Iterated Rationality Models and Conjunctive Readings of Disjunctions

Background. Scalar Implicatures (SIs), such as the inference from 'some' to 'some but not all' in (1), have been extensively studied in recent decades, both experimentally and theoretically.

(1) John did some of the homework ⇝ John did some but not all of the homework

In recent work, Fox & Katzir (2021; henceforth F&K21) present an argument in favor of a grammatical approach to SIs over a pragmatic approach based on conjunctive readings of disjunctive sentences. Our work presents a novel perspective on the initial (naive) speaker in this approach, that challenges their argument, thus contributing toward its evaluation as theoretical framework that account for SIs. **Conjunctive Readings of Disjunctions (CRDs).** Disjunction gives rise to a conjunctive interpretation in multiple configurations where the alternatives are not closed under conjunction, such as Warlpiri Connectives (Bowler 2014) and Free Choice inferences (Kamp 1974) like (2).

(2) John is allowed to eat an apple or a banana.

a. ⇝ John is allowed to eat an apple b. ⇝ John is allowed to eat a banana

Such inferences have been argued to be SIs (Kratzer & Shimoyama 2002, Alonso-Ovalle 2005) and have served as a central criterion for comparing two main approaches: (a) the *pragmatic approach*, according to which SIs are a pragmatic phenomenon that arises at the speech act level, driven by conversational principles (Horn 1972, Grice 1989, a.o); (b) the *grammatical approach*, according to which SIs are logical entailments derived compositionally within grammar, usually by a covert exhaustivity operator (Fox 2007, Chierchia et al. 2012, Bar-Lev and Fox 2017, a.o). While accounted for under the grammatical approach, CRDs were a major problem for early pragmatic theories, and as such were a central argument for the former in the literature (Fox 2007). However, a prominent body of work on *Iterated Rationality Models* (IRMs; following F&K21) changed this view, and offered a pragmatic derivation of these inferences (Franke 2009, Van Rooij 2010). **Iterated Rationality Models (IRMs).** The IRM approach bundles a variety of models that incorporate iterative process of pragmatic and probabilistic reasoning, including the Rational Speech Act model (Frank & Goodman 2012), Iterated Best Response (Franke 2009) a.o. Unlike previous pragmatic theories, some IRMs derive the conjunctive reading in cases like (2) (see F&K21). Simplifying somewhat, under such IRMs the speaker starts by assigning probabilities to every possible message $m \in M$, given the epistemic state they have in mind s, $P(m|s)$. These are computed based on a *naive speaker assumption*:

(3) if n messages make s true, then: $P(m|s) = \begin{cases} 1/n & m \text{ makes } s \text{ true} \\ 0 & \text{otherwise} \end{cases}$

Then, the listener has to compare the likelihoods, $P(s|m)$. Using Bayesian reasoning under the assumption that all prior probabilities $P(s)$ are flat, the listener instead compares $P(m|s)$ for all messages. A message identifies a state if its conditional probability is the strict maximum compared to all other messages. The identification process is iterative – in each iteration the listener pairs together messages and states according to this criterion. After being identified, those messages and states are eliminated, and the next iteration is applied to the remaining messages and states. Once convergence has been reached, or if no identification and elimination is possible, the process ends. **IRMs and CRDs: 2 disjuncts.** The IRM above achieves the desideratum in the simple case of 2 disjuncts: considering a 2-disjunct sentence 'A or B' where the possible epistemic states are all conjunctive combinations, the naive speaker's probabilities are (3). Based on them, in the first step the listener identifies the alternative messages 'A' and 'B' with $A \wedge \neg B$ and $\neg A \wedge B$ respectively. After eliminating these messages and states, 'A or B' is the only remaining message, as is the state $A \wedge B$, so they are necessarily paired together. **Beyond 2 disjuncts.** Van Rooij (2010), Franke (2011) and later F&K21 observe that the IRM fails with what appears to be a simple extension – constructions involving more than 2 disjuncts, like (5). Considering the 3-disjunct counterpart, 'A or B or C', the probabilities assigned by the naive speaker are as in (6). Based on these probabilities, in the first step the listener identifies the alternative messages 'A', 'B', 'C' with $A \wedge \neg B \wedge \neg C$, $\neg A \wedge B \wedge \neg C$ and $\neg A \wedge \neg B \wedge C$ respectively. After eliminating these messages and states, the probabilities are as in (7). The remaining states involve tie for the first place by multiple messages. Hence, in the second step, no identification is possible and the procedure ends.

(4)

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

(5) John is allowed to eat an apple, a banana, or cherries

a. $\rightsquigarrow$ J is allowed to eat an apple b. $\rightsquigarrow$ J is allowed to eat a banana c. $\rightsquigarrow$ J is allowed to eat cherries

F&K21 observe that this IRM fails twice: neither the full message ‘A or B or C’ nor the partial 2-way disjunctions identify any state. These failures are shared by all existing IRMs and attempts to fix them at most address the former only (van Rooij 2010, Franke 2011). F&K21 conclude that these results constitute an argument in favor of the grammatical approach over IRMs. **IRM with weighted probabilities.** The question underlying this work is whether a more nuanced IRM can address the challenge facing the existing IRMs. Specifically, it proposes to re-examine the assumption of a naive speaker in (3). Instead of a uniform distribution among all messages compatible with a certain state, we propose using weighted probabilities, which intuitively reflect the degree of overlap between messages and states rather than a binary value indicating only whether they are consistent or not. A weight of a message m given a state s, $w_s(m)$, is defined as the number of m’s deletion alternatives that are consistent with s, $Alt_{del}(s)$, based on Katzir’s (2007) definition of alternatives (the other components remain as in the previous IRM):

(6)

$P(m\|t)$	‘A’	‘B’	‘C’	‘A or B’	‘A or C’	‘B or C’	‘A or B or C’
$A \wedge \neg B \wedge \neg C$	1/4	0	0	1/4	1/4	0	1/4
$\neg A \wedge B \wedge \neg C$	0	1/4	0	1/4	0	1/4	1/4
$\neg A \wedge \neg B \wedge C$	0	0	1/4	0	1/4	1/4	1/4
$A \wedge B \wedge \neg C$	1/6	1/6	0	1/6	1/6	1/6	1/6
$A \wedge \neg B \wedge C$	1/6	0	1/6	1/6	1/6	1/6	1/6
$\neg A \wedge B \wedge C$	0	1/6	1/6	1/6	1/6	1/6	1/6
$A \wedge B \wedge C$	1/7	1/7	1/7	1/7	1/7	1/7	1/7

(7)

$P(m\|t)$	‘A or B’	‘A or C’	‘B or C’	‘A or B or C’
$A \wedge B \wedge \neg C$	1/6	1/6	1/6	1/6
$A \wedge \neg B \wedge C$	1/6	1/6	1/6	1/6
$\neg A \wedge B \wedge C$	1/6	1/6	1/6	1/6
$A \wedge B \wedge C$	1/7	1/7	1/7	1/7

(8) $w_s(m) = |\{a : a \in Alt_{del}(m) \wedge a \text{ makes } s \text{ true}\}|$ $\qquad$ $P(m|s) = w_s(m)/\Sigma_{m^* \in M} w_s(m^*)$

In other words, the current speaker, unlike the naive speaker, prefers some messages over others to convey a certain state, since they have a greater overlap with this state in terms of alternatives, thus considered better. With this speaker, the revised IRM derives the desiderata in the 3-disjuncts case, covering both the 2-out-of-3 and 3-out-of-3 messages. Based on the weighted probabilities in (9), in the first step ‘A’, ‘B’, and ‘C’ identify $A \wedge \neg B \wedge \neg C$, $\neg A \wedge B \wedge \neg C$ and $\neg A \wedge \neg B \wedge C$ (resp.). After eliminating these messages and states, the remaining probabilities are as in (10). In the second step, ‘A or B’, ‘B or C’, ‘A or C’ identify $A \wedge B \wedge \neg C$, $\neg A \wedge B \wedge C$ and $A \wedge \neg B \wedge C$ (resp.). After elimination, ‘A or B or C’ is the only remaining message, as is the state $A \wedge B \wedge C$, so they are necessarily paired together. Besides 3-disjunctions, these results generalize to any $n \geq 2$ disjunctions, both for n-out-of-n and k-out-of-n cases, thus significantly broadening the scope of the model’s success. Hence, the grammatical approach has no advantage over this IRM with respect to CRDs. **Implications.** The IRM proposed in this work eliminates F&K21’s argument against IRMs based on CRDs. It however does not alleviate other concerns raised for IRMs. This model is influenced by actual priors, like other IRMs, and therefore faces the challenges pointed out in the literature regarding prior sensitivity in SIs (e.g., F&K21, Cremers et al. 2023). Therefore, the present work renders IRMs a potentially viable theory of SIs, to the extent that other concerns for IRMs, as the one mentioned above, can be eliminated as well.

(9)

$P(m\|t)$	‘A’	‘B’	‘C’	‘A or B’	‘A or C’	‘B or C’	‘A or B or C’
$A \wedge \neg B \wedge \neg C$	1/9	0	0	2/9	2/9	0	4/9
$\neg A \wedge B \wedge \neg C$	0	1/9	0	2/9	0	2/9	4/9
$\neg A \wedge \neg B \wedge C$	0	0	1/9	0	2/9	2/9	4/9
$A \wedge B \wedge \neg C$	1/15	1/15	0	1/5	2/15	2/15	2/5
$A \wedge \neg B \wedge C$	1/15	0	1/15	2/15	1/5	2/15	2/5
$\neg A \wedge B \wedge C$	0	1/15	1/15	2/15	2/15	1/5	2/5
$A \wedge B \wedge C$	1/19	1/19	1/19	3/19	3/19	3/19	7/19

(10)

$P(m\|t)$	‘A or B’	‘A or C’	‘B or C’	‘A or B or C’
$A \wedge B \wedge \neg C$	1/5	2/15	2/15	2/5
$A \wedge \neg B \wedge C$	2/15	1/5	2/15	2/5
$\neg A \wedge B \wedge C$	2/15	2/15	1/5	2/5
$A \wedge B \wedge C$	3/19	3/19	3/19	7/19