's motivation for (internal) probabilities is presented within a complex model that belongs to a family of recent approaches to pragmatics that we will refer to as *Iterated Rationality Models* (IRMs). Other prominent IRMs include Goodman and Stuhlmüller 2013, Franke and Jäger 2014, and Bergen et al. 2016.[5] These proposals differ among themselves in various ways but share two central properties that we will focus on here and that set them apart from familiar alternative accounts in the literature. First, like Franke's model, the other IRMs in the literature are heavily probabilistic, which differs from other approaches, where the computation of SIs is sensitive to logical relations (in particular, entailment) but not to probabilities. Second, IRMs derive SIs through iterations of reasoning, typically going back and forth between hearers and speakers at increasing levels of sophistication, that attempt to gradually sort messages into (unions of) cells of the partition induced by the set of messages. This differs from other approaches (from Horn 1972 onward) in which SIs are typically derived without alternating between hearer and speaker perspectives and using only one or two rounds where alternatives are negated (or asserted).

IRMs are usually presented as bundles of these two components – probabilistic reasoning and back-and-forth iterations – and various other assumptions. For example, particular further commitments are needed to specify the use of probabilities to model hearers and speakers, with common choices in the literature reflecting the origins of existing IRMs in Bayesian and game-theoretic models of rationality. This bundling of assumptions makes it hard to see the motivation for each element and to tease apart the essential predictions of the system from the consequences of arbitrary decisions at various choice points. In what follows, we will try to unbundle the main components of IRMs mentioned above in order to understand Franke's argument for probabilities.

To this end, we start with a highly simplified iterative model where reasoning takes place from one participant's viewpoint (rather than going back and forth between speaker and hearer) and without probabilities. We then introduce the requisite complications where they might be needed: we provide a weak motivation for both probabilities and back-and-forth iterations between speaker and hearer in section 2.2; and we provide a potentially stronger argument for probabilities, extracted from Franke 2011,

[5]For a recent overview of this literature and its broader game-theoretic context see Benz and Stevens 2018.

in section 3.2. This argument, though very interesting, is not as simple as it appears at first sight as we discussion in section 4. Furthermore, if correct, it only serves to motivate internal probabilities and can in fact serve to argue that external probabilities are never relevant for the computation of SIs as discussed in section 5.

2 A Toy System

The present section begins our presentation of the type of reasoning involved in IRMs, and, as just mentioned, this is done with a very minimal IRM. As a first step we set aside probabilities and back-and-forth reasoning between speaker and hearer and focus only on the basic iterative aspect of IRMs. This will help show a central property of such models – namely, iterations – within a simple setting, and it will also help in highlighting why and where probabilities might become useful. In particular, we will see how iterated disambiguation derives SIs in cases of finite scalar alternatives. This can be done in a non-probabilistic setting and will then allow us to identify potential motivations for probabilities coming from cases that are more complex than the finite scalar one.

The following idealization, modeled on assumptions made by current IRMs, will be assumed throughout the discussion:

(1) Idealization of conversational setting:

a. The context set (the set of worlds consistent with whatever is common belief) is partitioned into cells

b. It is common belief that the epistemic state of the speaker entails one of the cells (the assumption of *an opinionated speaker*, sometimes called *the speaker competence* assumption)

c. It is common belief that the goal of the speaker is to convey that cell (the speaker's cell) to the hearer

d. It is common belief that the speaker is truthful – that is, they only say what they believe to be true in accordance with Grice's Maxim of Quality

Throughout the discussion, the partition of the context set, Π, will be induced by a set of alternatives, M. We will consider the elements of M to be syntactic objects and refer to them as *messages*.[6,7]

[6]The cells in the partition reflect all the consistent ways to assign truth values to the different alternatives in M. These can be characterized by defining an equivalence relation $\sim$ over the context set C such that for any $w, w' \in S$, $w \sim w'$ if for every $m \in M$, $[\![\text{ m }]\!](w) = [\![\text{ m }]\!](w')$. The partition of the context set is then the quotient set of the context set by $\sim$, $\Pi = C/\sim$.

[7]As has been discussed in detail in the literature, SIs cannot be derived if the alternatives are the set of all possible messages (all sentences of the language) and if there is no further way to differentiate between them. (This is so because of the symmetry problem mentioned in section 1.2.) One common method of differentiation, which we adopt here, is to offer a restrictive definition of alternatives (see Horn 1972, Katzir 2007, Fox and Katzir 2011, Trinh and Haida 2015, and Trinh 2018 for concrete proposals). This is convenient in the current context also in providing the basis for the partition of the context set. We note, however, that some of the literature on IRMs, such as Bergen et al. 2016, prefers to allow all possible messages to serve as alternatives and to differentiate between them in terms of costs. On this view, which we do not adopt here, the partition of the context set must come from some other source.

2.1 Scalar alternatives (the finite case)

In the scalar case, the set of alternatives is linearly ordered by entailment. We start with the simplest of the scalar cases, where there is only one alternative other than the assertion itself. Consider, for example, an assertion of 'some' with just 'all' as an alternative. Given these alternatives, the context set is partitioned into three cells: $\neg\exists$, $\exists \wedge \neg\forall$, $\forall$.[8] The cell $\neg\exists$ is inconsistent with the assertion (and with its alternative) so it will never be conveyed by any of the messages in the set of alternatives {some, all} (given (1d)); consequently, we will be able to set aside $\neg\exists$ for the discussion and focus on the cells in $\{\exists \wedge \neg\forall, \forall\}$.

Here is a possible strategy for the speaker to convey their cell:

- Step I: Suppose that the speaker's cell is $\forall$. In this case, the speaker can say 'all', and the hearer will easily be able to identify the correct cell. The hearer can do so since only the cell $\forall$ is consistent with the assertion and since the speaker is assumed to be truthful.

- Step II: Suppose now that the speaker's cell is $\exists \wedge \neg\forall$. In this case, the speaker cannot meet their goal directly. Suppose, however, that the speaker can rely on the hearer's knowledge of Step I above (namely, that if the speaker's cell were $\forall$ they would have uttered 'all') to rule out the cell $\forall$ upon an utterance of 'some'. This leaves the hearer with cell $\exists \wedge \neg\forall$ for 'some', as desired.

In other words, the first step allowed the conversational participants to pair the message 'all' and the cell $\forall$ and in effect peel them off. Following this step, the participants remain with just the message 'some' and the cell $\exists \wedge \neg\forall$ which can now be paired as well.[9]

Note that in the reasoning above, the speaker was assumed to follow the Maxim of Quality but was not explicitly assumed to follow the Maxim of Quantity. Elimination and iteration were able to derive what Quantity is typically used for (e.g., in Horn 1972).

The above generalizes to all SIs that rely on finite scales (as well as certain additional cases, some of which we will briefly mention below). Take, for example, the case where 'some' has not just 'all' but also 'many' as an alternative. The induced partition in this case is $\{\neg\exists, \exists \wedge \neg many, many \wedge \neg\forall, \forall\}$.

We can reason as before, adding one further step of peeling:

- Step I: If the speaker's cell is $\forall$, they will utter 'all', which will directly lead the hearer to the correct cell (since the message is inconsistent with any other cell). The cell (and possibly the message) are now peeled off.

[8] Concretely, 'some' might stand for an assertion of "John did some of the homework", with "John did all of the homework", notated as 'all' as its sole alternative. In this case, $\exists \wedge \neg\forall$ stands for the cell in which John did some but not all of the homework (with analogous interpretations for the remaining cell labels). For ease of presentation, we will stick with schematic shorthand such as 'some' and $\exists \wedge \neg\forall$ where no confusion is likely to arise.

[9] Note that what matters for the process just described is that the cell $\forall$ be peeled off after the first step. The message 'all' would do no damage if it stayed for the second step: that message is not true in any of the remaining cells (it was only true in the cell $\forall$, and that cell is now removed), so it cannot affect the identification of any other cell. For convenience, however, and for uniformity with a different notion of identification that we will see in (4) below, we will keep talking about peeling off both cells and messages.

- Step II: If the speaker's cell is $many \wedge \neg\forall$, they will utter 'many', which – given Step I – will lead the hearer to the correct cell (since the message is inconsistent with any of the remaining cells). The cell (and possibly the message) are now peeled off.
- Step III: If the speaker's cell is $\exists \wedge \neg many$, they will utter 'some', which – given Steps I, II – will lead the hearer to the correct cell (since the message is inconsistent with any of the remaining cells). The cell (and possibly the message) are now peeled off.
- Thanks to Step III, 'some' obtains its reading of 'some and not many'

To be able to refer to iterative peeling more easily, both using the current idea of cell identification and with certain variants thereof, we state the current criterion for cell identification in (2) and a general recipe for iterative peeling in (3).[10]

(2) CELL IDENTIFICATION (first version; further versions to be stated in (4) and (8)): Message m *identifies* a cell t given a set of cells Π if m is true in t and there is no distinct $t' \in \Pi$ such that m is true in t'

(3) PEELING STRATEGY: Given a set of messages M, a partition Π, and a criterion C for cell identification, we build a set X of message-cell pairings as follows:

a. Initialize $X = \emptyset$, as well as $M' = M$ and $\Pi' = \Pi$

b. Collect all message-cell pairs where the message identifies the cell according to C into a temporary set U. That is, $U = \{< m, t > \in M' \times \Pi' : m \text{ identifies } t \text{ according to } C\}$

c. If $U = \emptyset$, break and return X

d. Otherwise:

i. Update X with U. That is, $X = X \cup U$

ii. Remove from M' every message m that appears in the left-hand side of some pair in U

iii. Remove from Π' every cell t that appears in the right-hand side of some pair in U

iv. Go to step (3b)

The reasoning described in the scalar examples above follows the use of the recipe in (3) with the identification criterion in (2). The same reasoning extends to some non-scalar cases. For example, suppose that the set of messages is {A, B, A and B}, which induces the partition $\{\neg A \wedge \neg B, A \wedge \neg B, \neg A \wedge B, A \wedge B\}$.

[10] In keeping with the IRM literature, we present the model in procedural terms. As mentioned above, however, we believe that it can also be useful to view the model more abstractly, in terms of its mapping of inputs to outputs. This mapping, as in the discussion of the function f above, makes it easier to evaluate the different parameters that the system must be sensitive to than the procedural presentation. We take up this point in section 5 below, where we note that the abstract perspective of mapping inputs to outputs highlights a challenge to the idea that the computation of SIs has access to external probabilities.

Characterizing SIs in terms of inputs and outputs is also of help in understanding what an actual semantic operator in the grammar would look like if it were to behave like an IRM. This is particularly significant given the evidence in the literature that the operator 'only' is very similar to SIs (see Fox 2003), which suggests a unified treatment of the two.

- Step I: The message ‘A and B’ identifies the cell $A \wedge B$, so the cell (and possibly the message) can be peeled off. (Note that neither ‘A’ nor ‘B’ identifies a cell in this step.)
- Step II: Given Step I, the message ‘A’ identifies the cell $A \wedge \neg B$, and the message ‘B’ identifies the cell $\neg A \wedge B$

Crucially, the presence of the conjunctive alternative ‘A and B’ allowed the peeling process to start. As we will shortly see, when such an alternative is absent, peeling cannot start, and this challenge will motivate the introduction of probabilities into the system.

2.2 Infinite scales, back-and-forth reasoning, and a first motivation for internal probabilities

We defined cell identification in terms of a message that guarantees a particular cell. Imagine we approached cell identification from the opposite direction as well, by looking for a cell that guarantees a particular message. In the ‘some’/‘all’ case, for example, the hearer can reason that if a speaker is in cell $\exists \wedge \neg\forall$, the only message they can use is ‘some’. This reasoning might result in ‘some’ identifying the cell $\exists \wedge \neg\forall$, and both message and cell will be peeled off, which allows the remaining messages and cells to be paired in the next step. We could state this new (reverse) sense of identification as follows (using the same peeling strategy defined in (3) as before):[11]

(4) MIRROR-IMAGE CELL IDENTIFICATION (to be revised in (8)): Message m *mirror-identifies* a cell t given a set of messages M if m is true in t and there is no distinct message $m' \in M$ that is true in t

The motivation for our first notion of cell identification, as stated in (2) was quite straightforward: if m identifies t then, by definition (along with the assumption in (1d) that the speaker is truthful), a hearer who receives m knows that the speaker's cell is t. The motivation for the mirror-image notion in (4) seems less obvious: why should it matter if the speaker can use only m in a particular cell t? After all, if the same m is also true in a different cell t', the hearer will not be able to use m (without further assumptions) as a reliable indicator that the speaker's cell is t.

To see a potential conceptual motivation for (4) in terms of the goal of indicating the speaker's cell to the hearer, we propose thinking of the hearer's inference about the speaker's cell in terms of a best guess about that cell rather than full certainty

[11] Note that (4) makes it possible in principle for a message to identify multiple cells. Consider, as a schematic example, a partition $\Pi = \{1, 2, 3, 4\}$ induced by $M = \{A, B, C\}$, where $[\![\text{A}]\!] = 1 \cup 2$, $[\![\text{B}]\!] = 2 \cup 3$, and $[\![\text{C}]\!] = 3 \cup 4$. In Step I, ‘A’ mirror-image identifies 1 (since it is the only message true in that cell) and, similarly, ‘C’ mirror-image identifies 4. Then, in Step II, message ‘B’ mirror-image identifies both cell 2 and cell 3, since for each of them, this is the only remaining message that is true in that cell. This is a somewhat counter-intuitive state of affairs, and it is potentially problematic for communication, but we will not attempt here to explore its possible impact on the conversational setting or potential modifications that could avoid the issue. (For the original notion of cell identification in (2), a parallel issue can arise with multiple messages identifying the same cell. That, however, seems more natural and less problematic.) With our final statement of identification, in (8) below, this issue does not arise.

(perhaps because a message that conveys a cell with full certainty is not available, as will be the case in an example we will consider shortly). This, in turn, invites thinking about inference within a probabilistic setting, the informal idea being that if m is the only message true in cell t then (on certain assumptions) hearing m will make t more probable than any other cell in which additional messages are true. Consider again the case of 'some'/'all', and suppose that the hearer has received the message 'some'. The hearer's goal is to identify the speaker's cell. Assuming, as we do, that the speaker is truthful, this cell is either $\exists \wedge \neg\forall$ or $\forall$. The hearer can use Bayes' Rule to compare how likely each cell is given the message:[12]

$$\begin{aligned} P(\exists \wedge \neg\forall|\text{`some'}) &= \frac{P(\text{`some'}|\exists \wedge \neg\forall)P(\exists \wedge \neg\forall)}{P(\text{`some'})} \\ P(\forall|\text{`some'}) &= \frac{P(\text{`some'}|\forall)P(\forall)}{P(\text{`some'})} \end{aligned}$$

The denominator on the right-hand side of both cases is identical and does not affect the comparison.[13] As to the numerator, let us start by making the assumption that the prior probability of the two relevant cells is the same: $P(\forall) = P(\exists \wedge \neg\forall)$. This can follow from a general assumption of flat priors:

(5) Flat priors (tentative): The prior distribution is uniform, so that if Π is finite, for any $t \in \Pi$, $P(t) = \frac{1}{|\Pi|}$.

The assumption of a uniform prior distribution cannot hold if (as in an example we will consider immediately below) the partition is countably infinite. There might also be other reasons to abandon (5), such as connecting priors to what we referred to above as *external* probabilities – that is, to actual probabilistic assessments (that might be part of the common ground), a possibility that we discuss in section 5 below (though, as mentioned in the introduction, we will see an empirical argument against such a move). In the case of 'some'/'all', however, flat priors have the advantage of allowing us to focus entirely on the likelihood component for the purpose of the comparison of the two probabilities under consideration, and we will tentatively make this assumption here.

With both the denominator and the priors out of the way, the comparison of $P(\exists \wedge \neg\forall|\text{`some'})$ and $P(\forall|\text{`some'})$ boils down to a comparison of the likelihoods, $P(\text{`some'}|\exists\wedge \neg\forall)$ and $P(\text{`some'}|\forall)$. And it is here that the number of messages that are true in a

[12] According to Bayes' Rule, $P(t|m)$ can be written as

$$P(t|m) = \frac{P(m|t) \cdot P(t)}{P(m)}$$

Of the two factors in the numerator, $P(m|t)$ is referred to as the *likelihood*, and $P(t)$ is the *prior*. The denominator $P(m)$ can be ignored if, as in the discussion below, we are only interested in comparing the probability of various cells given the same message. If the denominator cannot be ignored (for example, if we wish to compute the actual probability of a cell given a message and not just make the relevant comparisons), it can be rewritten again as $\sum_{t' \in P} P(m|t') \cdot P(t')$.

[13] If it needs to be computed explicitly it can be done in the usual way by writing it as $P(\text{`some'}) = P(\text{`some'}|\exists \wedge \neg\forall)P(\exists \wedge \neg\forall) + P(\text{`some'}|\forall)P(\forall)$.

given cell can become relevant. More specifically, (4) would be explained as a consequence of probabilistic reasoning if we can assume that the conditional probability of a message given a cell is maximal whenever this message is the only one that is true in the cell.

Here is a way to justify this assumption. Recall from (1d) that we are assuming that the speaker is always truthful. If we further assume that the speaker always sends some message, as stated in (6a), then $P(\text{‘some’}|\exists \wedge \neg\forall) = 1$, since ‘some’ is the only message that is true in the cell $\exists \wedge \neg\forall$. If we also assume that every message that is true in a cell has a positive probability of being sent, as stated in (6b), then $P(\text{‘some’}|\forall) < 1$, since in the cell $\forall$ there are two true messages, ‘some’ and ‘all’ (so if each message has a positive probability, neither can have probability 1).

(6) Additional conversational assumptions:

a. The speaker always sends a message (and cannot remain silent)

b. If the speaker's cell is t and m is true in t, the speaker has a positive probability of uttering m

From the above, we obtain $P(\text{‘some’}|\exists \wedge \neg\forall) > P(\text{‘some’}|\forall)$, which in turn (on our current assumptions) means that $P(\exists \wedge \neg\forall|\text{‘some’}) > P(\forall|\text{‘some’})$. So, on the assumptions above, a hearer who receives ‘some’ can conclude that it is most probable that the speaker's cell is $\exists \wedge \neg\forall$. Consequently, if it is reasonable to take a message m as indicating the cell that m makes most probable (when such a cell exists), something like (4) can serve as a sensible criterion for cell identification.[14]

This kind of probabilistic reasoning, then, can motivate adopting something like (4).[15] But is there also an empirical reason to think that (2) is insufficient and that something like (4) needs to be added to the system (perhaps supporting a system that goes back and forth between the two notions of identification)? Relevant cases would be SIs in which at some point during iterative peeling the following hold: (a) there is no alternative that is true in just one cell (otherwise (2) could be used); and (b) there is a cell in which just one message is true (so (4) holds). We discuss two potential cases of this kind, though we conclude that neither provides stong support for (4). We present these cases, as an illustration of the type of consideration that might support (4) and as a way of indicating that we were unable to find stronger support.

As a first example, suppose that the set of messages is $\{A, B\}$, which induces the partition $\{\neg A \wedge \neg B, A \wedge \neg B, \neg A \wedge B, A \wedge B\}$. Clearly, an utterance of ‘A’ does not identify any cell: it is compatible with two distinct cells, $A \wedge \neg B$ and $A \wedge B$. Similarly for an utterance of ‘B’. And in the absence of a cell identifier, peeling using (2) cannot start. The situation changes when mirror-image identification in (4) is available: ‘A’ is the only message that is true in the cell $A \wedge \neg B$, and similarly for the message ‘B’ and the cell $B \wedge \neg A$, so mirror-image identification succeeds. The problem with using

[14]The idea of maximizing the probability of a cell given a message is very similar to Franke (2011)'s proposal. One difference between the two notions is that Franke's proposal maximizes also the speaker's probabilities of messages given a cell.

[15]Note, however, that while the motivation for (4) is probabilistic, its statement is not. Empirical evidence in favor of (4), then, might suggest a role for probabilistic reasoning in conventionalizing the non-probabilistic (4), but it will not necessarily support a role for probabilities within the IRM itself. In later sections we will consider revising this so as to give probabilities an actual role in the computation of SIs.

this case to support (4) is that it is not obvious that SIs based on the relevant sets of alternatives are ever computed. Specifically whenever an utterance of 'A' identifies the cell $A \wedge \neg B$ there are other plausible ways of accounting for this: either the conjunctive alternative is present as well, in which case peeling can proceed using (2), as discussed earlier, or the common-ground already excludes $A \wedge B$ (see Fox 2018).

For our second example, consider the possibility of infinite scales (a matter of some debate in the literature). One central candidate is expressions denoting the natural numbers. For example, for an assertion of "John has three children", the alternatives, one might claim, are all the sentences of the form "John has n children", $[\![n]\!] \in \mathbb{N}$. This set does not have a strongest element, so the peeling process based on (2) cannot start. On the other hand, the mirror-image notion of cell identification, as stated in (4), straightforwardly allows for peeling:[16]

- Step I: the message "John has one child(ren)" mirror-identifies the cell exactly-one (since in that cell no other alternative is true)
- Step II: the message "John has two children" mirror-identifies the cell exactly-two (since after peeling off the message "John has one child(ren)" in Step I, there is no message other than "John has two children" that is true in the cell exactly-two)
- Step III: the message "John has three children" mirror-identifies the cell exactly-three (for the same reasoning as in Step II), and the SI for the assertion is computed

So for cases such as the above, the original (2) fails and (4) succeeds. For natural numbers in a downward-entailing context, on the other hand, it is the original (2) that succeeds and (4) that fails. For example, "If John has three children, he is eligible for a tax break" has a strongest alternative (namely, "If John has one child(ren), he is eligible for a tax break"), from which peeling using the original (2) can start. However, there is no cell in which just one alternative is true, so peeling using the mirror-image notion in (4) has no starting point.

Given the above, it might seem reasonable to use both (2) and (4), perhaps going back and forth between the two. As mentioned briefly in the introduction, the idea of going back and forth between the perspective of the speaker and that of the hearer is central to the IRMs in the literature. Based on our present discussion, we suggest infinite scales as a possible empirical motivation for this choice. And since, as discussed above, mirror-image identification receives its motivation from probabilistic reasoning, we can take infinite scales to also indirectly support a role for probabilities in IRMs.[17]

[16] Above we illustrated a probabilistic motivation for the notion of mirror-image identification in (4) using the case of 'some'/'all' and assuming flat priors. For the countably infinite case we are currently considering, a uniform distribution over cells is of course not possible. As mentioned in note 15, however, (4) itself makes no mention of probabilities and does not depend on there being any particular kind of distribution over the cells in the partition.

[17] Note that the putative support for probabilities and back-and-forth reasoning based on infinite scales depends heavily on the basic setup: within alternative approaches to SIs, such as the neo-Gricean and the grammatical approaches, discrete infinite scales do not require an appeal to probabilities or back-and-forth reasoning. Dense scales, if they exist, would be problematic for both the neo-Gricean approach and the type of IRMs we are discussing (see Fox and Hackl 2006).

However, since it remains unclear whether SIs are ever truly based on infinite scales – numeric scales such as the above, for example, might be based on finite scales resulting from truncation of infinite ones – the support for both back-and-forth reasoning and for probabilities based on such scales is weak. For probabilities, we will see potentially stronger motivation below. As with the motivation for probabilities in the current section, what we will see will be at best an argument for internal probabilities: in both cases, actual probability assessments from outside of where SIs are computed will not play a role. For mirror-image identification and back-and-forth reasoning we have not been able to identify additional arguments: as far as we can tell, this remains one of the arbitrary choices made by IRMs in the literature and highlighted by the unbundling in the current paper. To keep the presentation simple, we now return to a one-sided notion of cell identification.

3 Non-scalar alternatives, conjunctive readings of disjunction, and a second motivation for internal probabilities

3.1 Conjunctive readings of disjunction

In the cases of SI discussed earlier, there was always either a message that was true in just one cell or a cell in which just one message was true, so peeling could always proceed using at least one of (2) or (4). In other cases, however, neither of the two notions of identification above are of help. A particularly relevant configuration in which no message is true in just one cell (so the original notion of cell identification fails) and in which mirror-image identification is unhelpful as well is that of disjunctions in the absence of a conjunctive alternative. In this case, the alternatives are {A or B, A, B}, and the induced partition is the same as in the case of {A and B, A, B} and of {A, B} discussed earlier: $\{\neg A \wedge \neg B, A \wedge \neg B, \neg A \wedge B, A \wedge B\}$. Instantiations of this schematic setting have been argued to exist and to give rise to conjunctive readings for disjunctive sentences, with examples including Warlpiri connectives (Bowler 2014), or-else disjunction (Meyer 2015), and disjunctions in child language (Singh et al. 2016). Somewhat more broadly, conjunctive reading of disjunctions has been observed for embeddings under particular environments such as existential operators, as in so-called Free Choice disjunctions: a sentence such as “You may eat the ice cream or the cake” has the implication that you may eat the ice cream and you may eat the cake. Moreover, this implication has been argued to be an SI (see Kratzer and Shimoyama 2002, Alonso-Ovalle 2005, and Fox 2007) and to be generated by a general mechanism that also derives the conjunctive SI in the case of {A or B, A, B} (see Fox 2007, Franke 2011, and Bar-Lev and Fox 2017). Other cases that have been argued to follow a similar logic include the further embedding of Free Choice disjunction under a universal operator (argued to be an SI in Bar-Lev and Fox 2017) and the embedding of disjunction in the antecedent of conditionals (see van Rooij 2010 and Franke 2011). Based on these works, we can formulate the following desideratum for theories of Scalar Imlicatures:[18]

(7) DESIDERATUM FOR THEORIES OF SCALAR IMPLICATURES: In certain cases (depending on the properties of ϕ), when a disjunction of the form $\phi(A \text{ or } B)$ has $\phi(A)$ and $\phi(B)$ as alternatives but not $[\phi(A) \text{ and } \phi(B)]$, the theory must provide the means for generating $[\![\, \phi(A) \,]\!] \wedge [\![\, \phi(B) \,]\!]$ as an SI.

However, the procedures of iterative peeling that we have seen fail to account for such conjunctive interpretations of disjunction. The reason for this is that no message is true in just one cell, so no message will identify a cell by the simple notion of cell identification in (2). And the same holds for mirror-image identification since there is also no cell in which just one message is true.

3.2 A role for internal probabilities in cell identification

To see how the challenge of conjunctive readings of disjunctions might be addressed within an IRM, imagine that there was a way for 'A' and for 'B' to be strengthened so that they would each identify a different cell. Suppose, more specifically, that 'A' identified $A \wedge \neg B$, and 'B' identified $B \wedge \neg A$ (as in the simpler case that involves just the two messages $\{A, B\}$). If that were possible, 'A or B' (after peeling) would identify the third cell in which it is true, namely $A \wedge B$, thus accounting for the conjunctive reading of disjunction in this case. The question, of course, is how to strengthen the messages 'A' and 'B' in the first place so that they can identify the relevant cells. Franke (2011)'s insight is that this can be done through the use of probabilities. In particular, while we just saw that the notion of mirror-image identification in (4) does not solve the problem of strengthening 'A' and 'B' (once the disjunctive message is added to the set of alternatives), the probabilistic components used to provide the motivation for (4) do allow the relevant messages to be strengthened once one further assumption is made, as we now show.

To facilitate the discussion, we start by restating the probabilistic motivation for (4) as an actual probabilistic criterion for cell identification. Recall that in the probabilistic setting under consideration, the hearer asks if there is a t that is more likely than any other cell t' given the message m: $P(t|m) > P(t'|m)$ for all $t' \neq t$. Using Bayes' Rule, this amounts to comparing $\frac{P(m|t)P(t)}{P(m)}$ and $\frac{P(m|t')P(t')}{P(m)}$, which, since the denominator is the same for both elements, amounts to comparing the numerators $P(m|t)P(t)$ and $P(m|t')P(t')$. The criterion, then, can be stated as follows:[19]

[18] We will not attempt here to characterize either the kinds of environments ϕ in which such strengthening occurs or the precise sets of alternatives that are involved. Different proposals in the literature make different predictions based on the choices of ϕ and the alternatives, and we will frame the discussion in terms of cases such as the simple {A or B, A, B} and Free Choice disjunction where both ϕ and the alternatives seem straightforward.

[19] While (8) and (4) are conceptually related, neither is stronger than the other, even if we assume flat priors. Below we will focus on cases in which a message m identifies a cell t according to (8) but not according to (4). But in principle there can be cases in which a message m identifies a cell t in the sense of (4) but not in the sense of (8). This is so since there might be another cell t' in which only m is true. In this case, m identifies both t and t' according to (4), but it identifies neither according to the probabilistic (8) (since $P(m|t) = P(m|t') = 1$, so on the assumption of flat priors $P(m|t)P(t) = P(m|t')P(t')$).

(8) MIRROR-IMAGE CELL IDENTIFICATION (final version, revised from (4)): Message m *probabilistically mirror-identifies* a cell t given a set of cells Π and messages M if for any other t' in Π, $P(t|m) > P(t'|m)$ (or, using Bayes' Rule, $P(m|t)P(t) > P(m|t')P(t')$).

The original notion of cell identification in (2) corresponds to certainty: when a message m is only true in cell t, the hearer who receives m knows (on the assumption of truthfulness, (1d)) that the speaker's cell is t. From the current probabilistic perspective, this means that $P(t|m) = 1$ and that for any $t' \neq t$, $P(t'|m) = 0$. Mirror-image cell identification in (4) does not rely on certainty but still guarantees, once various additional assumptions were made (in particular, (6a), (6b), and flat priors), that for any $t' \neq t$, $P(t|m) > P(t'|m)$ in cases where (a) only m is true in t, and (b) there is no other cell t' in which only m is true. For the case of conjunctive readings of disjunction, these assumptions are insufficient, since none of the speaker's probabilities is 1, as summarized in the following table (where ? stands for an unknown value that is greater than 0 but less than 1):

(9) Speaker's probability assignment (A/B/A or B):

$P(m|t)$	'A'	'B'	'A or B'
$\neg A \wedge \neg B$	0	0	0
$A \wedge \neg B$	?	0	?
$\neg A \wedge B$	0	?	?
$A \wedge B$	?	?	?

There is a simple way to ensure that $P(\text{'A'}|A \wedge \neg B) > P(\text{'A'}|A \wedge B)$ (so that, with the assumption of flat priors, we will have $P(A \wedge \neg B|\text{'A'}) > P(A \wedge B|\text{'A'})$, and 'A' will be strengthened to $A \wedge \neg B$ as desired) and that $P(\text{'B'}|\neg A \wedge B) > P(\text{'B'}|A \wedge B)$ (so that 'B' will be strengthened to $\neg A \wedge B$). This way involves counting: of the two states that make 'A' true, $A \wedge B$ makes three messages true while $A \wedge \neg B$ makes only two messages true; similarly for the cells that make 'B' true. On its own, this does not give us enough information to compare $P(\text{'A'}|t)$ for its two relevant cells (or $P(\text{'B'}|t)$ for its two cells), but it does if we make the further assumption that the speaker has no preference among the messages that are true in their cell, so that each is used with equal probability. That is, we add the following assumption, which following the IRM literature we refer to as that of a *naive speaker* and which subsumes (1d), (6a), and (6b):[20]

One can bring the two notions closer by demanding only $P(m|t)P(t) \geq P(m|t')P(t')$, as done by Franke (2011), but this complicates the empirical derivation of conjunctive readings of disjunction, so we do not adopt this move here.

[20] As in our earlier discussion of (4), we could treat the probabilistic setting – now also including (10) – as motivating background but state a non-probabilistic identification criterion for the actual peeling process. In the present case, we would need to incorporate the idea of comparing how many messages each relevant cell makes true. We could do so by saying that m identifies t if it is true in t and if any other t' that makes m true makes more additional messages true than t does. Note, however, that differently from (4), such a statement makes reference within the IRM itself to a comparison of cardinalities, a notion that is quite different from those typically used in accounts of SIs in the literature. For space considerations, we have chosen to skip the intermediate cardinality-based statement and to state the discussion here and below directly in terms of probabilities.

(10) Naive speaker: if cell t makes n different messages true, the speaker will choose each of them with probability $\frac{1}{n}$.

With (10), the unhelpful table in (9) becomes the following:

(11) Naive speaker's probability assignment (A/B/A or B):

$P(m\|t)$	'A'	'B'	'A or B'
$\neg A \wedge \neg B$	0	0	0
$A \wedge \neg B$	$\frac{1}{2}$	0	$\frac{1}{2}$
$\neg A \wedge B$	0	$\frac{1}{2}$	$\frac{1}{2}$
$A \wedge B$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Table (11) allows us to obtain the desired conjunctive reading of disjunction. Thanks to (10), $P(\text{'A'}|A \wedge \neg B) = \frac{1}{2} > \frac{1}{3} = P(\text{'A'}|A \wedge B)$. On our earlier assumption of flat priors, this means that, according to (8), 'A' is an identifier for the cell $A \wedge \neg B$ (ultimately, since $P(A \wedge \neg B|\text{'A'}) > P(A \wedge B|\text{'A'})$). Similarly, 'B' becomes an identifier for the cell $\neg A \wedge B$. No further identification is possible in the first stage, but after peeling, we have only the message 'A or B' and the cells $\neg A \wedge \neg B$ and $A \wedge B$, so 'A or B' becomes an identifier for the cell $A \wedge B$ as desired. The same holds in appropriate environments ϕ such as embedding under an existential operator or in the antecedent of a conditional. We can conclude that (8) allows our IRM to meet the desideratum in (7).

Before proceeding, note that the move to probabilistic identification in (8) supports simpler identification in finite scalar cases. For example, consider again the case of $M = \{\text{'some', 'many', 'all'}\}$, where the partition is $\Pi = \{\neg\exists, \exists \wedge \neg many, many \wedge \neg\forall, \forall\}$. We have the following probabilities for the naive speaker:

(12) Naive speaker's probability assignment (some/many/all):

$P(m\|t)$	'some'	'many'	'all'
$\neg\exists$	0	0	0
$\exists \wedge \neg many$	1	0	0
$many \wedge \neg\forall$	$\frac{1}{2}$	$\frac{1}{2}$	0
$\forall$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Given the probabilities in (12), and assuming flat priors (that is, $P(t) = \frac{1}{4}$ for all t), we can see that 'some' identifies $\exists \wedge \neg many$, 'many' identifies $many \wedge \neg\forall$, and 'all' identifies $\forall$. Note that, while the earlier notions of identification in (2) and (4) require three steps to complete cell identification in the present case, (8) accomplishes this in a single step and does not require peeling.

3.3 Interim discussion

As mentioned, the use of probabilistic IRMs to derive conjunctive readings of disjunctions (when a conjunctive alternative is absent) is due to Franke (2009, 2011) and originally presented within his particular system (see also van Rooij 2010). And as we just saw, the same solution is available within our simplified IRM.[21] Other IRMs in the literature, however, have not been able to derive conjunctive readings of disjunctions. In particular, IRMs from the Rational Speech Act family (such as Bergen et al. 2016) have not been able to derive such readings and thus fail with respect to the desideratum in (7). Given the evidence in the literature that these readings are SIs, the challenge for such IRMs is to either find a modification that succeeds in deriving the relevant readings or to offer an argument against the analysis of these readings as SIs and in favor of an alternative account of them.

Before proceeding, let us briefly review how the grammatical approach derives conjunctive readings of disjunctions.[22] Recall from the introduction that on this approach, SIs are derived via a covert exhaustivity operator, akin to 'only' and sometimes written as *Exh*, that can attach to various positions in the parse tree.[23] Here we will focus on the recent variant of the grammatical approach in Bar-Lev and Fox 2017, where the SIs discussed in the current paper are derivable using a single instance of *Exh* attached at the root and where exhaustification follows a two-step procedure. In the first step, those alternatives that can be safely negated while affirming the assertion are negated. This is done using the notion of *innocent exclusion* (Fox 2007): an alternative m is innocently excludable given an assertion S and a set of alternatives M if m is in all maximal sets of alternatives that can be negated without contradicting S. This ensures that alternatives that are negated do not lead to arbitrary entailments. In the case of $M = \{$A or B, A, B$\}$, for example, if S='A or B' then negating 'A' would entail that 'B' is true, which seems arbitrary; similarly, negating 'B' would entail 'A', which again seems arbitrary. Innocent exclusion formalizes this sense of arbitrariness: the maximal sets of alternatives that can be negated consistently with an affirmation of the assertion are $\{$A$\}$ and $\{$B$\}$, and no alternative is a member of both, so no alternative is innocently excludable in this case. After the negation of some or all of the innocently excludable alternatives, a second step of *innocent inclusion* (Bar-Lev and Fox 2017) determines which of the remaining alternatives can be affirmed consistently with the assertion, again while avoiding arbitrary choices (which here, too, is done by considering maximal sets of alternatives that can be affirmed consistently and choosing those that appear in all such sets). In our current example, both 'A' and 'B' can be asserted consistently with the assertion 'A or B', so both are affirmed, and the result is the desired conjunc-

[21]As stated, the procedure for strengthening disjunctions to conjunctions overgenerates (both as stated by us and Franke). For example, assuming that artists can be musicians or painters (possibly both), strengthening along the lines just discussed incorrectly predicts that "John is an artist" will be strengthened to imply that John is both a musician and a painter. However, this prediction is shared among all current proposals that can derive conjunctive meanings of disjunctions in cases where this is needed. Moreover, similar ways seem to be open to all such proposals to avoid the overgeneration problem, for example by making SIs blind to world knowledge (as has been argued in Fox and Hackl 2006 and Magri 2009 for reasons that are independent of the current matter). If the computation of SIs cannot see the contextual relations between 'artist', 'painter', and 'musician', all the relevant theories of SIs will be able to avoid the empirically unattested strengthening of 'artist' to mean both painter and musician. (We note, however, that a solution in terms of blindness does seem at odds with an interpretation of IRMs in terms of general, non-modular reasoning, a point that we discuss in section 5.)

[22]For other (non-probabilistic, non-IRM) proposals for the derivation of conjunctive readings of disjunctions see Klinedinst 2007 and Chemla 2009.

[23]See Fox 2007, Magri 2011, and Fox and Spector 2018 for proposals regarding the distribution of *Exh*, all very much inspired by Chierchia (2004).

tive reading.[24,25]

4 Challenge: more than two disjuncts

When there are more than two disjuncts, both Franke (2011)'s IRM and our own fail to derive conjunctive readings of disjunctions.[26] Consider the set of messages {A or B or C, A or B, A or C, B or C, A, B, C}, and the induced partition $\{\neg A \wedge \neg B \wedge \neg C, A \wedge \neg B \wedge \neg C, \neg A \wedge B \wedge \neg C, \neg A \wedge \neg B \wedge C, A \wedge B \wedge \neg C, A \wedge \neg B \wedge C, \neg A \wedge B \wedge C, A \wedge B \wedge C\}$. On the assumption of a naive speaker, as stated in (10), we have the following speaker probabilities:

(13) Naive speaker's probability assignment (three disjuncts):

$P(m\|t)$	'A'	'B'	'C'	'A or B'	'A or C'	'B or C'	'A or B or C'
$\neg A \wedge \neg B \wedge \neg C$	0	0	0	0	0	0	0
$A \wedge \neg B \wedge \neg C$	$\frac{1}{4}$	0	0	$\frac{1}{4}$	$\frac{1}{4}$	0	$\frac{1}{4}$
$\neg A \wedge B \wedge \neg C$	0	$\frac{1}{4}$	0	$\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{4}$
$\neg A \wedge \neg B \wedge C$	0	0	$\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
$A \wedge B \wedge \neg C$	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
$A \wedge \neg B \wedge C$	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
$\neg A \wedge B \wedge C$	0	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
$A \wedge B \wedge C$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$

Using the probabilities in (13), we can try to proceed by iterations of peeling as before, but as we will now show, the process does not yield the desired results.

[24] The definition of *Exh* in terms of the two steps of innocent exclusion and innocent inclusion can be seen as a grammatical way to support the speaker's goal of conveying a cell in the partition, as stated in (1c) above: assuming, as we do here, that the partition is induced by the set of alternatives, the cells are defined by the different consistent truth-value assignments to the alternatives; *Exh* provides a non-arbitrary way to set the truth value of a large number of alternatives, thus attempting to take an assertion close to a cell in the partition.

[25] In earlier variants of the grammatical approach, *Exh* was defined in terms of innocent exclusion alone. In the case of $M = \{$A or B, A, B$\}$, as we saw, no alternative is innocently excludable given $S =$'A or B'. That is, Exh_M(A or B) $= A \vee B$. For the other two alternatives we have Exh_M(A) $= A \wedge \neg B$ and Exh_M(B) $= B \wedge \neg A$. To obtain the conjunctive reading, a second occurrence of *Exh* is attached. Assuming that for this higher *Exh* the alternatives are $M' = \{Exh_M(\text{A or B}), Exh_M(\text{A}), Exh_M(\text{B})\}$, we now have both $Exh_M(A)$ and $Exh_M(B)$ as innocently excludable alternatives to Exh_M(A or B), so when $Exh_{M'}(Exh_M(\text{A or B}))$ is computed, both alternatives can be negated while Exh_M(A or B) is affirmed, which in turn amounts to the conjunctive reading. This accounts for conjunctive readings of disjunctions in a variety of cases (see Fox 2007, Bowler 2014, Meyer 2015, and Singh et al. 2016), and it does so in a way that resembles the direction followed earlier for cell identification: the first occurrence of *Exh* provides a way to identify the cells $A \wedge \neg B$ and $B \wedge \neg A$; and while this first occurrence does not directly identify a cell for 'A or B', a second occurrence can eliminate the cells that were identified using the first occurrence and leave only the conjunctive cell for the disjunction. In most of the cases discussed in the present paper the choice between the two variants of the grammatical theory does not make a difference. However, the variant that relies on innocent inclusion straightforwardly derives conjunctive readings for disjunctions in the antecedent of a conditional (see Bar-Lev 2018), a case that, as Franke (2011) notes, is not derived by the earlier variant that relies on innocent exclusion alone. See Bar-Lev and Fox 2017 for further arguments in favor of innocent inclusion.

[26] As mentioned, other IRMs in the literature fail already with two disjuncts.

- Step I. Assuming (8), message A probabilistically identifies the cell $A \wedge \neg B \wedge \neg C$, and both are peeled off. This is so since $P(\text{'A'}|A \wedge \neg B \wedge \neg C) = \frac{1}{4}$, which is greater than $P(\text{'A'}|t')$ for all $t' \neq A \wedge \neg B \wedge \neg C$, so assuming flat priors we obtain $P(\text{'A'}|A \wedge \neg B \wedge \neg C)P(A \wedge \neg B \wedge \neg C) > P(\text{'A'}|t')P(t')$ for all $t' \neq A \wedge \neg B \wedge \neg C$. Similarly, message B probabilistically identifies the cell $B \wedge \neg A \wedge \neg C$ and both are peeled off, and message C probabilistically identifies the cell $C \wedge \neg A \wedge \neg B$ and both are peeled off. No further messages identify cells at this step.

- Step II. At this point, all remaining messages are compatible with all remaining cells (other than $\neg A \wedge \neg B \wedge \neg C$), so no further identification is possible, and the process of identification stops.The remaining messages – {A or B or C, A or B, A or C, B or C} – can perhaps be treated as indicating uncertainty, or they might just lead to an anomaly. In any event, these messages do not get strengthened to conjunctive readings.

The inability of the peeling process to derive a conjunctive reading for 'A or B or C' is problematic since conjunctive readings of disjunctions are empirically attested regardless of the number of disjuncts. For example, (14), with three disjuncts, can be read as giving permission to choose between cake, ice cream, and fruit (rather than indicating uncertainty about the allowed choices).

(14) You are allowed to eat the cake, the ice cream, or the fruit
(Possible reading: You may choose between the three)

We presented the problem within our simplified system, but as Franke (2011) notes, a similar problem holds for his version of IRM.[27] The same problem arises already for two disjuncts in the presence of additional alternatives:

(15) Which of these three desserts (cake, ice cream, and fruit) are we allowed to eat?
a. You are allowed to eat the cake or the ice cream

In the given context, the answer in (15a) leads to the inference that we are allowed to eat the cake, we are allowed to eat the ice cream, and we are not allowed to eat the fruit. Within the peeling system, this – seemingly straightforward – extension of the basic free-choice case is not derived: the message in (15a) fails to identify a cell for the same reasons as in the discussion of three disjuncts above. Again, the problem carries over also to Franke (2011)'s system.[28]

We do not mean to imply that this problem cannot be addressed. In fact, Franke (2011) makes some preliminary suggestions that we believe deserve closer scrutiny.

[27] While both systems fail to obtain the correct strengthening of 'A or B or C', the actual outcome in the two systems is different. In our system, as we just saw, 'A or B or C' does not get strengthened at all. In Franke (2011)'s system, the outcome depends on whether the back-and-forth iterations start with a naive hearer or a naive speaker: for the former, 'A or B or C' becomes a surprise message that a speaker is expected not to use (and that a hearer can interpret as compatible with any of the different cells), while for the latter it gets strengthened to mean that two of the three disjuncts are true but not all three.

[28] In this case, the same incorrect prediction is made in Franke's system regardless of whether iterations start with a naive speaker or a naive hearer: in both cases, a disjunction such as 'A or B' is incorrectly predicted to be strengthened to mean that two of the three disjuncts are true but not all three.

We note, however, that under the grammatical approach, the problem does not arise in the first place: the same system that was designed to handle two disjuncts also derives the case of more than two disjuncts. Using the proposal of Bar-Lev and Fox 2017, strengthening proceeds as follows. First, none of the alternatives to ‘A or B or C’ is innocently excludable (there is no alternative that is in all maximal sets that can be negated consistently with ‘A or B or C’), so none are excluded and all remain as possible alternatives for inclusion. Next, the alternatives are all innocently includable – that is, they can all be affirmed together consistently with ‘A or B or C’ – so all of them are affirmed, which yields the conjunctive reading that A, B, and C are all true. For the case of two disjuncts in the presence of a third alternative, the difference is that in the first step, the third alternative – for example, ‘C’ if ‘A or B’ is asserted – is innocently excludable and is therefore negated. The other alternatives remain innocently includable, which yields the reading that A and B are true (because of inclusion) but C is false (because of exclusion in the first step).[29]

5 External probabilities

We are now ready to return to the question of modularity with which we began. Our discussions in sections 2–4 was stated in terms of internal probabilities and was agnostic to the modularity question pertaining to the source of the probabilities used for cell identification. Under a non-modular approach, those would be related to real assessments of probabilities of cells and messages – that is, to external probabilities. But they could also be thought of as an entirely formal construct that is used within the IRM, a possibility that would support a modular approach. Recall from section 1.2 that the difference between these two possibilities can be thought of as a difference in the type of input that the IRM needs to take. On the view that probabilities are a formal construct that is used modularly within the IRM, the input to the IRM would be just a message and a set of alternatives – that is, the same as with more traditional approaches to SIs – and the IRM itself would derive the probabilities from the alternatives and the cells that they induce according to some recipe (for example, using something like the assumptions of *flat priors* in (5) and *naive speaker* in (10)). On this view, IRMs would look very much like traditional approaches to SIs in terms of their input and output.

If, on the other hand, the probabilities are real assessments of likelihood (determined in a non-modular way, externally to where SIs are computed), the mapping from inputs to outputs would be quite different from what it is in traditional accounts. Specifically, it would take as input not only a message and a set of alternatives (as in traditional accounts) but also a probability distribution over cells. Of course, adding another input – especially one that has the potential to make the theory more expressive – requires justification. But if it can be shown that external probability assessments interact in interesting ways with SIs (and perhaps related phenomena), this would provide justification for the probabilistic input parameter and, moreover, will give IRMs an advantage over traditional accounts in which only logical relations (such as entailment)

[29]The correct result is obtained for both cases also if we use the proposal of Fox 2007, in which exhaustification uses only the negation of innocently excludable alternatives and in which conjunctive readings of disjunction are derived using the recursive application of the exhaustivity operator.

are used for deriving SIs while probabilities do not play a role.[30]

As we show in the present section, however, this is not the case: when the probabilities in IRMs are allowed to reflect general cognitive probabilistic assessments, they lead to the incorrect prediction that as the prior probability of the cells varies, so do the SIs of various messages.

5.1 Do priors affect SIs?

Consider again the case of $M = \{$'some', 'many', 'all'$\}$, where the partition is $\Pi = \{\neg\exists, \exists \wedge \neg many, many \wedge \neg\forall, \forall\}$. Earlier we showed how probabilistic cell identification as stated in (8) derives the correct results in this case on the assumption of flat priors. If priors are a real input to the system, however, we need to examine other possibilities.

First, recall the naive speaker's probabilities from (12), repeated here:

(16) Naive speaker's probability assignment (some/many/all):

$P(m\|t)$	'some'	'many'	'all'
$\neg\exists$	0	0	0
$\exists \wedge \neg many$	1	0	0
$many \wedge \neg\forall$	$\frac{1}{2}$	$\frac{1}{2}$	0
$\forall$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Suppose now that, differently from our earlier discussion of this example, the cell $many \wedge \neg\forall$ is more than twice as likely as any of the other cells, all of which have the same prior probability. That is, $P(\neg\exists) = P(\exists \wedge \neg many) = P(\forall) < \frac{1}{5}$ and $P(many \wedge \neg\forall) > \frac{2}{5}$. If those are the priors, 'some' will identify the incorrect $many \wedge \neg\forall$ rather than the correct $\exists \wedge \neg many$. To see why, recall that we are looking for the cell t that maximizes $P(t|\text{'some'})$, or, equivalently (given Bayes' Rule) the cell that maximizes $P(\text{'some'}|t) \cdot P(t)$. Recall further that we are assuming naive speaker's probabilitieis, as summarized in (16). Suppose now a sufficiently biased prior such as $P(\neg\exists) = P(\exists \wedge \neg many) = P(\forall) = \frac{1}{6}$ and $P(many \wedge \neg\forall) = \frac{1}{2}$. Then $P(\text{'some'}|many \wedge \neg\forall) \cdot P(many \wedge \neg\forall) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$ while $P(\text{'some'}|\exists \wedge \neg many) \cdot P(\exists \wedge \neg many) = 1 \cdot \frac{1}{6} = \frac{1}{6}$ (and $P(\text{'some'}|\forall) \cdot P(\forall) = \frac{1}{3} \cdot \frac{1}{6} = \frac{1}{18}$). In other words, while the assumption of naive speaker gives an advantage in the case of 'some' to the correct cell, $\exists \wedge \neg many$, over the alternative cells, this advantage is too small to overcome a sufficiently biased prior such as the one considered here. Other biased priors lead to additional incorrect identifications. For example, if $P(\neg\exists) = P(\exists \wedge \neg many) = P(many \wedge \neg\forall) < \frac{1}{6}$, both 'some' and 'many' will incorrectly identify $\forall$.

A possible response to the problem just noted might be to say that when simple, non-probabilistic identification (or mirror-image identification) is possible, that is the notion that is used rather than the probabilistic one in (8). This would be useful for $M = \{$'some', 'many', 'all'$\}$ since, as we saw earlier, this case is handled straightforwardly by non-probabilistic identification. Unfortunately for the idea of external

[30] Recall from the introduction that the relevant interactions must go beyond the observation that probabilities can affect the presence or absence of an SI, an observation that is handled by all accounts of SIs. To provide support for IRMs, probabilities should not just be relevant for the question of whether or not strengthening of a message applies, they should also affect the nature of strengthening.

probabilities, however, biased priors are a problem for probabilistic identification also in the case of conjunctive readings of disjunctions, where non-probabilistic notions of identification are of little help. Consider again the naive speaker's probabilities for this case, repeated here from (11):

(17) Naive speaker's probability assignment (A/B/A or B):

$P(m\|t)$	'A'	'B'	'A or B'
$\neg A \wedge \neg B$	0	0	0
$A \wedge \neg B$	$\frac{1}{2}$	0	$\frac{1}{2}$
$\neg A \wedge B$	0	$\frac{1}{2}$	$\frac{1}{2}$
$A \wedge B$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Suppose that, rather than flat priors as in our earlier discussion of disjunctions, the priors are instead $P(\neg A \wedge \neg B) = P(\neg A \wedge B) = P(A \wedge B) = \frac{1}{5}$ and $P(A \wedge \neg B) = \frac{2}{5}$. If this is the case, we have 'A or B' incorrectly identifying the cell $A \wedge \neg B$. If, on the other hand, $P(\neg A \wedge \neg B) = P(\neg A \wedge B) = P(A \wedge \neg B) = \frac{1}{5}$ and $P(A \wedge B) = \frac{2}{5}$, we have 'A' incorrectly identifying $A \wedge B$.

5.2 Other IRMs and biased priors

Above we illustrated the sensitivity to priors of the IRM developed in the present paper. As far as we can tell, however, the problem extends to other IRMs in the literature. For example, Franke (2009, pp. 71–6) notes the sensitivity of his IRM to priors in scalar cases such as 'some'/'all'.[31]For IRMs of the Rational Speech Act family, such as Bergen et al. 2016, biased priors can also lead to incorrect predictions in the scalar case, as discussed in Spector 2018, though the precise predictions depend on further intricacies of that family of models.[32]

Sensitivity to biased priors has not, to our knowledge, been discussed for non-scalar cases such as $M = \{\text{A or B, A, B}\}$. As in our own system, however, IRMs in the literature make incorrect predictions in this case as well. In Franke (2011)'s IRM, for example, if $P(A \wedge \neg B) = P(\neg A \wedge B) < \frac{2}{3} \cdot P(A \wedge B)$, both 'A' and 'B' will incorrectly be strengthened to mean $A \wedge B$. More dramatically, the slightest bias in favor of either $A \wedge \neg B$ or $B \wedge \neg A$ can lead to an incorrect strengthening of 'A or B' to mean the favored cell. For example, if $P(A \wedge \neg B) = P(A \wedge B)$ and each was just a little bit smaller than $P(\neg A \wedge B)$ – perhaps $P(A \wedge \neg B) = P(A \wedge B) = 0.9999 \cdot P(\neg A \wedge B)$ – the message 'A or B' will be strengthened to mean $\neg A \wedge B$.The Rational Speech Act family, too, is sensitive to biased priors in this case, though again the predictions depend on additional parameters within that approach.

We conclude that giving IRMs access to external probabilities is problematic. Not only do we not seem to see any meaningful interactions between such probabilities and SIs (which, if found, could have provided an argument in favor of IRMs) but in cases such as those seen above, external probabilities lead to incorrect results. This argues

[31]See Rothschild 2013 for further discussion of prior sensitivity of this IRM in the case of 'some'/'all'.

[32]In particular, the implications of biased priors within the Rational Speech Act approach depend on a so-called rationality parameter that determines the relative weights of the hearer's beliefs when incorporated into a more sophisticated speaker's probabilities. See Spector 2018 for pertinent discussion.

for a more modular approach, where probabilities are simply not part of the input to the mechanism that computes SIs.

6 Conclusion

In this paper we presented the strongest argument we are aware of for the introduction of probabilities to the system that computes SIs. We saw that the argument, if successful, is incompatible with external probabilities and thus requires a highly modular approach to the computation. As such it is conceptually on a par with other modular approaches and its evaluation will have to based on overall theory comparison. We focused a very narrow dimension of comparison pertaining to the theory of the conjunctive interpretation of disjunction and its insensitivity to the number of disjuncts involved. But the comparison will have to be broadened to touch on the full range of phenomena that have been argued to bear on the correct theory of SIs.[33]

References

Alonso-Ovalle, Luis. 2005. Distributing the disjuncts over the modal space. In *Proceedings of NELS 35*, ed. Leah Bateman and Cherlon Ussery.

Bar-Lev, Moshe E. 2018. Free choice, homogeneity, and innocent inclusion. Doctoral Dissertation, Hebrew University of Jerusalem.

Bar-Lev, Moshe E, and Danny Fox. 2017. Universal free choice and innocent inclusion. In *Semantics and Linguistic Theory*, ed. Dan Burgdorf, Jacob Collard, Sireemas Maspong, and Brynhildur Stefánsdóttir, volume 27, 95–115.

Benz, Anton, and Jon Stevens. 2018. Game-theoretic approaches to pragmatics. *Annual Review of Linguistics* 4:173–191.

Bergen, Leon, Roger Levy, and Noah Goodman. 2016. Pragmatic reasoning through semantic inference. *Semantics and Pragmatics* 9.

Bowler, Margit. 2014. Conjunction and disjunction in a language without 'and'. In *Semantics and Linguistic Theory*, volume 24, 137–155.

Buccola, Brian, and Andreas Haida. 2018. Obligatory irrelevance and the computation of ignorance inferences. Ms., ENS and HUJI, September 2018.

Chemla, Emmanuel. 2009. Similarity: Towards a unified account of scalar implicatures, free choice permission and presupposition projection. Under revision for *Semantics and Pragmatics*.

Chemla, Emmanuel, and Raj Singh. 2014a. Remarks on the experimental turn in the study of scalar implicature, part I. *Language and Linguistics Compass* 8:373–386.

Chemla, Emmanuel, and Raj Singh. 2014b. Remarks on the experimental turn in the study of scalar implicature, part II. *Language and Linguistics Compass* 8:387–399.

[33]Relevant phenomena of this kind include the distribution of embedded SIs (Chierchia et al. 2012), the theory of NPI licensing (Chierchia 2013 and Crnič (2013)), the connection between SIs and various grammatical phenomena (Fox and Hackl 2006, Fox and Katzir 2011, and Katzir 2014), and various interactions between SIs and ignorance inferences (Sauerland 2004, Spector 2006, and Fox 2014).

Chierchia, Gennaro. 2004. Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In *Structures and beyond*, ed. A. Belletti, 39–103. Oxford University Press.

Chierchia, Gennaro. 2013. *Logic in grammar*. Oxford University Press.

Chierchia, Gennaro, Danny Fox, and Benjamin Spector. 2012. Scalar implicature as a grammatical phenomenon. In *Semantics: An international handbook of natural language meaning*, ed. Paul Portner, Claudia Maienborn, and Klaus von Heusinger, volume 3, chapter 87, 2297–2331. Mouton de Gruyter.

Crnič, Luka. 2013. Focus particles and embedded exhaustification. *Journal of Semantics* 30:533–558.

Fox, Danny. 2003. Implicature calculation, only, and lumping: Another look at the puzzle of disjunction. Talk presented at Yale University, November 2003.

Fox, Danny. 2007. Free choice disjunction and the theory of scalar implicatures. In *Presupposition and implicature in compositional semantics*, ed. Uli Sauerland and Penka Stateva, 71–120. Palgrave-Macmillan.

Fox, Danny. 2014. Cancelling the Maxim of Quantity: Another challenge for a Gricean theory of scalar implicatures. *Semantics and Pragmatics* 7:1–20.

Fox, Danny. 2016. On why ignorance might not be part of literal meaning. commentary on Marie-Christine Meyer. Handout of a talk presented at the MIT Workshop on Exhaustivity, September 2016.

Fox, Danny. 2018. Partition by exhaustification: comments on Dayal 1996. To Appear in Proceedings of Sinn und Bedeutung, June 2018.

Fox, Danny, and Martin Hackl. 2006. The universal density of measurement. *Linguistics and Philosophy* 29:537–586.

Fox, Danny, and Roni Katzir. 2011. On the characterization of alternatives. *Natural Language Semantics* 19:87–107.

Fox, Danny, and Benjamin Spector. 2018. Economy and embedded exhaustification. *Natural Language Semantics* 26:1–50.

Franke, Michael. 2009. Signal to act: Game theory in pragmatics. Doctoral Dissertation.

Franke, Michael. 2011. Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics* 4:1–82.

Franke, Michael, and Gerhard Jäger. 2014. Pragmatic back-and-forth reasoning. In *Pragmatics, semantics and the case of scalar implicatures*, ed. Salvatore Pistoia-Reda, 170–200. Springer.

Goodman, Noah D., and Andreas Stuhlmüller. 2013. Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science* 5:173–184.

Grice, H. P. 1975. Logic and conversation. In *Syntax and semantics*, ed. Peter Cole and James L. Morgan, volume 3, 41–58. Academic Press.

Grice, Paul. 1989. *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Hirschberg, Julia. 1985/1991. *A theory of scalar implicature*. New York: Garland.

Horn, Laurence. 1972. On the semantic properties of the logical operators in English. Doctoral Dissertation, UCLA.

Horn, Laurence. 1989. *A natural history of negation*. Chicago: University of Chicago

Press.
Katzir, Roni. 2007. Structurally-defined alternatives. *Linguistics and Philosophy* 30:669–690.
Katzir, Roni. 2014. On the roles of markedness and contradiction in the use of alternatives. In *Semantics, pragmatics, and the case of scalar implicatures*, ed. Salvatore Pistoia Reda, 40–71. Palgrave-Macmillan.
Klinedinst, Nathan. 2007. Plurality and possibility. Doctoral Dissertation, UCLA.
Kratzer, Angelika, and Junko Shimoyama. 2002. Indeterminate pronouns: The view from Japanese. In *Proceedings of the Tokyo conference on psycholinguistics*, ed. Yukio Otsu, volume 3, 1–25. Tokyo: Hituzi Syobo.
Magri, Giorgio. 2009. A theory of individual-level predicates based on blind mandatory scalar implicatures. *Natural Language Semantics* 17:245–297.
Magri, Giorgio. 2011. Another argument for embedded scalar implicatures based on oddness in downward entailing environments. *Semantics and Pragmatics* 4:1–51.
Magri, Giorgio. 2017. Blindness, short-sightedness, and Hirschberg's contextually ordered alternatives: a reply to Schlenker (2012). In *Linguistic and psycholinguistic approaches on implicatures and presuppositions*, ed. Salvatore Pistoia-Reda and Filippo Domaneschi, 9–54. Palgrave Macmillan.
Meyer, Marie-Christine. 2013. Ignorance and grammar. Doctoral Dissertation, MIT, Cambridge, MA.
Meyer, Marie-Christine. 2014. Deriving Hurford's constraint. In *Proceedings of SALT 24*, 577–596.
Meyer, Marie-Christine. 2015. Generalized free choice and missing alternatives. *Journal of Semantics* .
van Rooij, Robert. 2010. Conjunctive interpretation of disjunction. *Semantics and Pragmatics* 3:1–28.
Rothschild, Daniel. 2013. Game theory and scalar implicatures. *Philosophical Perspectives* 27:438–478.
Sauerland, Uli. 2004. Scalar implicatures in complex sentences. *Linguistics and Philosophy* 27:367–391.
Singh, Raj, Ken Wexler, Andrea Astle-Rahim, Deepthi Kamawar, and Danny Fox. 2016. Children interpret disjunction as conjunction: Consequences for theories of implicature and child development. *Natural Language Semantics* 24:305–352.
Spector, Benjamin. 2006. Aspects de la pragmatique des opérateurs logiques. Doctoral Dissertation, Université de Paris 7, Paris.
Spector, Benjamin. 2018. What do we learn from game-theoretic pragmatics? Talk given at SALT 28, May 2018.
Trinh, Tue. 2018. Keeping it simple. *Natural Language Semantics* 26:111–124.
Trinh, Tue, and Andreas Haida. 2015. Constraining the derivation of alternatives. *Natural Language Semantics* 23:249–270.