Notes on iterated rationality models of scalar implicatures*

Danny Fox and Roni Katzir

October 27, 2020

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 investigates what at first sight seems like a very interesting argument for IRMs coming from the conjunctive interpretation of disjunctive sentences. We then outline challenges for the argument based on a theoretical comparison with the grammatical theory of scalar implicatures. The comparison focuses on the full distribution of conjunctive interpretation, on the question of sensitivity to priors, and on other results achieved within the grammatical theory that the IRM literature does not engage with.

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. One approach to the problem, which has recently gained popularity, is embodied in a set of models involving iterated steps of probabilistic reasoning, taken separately by a speaker and an addressee, each making specific assumptions about the workings of the other, and eventually converging on an enriched meaning of an utterance. Our goal in this paper is to contribute toward the evaluation of such models, which we will call iterated rationality models (IRMs), and which include the proposals in Franke 2009, 2011, Goodman and Stuhlmüller 2013, Franke and Jäger 2014, and Bergen et al. 2016, among others.[1] We will work toward the evaluation of IRMs as theories of SIs against the background of a very specific alternative, the so-called

*Acknowledgments: To be added.

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

grammatical approach (see Chierchia 2004) and in particular a variant of this approach where SIs are derived using a covert operator, sometimes 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 (see Fox 2007 and Chierchia et al. 2012).[2,3]

With this goal in mind, we will attempt to identify the best arguments in favor of IRMs as accounts of SIs. This, as we will explain in the rest of this introduction, leads us to Franke (2009, 2011), which, in our mind, is unique in attempting to derive a rich set of SIs that are outside the reach of many theories, and thus has the potential to yield an important argument. Elaborating on an observation already in Franke 2011, however, we will conclude that the argument fails, and that a theory based on a modular computation of SIs within grammar is at present superior. Our second conclusion will be that Franke (2011)'s theory, to the extent it can be made successful, will require a modular architecture and will thus share important architectural properties with the grammatical theory, somewhat unexpectedly, perhaps.

The arguments in favor of IRMs that we will extract from Franke (2009, 2011) will be empirical. But before we get there, we think we should put aside a possible conceptual argument in favor of IRMs – one that we feel might contribute to the recent popularity of the approach – namely the hope that they could serve as a step toward a unified theory of language interpretation and other aspects of human action and thought (as suggested, e.g., in Goodman and Stuhlmüller 2013). It is our feeling that a hope of this sort can be turned into an argument only with concrete unified proposals that can be evaluated. Since such proposals do not exist, to the best of our knowledge, we will put this aside and focus on other considerations.

One such consideration that we would like to reject as well is based on correlations between the distribution of SIs and the nature of probabilistic epistemic states (i.e., the prior probabilities that participants in the conversation assign to various states of affairs). Correlations have indeed been reported and sometimes been presented as favoring IRMs and their use of probabilities in the derivation of SIs. However, as far as we can tell, the attested correlations are irrelevant for theoretical adjudication, as they have no bearing on the nature of the systems that enter into the computations of 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 the alternative "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 expected by all major approaches to SIs, including those that are modular and non-proababilistic. To see this, take the variant

[2]Specifically, $Exh(A)(p)$ asserts that p is true and that a certain subset of alternatives from the set A are false. This subset is chosen in a way that avoids contradictions and arbitrary choices from among the alternatives. In a more recent development (Bar-Lev and Fox 2017; Bar-Lev 2018), Exh also affirms various alternatives, again while avoiding contradictions and arbitrary choices.

[3]IRMs have been proposed for purposes other than the derivation of SIs. Among many other things, there is recent work that reinterprets IRMs – in particular, developments of the lexical uncertainty model of SIs (Bergen et al. 2016) – as theories of disambiguation, presupposing a grammatical derivation of SIs (see Champollion et al. 2019 and Franke and Bergen 2019). This paper does not bear on such proposals and is in fact totally compatible with them.

of the grammatical approach mentioned above, where SIs are derived from syntactic representations in which a silent operator *Exh* is present. Under this approach the correlation can be derived through the probability that the potential alternative "Many of the students did their homework" will be used, i.e., that it 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.[4] We therefore put aside this putative argument as well.[5]

What remains at the moment, in our opinion, are empirical facts pertaining to the family of SIs that can be derived from a given sentence. Given our current state of knowledge, this is the only area where important considerations exist that can bear on theory choice. And since, in general, arguments must involve theory comparison, we will have to ask whether there are facts in this domain that favor IRMs relative to prominent alternatives, and in particular relative to the grammatical theory which we will use as our frame of reference.

Most work on IRMs focuses on the simplest kinds of exhaustivity inferences (e.g., the strengthening of 'some' to 'some but not all') and since such inferences are captured

[4]For this reason, we fail to see the relevance of the following quote from Goodman and Stuhlmüller (2013): "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). Theories of SIs, like theories of other linguistic phenomena, will of course have to interact with theories of disambiguation, and these are very likely to be probabilistic in nature. So the interatcion that Goodman and Stuhlmüller (2013) refer to is not predicted by many theories of SIs in exactly the same way that parallel interactions are not predicted in the case of structural ambiguities, but this can hardly be viewed as relevant for the evaluation of non-probabilistic grammars. This matter becomes particularly clear when one discusses explicit theories of disambiguation (as in the recent reinterpretation of IRMs as theories of disambiguation that help choose between parses that differ in the distribution of *Exh*; see Champollion et al. 2019 and Franke and Bergen 2019).

[5]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 challenge for the grammatical theory is not related to correlations with probabilistic assessments, as it is dealt with straightforwardly by approaches that are blind to probabilistic considerations, such as that of Sauerland (2004). Furthermore the challenge for the grammatical theory is met 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, which 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" ($Exh \gg K$), 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.

straightforwardly by many competitors, they are not likely to be useful. We will thus focus on Franke (2009, 2011) who develops a particular view on run-of-the-mill SIs that can be extended, as he shows, to capture an intricate set of inferences that do not follow under many existing alternatives.[6]

Assuming that Franke (2011)'s proposal is successful, we will want to ask which assumptions of IRMs are crucial for this success. In order to address this question, we think it is useful to derive the results with a primitive version of Franke (2011)'s proposal, which highlights those assumptions that are crucial for the model's success. This will allow us to appreciate the results of the model along with its limitations.

We will see, through our primitive version, how Franke (2011)'s insights yield an account for a large set of places where disjunctive sentences are strengthened to yield a conjunctive meaning. This gives it an edge over many approaches to SIs, but a very serious limitation becomes apparent. This limitation was already noted by Franke (2011); here we highlight its significance by drawing an explicit comparison with the grammatical approach, which does not suffer from this limitation. This comparison, together with a brief reminder of key results of the grammatical approach that the IRM literature has yet to engage with, present important arguments, we think, in favor of the grammatical approach.[7]

So our first conclusion will be that, at present, there is no argument that favors IRMs over alternative approaches to SIs. In fact, if we are right, the best IRM is inferior to existing alternatives. But assuming that the limitations we will bring to light can be dealt with, we can ask which component of IRMs might be supported by Franke (2011)'s results. We will see, through our primitive model, that Franke (2011)'s achievements depend on iterated probabilistic considerations. This might lead to the impression that Franke (2011)'s system could end up arguing that the computation of SIs is dependent on the prior probabilities that speakers and hearers bring with them to the interpretive task. This impression – which, we wish to stress, is not made in Franke (2009, 2011)'s own work – is incorrect. In fact, the opposite is true. As we will see the system gives correct results only if priors are flat. That priors in his system should not faithfully reflect speaker's beliefs has already been pointed out by Franke (2009). We will add to this the novel observation that the situation is much more dramatic in

[6]Again, our focus here is exclusively the adequacy of IRMs for SIs. This paper does not attempt to evaluate the ability of IRMs to handle other phenomena such as disambiguation, hyperbole, and other cases that have been discussed in the IRM literature. We think the case of so called reference games (Frank and Goodman 2012) is potentially very interesting in that it provides what at first sight might look like an argument for IRMs. However, as far as we can see, the basic results can be captured by a variety of approaches to SIs. In work in progress (Asherov, Fox, and Katzir 2020) we build on the logic of cell identification discussed below to uncover cases where the predictions of an exhaustivity based theory and those of existing IRMs diverge in this domain.

[7]A reviewer suggests that earlier work in Bidirectional Optimality Theory (Blutner 1998, 2000; Jäger 2002) can be reinterpreted as an IRM that is similar in many ways to the one that we develop here and that this interpretation allows one to see that it not only matches Franke (2011)'s results but in fact surpasses them (accounting for the simple case of three-place disjunction in (13) below). This reconstructed IRM, as far as we can see, still does not derive the correct pattern of conjunctive readings of disjunctions (as it does not account for 2-place disjunction in the presence of a third alternative; see our discussion of (17a)). It also does not, at present, offer responses to the arguments in the literature in favor of the grammatical approach. It therefore does not affect the analysis in the present paper or the – largely negative – conclusions that we reach. We therefore set it aside in what follows.

the area that constitute his most important achievement, namely disjunctive sentences, the only area we are aware of that provides a potential argument that probabilistic considerations might be relevant for the computation of SIs. If Franke (2009, 2011)'s approach is correct and, for that matter, if there are any reasons at all to believe that probabilistic considerations enter into the computation of SIs, then the system that computes SIs would have to be blind to actual priors. We will discuss the significance of this point to debates about the modularity of SI computation.

2 Cell identification (first attempt)

2.1 General assumptions

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*.[8,9]

2.2 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

[8] 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 $\sim$, $\Pi = C/\sim$.

[9] 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 so-called *symmetry problem*.) 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.

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: $\neg\exists$, $\exists \wedge \neg\forall$, $\forall$.[10] 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.[11]

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.

- 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

[10] 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.

[11] 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.

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).

(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 $\{\neg A \wedge \neg B, A \wedge \neg B, \neg A \wedge B, A \wedge B\}$.

- 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 could be taken as motivation for introducing probabilities into the system.

3 Infinite scales, back-and-forth reasoning, and a first motivation for 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):[12]

(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 might 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 (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

[12] 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 $[\![A]\!] = 1 \cup 2$, $[\![B]\!] = 2 \cup 3$, and $[\![C]\!] = 3 \cup 4$. (To see that Π is induced by M, note that $1 = [\![A]\!] \cap [\![B]\!]^c \cap [\![C]\!]^c$, $2 = [\![A]\!] \cap [\![B]\!] \cap [\![C]\!]^c$, $3 = [\![A]\!]^c \cap [\![B]\!] \cap [\![C]\!]$, $4 = [\![A]\!]^c \cap [\![B]\!]^c \cap [\![C]\!]$, and that all other attempts to assign truth values to the different alternatives in M are inconsistent.) 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, as explained in footnote 19.

likely each cell is given the message:[13]

$$P(\exists \wedge \neg\forall|\text{‘some’}) = \frac{P(\text{‘some’}|\exists \wedge \neg\forall)P(\exists \wedge \neg\forall)}{P(\text{‘some’})}$$

$$P(\forall|\text{‘some’}) = \frac{P(\text{‘some’}|\forall)P(\forall)}{P(\text{‘some’})}$$

The denominator on the right-hand side of both cases is identical and does not affect the comparison.[14] 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 actual probabilistic assessments (that might be part of the common ground), a possibility that we discuss in section 6 below (though 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 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

[13] 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 $\sum_{t' \in \Pi} P(m|t') \cdot P(t')$.

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

$P(\text{'some'}|\forall) < 1$, 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.[15]

This kind of probabilistic reasoning, then, can motivate adopting something like (4).[16] 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 strong 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 (though in the next section we will consider potential support for a probabilistic variant of (4)).

As a first example, suppose that the set of messages is $\{A, B\}$, which – assuming that A and B are logically independent – 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 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 or the common-ground already excludes $A \wedge B$ (see Fox 2019); in either case peeling can proceed using (2).

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

[15]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.

[16]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.

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:[17]

- 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 can be motivated through probabilistic reasoning, we can take infinite scales to also indirectly support a role for probabilities in IRMs. 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 formal probabilistic computations: in both cases, actual probability assessments about states of the world will not play a role. We will discuss the significance of this distinction in section 6.

[17] 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 16, 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.

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

4.1 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 (which we will generalize in (14) below) 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.

[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.

4.2 A role for 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 discussed above 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', the probabilistic components that we 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 every 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]

(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 every 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):

[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 $P(m|t)P(t) = P(m|t')P(t')$). 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.

(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]

(10) Naive speaker: if cell t makes n different messages true, the speaker will choose each of them with probability $\frac{1}{n}$.[21]

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

[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 every 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 omit a cardinality-based statement and to state the discussion here and below directly in terms of probabilities.

[21] As mentioned in fn. 9, the IRM literature sometimes assumes that messages are associated with costs (for example, correlating with phonetic effort) and that these costs can affect the speaker's choice. Such assumptions are not necessary for the present derivation, so we do not make them here. We were also not able to see places where costs would affect our conclusions in the discussion below.

the cell $\neg A \wedge B$. No further identification is possible in the first stage.[22] However, 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 various cases that could already be handled with non-probabilistic identification. For example, consider again the case of $M = \{$'some', 'many', 'all'$\}$, where the partition is $\Pi = \{\neg\exists, \exists \wedge \neg many, many \wedge \neg\forall, \forall\}$. 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.

4.3 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.[23] To our knowledge, the literature does not provide derivations of conjunctive readings of disjunctions within other kinds of IRMS (including those of the Rational Speech Act family; in particular, the specific system of Bergen et al. 2016 does not yield these readings). Given

[22] This is so since (8) requires $P(t|m) > P(t'|m)$, using the strict $>$, rather than requiring $P(t|m) \geq P(t'|m)$. Choosing $\geq$ would have made 'A or B' identify both $A \wedge \neg B$ and $\neg A \wedge B$, an incorrect result.

[23] 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, along the lines suggested in works such as Goodman and Stuhlmüller 2013.)

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.

5 Theory comparison

By deriving conjunctive interpretations of disjunction, Franke (2009) provides a non-trivial accomplishment for IRMs that could serve as part of an argument in favor of the overall approach. The question, of course, is how well Franke's results fare compared to non-IRM alternatives. The present section provides a preliminary comparison of this kind, pitting the IRM we have developed above against a specific non-IRM alternative, and arrives at the conclusion that the IRM solution does not compare favorably with the non-IRM one.

The non-IRM alternative we will focus on is the grammatical approach, and before proceeding, let us briefly review how this approach derives conjunctive readings of disjunctions.[24] On this approach, as mentioned briefly in the introduction, 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.[25] 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 conjunctive read-

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

[25] See Fox 2007, Magri 2011, and Fox and Spector 2018 for proposals regarding the distribution of *Exh*, with roots going back to Chierchia (2004).

ing.[26,27]

5.1 More than two disjuncts

We have now seen how conjunctive readings of disjunction can be derived both in an IRM and in a non-IRM, focusing on cases in which there are two disjuncts. We now turn to what should, to our mind, be a simple extension: in all cases where a conjunctive interpretation arises for a disjunctive construction with two disjuncts it also arises for a minimally different construction in which more than two disjuncts are involved. The following illustrates:

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

We can state this as a number-independent variant of (7):

(14) NUMBER INDEPENDENCE OF CONJUNCTIVE READINGS OF DISJUNCTION:
If $\phi(\cdot)$ is an environment that gives rise to conjunctive readings in the case of $\phi(A_1 \text{ or } A_2)$, it also does so in the case of $\phi(A_1 \text{ or } A_2 \text{ or} \ldots \text{or } A_n)$.

Under the grammatical approach, the relevant inference for (13) is derived straightforwardly (as is the broader generalization in (14)) by the same system that was designed to handle 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

[26] 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.

[27] In earlier variants of the grammatical approach, *Exh* was defined in terms of innocent exclusion alone. In the case of $M = \{\text{A or B, A, B}\}$, as we saw, no alternative is innocently excludable given $S =$'A or B'. That is, $Exh_M(\text{A or B}) = A \vee B$. For the other two alternatives we have $Exh_M(\text{A}) = A \wedge \neg B$ and $Exh_M(\text{B}) = B \wedge \neg A$. 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 $Exh_M(\text{A or B})$, so when $Exh_{M'}(Exh_M(\text{A or B}))$ is computed, both alternatives can be negated while $Exh_M(\text{A or B})$ 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.

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.[28]

While the generalization to more than two disjuncts is handled by the grammatical approach without any modification, the same is not true for the IRM approaches discussed above. When there are more than two disjuncts, both Franke (2011)'s IRM and our own fail to derive conjunctive readings of disjunctions.[29] 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 $\{\neg A \wedge \neg B \wedge \neg C, A \wedge \neg B \wedge \neg C, \neg A \wedge B \wedge \neg C, \neg A \wedge \neg B \wedge C, A \wedge B \wedge \neg C, A \wedge \neg B \wedge C, \neg A \wedge B \wedge C, A \wedge B \wedge C\}$.[30] On the assumption of a naive speaker, as stated in (10), we have the following speaker probabilities:

(15) 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 (15), 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.

- 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. This leaves (16) as the relevant part of the table for the rest of the peeling process.

(16) The portion of (15) that is relevant after the first round of peeling:

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

[28] The correct result is obtained 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.

[29] As mentioned, some IRMs in the literature fail already with two disjuncts.

[30] Note that the same partition can also be induced by a smaller set of messages (in particular, any subset that includes 'A', 'B', and 'C'). This, however, will not help with cell identification.

- Step II. At this point, as seen in (16), all remaining messages are compatible with all remaining cells, 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 clearly problematic given examples such as (13) and the broader generalization in (14). We presented the problem within our simplified system, but as Franke (2011) notes, a similar problem holds for his version of IRM.[31]

The peeling process cannot derive the conjunctive meaning for three disjuncts because the extra disjunct creates a situation where there are too many cells that cannot be peeled off. The same problem arises already for two disjuncts in the presence of additional alternatives (and consequently additional cells):

(17) 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 (17a) 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 (17a) 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.[32]

Comments by a reviewer highlight the significance of considering the case of (17a) above, rather than just the basic conjunctive reading of 3-way disjunctions. The reviewer observes that a certain kind of modification of existing IRMs can yield the conjunctive reading of 3-way disjunctions. The implementation of this modification can vary between frameworks, but the general idea is as follows. There are various criteria that one can introduce that would favor the alternatives in the cases we are dealing with which contain fewer disjuncts. Single disjuncts will be best where possible, and as in

[31] While both IRMs 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.

[32] 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 (any two) but not all three. As before, the problem does not arise with the grammatical approach, where the case of two disjuncts in the presence of an additional alternative is derived straightforwardly using the same mechanism that derived the basic cases. Referring back to the steps in the derivation listed above, the only 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).

our IRM they will identify their respective exactly-1-positive cells (with ‘A’ identifying $A \wedge \neg B \wedge \neg C$, and so on). At this point, the 2-way disjunctions will all be better than the full 3-way disjunction, but the 2-way disjunctions will still not be paired with the appropriate states for the reasons discussed in reference to the table in (16) during step II. However, if we now also adopt a notion of cell identification that allows for many-to-many pairings, each of the 2-way disjunctions can be paired with all of the exactly-2-positive cells $A \wedge B \wedge \neg C$, $A \wedge \neg B \wedge C$, and $\neg A \wedge B \wedge C$.[33] This leaves just the 3-way disjunction ‘A or B or C’ and the 3-positive cell $A \wedge B \wedge C$, which leads to the correct pairing for the 3-way disjunction, thus avoiding the impasse we ran into in our peeling process above and directly addressing the challenge posed by cases like (13). We note, however, that regardless of what one thinks of the criteria that need to be introduced, the approach leads to failure in the 2-out-of-3 disjunct case: it incorrectly pairs a message such as ‘A or B’ with all three of the exactly-2-positive cells rather than just with $A \wedge B \wedge \neg C$.

We certainly do not mean to imply that these problems cannot be addressed. 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.

5.2 A broader comparison

In the previous sections, we presented the most interesting argument we found in the literature that IRMs might offer adequate accounts of SIs. We saw that an initially promising theory of SIs runs into non-trivial challenges, not generalizing in the way it should beyond the simplest cases. But, of course, the conjunctive meaning of disjunction is just one point of comparison among different theories of SI. The question of theory comparison needs to be be much broader in scope, taking into account, as usual, any observation, empirical or theoretical, that might bear on theory choice.

Though we will not attempt to go into the various issues involved in any detail here, we would like to mention the growing body of observations that have been presented in favor of the grammatical approach. Probably the most discussed are observations pertaining to the distribution of SIs, and in particular to the presence of strengthening in embedded positions (see Cohen 1971, Landman 2000, Chierchia 2004, Chierchia, Fox, and Spector 2012, among other work). Strenghtening in embedded positions is expected if SIs result from a grammatical representation containing an occurrence of *Exh*, but not if strengthening is the result of attempts to retrieve the intentions of a speaker given certain assumptions about the nature of the common ground, as in the derivation of SIs within IRMs or other Gricean approaches.[34]

[33] To see why this might be a reasonable mapping note that each of the 2-way disjunctions receives a higher probability given the exactly-2-positive cells ($\frac{1}{6}$) than it does given the 3-positive cell ($\frac{1}{7}$). So if we have a criterion that favors the 2-way disjunctions, we can appeal to probabilities to map these disjunctions to the exactly-2-positive cells.

[34] Initially it was thought that strengthening in embedded positions could be derived within an extension of IRMs that involves lexical uncertainty, as proposed by Bergen et al. (2016). As mentioned in fn. 3, however, more recent work has pointed out empirical shortcomings of this approach and moved toward a reinterpretation of IRMs as a disambiguation framework that reasons about *Exh* and is thus very much in line with the grammatical approach (see Champollion et al. 2019 and Franke and Bergen 2019).

But, although these observations about embedded exhausitification have received some attention in literature on IRMs, the body of pertinent work is much broader, including arguments that *Exh* plays a role in grammar: that it has the properties of a focus associating particle (see Fox and Katzir 2011 and Katzir 2014) and that it is relevant for polarity phenomena (see Chierchia 2006, 2013 and Crnič 2013, 2020), for the semantics of questions (see Groenendijk and Stokhof 1984 and Fox 2019), and for the licensing of ellipsis (see Crnič 2017, 2018), arguments for modularity in the computation of SIs (see Fox and Hackl 2006 and Magri 2009, 2011), and arguments that SIs can persist even in conversational situations that do not satisfy the assumptions about the common ground needed for their computation within IRMs or other Gricean approaches (see Fox 2014). It seems to us that the this body of observations needs to be part of the conversation.

6 Probabilities and modularity

Suppose, however, that the challenges discussed for IRMs in the previous section can be overcome and that a variant of the IRMs under consideration becomes a leading contender in the account of SIs. Given the crucial reliance of current IRMs on probabilistic reasoning, we would then have an argument in favor of incorporating probabilities into the mechanisms that derive SIs. Since probabilistic reasoning is often bundled together with representations of world knowledge and beliefs of discourse participants, it might seem that such an argument for probabilities will also become an argument for a non-modular system in which the computation of SIs is aware of various aspects of world knowledge and probability assessments about states of affairs.

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. We will argue that if IRMs of the kind discussed above turn out to provide the correct account of SIs, the probabilities involved should probably be thought of as formal constructs, internal to the mechanism that computes SIs, rather than an external input that reflects actual beliefs about the world.

6.1 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:

(18) 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 probabilities, as summarized in (18). 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 $\forall$. Clearly these are incorrect predictions. For example, if John says "I like some of my children," you will be justified in thinking that he does not like all of his children; the fact that parents usually like all their children is entirely irrelevant.[35]

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 $M = \{\text{'some', 'many', 'all'}\}$ since, as we saw earlier, this case is handled straightforwardly by non-probabilistic identification. But 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. In fact, the problem posed by biased priors is particularly striking in this case. Consider again the naive speaker's probabilities for this case, repeated here from (11):

(19) 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	$\frac{1}{2}$	$\frac{1}{2}$
$A \wedge B$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

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

[35] To see that the inference from 'some' to 'not all' holds in this case and that no similar inference from 'some' to 'all' is available, note that you may respond to John's assertion with "That can't be true–I'm convinced you like all of them" but not with "That can't be true–I'm convinced you like just some of them".

$A \wedge \neg B$ 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 $\neg A \wedge B$. This is a blatantly wrong prediction. If Mary tells her daughter Kim "You may take the Ford or the Porsche tonight", Kim will be justified in thinking that she can choose; the fact that permission to take the Ford but not the Porsche is somewhat likelier than permission to take the Porsche is simply irrelevant. Additional examples illustrating this problem are readily constructed. In (20a), for example, we can conclude that we can buy Italian newspapers and that we can buy English newspapers, even if the latter is likelier than the former (for example, if the store is in Boston). In (20b) we can conclude that if we ignore our neighbor we will regret it and that if we ignore our best friend we will regret it, even though the latter might seem likelier than the former. And in (20c) we can conclude that if we do the reading we will get an A and that if we answer all of the questions correctly we will get an A, again despite an imbalance between the likelihoods.

(20) a. In this store you can buy Italian or English newspapers
b. If you ignore your neighbor or your best friend, you are going to regret it
c. If you do the reading or answer all of the questions correctly you will get an A

6.2 Approaches to biased priors in the IRM literature

Above we illustrated the problem of sensitivity to priors faced by the IRM we developed. The exact same problem arises in Franke's system. Specifically, like our simplified IRM it makes wrong predictions in the scalar case (such as 'some'/'all') when heavily skewed priors are allowed. And likewise it makes wrong predictions for disjunctive sentences even when the slightest deviation from flat priors is permitted. Here we would like to discuss ideas entertained in the IRM literature for issues of this sort and explain why we do not think they provide a satisfactory answer to the questions that arise, at least not at present.

Franke (2009, 2011), who noted the problem in the scalar case, considers various responses. In one place (Franke 2009, p. 75), noting that the problem in the scalar case is confined to heavily skewed priors, he writes "I am not worried if a theory makes unintuitive predictions for unnatural, non-occurring parameter settings, as long as there is some sufficient margin around natural parameter settings in which predictions are robust." We disagree with this approach as will become evident in the next section, but in any event it is not helpful in the case of the conjunctive meaning of disjunction.[36] In this case, as we mentioned, the slightest skew in the prior distribution leads

[36]If a new theoretical proposal makes new and unique predictions for certain situations, the correct approach, we would think, is to figure out ways of bringing about these situations (by an appropriate experimental setup, formal or informal). Simply asserting that the situations are "non-occuring" is unhelpful. The question one needs to ask, we think, is whether the situations can be made to occur. If not, this is of course an unfortunate situation – a missed opportunity to test what is unique about the new theoretical approach. But demonstrating that the situations cannot be made to occur *in principle* is not going to be simple, and it is completely irrelevant to observe that they do not occur in run-of-the-mill circumstances.

to unattested results, and since slightly skewed priors are ubiquitous, the problem is quite dramatic in this domain, as observed by the examples above.

Franke (2009, pp. 75–6) also considers a "technical solution" to the problem in the scalar case based on the idea that "the sender considers it very unlikely, but still possible, that the priors are not heavily skewed".[37] This technical solution will not be helpful in the case of disjunction, where as we have seen priors need to be perfectly flat rather than simply not heavily skewed. We conclude that Franke 2011's claim (fn. 20, p. 27) that flat priors are just a convenience and not a necessity is incorrect.

The extreme sensitivity to priors in the case of conjunctive readings of disjunctions is important because, as discussed in the previous section, these readings provide us with the only argument that probabilistic reasoning plays any role in the computation of SIs. If we only had the scalar case, we could live happily with the simple, non-probabilistic version of cell identification. Likewise, if the problem of biased priors arose only in the scalar case, there would be numerous ways to avoid the problem. In addition to Franke's proposals we could also simply assume that the non-probabilistic version of cell-identification applies whenever it works (as entertained briefly above). It is precisely because the conjunctive interpretation might require probabilistic identification that the problem of biased priors in this domain is so pressing. We therefore conclude that if there is any evidence that probabilistic considerations enter into the computation of SIs, the system needs to be al priors. This possibility is in line with yet another possibility that Franke (2009) entertains (p. 132).

6.3 Modularity

In the previous section we have discussed the problem of prior sensitivity that arises if one assumes that the priors in an IRM reflect actual probability assessments. We have considered various responses to this problem and their inapplicability to the most important case, namely the conjunctive interpretation of disjunction – the only area where a simple non-probabilistic IRM fails. We are, of course, not claiming that there cannot be a way out of this problem. For example, we think that it is possible to identify a property of the computation that yields a conjunctive interpretation and to assert that whenever a hearer encounters this property they stipulate flat priors. But proposals of this sort are plainly excuses for an inability to corroborate an expectation of anyone who believes in sensitivity to actual priors, namely that this sensitivity could reveal itself. In other words, it seems to us that a theory that assumes sensitivity to actual priors should be supported by areas where this sensitivity is attested rather than defended against counter-examples by various band-aid solutions. The fact that the relevant literature is full of the latter, but lacks any of the former calls for an explanation.

We believe that there is one very natural explanation that should be taken seriously. It is possible that we haven't found any evidence for sensitivity to actual priors because such sensitivity does not exist. Specifically, it is possible that the system is modular and

[37] A different technical fix is considered in Degen et al. 2015, who present their idea within an IRM that does not at present account for the conjunctive readings of dijunction. Checking whether it can be incorporated within Franke's IRM or our own requires investigating various decision points and will have to be left for some other occasion, though we think the general comments made in the next section will remain relevant.

if it makes any probabilistic computations – a questionable premise given the empirical challenge discussed in section 5 – the priors involved are formal constructs defined internally to the system. This conclusion should be taken particularly seriously given independent evidence for modularity in the computation of SIs even in domains that do not involve probabilistic considerations (see Fox and Hackl 2006 and Magri (2009)).

7 Conclusion

We considered what to our mind is the strongest argument for the viability of IRMs as accounts of SIs: their ability to derive certain conjunctive readings of disjunction. Upon closer inspection, we saw that the current IRMs that accomplish this result – Franke (2011)'s and ours – do not compare well to the grammatical approach. In particular, they struggle with extending the basic result to cases with more than two disjuncts (which the grammatical approach does without modification), and we also noted a broader comparison with established results of the grammatical approach that remains to be made. As far as we can see, this currently leaves the grammatical approach as the front-runner as an account of SIs. We further noted that, even if one sets aside the problems, current IRMs furnish no argument for a non-modular architecture. In fact, the reverse is true: the lack of evidence for a role of belief assessments in the computation of SIs and the extreme sensitivity of the derivation of conjunctive interpretations of disjunctions to biased priors suggest a highly modular computation in which probabilities, if needed, are a purely formal construct.

References

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

Asherov, Daniel, Danny Fox, and Roni Katzir. 2020. A problem for iterated rationality approaches to scalar implicatures. Ms., MIT and TAU.

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.

Blutner, Reinhard. 1998. Lexical pragmatics. *Journal of Semantics* 15:115–162.

Blutner, Reinhard. 2000. Some aspects of optimality in natural language interpretation. *Journal of Semantics* 17:189–216.

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.

Champollion, Lucas, Anna Alsop, and Ioana Grosu. 2019. Free choice disjunction as a rational speech act. Handout for poster presented at SALT 29, May 2019.

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.

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. 2006. Broaden your views: Implicatures of domain widening and the "logicality" of language. *Linguistic Inquiry* 37:535–590.

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.

Cohen, Laurence Jonathan. 1971. Some remarks on Grice's views about the logical particles of natural language. In *Pragmatics of natural language*, ed. Yehoshua Bar-Hillel, 50–68. Reidel.

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

Crnič, Luka. 2017. Free choice under ellipsis. *The Linguistic Review* 34:249–294.

Crnič, Luka. 2018. A note on connected exceptives and approximatives. *Journal of Semantics* 35:741–756.

Crnič, Luka. 2020. Number in NPI licensing. Ms., HUJI, available at lingbuzz/004954, January 2020.

Degen, Judith, Michael Henry Tessler, and Noah D Goodman. 2015. Wonky worlds: Listeners revise world knowledge when utterances are odd. In *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, ed. D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimini, T. Matlock, C. J. Jennings, and P. P. Maglio, 548–553.

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. 2019. Partition by exhaustification: Comments on Dayal 1996. In *Proceedings of Sinn und Bedeutung*, volume 22.

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.

Frank, Michael C., and Noah D. Goodman. 2012. Predicting pragmatic reasoning in language games. *Science* 336:998.

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 Leon Bergen. 2019. Pragmatic reasoning about silent only. Ms., University of Osnabrück and UCSD.

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.

Groenendijk, Jeroen, and Martin Stokhof. 1984. Studies in the semantics of questions and the pragmatics of answers. Doctoral Dissertation, Universiteit van Amsterdam, Amsterdam.

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

Jäger, Gerhard. 2002. Some notes on the formal properties of bidirectional optimality theory. *Journal of Logic, Language and Information* 11:427–451.

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.

Landman, Fred. 2000. *Events and plurality: The Jerusalem lectures*. Dordrecht: Kluwer Academic Publishers.

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.

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.

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.

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.