October 2021

Reference games and the nature of exhaustification*

Daniel Asherov, Danny Fox, & Roni Katzir

Abstract: The literature on scalar implicatures (SIs) varies in its views on the division of labor between grammar and general reasoning in the derivation of SIs. According to the grammatical approach, the SIs of a given sentence are logical entailments of particular parses of that sentence – specifically, parses with a silent exhaustivity operator, notated as *Exh*, a covert counterpart of 'only'. According to a competing view, there is no need for anything like *Exh* in the grammar; rather, the dynamics of conversation suffice to derive SIs. The present squib compares the two views. We will consider the *Exh*-free approach in the context of what we will refer to as iterated rationality models (IRMs). According to this prominent approach, SIs arise from iterated steps of reasoning by discourse participants about the goals and means available to other discourse participants. Focusing on one-shot reference games (Rosenberg & Cohen 1964, Frank & Goodman 2012), we will argue that in this setting SIs can arise only in circumstances where a speaker's utterance can be parsed with *Exh*. This observation, which is a direct prediction of the grammatical theory, is problematic for IRMs that are able to derive SIs from non-exhaustified representations.

1. Introduction

A listener who hears a sentence like (1a) is likely to infer that (1b) is false. This inference is interesting because there are well known arguments that (1a) has a meaning that does not entail the negation of (1b). Practically all responses to these arguments share the assumption that the inference comes about from a process that takes into account not just the basic linguistic expression in (1a) but also alternative expressions. This process leads to what is called a *scalar implicature* (SI). (1c) states the *strengthened meaning* of (1a) – the conjunction of (1a) with its SI that (1b) is false.

(1) a. *Some* of the bananas are green.
    b. *All* of the bananas are green.
    c. *Some but not all* of the bananas are green.

*We thank Adam Albright, Anton Benz, Emmanuel Chemla, Luka Crnič, Michel DeGraff, Alon Fishman, Yosef Grodzinsky, Martin Hackl, Joshua Hartshorne, Philippe Schlenker, Roger Schwarzschild, Greg Scontras, Jesse Snedeker, Benjamin Spector, and Donca Steriade for their comments, as well as the audiences at the LSA 2021 annual meeting, MIT Experimental group, MIT LingLunch, TAU Computational Linguistics Lab, the Computational Psycholinguistics Laboratory at MIT, ZAS, ESSLLI 2021, and Harvard LangCog.

According to one view, the so-called *grammatical* approach (see Fox 2007 and Chierchia, Fox, & Spector 2012),[2] the SI of (1a) arises because, in addition to the parse that does not entail the negation of (1b), there is another parse that does. More generally, the grammatical approach treats SIs as logical entailments of parses that the relevant sentences can be associated with – specifically, parses with a silent exhaustivity operator (notated as *Exh*), a covert counterpart of the overt alternative-sensitive operator 'only'. According to a competing view, there is no need for anything like *Exh* in the grammar; rather, the dynamics of conversation suffice to derive SIs. The present squib compares the two views and provides novel evidence that *Exh* is needed.

For the purposes of the discussion we will consider the *Exh*-free approach in the context of what we will refer to as *iterated rationality models* (IRMs; Benz 2006, Benz & van Rooij 2007, Bergen & Goodman 2015, Frank & Goodman 2012, Frank et al. 2016, Franke 2009, 2011, Goodman & Stuhlmüller 2013, Rothschild 2013, Scontras et al. 2018, among others). According to this prominent approach, SIs arise from a general-purpose decision-making mechanism available to humans and specifically from iterated steps of reasoning by discourse participants about the goals and means available to other discourse participants.[3]

In what follows we will consider the simple setting of one-shot *reference games* (Rosenberg & Cohen 1964, Frank & Goodman 2012). We will argue that in this setting SIs can arise only in circumstances where a speaker's utterance can be parsed with *Exh*. This observation, which is a direct prediction of the grammatical theory, is problematic, as we will see, for IRMs that are able to derive SIs from non-exhaustified representations.

2. Baseline: Overlapping predictions of Exh and Exh-free IRMs

Consider Scenario A in Figure 1 (Stiller et al. 2011, 2015, Vogel et al. 2014). Given Scenario A, a listener might understand (2a) to convey that the relevant crate *is not with an apple*, arriving at the strengthened meaning in (2b).

---

[2] The approach has roots in earlier work, notably Groenendijk & Stokhof 1984, Krifka 1995, Landman 2000, and Chierchia 2004.

[3] Given our goal of providing an argument for *Exh*, most of our discussion of IRMs will focus exclusively on models that do not assume *Exh*. Recent work in the IRM literature has argued that IRMs work better when the grammar does, in fact, include *Exh* (Champollion et al 2019, Franke & Bergen 2020). Nevertheless, we will argue in §7 that such models are problematic as well, insofar as they are capable of deriving a strengthened meaning also when *Exh* is absent.

We further note that much of the earlier *Exh*-free literature was framed not within IRMs but rather within the so-called Neo-Gricean approach (see Horn 1972, Gazdar 1979, Sauerland 2004, among others). That approach, however, was argued against based on a variety of considerations (see Fox 2007, 2014, Magri 2009, 2011, Chierchia, Fox, & Spector 2012, and Rothschild 2013, among others), and we will set it aside in what follows.

(2) a. Pick the crate *with a banana*.
  b. Pick the crate *with a banana and with no apple*.

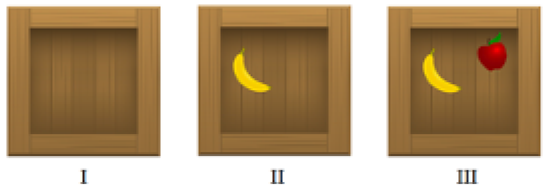

FIGURE 1: SCENARIO A

Stiller et al. (2011, 2015) tested listeners' intuitions in reference games with a context like Scenario A with three crates ('states' I-III).[4] The task of a listener is the following: given Scenario A, detect the object which the speaker intends to refer to, given that the speaker *can only use one word*, and that the speaker *uttered the word 'banana'*. Although both state II and state III contain a banana, listeners seem to have the intuition that the expression 'banana' refers to state II. Stiller et al.'s participants chose state II 96% of the time. This result was replicated in an experiment by Vogel et al. (2014), who gave only one trial per participant; 75% of their participants chose state II.

In the simple case of Scenario A, taking the set of alternatives to consist of single-word fruit names, both the grammatical approach and IRMs successfully predict listeners' intuition. As mentioned in the introduction, the grammatical approach accounts for the intuition in (1) within grammar. Specifically it assumes that to generate a strengthened meaning for a sentence, a silent alternative-sensitive operator *Exh* must be inserted in the grammar. The effect of adding *Exh* (or overt *only*) which we label *exhaustification*, is as follows.

(3) EXHAUSTIFICATION: Given a linguistic expression λ and a set of linguistic alternatives ALT, the strengthened meaning of λ is the conjunction of λ and the negation of all members of ALT which are not entailed by λ.[5]

---

[4] The scenarios in Stiller et al. (2011, 2015) and Vogel et al. (2014) were based on other, structurally identical, stimuli. An example counterpart for Scenario A in FIGURE 1 was the following: one smiling face without an article of clothing, one with only glasses, and one with both glasses and a hat. In FIGURE 1 these states correspond to state I, state II, and state II, respectively. We chose to use crates with fruit based on a suspicion that this raises the likelihood of balanced salience among the identifying features. The intuition is that salience is more balanced among names of different fruit (e.g. *pear*, *apple* and *banana*) than among names of different features of the same character (e.g. *tie*,. *hat* and *glasses*).

We also modified the utterances. In Stiller et al. (2011, 2015) and Vogel et al. (2014) the stimuli are single word utterances. We decided to modify those to full sentences so as to have greater control over the types of mental representations that a listener might construct internally in response to a single word utterance.

[5] This is an oversimplification that does not take into account the possibility of multiple alternatives that, when negated together, would be inconsistent with the assertion (see Sauerland 2004, Fox 2007). It also ignores the

For now, we assume that the set of alternatives (ALT) of 'banana' includes similar one-word utterances, such as 'apple', as stated in the description of the task.[6] For (2a), this means that exhaustification can yield the meaning (2b), paraphrasable as *Pick the crate with only a banana*. Since in Scenario A this is only consistent with state II, the grammatical approach predicts correctly that listeners will identify II as the intended crate.

(4) Exhaustification of 'banana' given Scenario A

a. Given: λ = 'banana', ALT = {'banana','apple'}
b. Get strengthened meaning, Exh(ALT, banana) = banana & not apple

IRMs derive a similar inference. Recall that in this approach, listeners take into account the strategies available to the speaker for achieving a communicative goal. A strategy involves the selection of an utterance from a set of available alternatives, often referred to within the IRM world as *messages*, and the communicative goal, at least in our case, is conveying an intended state. Speakers, in turn, take into account the fact that listeners can use this type of reasoning. In the case of Scenario A, this reasoning can proceed as follows (presented first in schematic terms and elaborated further below):[7]

(5) Iterated rationality reasoning about Scenario A (schematic)

Step 1: Although the message 'banana' is initially ambiguous, the message 'apple' is not. Specifically, 'apple' would lead the listener directly to the state 𝔹𝔸, since it is inconsistent with the other two states. Thus, if a rational speaker had intended the state 𝔹𝔸, they would use the message 'apple'.

Step 2: Given step 1, the speaker who used the message 'banana' did not intend the state 𝔹𝔸. Once 𝔹𝔸 is eliminated as the intended state, the only remaining option is state 𝔹.

Each model in the IRM family implements this idea slightly differently. In what follows, we present a simplified algorithm from Fox & Katzir (2021), in which the core idea is transparent. As such, we think that it conveys the rationale behind existing models rather clearly while de-emphasizing implementational differences. The algorithm, stated in (6B), involves iterations in which messages and states are paired together according to an identification criterion and peeled off. The identification procedure we start with, *Semantic Identification*, is stated in (6A).

---

possibility that alternatives under certain circumstances could be affirmed rather than negated (see Bar-Lev & Fox 2020, Fox & Katzir 2021). For our immediate purposes we can set these complications aside.

[6] We will come back to this question in §6, where we show that adding richer alternatives does not change the overall results of either the grammatical approach or IRMs.

[7] Here and below, messages will be referred to with strings of words between quotation marks, e.g. 'banana'. States will be labeled by their content with capital initials: in Scenario A state I is labeled EMP (for empty), state II is labeled 𝔹 (for banana), and state III is labeled 𝔹𝔸 (for banana & apple).

(6) State Identification

(A) IDENTIFICATION (SEMANTIC): Given a set of messages *M* and a set of states *T*, a message identifies a state if it is true in that state and there is no other state in which it is true.

(B) PEELING PROCEDURE: Given a set of messages *M,* a set of states *T*, and an identification criterion *C*:

a. Collect all message-state pairs where the message identifies the state according to *C*
b. If at least one state was identified in the most recent application of (a):
    i. Peel off (=remove) all messages and all states that were collected into pairs in the most recent application of (a)
    ii. Repeat (a)

(5) above can be seen as a simple application of the peeling procedure (6B) with semantic identification (6A) to Scenario A, assuming that the only possible messages are 'apple' and 'banana'. We can conclude that this simple IRM is successful in capturing the fact that given Scenario A, listeners understand the message 'banana' as referring to the state that contains banana and does not contain an apple (state II).[8]

3. A problem for iterated rationality

We saw that a rather simple model of rational reasoning captures listeners' intuitions in a non-trivial mapping between a message and a state. In this section and the next, we will modify Scenario A so as to avoid having a crate with a banana and no other fruit. We will see that in such cases "Pick the crate with a banana" is infelicitous, a fact that is correctly predicted by the grammatical approach. IRMs, on the other hand, allow for iterated and/or probabilistic reasoning to identify crate II.

As our first illustration, consider Scenario B in FIGURE 2, which is based on a scenario in Vogel et al. (2014) (see also Frank & Goodman 2012). It differs from Scenario A in the following way: a pear is added to both crate I and crate II. This scenario is important for the comparison between the grammatical approach and the IRM approach because, unlike Scenario A, here there is no state with *only* a banana.

---

[8] Given our choice of utterances, one needs to also worry about the use of the definite article *the,* which presupposes that the predicate it combines with is true of only one individual. Given this presupposition, strengthening in our scenarios must occur before the application of the definite article *the*. (This is not necessarily the case in the reference games described in the original literature (see note 4).) For the purpose of this discussion we will assume that both approaches have a way to deal with this worry. On the grammatical approach, the strengthened meaning would be derived by an embedded occurrence of *Exh*, e.g.: *The crate Exh with a banana.* The IRM literature could appeal to Horn's (1989) claim that the definite article can be used "meta-linguistically": instead of presupposing that there is a unique individual that the predicate is true of, the article – on its meta-linguistic use – presupposes that there is a unique individual that can be identified by the predicate.

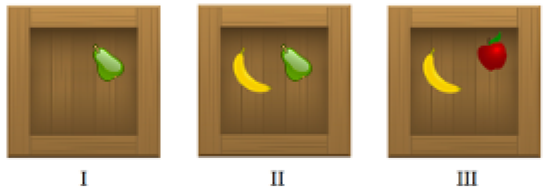

FIGURE 2: SCENARIO B

Listeners' intuition in Scenario B is strikingly different from that in Scenario A. Here, given the message 'banana', listeners typically do not know which state the speaker has in mind. This intuition is stated in (7). Indeed, Vogel et al. (2014) elicited listeners' judgments in a reference-game experiment and found that only 52% of the participants chose state II (cf. 75% in Scenario A).

(7) Intuition in Scenario B
If the speaker utters the sentence 'pick the crate with a banana', it is unclear which crate they are referring to.[9]

The grammatical approach makes the correct prediction, as seen in 8. In its literal meaning, 'banana' is true in both state $\mathbb{BP}$ (state II) and state $\mathbb{BA}$ (state III). In its exhaustified meaning, i.e. banana and no other alternative, the message 'banana' is not true in any of the states. Thus, the intuition that no single state fits the message 'banana' is correctly predicted.

(8) Exhaustification of 'banana' given Scenario B
a. Given: λ = 'banana', ALT = {'banana','apple','pear'}
b. Apply exhaustification to λ: not apple & not pear
c. Get strengthened meaning: banana & not apple & not pear

In contrast, iterated rational reasoning predicts that listeners would understand the message 'banana' as referring to state $\mathbb{BP}$ (state II). The derivation of the meaning in State Identification is spelled out in (9).

(9) Iterated rationality reasoning about Scenario B

Step 1: Although the message 'banana' is initially ambiguous (as is the message 'pear'), the message 'apple' is not. Specifically, 'apple' would lead the listener directly to the state $\mathbb{BA}$, since it is inconsistent with the other two states. Thus, if a rational speaker had intended the state $\mathbb{BA}$, they would use the message 'apple'.

[9] We note, however, that Frank & Goodman (2012) conclude that in a similar scenario to B speakers do tend to be pushed by the parallel of the message 'banana' to the parallel of crate II, but this conclusion was based on rather indirect evidence, and this interpretation of the data has been challenged in Sikos et al. 2021.

Step 2: Given step 1, the speaker who used the message 'banana' did not intend the state 𝔹𝔸. Once 𝔹𝔸 is eliminated as the intended state, the only remaining option is state 𝔹.

It is easy to see that the reasoning in (9) is identical to that in (5), which was used for Scenario A. The IRM approach thus incorrectly predicts that listeners will choose crate II in scenario B as well.

4. Attempting to rescue IRMs through probabilistic identification and blocking iterations

Is there a difference between scenarios A and B that might be relevant for the workings of an IRM so as to overcome this problem of overgeneration? Vogel et. al. (2014) focus on a difference that pertains to the perspective of a speaker that needs to select from among the alternative messages that are true in a given state. In Scenario A, the message 'banana' is the only message that such a speaker can select for crate II but one out of two possible messages for crate III. In scenario B, on the other hand, `banana' is one out of two possible messages both for crate II and for crate III. This could be stated in probabilistic terms as follows on the assumption that the speaker chooses randomly between the messages that are possible given the state they wish to refer to: in scenario A, 'banana' will be uttered with probability 1 for crate II and with probability 0.5 for crate III, while in scenario B, 'banana' will be uttered with probability 0.5 both for crate II and for crate III. Suppose, then, that we changed the identification criterion so as to take this distinction into account:[10]

(10) Probabilistic State Identification (PSI)

IDENTIFICATION (PROBABILISTIC): Given a set of messages M and a set of states T, a message identifies a state if the likelihood that the speaker would use the message to describe that state is higher than for any other state.

In Scenario A, our IRM with PSI correctly finds that the message 'banana' refers to state 𝔹. Here, unlike the non-probabilistic State Identification, our IRM discovers this message-cell pairing <u>within one step.</u>.

In Scenario B, PSI still arrives at the incorrect result that the message 'banana' refers to state II (𝔹ℙ), just like non-probabilistic State Identification:

[10] This probabilistic notion of identification can be motivated in terms of Bayesian reasoning, as is commonly done in the IRM literature. Such motivation, however, leads to the problematic prediction that scalar implicatures should be sensitive to the prior probabilities of states. See Fox & Katzir (2021) for discussion. Note that whenever m identifies t by (6) it will also identify t by (9) (at least when the set of messages is finite). This means that the move to (9) does not block identification in Scenario B as we discuss below. The only thing this move does is to make identification "easier" in A than in B.

(11) Iterated rationality reasoning about Scenario B (Probabilistic)

Step 1: Although the message 'banana' does not identify a state (and similarly for the message 'pear'), the message 'apple' does. Specifically, 'apple' would lead the listener directly to the state 𝔹𝔸, since P('apple'|𝔹𝔸)=0.5>0=P('apple'|𝔹P). Thus, if a rational speaker who is trying to identify a state with PSI had intended the state 𝔹𝔸, they would use the message 'apple'.

Step 2: Given step 1, the speaker who used the message 'banana' did not intend the state 𝔹𝔸. Once 𝔹𝔸 is eliminated as the intended state, the only remaining option is state 𝔹P.

While PSI still makes the wrong prediction for Scenario B, there is now a distinction between Scenario A and Scenario B: while in Scenario A identification is achieved <u>within one iteration (in Step 1)</u>, in Scenario B this is only achieved <u>in the second iteration (only in Step 2)</u>.

Vogel et al. (2014) notice the discrepancy in listeners' behaviour across the two scenarios and point out a difference in the number of iterations required to arrive at a solution in each of them. They propose that this difference is responsible for listeners' inability to draw an inference in Scenario B (also see Frank et al. 2016). While the IRM can, in principle, derive the inference in Scenario B, humans on their proposal cannot proceed past the step of iteration needed for identification in Scenario A (past step 1 in our simplified system). The goal of the next section is to test Vogel et al.'s idea by constructing a scenario that, on the one hand, has no crate with a banana and no other fruit (so that grammatical exhausitification fails), but, on the other hand, makes it possible for IRMs to identify crate II even under Vogel et al.'s proposal (that is, even if humans are limited to no more than one iteration of PSI).[11] We will find that, even with this limitation, the IRM approach still suffers from an overgeneration problem.

5. The problem arises even within a single iteration

Consider Scenario C (FIGURE 3). Like Scenario B, there is no state with *only* a banana. Unlike Scenario B, we will see that PSI finds message-state pairings with a single iteration, including a pairing for the message 'banana'. If humans really are capable of deriving SIs whenever probabilistic reasoning can generate them within one iteration, listeners in Scenario C should be able to find the correct message-state pairing for the message 'banana'. Unlike Scenarios A and B, the intuitions for Scenario C have not been discussed in the literature. Nevertheless, we think the intuitions are rather sharp.

[11] The idea that humans cannot proceed past a first step of PSI clashes directly with the need for multiple steps of PSI in an IRM to account for other inferences such as conjunctive readings of disjunction (see Franke 2009, 2011, van Rooij 2010, and Fox & Katzir 2021). We will set aside this concern for the purposes of the present discussion.

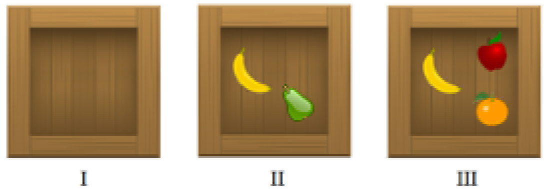

FIGURE 3: SCENARIO C

Our sentence 'Pick the crate with a banana', is infelicitous in scenario C and it seems unclear which crate the speaker has in mind, just as it was in scenario B and unlike scenario A. We take this to mean that as in scenario B an SI does not arise in this case. The intuition is stated in (12).

(12) Intuition in Scenario C
If the speaker utters the message 'banana', it is unclear which state they are referring to.

The grammatical approach arrives straightforwardly at (12). In its literal meaning, 'banana' is true in both $\mathbb{BP}$ (state II) and $\mathbb{BAO}$ (state III). In its exhaustified meaning, i.e. banana and no other alternative, the message 'banana' does not match any of the states.

In contrast, the IRM approach predicts that listeners would understand the message 'banana' as referring to state II ($\mathbb{BP}$), and it does so <u>within one iteration.</u> This is so since P('banana'|$\mathbb{BP}$)=0.5>0.33=P('banana'|$\mathbb{BAO}$).

While limiting iterated rationality to one iteration may block inference in Scenario B, it fails to do so in Scenario C. This is a problem for the IRM approach. Specifically, we've seen that this approach suffers from overgeneration, incorrectly predicting strengthening to be possible in scenario B. We saw that there is a way of dealing with this problem if we replace State Identification with PSI. But this leads to overgeneration in Scenario C.

6. Further attempts to avoid the overgeneration problem for IRMs

Before concluding this paper, we would like to discuss two further potential IRM responses to the overgeneration problem identified above, one based on the idea that a message that identifies a state must be no worse than other available messages that can serve the same purpose and another based on the idea that sensitivity to probabilities might be more limited than assumed in PSI. We present minimal variations on scenario C which will demonstrate that neither is able to solve the overgeneration problem.

6.1. No sub-optimal identifiers

The first response is based on the intuition that a speaker who wants to identify state II in scenario C has a more obvious strategy than using the message 'banana', namely using the message 'pear'. Assuming that this intuition can be grounded in a principle, then we can say that even though 'banana' is an identifier for crate II according to PSI, it is a suboptimal message. One can use this observation as the basis for a modification of our IRM so that the problem of overgeneration in scenario C will not arise.[12] This, together with restricting identification to only one iteration, could give us an IRM story for scenarios A-C.

But we think that this move will not be helpful. To see why, consider the following minimal variation on scenario C:

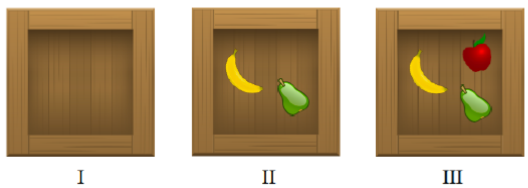

FIGURE 4: SCENARIO C'

In scenario C', 'banana' still identifies crate II using PSI, just like it did in scenario C. But differently from scenario C, in scenario C' 'pear' is not a better message than 'banana' in any conceivable sense. This means that 'banana' would identify crate II in scenario C' by any modification of PSI that avoids identification by suboptimal messages. And yet, in our judgment the utterance 'Pick the crate with the banana' remains as bad in scenario C' as it was in scenario C.

6.2 Granularity

The second response is based on the idea that our sensitivity to probabilities (at least when computing SIs) is not as fine-grained as assumed in PSI. In scenario A, where PSI successfully allowed for 'banana' to

---

[12] 'pear' identifies crate II by State Identification, so in probabilistic terms it gives the state probability 1. So the probability of crate II given the message pear (=1) is higher than the probability of this crate given the message 'banana' (<1). For this reason using the message 'banana' could be thought of as a *dominated strategy*, which would result in failure of identification under various modifications of our IRM. We thank Anton Benz for bringing up this point. One local modification is to replace PSI with the following:

> Given a set of messages *M* and a set of states *T*, a message *m* in M identifies a state *t* in T if:
> a. *t* has higher probability given *m* than any other state:
> b. There is no message that gives *t* a higher probability than *m*

identify crate II in the first iteration, there was a big difference between the probability of 'banana' in crates II and III: 1 in crate II and 0.5 in crate III. In scenarios C and C', where PSI incorrectly allowed for 'banana' to identify crate II (also in the first iteration), the difference between the probability of 'banana' in the two crates was much smaller: 0.5 in crate II and 0.33 in crate III. To block identification by 'banana' in scenarios C and C', then, one might propose a less fine-grained version of PSI that allows a message *m* to identify a state *t* only when P(m|t) is sufficiently bigger than P(m|t') for any other cell, and one can imagine various statements of what counts as sufficiently bigger.

We think, however, that no such statement will succeed. To see why, consider the following variant of scenario C': instead of using only single fruit names in our messages we will now allow also for *conjunctions* of fruit names. In this setting, 'Pick the crate with a banana and a pear' is a good message and is understood as referring to crate II. 'Pick the crate with a banana', on the other hand, is a bad message, just as it was when we considered only single fruit names. The problem for any attempt to modify PSI based on probability differences is that the two messages have the exact same probabilities in crates II and III: 0.33 in crate II and 0.14 in crate III. No way of restricting identification based on probability differences will therefore succeed in allowing 'Pick the crate with a banana and a pear' to identify crate II while preventing 'Pick the crate with a banana' from doing the same.[13]

7. Conclusions

We saw that, across a range of scenarios, strengthening in reference games is possible exactly when *Exh* yields strengthening. This is, of course, expected under the grammatical approach. On the other hand, it is surprising for IRMs that do not assume *Exh*. We take this as further evidence that IRMs cannot replace *Exh*, a conclusion that is very much in line with Champollion et al 2019, Franke & Bergen 2020, and Fox & Katzir 2021. But the scenarios above argue for a stronger conclusion. Even with a grammar that includes *Exh*, an IRM could in principle yield the same problematic strengthened meanings as before, simply by using parses without *Exh* and proceeding as in our discussion above. This would leave such *Exh*-enhanced IRMs with the same overgeneration problem as before. If IRMs are to be maintained, then, they must not just work with a grammar that has *Exh* but also be prevented from yielding strengthenings in the scenarios above other than through *Exh*.

If one adopts an IRM approach to model pragmatic interactions, one needs to understand why it is that this IRM relies on *Exh* for strengthening and cannot also derive strengthening in other ways.[14] One possibility is that the grammar offers only parses with *Exh* in the scenarios discussed above, though this would require further explanation, as strengthening is, of course, not in general obligatory (see, e.g., (1)). Another possibility is that, as Vogel et al. propose, humans happen to be limited to a single iteration for

[13] This property of IRMs persists even if one assigns different costs to different messages, e.g. by message length, as has been suggested for RSA (see Scontras et al. 2018).
[14] See Fox 2007, 2014 for an answer to a parallel question that arises when one adopts a maxim-based approach of the sort envisioned by Grice.

some reason. Differently from Vogel et al., this single iteration cannot be probabilistic, since that would incorrectly lead to identification in scenarios C and C', as discussed above. (Note that identification in Scenario A would be possible only with '*Exh*(banana)'.) A third possibility relates to a possible distinction between formal alternatives used by *Exh* and *only* in grammar on the one hand and the alternatives that people consider in pragmatic reasoning (when reasoning about each other's communicative goals and strategies) on the other hand. The alternatives used by *Exh* and *only* in grammar have been argued to be rather restricted (Rooth 1985, 1992, Fox and Katzir 2011). In our earlier presentation we assumed that the set of messages in an IRM can be similarly restricted. This assumption, which is shared with earlier work such as Frank & Goodman 2012, Stiller et al. 2011, and Vogel et al. 2014, can be questioned. It is possible that during pragmatic reasoning humans actually have access, by necessity, to all the sentences in their language as potential messages. (See Fox 2007, 2014, Fox and Katzir 2011 for conceptual and empirical motivation.) In this case, there will always be a message that will lead to identification already in the first iteration of an IRM, with the consequence that no further iterations could lead to any enrichment of meaning. Other possibilities can presumably be imagined as well. Whatever the correct explanation, and whatever is ultimately concluded about whether IRMs are part of the human cognitive system, we are left with the conclusion that when *Exh* does not yield a strengthening – as in scenarios B, C, and C' above – there is no other mechanism that can bypass *Exh* and yield the same type of strengthening.

References

Bar-Lev, Moshe E. and Danny Fox. 2020. Free choice, simplification, and Innocent Inclusion. *Natural Language Semantics* 28, 175–223.

Benz, Anton, and Robert Van Rooij. 2007. Optimal assertions, and what they implicate. A uniform game theoretic approach. *Topoi* 26(1), 63-78.

Benz Anton. 2006. Utility and Relevance of Answers. In Benz A., Jäger G., van Rooij R. (eds). *Game Theory and Pragmatics* (pp. 195-219). Palgrave Macmillan, London.

Bergen, Leon and Goodman, Noah D. 2015. The strategic use of noise in pragmatic reasoning. *Topics in cognitive science* 7(2), 336-350.

Champollion, L., Alsop, A., and Grosu, I. 2019. Free choice disjunction as a rational speech act. *Proceedings of SALT 29*, 238-257.

Chierchia, Gennaro. 2004. Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In A. Belletti (ed.), *Structures and Beyond 3*. Oxford: Oxford University Press, 39–103 .

Chierchia, Gennaro, Danny Fox, and Benjamin Spector. 2012. Scalar implicature as a grammatical phenomenon. In P. Portner, C. Maienborn, and K. von Heusinger (eds.), *Semantics: An international handbook of natural language meaning*, volume 3, 2297–233, Mouton de Gruyter.

Fox, Danny. 2007. Free choice disjunction and the theory of scalar implicatures. In U. Sauerland and P. Stateva (eds.), *Presupposition and implicature in compositional semantics*, 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, 5-1.

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

Fox, Danny and Roni Katzir. 2021. *Notes on iterated rationality models of scalar implicatures.* To appear, *Journal of Semantics*.

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

Frank, Michael C., Andrés G. Emilsson, Benjamin Peloquin, Noah D. Goodman, and Christopher Potts. 2016. *Rational speech act models of pragmatic reasoning in reference games*. Unpublished Ms.

Franke, Michael. 2009. *Signal to act: Game theory in pragmatics*. Amsterdam: Institute for Logic, Language and Computation.

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

Franke, Michael and Leon Bergen. 2020. Theory-driven statistical modeling for se- mantics and pragmatics: A case study on grammatically generated implicature readings. *Language* 96(2), e77–e96.

Gazdar, Gerald. 1979. *Pragmatics: Implicature, presupposition, and logical form*. New York: Academic Press.

Goodman, Noah. D., and Andreas Stuhlmüller. 2013. Knowledge and implicature: Modeling language understanding as social cognition. *Topics in cognitive science* 5(1), 173-184.

Groenendijk, J. and Stokhof, M. 1984. *Studies in the Semantics of Questions and the Pragmatics of Answers*. PhD thesis, Universiteit van Amsterdam, Amsterdam.

Horn, Laurence R. 1972. *On the semantic properties of logical operators in English*. UCLA PhD thesis.

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

Krifka, M. 1995. The semantics and pragmatics of polarity items. *Linguistic Analysis* 25, 1–49.

Landman, F. 2000. *Events and Plurality: The Jerusalem Lectures*. Kluwer Academic Publishers, Dordrecht.

Magri, Giorgio. 2009. A theory of individual-level predicates based on blind mandatory scalar implicatures. *Natural Language Semantics* 17(3), 245–297.

Magri, Giorgio. 2011. Another argument for embedded scalar implicatures based on oddness in downward entailing environments. *Semantics and Pragmatics* 4(6), 1–51.

Rooth, Mats. 1985. *Association with Focus*. PhD thesis, University of Massachusetts, Amherst.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1, 75–116.

Rosenberg, Seymour, and Cohen, Bertram D. 1964. Speakers' and listeners' processes in a word-communication task. *Science* 145(3637), 1201-1203.

Rothschild, Daniel. 2013. Game theory and scalar implicatures. *Philosophical Perspectives* 27, 438-478.

Sauerland, Uli. 2004. Scalar implicatures in complex sentences. *Linguistics and Philosophy* 27(3), 367–391.

Scontras, Gregory, Michael Henry Tessler, and Michael Franke. 2018. Probabilistic language understanding: An introduction to the Rational Speech Act framework. Retrieved 2021-7-8 from https://www.problang.org.

Sikos, Les, Noortje J. Venhuizen, Heiner Drenhaus, and Matthew W. Crocker. 2021. Reevaluating pragmatic reasoning in language games. *PloS one* 16(3), e0248388.

Stiller, Alex, Noah Goodman, and Michael C. Frank. 2011. Ad-hoc scalar implicature in adults and children. *Proceedings of the 33rd annual meeting of the Cognitive Science Society*, 2134-2139.

Stiller, Alex, Noah Goodman, and Michael C. Frank. 2015. Ad-hoc implicature in preschool children. *Language Learning and Development* 11(2), 176-190.

van Rooij, Robert. 2010. Conjunctive interpretation of disjunction. *Semantics and Pragmatics* 3, 1–28.

Vogel, Adam, Andreas G. Emilsson, Michael C. Frank, Dan Jurafsky, and Christopher Potts. 2014. Learning to reason pragmatically with cognitive limitations. *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, 3055–3060.