Are second language speakers more pragmatically tolerant? Explaining the differences in scalar implicature generation between L2 and L1

Irene Mognon, Amber L. Marree & Petra Hendriks*

Abstract. Children's difficulties with Scalar Implicature (SI) generation have been argued to stem from their tolerance towards pragmatic violations rather than from issues with the inferential process per se (Katsos & Bishop 2011). Ternary judgment tasks have been used to support this view. In these tasks, when presented with underinformative sentences, children, as well as adults, choose an intermediate option between acceptance and rejection, thus demonstrating sensitivity to underinformativeness. Some recent studies show that adult second language (L2) speakers also generate SIs at lower rates. In this work, we investigated whether pragmatic tolerance, possibly emerging because of limited language exposure, could explain the difference between (adult) L2 and L1 speakers. Contrary to our expectations, neither our L1 control group nor our L2 groups (L2 High and L2 Low Proficiency) consistently selected the intermediate option when judging underinformative sentences. However, the L2 Low Proficiency group showed a significantly higher tendency to accept underinformative sentences compared to the L1 group. Hence, our results do not support the hypothesis that L2 speakers are more pragmatically tolerant than L1 speakers. However, our findings show that, despite the adoption of a ternary judgment task, low-proficient L2 speakers display a strong tendency to interpret underinformative sentences literally. We argue that this tendency in the L2 can be attributed to the increased cognitive effort involved in SI generation.

Keywords. Scalar implicatures; pragmatics; pragmatic tolerance; L2

1. Introduction. Natural language utterances can often receive more than one interpretation. Consider, for instance, a sentence with the quantifier *some* like the one presented in (1):

(1) Some of my plants need water.

Given its syntactic and semantic features, the literal interpretation of (1) corresponds to (2):

(2) At least some and maybe all of my plants need water.

That (2) is the literal interpretation of (1) can be verified by observing that (1) does not appear to be incompatible with a context in which all the plants need water. Despite this, however, (1) can also be interpreted as in (3):

(3) Not all of my plants need water.

Unlike (2), the interpretation in (3) is argued to arise via Scalar Implicature (SI) generation (Grice 1989, Geurts 2010, Noveck 2018). According to the traditional approach, SIs are inferences based on the linguistic alternatives that the speaker could, but did not use in producing the utterance. Specifically, SI generation can be described as follows: In order to communicate that all the plants need water, sentence (4), rather than (1), should be used.

* Authors: Irene Mognon, University of Groningen and Goethe University Frankfurt (mognon@em.uni-frankfurt.de), Amber L. Marree, University of Groningen (a.l.marree@student.rug.nl) & Petra Hendriks, University of Groningen (p.hendriks@rug.nl).

(4) All of my plants need water.

Thus, when the speaker utters (1) and not (4), (1) can be taken to imply that, according to the speaker, not all plants need water. Enriching (1) with the negation of (4) allows for the interpretation presented in (3), which we will refer to as an SI, to emerge.

To explain the underlying process through which SIs arise, several theoretical accounts have been proposed. On the one hand, we find what can be called the Enrichment Approach: a family of accounts sharing the assumption that SIs emerge via an extra-linguistic, context-dependent pragmatic process based on reasoning about the context, the speaker's epistemic state, and the speaker's intentions (Sperber & Wilson 1986, Carston 2006, Huang & Snedeker 2009, Geurts 2010). Besides the Enrichment Approach, another prominent approach is the Default Approach (Levinson 2000, Chierchia 2004). Substantial differences exist between accounts within this latter family. At the same time, these accounts share the assumption that SIs do not necessarily involve a cognitive cost. Rather, according to the Default Approach, SIs are generated automatically on the basis of semantic rules (Chierchia 2004) or heuristics of language (Levinson 2000).

The Enrichment Approach and the Default Approach offer different predictions regarding SI generation and processing. According to the Enrichment Approach, SIs are secondary interpretations (i.e., they are generated after the literal meaning of the sentence is computed) and require extra processing effort compared to literal meanings. According to the Default Approach, especially according to Levinson's (2000) proposal, SIs are interpreted directly, and canceling the implicature in order to derive the literal interpretation of the sentence ("at least some and maybe all") incurs a cognitive cost.

A wealth of literature has been devoted to shed light on this debate. In the next section, we describe some of the relevant empirical evidence.

1.1. EMPIRICAL EVIDENCE ON SI GENERATION. In the typical SI task (sentence evaluation task with underinformative sentences, see Noveck 2001, Bott & Noveck 2004), participants are presented with *some*-underinformative sentences (e.g., *Some elephants have trunks*). The rejection of such sentences implies that the SI (corresponding to the "not all"-meaning of *some*) has been derived, because if one interprets the sentence as "Not all elephants have trunks", then the sentence is false. On the other hand, acceptance of *some*-underinformative sentences naturally follows from the derivation of the literal interpretation of the sentence.

The available experimental evidence shows that typical adult language users can readily access both the literal interpretation of sentences and their SIs. This is confirmed by the fact that adult individuals are easily capable of interpreting *some* with both its literal meaning ("at least some and maybe all") and with its SI-derived interpretation ("not all") when instructed to do so in an experimental setting (Bott & Noveck 2004, Exp. 1). At the same time, in experimental tasks in which adult individuals are not explicitly instructed to opt for one or the other interpretation, their interpretations distribute bimodally: In many experiments (e.g., Bott & Noveck 2004, Exp. 3), roughly 40% of individuals consistently accept *some*-underinformative statements, thus opting for the literal interpretation; the remaining 60% of participants reject *some*-underinformative sentences, thus demonstrating a preference for the SI-derived meaning. The observed variation in individuals' tendency to generate SIs has been attributed to various factors, including personality traits (Feeney & Bonnefon 2013) and cognitive skills (Khorsheed & Gotzner 2023). Among those, working memory has emerged, at least in some studies, as a significant predictor of SI generation tendencies (Antoniou et al. 2016, Nys et al. 2024, cf. Heyman & Schaeken 2015).

Furthermore, when SIs are derived in a sentence evaluation task, participants' reaction times invariably emerge as significantly slower compared to the reaction times associated with the literal interpretation of different types of sentences (Bott & Noveck 2004, Tomlinson et al. 2013, Spychalska et al. 2016, van Tiel et al. 2019, Ronderos & Noveck 2023). This slower processing is seen as an indication of an additional process taking place after the computation of the literal interpretation of the sentence. Thus, a large body of literature supports the hypothesis that, as the Enrichment Approach suggests, SIs are secondary interpretations and are associated with cognitively effortful processing.

Data from first language acquisition appear to point in the same direction. Despite demonstrating good knowledge of the semantics of quantifiers, when presented with underinformative sentences like *Some elephants have trunks*, 7- to 11-year-old children tend to accept the statements (Noveck 2001). Likewise, when presented with a scenario in which all three out of three horses jump over a fence, 5-year-olds accept underinformative statements such as *Some of the horses jumped over the fence*. Results from both paradigms, therefore, indicate that children strongly prefer the literal meaning of *some* ("at least some and maybe all"). Moreover, these results suggest that the ability to derive SIs develops gradually in language acquisition, and is not yet fully adult-like even in adolescence (Noveck 2001, Porrini 2024). From the perspective of the Enrichment Approach, this may be unsurprising: If it is true that SIs are effortful even for adults, albeit being within the reach of their cognitive skills, children's still-developing cognitive abilities may explain why they do not access SIs early in language acquisition and why they become adult-like only later, as their cognitive system matures. Accounts attributing children's difficulties with SIs to limited cognitive skills or resources have been explicitly offered by Reinhart (2004), Pouscoulous et al. (2007), and Mognon et al. (2021a, 2021b).

Second language (L2) learning represents another interesting testing ground for theories of SIs and, in particular, for the hypothesized cognitive cost associated with SI generation. Being modulated by many factors, L2 processing appears generally less automatic, more effortful, and more reliant on cognitive resources such as working memory (Reichle et al. 2016) than L1 processing. Because of this, if deriving SIs is cognitively effortful, it is expected that L2 speakers derive a reduced rate of SIs compared to L1 speakers. In this regard, however, results are mixed: In some studies, no difference between L1 and L2 has been found (Antoniou et al. 2019). In other studies, fewer SIs were generated by L2 learners compared to L1 speakers (Mazzaggio et al. 2021), and the SI rate was found to be modulated by L2 proficiency (Khorsheed et al. 2022).

Besides difficulties due to limited cognitive skills and resources, a reduced rate of SIs could also stem from other causes. Interestingly, in the realm of language acquisition, other explanations have been proposed for children's non-adult-like SI generation rates, a prominent one being the Pragmatic Tolerance Account (Katsos & Smith 2010, Katsos & Bishop 2011, Katsos 2014). As an account of children's inferential skills, the Pragmatic Tolerance Account is not often discussed in the literature on adult SI generation. However, as we clarify below, we believe this account to be potentially relevant for explaining SI generation in adult L2 speakers. In the remainder of the paper, we present an experimental study aimed at investigating the hypothesis that adult L2 speakers may derive fewer SI interpretations compared to adult L1 speakers, not because of limited cognitive resources, but–as it has been argued for child (L1) speakers–for reasons related to pragmatic tolerance.

1.2. THE PRAGMATIC TOLERANCE ACCOUNT. According to the Pragmatic Tolerance Account (Katsos & Bishop 2011), children appear to generate fewer SIs than adult (L1) speakers because, unlike adults, they are generally more tolerant towards pragmatic violations. The reasoning is as

follows: A sentence like *Some elephants have trunks* is, in fact, not literally false, and children simply are more lenient than adults when judging this type of sentence. Because of this, the methodology customarily used in SI studies (in which participants are asked to either reject or accept *some*-underinformative sentences) misrepresents children's actual pragmatic competence. However, when the appropriate methodology is used, children can perform adult-like, showing their ability to recognize the infelicity of *some*-underinformative sentences.

In support of this claim, Katsos & Bishop (2011) designed a ternary truth-value judgment task. Exactly as in binary truth-value judgment tasks, in ternary truth-value judgment tasks, the task requires evaluating different types of sentences. However, instead of binary response options (rejection vs. acceptance), participants are given a ternary scale including an intermediate option (corresponding to "true but sub-optimal"). Katsos & Bishop (2011) showed that, when asked to judge underinformative *some*-sentences (*The mouse picked up some of the carrots* in a context in which all five out of five carrots have been picked up by the mouse), children and adults performed alike, both overwhelmingly selecting the intermediate option. According to the authors, this result suggests that children are as sensitive to underinformativeness as adults. The difference between children and adults attested in the classical binary tasks is simply due to a different attitude towards pragmatic infelicity.

Importantly, according to Katsos & Bishop (2011), children's disposition to tolerate underinformativeness may stem from their reduced language input. Having had less exposure to language than adults, children may be less confident in their linguistic judgments. This would make them more tolerant and more prone to accept underinformative utterances.

In light of this, could pragmatic tolerance also play a role in adult L2 speakers' SI generation? In other words, do L2 speakers accept underinformative sentences to a larger degree than adult L1 speakers do, and if so, can this be explained by their higher pragmatic tolerance due to their limited L2 exposure? In what follows, we present our study aimed at investigating these research questions.

2. Current study

2.1. Methods

2.1.1. Participants.

Ninety-one adult speakers, all native speakers of Dutch, participated in the experiment. Participants were recruited via a local news bulletin in a small rural town in the Netherlands.

2.1.2. Design and procedure.

The 91 participants were randomly divided in two groups: 43 participants were included in the L1 group and were tested in their L1 (Dutch); 48 participants were included in the L2 group and were tested in their L2 (English).

The SI task was a ternary sentence evaluation task. In each trial, participants heard a recorded sentence and had to rate it as quickly as possible, choosing between three options: *False*, *A bit true*, *True* (for the L2 group) and *Onwaar*, *Een beetje waar*, *Waar* (i.e., the Dutch translations of the English terms, for the L1 group). Following Bott & Noveck (2004), the task included six conditions (see Table 1): the critical *some*-Underinformative condition and five control conditions with the quantifiers *some* and *all* (*some*-True, *some*-False, *all*-True, *all*-False, *all*-FalseAbsurd). Participants saw 6 items for the *some*-Underinformative condition and 2 items for each of the control conditions. The total number of items seen by each participant was, therefore, 16.

The experiment was run as an online questionnaire using the software Qualtrics. To ensure that all participants would correctly understand the task, instructions were given to both groups in their

native language (Dutch). During the experimental phase, the same sentences were presented in Dutch to the L1 group, and in their English translation to the L2 group.

After the SI task, the L2 group was asked to report on their perceived English proficiency on a 7-point Likert scale and on their usage of English in daily life (in reading, speaking, and listening). Furthermore, the L2 group had their L2 English proficiency assessed by means of the standardized Lexical Test for Advanced Learners of English (LexTALE, Lemhöfer & Broersma 2012).

Condition	**Example sentences**		**Expected response**
	English (L2 group)	**Dutch (L1 group)**	
some-Underinformative	*Some cats are mammals.*	*Sommige katten zijn zoogdieren.*	*A bit true* (intermediate option)
some-True	*Some pets are dogs.*	*Sommige huisdieren zijn honden.*	*True* (acceptance)
some-False	*Some insects are lions.*	*Sommige insecten zijn leeuwen.*	*False* (rejection)
all-True	*All lions are mammals.*	*Alle leeuwen zijn zoogdieren.*	*True* (acceptance)
all-False	*All birds are chickens.*	*Alle vogels zijn kippen.*	*False* (rejection)
all-FalseAbsurd	*All ducks are insects.*	*Alle eenden zijn insecten.*	*False* (rejection)

Table 1: Overview of the materials. The expected responses are based on the Pragmatic Tolerance Account

3. Results

3.1. L2 PROFICIENCY. LexTALE scores, calculated using the method recommended by Lemhöfer & Broersma (2012), range from 0 to 100. Table 2 presents the summary statistics for our participants (L2 group only), while Figure 1 visually displays the individual scores. These scores correspond to different levels of the Common European Framework (CEF) for language levels (Council of Europe 2001). As reported in Lemhöfer & Broersma (2012), a LexTALE score below 60 corresponds to the CEF level B1 or lower (lower intermediate user and lower), whereas a score above 80 corresponds to C1/C2 (advanced/proficient user). In light of the wide range of proficiency levels among our participants, we divided the L2 group into two subgroups using a median split. Participants with LexTALE scores of 70 or below were classified as L2 Low Proficiency (N = 25 participants), and those with scores above 70 were classified as L2 High Proficiency (N = 23 participants).

Range	Mean	Median
53.75 - 92.50	71.98	70.00

Table 2: Summary statistics of LexTALE scores (L2 group, N = 48)

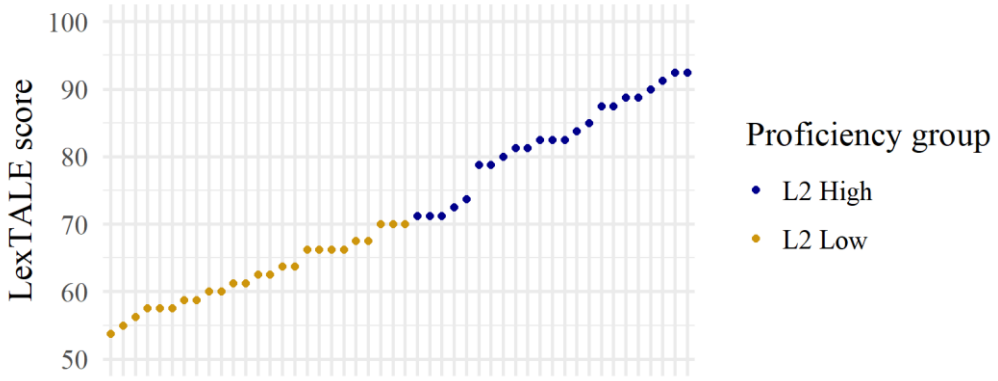

Figure 1: LextTALE score distribution (L2 group, each dot presenting a participant)

Participants' self-reported usage (in the domains of reading, listening, and speaking) and self-rated proficiency showed only weak correlations with the LexTALE score ($\tau_b < .34$). Compared to self-ratings, the LexTALE task has been shown to provide a finer-grained measure of English proficiency (Lemhöfer & Broersma 2012). Thus, we decided to focus exclusively on LexTALE scores and not to consider the other measures further.

3.2. TERNARY SENTENCE EVALUATION TASK. Nine participants (8 from the L2 Low Proficiency group and 1 from the L2 High Proficiency group) gave fewer than 80% correct responses in the control conditions. Therefore, these participants were excluded from further analysis. After excluding these participants, performance on the control conditions was as expected: False sentences were overwhelmingly rejected, true sentences were overwhelmingly accepted, and the intermediate option was hardly ever selected (Fig. 2). The control conditions were, therefore, not analyzed further. Participants' responses in the critical condition *some*-Underinformative are shown in Figure 3.

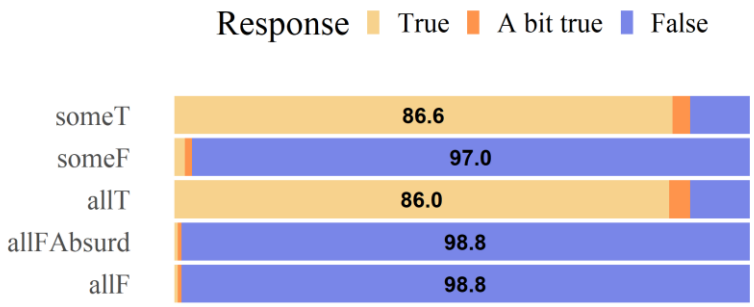

Figure 2: Responses (%) on the five control conditions by all participants (N = 82)

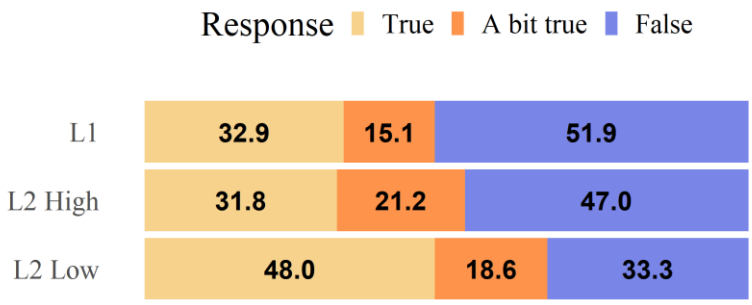

Figure 3: Responses (%) on the *some*-Underinformative condition by the three participant groups (L1 vs. L2 High vs. L2 Low) (N = 82)

Data analysis for the critical *some*-Underinformative condition was conducted in R (R Core Team 2024, R version 4.2.3) and aimed to address three primary objectives. Our first objective was to investigate whether the three groups responded differently when presented with a ternary scale. Therefore, given the ordered nature of our response variable (*True, A bit True, False*), we modeled the cumulative probability of the response falling into a particular level or lower, based on the proficiency group (L1, L2 High Proficiency, L2 Low Proficiency). Our second objective was to examine whether participants' tendency to interpret sentences literally (as indicated by their full acceptance of *some*-Underinformative sentences) varied across proficiency groups. To address this issue, we analyzed our data focusing on the response *True*. Our third objective was to assess whether the probability of displaying pragmatic tolerance (selecting the intermediate option in response to underinformativeness) was influenced by the proficiency group. To address this issue, we focused on the response *A bit true*.

3.2.1. RESPONSES ON THE TERNARY SCALE. First, to determine whether the three participant groups responded differently in the ternary judgment task, we analyzed our data using ordinal mixed-effect regression (R package *ordinal*, Christensen 2019). As outcome variable, we included in our model the ordered variable RESPONSE (*True, A bit True, False)* and as predictor the categorical variable PROFICIENCY with three levels: L1, L2 High Proficiency, and L2 Low Proficiency. Because the model violated the proportional odds assumption (assessed via Harrell's 2001 graphical method), we re-ran the analysis using the function *clmm2* instead of *clmm*. This allowed for the inclusion of a scale effect for the predictor variable PROFICIENCY. We also included a by-participant random effect (note that *clmm2* does not allow for more than one random effect and thus the by-item random effect was not included). The predictor PROFICIENCY turned out not to be significant ($\beta_{\text{HighProficiency}} = 0.25$, $p > .5$; $\beta_{\text{LowProficiency}} = 1.19$, $p > .5$), indicating that the three groups did not respond significantly differently when selecting one of the three response options (acceptance, intermediate option, rejection) in their judgment of *some*-Underinformative sentences.

3.2.2. LITERAL INTERPRETATIONS OF UNDERINFORMATIVE SENTENCES. Our second objective was to determine whether the three participant groups showed a different tendency to interpret sentences literally by opting for the acceptance option (*True*) as opposed to the intermediate (*A bit true*) or the rejection (*False*) options. Therefore, we dichotomized the responses by creating a binary outcome variable with two levels: acceptance (*True* responses) vs. other responses (*A bit true* + *False* responses together). We then used binary logistic regression modeling (R package *lme4*,

Bates et al. 2015) to predict the likelihood of selecting *True* vs. other responses based on the proficiency group (L1 vs. L2 High Proficiency vs. L2 Low Proficiency). Our model, therefore, included PROFICIENCY as the main predictor and the maximum random effect structure allowed by our data (a by-participant random slope for PROFICIENCY and a by-trial random effect). No difference was found between our reference level (the L1 group) and the L2 High Proficiency group ($\beta_{\text{HighProficiency}}$ = -0.09, p = .90). However, a difference emerged between the reference level (the L1 group) and the L2 Low Proficiency group, such that the latter group was significantly more likely to select the *True* response ($\beta_{\text{LowProficiency}}$ = 1.25, p =.03).

3.2.3. INTERMEDIATE RESPONSES TO UNDERINFORMATIVE SENTENCES. For our third objective, we ran a similar logistic regression analysis to assess participants' tendency to select the intermediate response option. Thus, we further dichotomized the responses by creating another outcome variable with two levels: intermediate option (*A bit true* responses) vs. other responses (*True* + *False* responses together). This logistic regression analysis was used to predict the likelihood of selecting *A bit true* vs. one of the other responses based on the proficiency group (L1 vs. L2 High Proficiency vs. L2 Low Proficiency). Thus, we included PROFICIENCY as main predictor and a by-participant random effect (because of convergence issues, no slopes and no by-trial random effects were included). The model showed that the intermediate option was less likely to be selected by the L1 group (our reference level) compared to the other two responses (β = -3.7, p < .001). Importantly, the predictor PROFICIENCY was not significant ($\beta_{\text{HighProficiency}}$ = 0.73, p = .43; $\beta_{\text{LowProficiency}}$ = 0.81, p = .42). This indicates that the three groups did not differ in their preferences towards the intermediate option as opposed to the other choices.

In summary, our ordinal regression analysis suggested that the three groups did not use the ternary scale significantly differently when judging *some*-Underinformative sentences. The binary logistic regression analyses provided us with two additional findings: First, the probability of selecting the acceptance option was modulated by language proficiency; second, the probability of selecting the intermediate option was equally low in the three groups.

4. Discussion and conclusions. With this study, we aimed to investigate the hypothesis that L2 speakers may generate fewer SIs compared to L1 speakers because of a higher tolerance towards pragmatic violations, possibly stemming from their reduced language exposure compared to L1 speakers. Following the Pragmatic Tolerance Account, we expected L1 and L2 speakers to behave alike in our ternary sentence evaluation task and to opt for the intermediate option (*A bit true*) when judging *some*-underinformative sentences. Contrary to our expectations, our L2 participants did not prefer to judge these sentences as "true but sub-optimal" using the intermediate option of the ternary scale. This is in contrast with the hypothesis that L2 speakers might be sensitive to underinformativeness and accept *some*-underinformative sentences in binary tasks only because they are more tolerant towards pragmatic infelicity. In other words, our results do not support the idea that the Pragmatic Tolerance Account can be extended to L2 learning.

Importantly, however, if the Pragmatic Tolerance Account provides a viable account of SI generation in experimental tasks and if adult L1 speakers are fully capable of recognizing the infelicity of *some*-underinformative sentences, we would also expect L1 control speakers to preferentially select the intermediate option when judging underinformative sentences like *Some cats are mammal*s. This, however, was not the case. Like the L2 speakers, the L1 speakers did not prefer the intermediate option; in fact, they selected the intermediate option in only 15% of the cases.

Our findings, we believe, underscore three fundamental issues related to, (1) the role of language proficiency in SI generation, (2) the cognitive costs involved in SI generation and the conflicting findings emerging in previous L2 literature, and (3) the reliability of ternary judgment tasks for the study of pragmatics.

4.1. SCALAR IMPLICATURE GENERATION AND LANGUAGE PROFICIENCY. The first issue warranting discussion concerns the fact that, in spite of the unexpected results in relation to the intermediate response in our study, a significant difference between groups did emerge in our data and was linked to language proficiency. The L2 Low Proficiency group tended to fully accept *some*-underinformative sentences significantly more often than the L1 group; in contrast, the L2 High Proficiency group did not differ from the L1 group. On the one hand, the fact that the L2 High Proficiency group showed a similar pattern of responses compared to the L1 group is not surprising, as the mastery of English of the L2 High Proficiency group, based on the scores of the LexTALE task, was extremely high (equal to or above level C1 for most of the participants). The Dutch participants in the L2 High Proficiency group, in essence, were almost native-like in their L2 English. On the other hand, the L2 Low Proficiency participants accepted *some*-underinformative sentences more than the other two groups, thus displaying a stronger preference for the literal interpretation of such sentences. These results, we believe, align with recent findings showing the modulating effect of language proficiency on the tendency to accept the literal meaning of utterances, as opposed to generating SIs (Khorsheed et al. 2022, Mazzaggio et al. 2021).

Given that our results do not support the Pragmatic Tolerance Account, how can we explain the modulating role of language proficiency on the tendency to interpret sentences literally or pragmatically? As mentioned in the Introduction, according to the Default Approach to SIs, accepting the literal interpretation of *some*-underinformative sentences should come at a cost. Our results do not support this idea, given that the only group that preferentially accepted *some*-underinformative sentences in our experiment was the Low Proficiency group, that is to say, the group of participants whose cognitive resources were arguably more limited. On the other hand, according to the Enrichment Approach, SIs should incur a cognitive cost. Our results support this hypothesis: Participants with fewer cognitive resources available (i.e., participants in the L2 Low Proficiency group) opted for the cognitively less demanding option, that is to say, they interpreted *some*-underinformative sentences literally.

4.2. THE COGNITIVE COST OF SCALAR IMPLICATURE GENERATION. As a second issue, we would like to discuss the cognitive costs of SI generation in the L2 and the mixed findings of previous literature. Previous studies on SIs in the L2 do not unequivocally show that L2 speakers generate fewer SIs compared to L1 speakers. In some experiments, no difference emerged between L1 and L2 groups (e.g., Antoniou & Katsos 2017). On the other hand, Slabakova (2010) found that L2 speakers are more likely to generate SIs than L1 speakers. How to explain the mixed evidence? One important factor seems to be language immersion. In the study of Slabakova (2010), at the time of testing all L2 participants were immersed in an L2 environment, and were attending university courses in their L2. Arguably, therefore, Slabakova's (2010) participants were accustomed in their daily lives to invest extra cognitive resources in L2 processing. This factor may have made them prone to derive SIs, even more than the L1 speakers included in Slabakova's (2010) study. In line with this, as already noted by Mazzaggio et al. (2021), immersion in the L1 is likely to give rise to the opposite effect: Both in Mazzaggio et al. (2021) and in Khorsheed et al. (2022), participants were immersed in their L1 environment, and a reduced rate of SIs was found in the L2 compared

to the L1. Our study confirms this pattern of results by finding a reduced rate of SIs in a group of L2 Low Proficiency participants who were also immersed in their L1 environment.

Besides language immersion, a second factor that could explain the mixed findings in the literature is depth of processing. As suggested by Mazzaggio et al. (2021) and discussed by Khorsheed & van Tiel (2024), the difficulties generally experienced by L2 speakers in SI generation may disappear when participants are given the chance to engage in deeper processing. In experimental paradigms in which participants are put under time constraints or stimuli are presented auditorily, L2 speakers struggle to generate SIs because SI generation imposes a burden on their already reduced cognitive resources. In line with this hypothesis, Khorsheed & van Tiel (2024) demonstrated experimentally that when the task grants participants more time to process the sentences, L2 speakers are able to overcome the difficulties connected to SI generation and show native-like performance. Again, our results are in line with these observations. In our experiment, in which the stimuli were presented auditorily and participants were asked to answer as quickly as possible, L2 speakers (at least those with lower proficiency levels) were more likely to fully accept *some*-underinformative sentences. Arguably, this is because they did not have sufficient time to engage in deeper processing and to go beyond the literal meaning of the utterances.

4.3. TERNARY JUDGMENT TASKS. Finally, our study suggests that the reliability of ternary judgment tasks for gauging the pragmatic-inferential skills of different populations should not be taken for granted. In fact, our adult L1 speakers, despite presumably being able to recognize underinformativeness, failed to select the intermediate response when judging *some*-underinformative sentences. How can we explain their unexpected behavior? Could these puzzling results be due to a flaw in our experimental design?

Our study is not the first SI experiment based on a ternary judgment task in which an unexpected pattern of results emerged. Wampers et al. (2018), for instance, tested the SI generation skills of patients with psychosis and adult L1 controls. The experiment included a task that was virtually identical to the one used in our experiment, namely a sentence evaluation task with underinformative sentences and patently false and true control sentences with *some* and *all*. The adult L1 control group in Wampers et al. (2018) preferentially rejected *some*-underinformative sentences, and both groups failed to show a preference for the intermediate option. Intermediate answers were only given in 22% and 28% of the cases by patients and controls, respectively. Another study in which the intermediate option was hardly ever selected by participants is Schaeken et al. (2021). In this study, using a different design and a 5-point Likert scale, the authors found that the selection rate of the intermediate option as a response to *some*-underinformative sentences was invariably below or equal to 15%. This was true both for the control group (adult L1 language users) and for the group of patients under investigation (i.e., individuals with schizophrenia and other psychotic disorders). This pattern of results (i.e., no preference for the intermediate option) is surprising if one considers that in the original experiment of Katsos & Bishop (2011), the control group of adults was reported to "invariably" select intermediate responses, suggesting a rate of selection of intermediate responses close to 100%.

At this point, it is worth highlighting an important difference in experimental design between the study of Katsos & Bishop (2011) and the other studies mentioned above. Whereas, in line with Wampers et al.'s (2018) study, we used the classical sentence evaluation task (requiring participants to judge sentences like *Some cats are mammals* in isolation, without visual context), Katsos and Bishop (2011) adopted a truth-value judgment task (requiring participants to consider, for instance, a visual scene in which a mouse picked up five out of five carrots and judge the sentence *The mouse picked up some of the carrots*). Speculatively, we suggest that the inclusion of a visual

context in the task motivates participants to evaluate the visual context instead of the sentence. In truth-value judgment tasks, selecting the intermediate option could be the participants' strategy to suggest that a small modification of the visual context (e.g., the mouse picking up four instead of five carrots) could suffice to make the sentence felicitous in that context. The same reasoning is, however, not applicable to sentence evaluation tasks: The very meaning of the words *cats* and *mammals*, for instance, is semantically determined and hence no small modification to the test materials suffices to make the sentence felicitous.

Be as it may, the fact that a preference for the intermediate option cannot be replicated in experiments using different tasks casts doubts on the primary assumption of the Pragmatic Tolerance Account. That is, if language users are sensitive to underinformativeness (which is a prerequisite for SI generation), they are assumed to select the intermediate option to signal their sensitivity towards underinformativeness, while rejecting patently false utterances and accepting patently true utterances. Therefore, we argue that caution should be exercised when using ternary judgment tasks. In particular, the idea that ternary judgment tasks can be used, as is sometimes claimed, as a finer measure of the pragmatic skills of various populations deserves to be reconsidered.

In summary, our study does not bring support to the idea that the Pragmatic Tolerance Account can be extended to SI generation in the L2. However, our findings add to previous research by suggesting that SIs are effortful. In adult L2 speakers, this effort leads to a greater preference for the literal interpretation of sentences when the task does not grant participants sufficient time to engage in deeper processing and when the L2 proficiency is relatively low.

References

Antoniou, Kyriakos, Chris Cummins & Napoleon Katsos. 2016. Why only some adults reject under-informative utterances. *Journal of Pragmatics* 99. 78–95. https://doi.org/10.1016/j.pragma.2016.05.001.

Antoniou, Kyriakos & Napoleon Katsos. 2017. The effect of childhood multilingualism and bilectalism on implicature understanding. *Applied Psycholinguistics* 38(4). 787–833. https://doi.org/10.1017/S014271641600045X.

Antoniou, Kyriakos, Alma Veenstra, Mikhail Kissine & Napoleon Katsos. 2019. How does childhood bilingualism and bi-dialectalism affect the interpretation and processing of pragmatic meanings? *Bilingualism: Language and Cognition* 23(1). 186–203. https://doi.org/10.1017/S1366728918001189.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1). 1–48. https://doi.org/10.18637/jss.v067.i01.

Bott, Lewis & Ira A. Noveck. 2004. Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language* 51(3). 437–457. https://doi.org/10.1016/j.jml.2004.05.006.

Carston, Robyn. 2006. Relevance Theory and the Saying/Implicating Distinction. In: Laurence R. Horn & Gregory Ward (eds.), *The Handbook of Pragmatics*. 633–656. 1st edn. Wiley. https://doi.org/10.1002/9780470756959.ch28.

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

Christensen, Rune Haubo B. 2019. ordinal—Regression Models for Ordinal Data. R package version 2019.12-10. https://cran.r-project.org/package=ordinal.

Council of Europe. 2001. *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Feeney, Aidan & Jean-François Bonnefon. 2013. Politeness and Honesty Contribute Additively to the Interpretation of Scalar Expressions. *Journal of Language and Social Psychology* 32(2). 181–190. https://doi.org/10.1177/0261927X12456840.

Geurts, Bart. 2010. *Quantity implicatures*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511975158.

Grice, Herbert Paul. 1989. *Studies in the Way of Words*. Cambridge: Harvard University Press.

Harrell, Frank E. 2001. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer.

Heyman, Tom & Walter Schaeken. 2015. Some diferences in some: Examining variability in the interpretation of scalars using latent class analysis. *Psychologica Belgica* 55(1). 1–18. https://doi.org/10.5334/pb.bc.

Huang, Yi Ting & Jesse Snedeker. 2009. Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive Psychology* 58(3). 376–415. https://doi.org/10.1016/j.cogpsych.2008.09.001.

Katsos, Napoleon. 2014. Scalar Implicature. In: Danielle Matthews (ed.), *Pragmatic Development in First Language Acquisition*. 183–198. Amsterdam: John Benjamins.

Katsos, Napoleon & Dorothy V.M. Bishop. 2011. Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition* 120(1). 67–81. https://doi.org/10.1016/j.cognition.2011.02.015.

Katsos, Napoleon & Nafsika Smith. 2010. Pragmatic Tolerance or a Speaker-Comprehender Asymmetry in the Acquisition of Informativeness? In: Katie Franich, Kate M. Iserman, and Lauren L. Keil (eds.), *Proceedings of the 34Th Annual Boston University Conference on Language Development, Vols 1 and 2*. 221–232.

Khorsheed, Ahmed & Nicole Gotzner. 2023. A closer look at the sources of variability in scalar implicature derivation: a review. *Frontiers in Communication* 8. https://doi.org/10.3389/fcomm.2023.1187970.

Khorsheed, Ahmed & Bob Van Tiel. 2024. Why second-language speakers sometimes, but not always, derive scalar inferences like first-language speakers: Effects of task demands. *Language Acquisition* 1–19. https://doi.org/10.1080/10489223.2024.2383574.

Khorsheed, Ahmed, Sabariah Md. Rashid, Vahid Nimehchisalem, Lee Geok Imm, Jessica Price & Camilo R. Ronderos. 2022. What second-language speakers can tell us about pragmatic processing. *Plos One* 17(2). e0263724. https://doi.org/10.1371/journal.pone.0263724.

Lemhöfer, Kristin & Mirjam Broersma. 2012. Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior research methods* 44. 325–343.

Levinson, Stephen C. 2000. *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT press.

Mazzaggio, Greta, Daniele Panizza & Luca Surian. 2021. On the interpretation of scalar implicatures in first and second language. *Journal of Pragmatics* 171. 62–75. https://doi.org/10.1016/j.pragma.2020.10.005.

Mognon, Irene, Simone A. Sprenger, Sanne J. M. Kuijper & Petra Hendriks. 2021a. Complex Inferential Processes Are Needed for Implicature Comprehension, but Not for Implicature Production. *Frontiers in Psychology* 11. 1–18. https://doi.org/10.3389/fpsyg.2020.556667.

Mognon, Irene, Simone Sprenger, Sanne Kuijper & Petra Hendriks. 2021b. Balancing the (Horn) scale: explaining the production-comprehension asymmetry for scalar implicatures. In: Alexandros Kalomoiros & Lefteris Paparounas (eds.), *UPenn Working Papers In Linguistics* 27(1). Philadelphia, PA: University of Pennsylvania.

Noveck, Ira A. 2001. When children are more logical than adults: experimental investigations of scalar implicature. *Cognition* 78(2). 165–188. https://doi.org/10.1016/S0010-0277(00)00114-1.

Noveck, Ira A. 2018. *Experimental pragmatics: The making of a cognitive science*. Cambridge: Cambridge University Press.

Nys, Bojan Luc, Wai Wong & Walter Schaeken. 2024. Some scales require cognitive effort: A systematic review on the role of working memory in scalar implicature derivation. *Cognition* 242. 105623. https://doi.org/10.1016/j.cognition.2023.105623.

Porrini, Anna Teresa. 2024. *Cooperative Intentions and Epistemic Reasoning in Scalar Implicature Derivation: A Developmental Perspective*. Trento, Italy: Università degli studi di Trento dissertation. https://hdl.handle.net/11572/410450.

Pouscoulous, Nausicaa, Ira A. Noveck, Guy Politzer & Anne Bastide. 2007. A developmental investigation of processing costs in implicature production. *Language Acquisition* 14(4). 347–375. https://doi.org/10.1080/10489220701600457.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Reichle, Robert V., Annie Tremblay & Caitlin Coughlin. 2016. Working memory capacity in L2 processing. *Probus* 28(1). 29–55. https://doi.org/10.1515/probus-2016-0003.

Reinhart, Tanya. 2004. The Processing Cost of Reference Set Computation: Acquisition of Stress Shift and Focus. *Language Acquisition* 12(2). 109–155. https://doi.org/10.1207/s15327817la1202_1.

Ronderos, Camilo R. & Ira Noveck. 2023. Slowdowns in scalar implicature processing: Isolating the intention-reading costs in the Bott & Noveck task. *Cognition* 238. https://doi.org/10.1016/j.cognition.2023.105480.

Schaeken, Walter, Linde Van de Weyer, Marc De Hert & Martien Wampers. 2021. The Role of Working Memory in the Processing of Scalar Implicatures of Patients With Schizophrenia Spectrum and Other Psychotic Disorders. *Frontiers in Psychology* 12. 1447. https://doi.org/10.3389/FPSYG.2021.635724.

Slabakova, Roumyana. 2010. Scalar implicatures in second language acquisition. *Lingua* 120(10). 2444–2462. https://doi.org/10.1016/j.lingua.2009.06.005.

Sperber, Dan & Deirdre Wilson. 1986. *Relevance: Communication and cognition*. Vol. 142. Cambridge, MA: Harvard University Press.

Spychalska, Maria, Jarmo Kontinen & Markus Werning. 2016. Investigating scalar implicatures in a truth-value judgement task: Evidence from event-related brain potentials. *Language, Cognition and Neuroscience* 31(6). 817–840. https://doi.org/10.1080/23273798.2016.1161806.

Tiel, Bob van, Elizabeth Pankratz & Chao Sun. 2019. Scales and scalarity: Processing scalar inferences. *Journal of Memory and Language* 105. 93–107. https://doi.org/10.1016/j.jml.2018.12.002.

Tomlinson, John M., Todd M. Bailey & Lewis Bott. 2013. Possibly all of that and then some: Scalar implicatures are understood in two steps. *Journal of Memory and Language* 69(1). 18–35. https://doi.org/10.1016/j.jml.2013.02.003.

Wampers, Martien, Sofie Schrauwen, Marc De Hert, Leen Gielen & Walter Schaeken. 2018. Patients with psychosis struggle with scalar implicatures. *Schizophrenia Research* 195. 97–102. https://doi.org/10.1016/j.schres.2017.08.053.