

Article

Rapid Access to Scalar Implicatures in Adjacency Pair Contexts: Experimental Evidence in Spanish

Rodrigo Loredo ^{1,2,*} , Juan E. Kamienkowski ^{2,3} and Virginia Jaichenco ¹ 

¹ Instituto de Lingüística, Facultad de Filosofía y Letras, Universidad de Buenos Aires, Viamonte 430, Buenos Aires C1053ABJ, Argentina

² Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET), Godoy Cruz 2290, Buenos Aires C1425FQB, Argentina

³ Laboratorio de Inteligencia Artificial Aplicada, Instituto de Ciencias de la Computación (ICC), CONICET-Universidad de Buenos Aires, Viamonte 430, Buenos Aires C1053ABJ, Argentina

* Correspondence: rloredo@filo.uba.ar

Received: 7 May 2019; Accepted: 21 June 2019; Published: 27 June 2019



Abstract: A conversational implicature arises when there is a gap between the syntactically and semantically encoded meaning of a sentence and the pragmatic meaning that is inferred in an actual communicative situation. Several experimental studies have approached the processing of implicatures and examined the extent to which the derivation of the pragmatic meaning is effortful, especially in the case of generalized implicatures, where the inferred meaning seems to be the most frequent one. In this study, we present two experiments that explore the processing of scalar implicatures with *algunos* ‘some’ in adjacency pair contexts through an acceptability judgment task and a self-paced reading task. Our results support the claim that the access to the meaning of *some* as *only some* is context sensitive. Moreover, they also indicate that adjacency pair structure contributes to making that meaning rapidly available.

Keywords: generalized conversational implicatures; adjacency pair; relevance theory; sentence gestalt

1. Introduction

1.1. Conversational Implicatures and the Generalized/Particularized Distinction

The term conversational implicature was proposed by Grice (1975) to explain the meaning gap between logical meaning (i.e., the syntactically and semantically encoded meaning of a sentence) and the pragmatic meaning (i.e., the speaker meaning that is inferred in a particular communicative situation). For example, in (1), the sentence meaning of the answer could be paraphrased as (1a) and it could have one of the pragmatic meanings of (1b).

1. — John, are you ok?
— I’m freezing, Mary.
 - a. John is very cold.
 - b. John wants Mary to turn down the air conditioner or turn up the heat.

From a theoretical perspective, Grice recognized two types of conversational implicatures. On the one hand, (1) is a case of a particularized implicature (PCI) because contextual assumptions have a crucial role in sentence comprehension. In order to infer (1b), Mary needs to know if the air conditioner is high or the heat is low or if it is winter or summer. It could also be the case that they were walking in the street and Mary had an extra sweater and John wants Mary to give it to him, or that the temperature

is not the problem, but John is sick. Thus, knowing facts about the situation is decisive to derive the actual meaning of the answer in (1).

On the other hand, there are implicatures where specific lexical or syntactic structures work as triggers for the derivation of an implicature (e.g., quantifiers, numbers, conjunctions). Generalized implicatures (GCI), such as the one in (2), are relatively independent of context and can be generally derived without recurring to it.

2. — Did the packages arrive?

— Some of them did.

a. At least one package arrived (and it could be all of them).

b. Some, but not all packages arrived.

Example (2) is a subtype of GCI called scalar implicatures (Horn 1984). In this example, *some* functions as an implicature trigger and takes the pragmatic meaning of *some but not all* based on the assumption that the speaker is cooperating and, therefore, if she had more information (i.e., that *all* the packages arrived), she would provide it. Quantifiers such as *some* and *all* can be considered to form a scale, in which *all* is more informative or stronger than *some*. If it is assumed that the conversational maxim of quantity is being held (that is to say, that the speaker is being as informative as possible), when a member of the scale is asserted, the higher members of the scale are negated because they are more informative, despite the fact that the logical meaning of the trigger does not codify that negation (*some* is logically implied by *all*, it does not negate it). Implicature triggers can also appear in contexts where the GCI derived meaning is not relevant. For instance, in (3), the *only some* meaning of (3b) is not the most relevant for this context.

3. — What happened?

— Some packages arrived.

a. At least one package arrived (equivalent to ‘packages arrived’).

b. Some, but not all packages arrived.

In this case, the idea that *an indefinite number of packages has arrived* (the logical meaning paraphrased in (3a)) is more relevant for this context. These cases are called lower-bound contexts in opposition to contexts like (2), which are labelled as upper-bound (Horn 1984; Breheny et al. 2006). Thus, in lower-bound contexts, the logical meaning of the GCI trigger will be the preferred one, whereas in upper-bound contexts, the pragmatic one will be the most relevant. It is noteworthy that how and why a context acts as upper or lower bound is still an open question (Dupuy et al. 2016).

The idea that GCIs are derived in the presence of specific triggers and relatively independently to context led to the hypothesis that this type of implicature could be derived by default (Horn 1984; Levinson 2000). Thus, the defaultness approach to GCIs states that in the presence of one of these triggers, the implicated meaning is derived rapidly. If the trigger appears in a lower-bound context, the implicated meaning is also derived, but it is cancelled in a later stage, when the context is integrated to the interpretation process in order to retrieve the logical meaning. Under this approach, GCIs and PCIs differ in their derivation mechanisms; the first undergoes a default process that is triggered by specific words or structures, whereas the latter relies on a non-default process highly constrained by contextual assumptions. There is also another approach that states that the context is integrated early to the derivation (Sperber and Wilson 1995; Carston 1998), and that there are no substantial differences between the derivation of GCIs and PCIs. In the presence of any kind of linguistic stimuli, the context is integrated early and any access to implicatures is driven by the contextual assumptions. In this case, scalar implicatures will only be derived if the context is upper-bound, whereas in lower-bound contexts, they will not. Nevertheless, it is important to notice that not only the time course of the context integration is important for the distinction between PCIs and GCIs, but also the role that context has in the derivation of each kind of implicature. PCIs are characterized as being a case of vagueness,

because the set of possible meanings is, at least initially, indefinite. Conversely, GCIs tend to have one closed set of possible meanings available, and could be treated as a case of ambiguity resolution.

1.2. Experimental Studies of Conversational Implicatures

There have been several experimental approaches to conversational implicatures and, more specifically, to scalar implicatures (cfr. [Noveck and Reboul 2008](#) or [Breheny 2019](#) for a review). The majority of these studies focused on the nature of GCI processing, leaving aside the direct distinction with PCIs, and the main objective of these experiments was to find indicators of the processing cost of GCI triggers in the two contexts presented above.

Two hypotheses linked with the models previously outlined were proposed:

1. If GCIs are processed by default, the processing cost of GCI triggers in upper-bound contexts should be lower than in lower-bound contexts, because the default approach presumes that contextual assumptions are computed later in the derivation process, and that if the default meaning is the preferred one (upper-bound contexts), these assumptions will not add extra effort to the derivation. On the contrary, if the preferred meaning is the logical one (lower-bound contexts), there will be an extra step of implicature cancellation in which contextual assumptions are used. Thus, in this case, the processing cost will be higher.

2. If GCIs are not processed by default, the processing cost of GCI triggers in upper-bound contexts should be higher than in lower-bound contexts, as the guided-by-context approach states that the derivation of the GCI is effortful and it is only made if the context supports it. Because there is no derivation of implicatures in lower-bound contexts and contextual assumptions are not used, there is no extra processing cost.

Most of the early research on scalar implicatures with *some* and disjunctions concluded that the derivation of GCIs is effortful ([Breheny et al. 2006](#); [De Neys and Schaeken 2007](#); [Bott et al. 2012](#)), and that it is only triggered in the presence of contexts in which the enriched meaning is relevant (upper-bound contexts), whereas in contexts in which the logical meaning is relevant (lower-bound), the implicated meaning is not accessed. Nevertheless, contrasting experimental data showed that other types of GCIs do not exhibit the same processing cost ([Huang and Snedeker 2009](#); [Marty et al. 2013](#) for numerals, and [Chemla and Bott 2012](#) for free-choice implicatures). Furthermore, there have even been some studies that provided data in favor of the effortless derivation of the same type of GCIs that had been previously tested ([Grodner et al. 2010](#); [Politzer-Ahles and Fiorentino 2013](#); [Politzer-Ahles and Husband 2018](#) for scalar implicatures with *some*) and in favor of default derivation in maximal neutral contexts ([Barbet and Thierry 2018](#)).

In the particular case of the scalar implicature with *some*, there are two studies with controversial findings. On the one side, in [Breheny et al. \(2006\)](#), the scalar implicature was tested by assessing the reading times for the quantifier *some* in the two contexts mentioned before (lower-bound and upper-bound). They used reported speech stimuli, as can be seen in (4):

4. Upper-bound: Mary asked John whether he intended to host all of his relatives in his tiny apartment. John replied that he intended to host some of his relatives. The rest would stay in a nearby hotel

Lower-bound: Mary was surprised to see John cleaning his apartment and she asked the reason why. John told her that he intended to host some of his relatives. The rest would stay in a nearby hotel.

Their results showed higher reading times at the *some* segment in upper-bound contexts, in line with the guided-by-context approach. Furthermore, they found that the reading times at *the rest* segment were lower in the upper-bound contexts, and used this as a way to assess whether or not the implicature was derived.

On the other side, the replication of this study by [Politzer-Ahles and Fiorentino \(2013\)](#) showed that the issue is still open to debate. They presented the same task with changes in the stimuli, arguing

that the repetition of the noun that is being quantified could be the cause of the delay that [Breheny et al. \(2006\)](#) observed in their experiment. Thus, the repetition of relatives in the example above, where a pronoun would be expected, could be affecting reading times. In their study, the only difference between the two conditions was one word: *any* in lower-bound and *all* in upper-bound contexts:

5. Upper-bound: Mary was preparing to throw a party for John's relatives. She asked John whether *all* of them were staying in his apartment. John said that some of them were. He added that the rest would be staying in a hotel.

Lower-bound: Mary was preparing to throw a party for John's relatives. She asked John whether *any* of them were staying in his apartment. John said that some of them were. He added that the rest would be staying in a hotel.

With these stimuli, their results showed no significant differences in the reading times for the segments with *some*. They used the same method as [Breheny et al. \(2006\)](#) for assessing the realization of the implicature (i.e., measuring reading times for *the rest* segment) and they obtained the same results, with a higher latency in lower-bound contexts. Therefore, their data suggest a context sensitive, but effortless derivation of scalar implicatures.

However, there are two issues that were not addressed. Firstly, there are not sufficient experimental data to give a clear definition about the nature of the distinction between lower and upper bound contexts. Both [Breheny et al. \(2006\)](#) and [Politzer-Ahles and Fiorentino \(2013\)](#) tested that distinction indirectly with the reading time of *the rest* segment of the stimuli, and linked the higher latencies to the fact that in lower-bound contexts, the reference to *the rest* of the quantified noun phrase is not available. Secondly, there is no clear explanation of which factors of the context are actually constraining the reading times for the GCI trigger. As we describe in the next section, the use of an indirect speech structure could be imposing restrictions on the implicature processing that have not been tested yet.

1.3. The Present Study

Our study had three goals. The first goal was to test if the upper-bound contexts of our experiment make the reading of the GCI trigger *some* as *only some* more relevant and if the lower-bound contexts have the opposite effect. To do so, we designed a judgment task to contrast the acceptability of the combination of both contexts with *some* and *only some*. We hypothesize that upper-bound contexts will be acceptable with *some* and *only some*, whereas lower-bound contexts will only be acceptable with *some* and not with *only some*, because in the latter, the only some reading will be less relevant.

The second goal was to rule out the effects that indirect speech structures might impose. For that purpose, we used a self-paced reading task, as in the studies of [Breheny et al. \(2006\)](#) and [Politzer-Ahles and Fiorentino \(2013\)](#), but with some changes in the stimuli. Both of these studies used a text in which the conversation involved was presented in reported speech, as shown in (4) and (5). This kind of structure could be interfering with other processes that can be co-occurrent with implicature processing. In our experiment, we present a dialogical structure with a context, a question, and an answer, using the notion of adjacency pair proposed by [Schegloff and Sacks \(1973\)](#). An adjacency pair is a sequence of two utterances produced by different speakers with a functional relationship between the two parts (e.g., question–answer, offer–acceptance/rejection). As [Levinson \(1983\)](#) noted, adjacency pairs have conditional relevance, that is, given the first utterance, the following utterance is relevant and expectable. Thus, this unit, which works as the fundamental unit of conversation, simplifies the indirect speech structure and provides a reduced, but richer context for the stimuli.

The third goal was to assess how the implicature with *algunos* 'some' behaves in Spanish. It is worth mentioning that the semantics of the quantifier *algunos* 'some' overlaps with the meaning of the indefinite plural article *unos/unas* in lower-bound contexts ([Gutiérrez-Rexach 2001](#)). For example, in Spanish, the dialog in (3) could also be formulated with *unos* instead of *algunos* 'some'.

6. — ¿Qué pasó?

— Llegaron unos/algunos paquetes.

Nevertheless, the main distinction among the two is that only *algunos* ‘some’ can work, as it does in English, as an implicature trigger (Vargas-Tokuda et al. 2009). This raises the question about whether it is possible to access to the logical meaning of *algunos* ‘some’ in lower-bound contexts. If this is the case, we would expect the acceptability of the stimuli with lower-bound contexts + *algunos* ‘unos’ to not show significant differences when compared with upper-bound contexts. To avoid possible confounds, we did not include *unos* as a variable in our experiments.

2. Experiment 1: Context Assessment in an Acceptability Judgment Task

This experiment aimed to test the nature of the relationship between contexts and quantifiers in an adjacency pair structure.

2.1. Method

2.1.1. Participants

A total of 260 Argentinians (143 women and 117 men), native Spanish speakers, aged 18–39 took part in the experiment. All of them were university students or had finished their university education. They were contacted through social media in student groups of the University of Buenos Aires, and they did not receive compensation for their participation. All participants reported that they had been living in Argentina for the last five years and had Spanish as their first language. Their knowledge of other languages was not registered nor evaluated, as the experiment was presented and conducted entirely in Spanish. All participants gave their informed consent for inclusion before they participated in this experiment. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of the Facultad de Filosofía y Letras of the University of Buenos Aires (research project UBACYT 20020160100123BA).

2.1.2. Materials

Twenty quadruples as the one in (7), divided in four lists, were used as critical stimuli. They were composed of three parts: a context that introduced a situation, a question, and an answer to that question. The conditions combined a lower-bound (LB) or upper-bound (UB) context with a quantified phrase with *algunos* ‘some’ (S) or *solo algunos* ‘only some’ (OS). They also included a phrase with *el resto de* ‘the rest of’ that made reference to the quantified elements. In the lower-bound + *some* condition, this phrase was removed because its inclusion would have made the dialog less acceptable. Our stimuli also avoid the repetition of the noun noted by Politzer-Ahles and Fiorentino (2013). Both contexts presented words that were semantically related to the critical stimuli. UB contexts had a noun phrase co-referential to the element quantified in the quantifier segment, and LB contexts had a noun phrase co-referential to another element in the quantifier segment (not the quantified one). Therefore, possible priming effects between words in the context and in the critical segment were balanced across conditions. The number of words was preserved (+/−3), as the number of clauses for the previous contexts and questions. In (7), we present an example (both in Spanish and its translation) of the transformation based on the stimulus of (4) and (5).

7. Upper-bound:

John is cleaning his apartment. His high school mates are coming to the city to visit.

— Why is he cleaning?

— Some/only some friends are going to stay in his house. The rest is going to stay in a hotel.

Juan está limpiando su departamento. Sus compañeros de la secundaria van a venir de visita a la ciudad.

— ¿Por qué está limpiando?

— Van a quedarse en su casa algunos/solo algunos amigos. El resto se va a quedar en un hotel.

Lower-bound:

John is cleaning his house. His children are helping him to mow the grass and tidy up the garden.

— Why is he cleaning?

— Some/only some friends are going to stay in his house. The rest is going to stay in a hotel.

Juan está limpiando su casa. Sus hijos lo están ayudando a cortar el pasto y a ordenar el jardín.

— ¿Por qué está limpiando?

— Van a quedarse en su casa algunos/solo algunos amigos. El resto se va a quedar en un hotel.

We also added two types of control stimuli: highly acceptable stimuli (8) and highly unacceptable stimuli (9).

8. Highly acceptable stimuli (HAS):

Last Friday I bought books for my family.

— What did you buy?

— I bought two novels for my parents and a comic book for my brother.

El viernes compré unos libros para mi familia.

— ¿Qué compraste?

— A mis padres les compré dos novelas y a mi hermano un cómic.

9. Highly unacceptable stimuli (HUS):

The electrician is coming tomorrow.

— On what day is Christmas?

— We'll eat at three o'clock. I'll be here to cook.

Mañana viene el electricista.

— ¿Qué día cae Navidad?

— Comemos a las tres de la tarde. Yo voy a estar para cocinar.

The lists were organized with a Latin square structure to avoid repetitions, creating four lists of 20 critical stimuli (4 in each condition), 20 HAS, and 20 HUS (total $n = 60$). For the analysis, we used the same number of subjects for each of the four lists ($n = 65$). The complete list of stimuli can be found in Supplementary Material S1.

2.1.3. Procedure

The task consisted of indicating how natural the dialogue sounded, using a seven-point Likert scale (0 to 6) with two endpoints (less natural–very natural), as shown in Figure 1. The middle points were tagged only with numbers. This allowed us to treat the responses of the participants as interval, rather than multinomial, data (Coward 1997). The stimuli were presented in full, in three lines. They were fully randomized in blocks to ensure that none of the 260 participants saw the stimuli in the same order and that no condition appeared twice in a row. The context was presented in italics and the rest in regular font. The scale buttons were below the dialog. Between stimuli, we presented a resting screen that was also used to reset the mouse cursor position. The participants had a time limit

of 13,000 ms to complete the judgment of each dialog. Before starting the activity, the participants had to go through four practice dialogues (two HAS and two HUS).

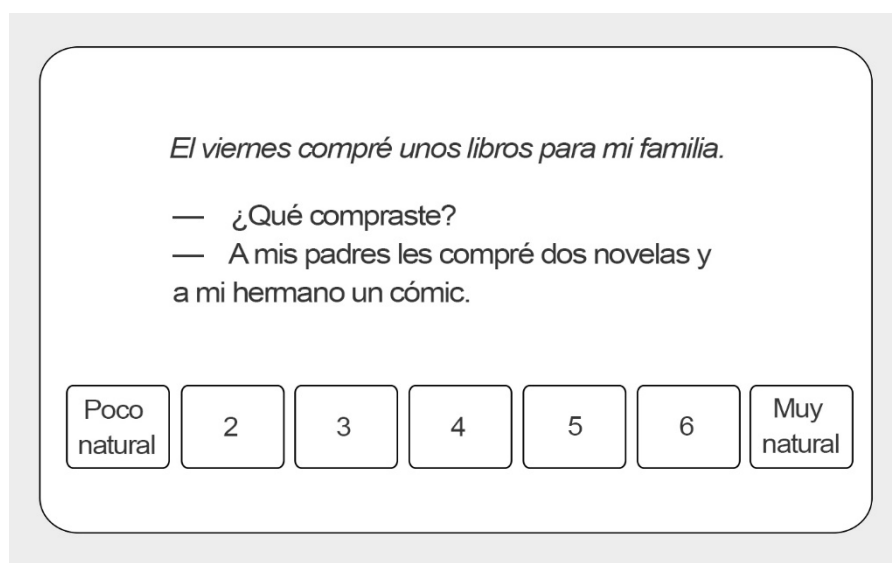


Figure 1. Procedure of experiment 1. Example of the screens that participants saw during one item of the experiment.

The experiment was presented using .html and the JavaScript library JsPsych (De Leeuw and Motz 2016). As the experiment was administered through the Internet, we used different control measures based on previous validated experiments (Sprouse 2011; Enochson and Culbertson 2015). Firstly, the maximum duration of the experiment was limited to 15 min to reduce possible distractions. Furthermore, after each answer, the participant had the chance to rest. Finally, the control stimuli were specially designed to validate the satisfactory completion of the task.

2.1.4. Statistical Analysis

Data points where no answer was given were deleted (284 for critical stimuli, 5.4%; and 248 for control, 4.7%), leaving a total of 4192 data points for critical stimuli and 10,152 for control stimuli. Data were analyzed with R 3.3.3 with {lme4} library for linear mixed models (LMM) (Bates et al. 2015a). A justification for the use of LMM for Likert scales can be found in Norman (2010). However, we also used a non-parametric analysis (Kruskal–Wallis and Wilcoxon tests) and a Generalized Linear Model, which showed similar results. We used a parsimonious model approach following Bates et al. (2015b) with condition as a fixed effect and participant and item as random effects. We used {multcomp} library (Hothorn et al. 2008) for multiple comparisons with the Tukey method. The R code is available in Supplementary Material S2.

2.2. Results and Discussion

2.2.1. Control Items

The acceptability mean for highly acceptable stimuli (HAS) (mean = 5.31, SD = 1.14) was higher than that for highly unacceptable stimuli (HUS) (mean = 0.23, SD = 0.71) (coeff = −5.08, SD = 0.018, $t = -269$). A post-hoc Tukey test showed that this difference is significant ($p < 0.0001$). We regard this as evidence that the experimental task worked, and that participants responded as expected.

2.2.2. Quantifiers

The means of acceptability for dialogs with quantifiers differ for lower-bound context with *only some* in comparison with the other conditions, as shown in Table 1.

Table 1. Mean of acceptability by condition (SD in parenthesis).

LB + OS	LB + S	UB + OS	UB + S
3.09 (2.03)	4.55 (1.70)	4.63 (1.59)	4.66 (1.58)

LB = lower-bound; UB = upper-bound; S = *some*; OS = *only some*.

The LMM and post-hoc Tukey test showed that the difference between LB + OS and the rest of the conditions is significant, whereas there is no difference between the other conditions (cfr. Table 2).

Table 2. Linear mixed models (LMM) and post-hoc Tukey test for acceptability.

Contrast	Standard Error	z	p
LB + OS × LB + S	0.06	−23.2	<0.0001
LB + OS × UB + OS	0.06	−24.31	<0.0001
LB + OS × UB + S	0.06	−24.91	<0.0001
UB + S × LB + S	0.06	1.83	0.2
UB + OS LB + S	0.06	1.24	0.5
UB + OS UB + S	0.59	−0.59	0.9

These results suggest that UB contexts create a relevant quantifiable reference that is compatible with both *some* and *only some*. In both cases, the context facilitates the meaning of *some and not all* for the segment with the quantifiers. On the contrary, LB contexts do not facilitate this interpretation. For LB + S, the acceptability scores suggest that the only relevant meaning that *some* can receive is the one of the existential quantifier (at least one *x*), given that the combination LB + OS is evaluated negatively. If the difference between UB and LB contexts is that the first makes the meaning of *some* as *only some* more relevant and that the second has the opposite effect, the expected result would be an incompatibility between LB contexts and the *only some* meaning of the quantifier. Our acceptability judgment shows that although the incompatibility is a matter of degree (it is not completely unacceptable), the combination of LB + OS is significantly less acceptable than the other conditions. Therefore, given these results, we could claim that in our LB contexts, the most expected (and thus more relevant) interpretation of the quantifier is the logical one, whereas for UB contexts, it is the pragmatically enriched one.

3. Experiment 2: Self-Paced Reading Task

In this experiment, we carried out a self-paced reading task in order to test the predictions presented in the introduction.

3.1. Method

3.1.1. Participants

The participants were 212 Argentinians (112 women and 100 men), native Spanish speakers, aged 18–39. All of them were university students or had finished their university education. They were contacted through social media in student groups of the University of Buenos Aires and they had not participated in the previous experiment. They did not receive compensation for their participation. All participants reported that they had been living in Argentina for the last five years and had Spanish as their first language. Their knowledge of other languages was not registered nor evaluated, as the experiment was presented and conducted entirely in Spanish. All participants gave their informed

consent for inclusion before they participated in this experiment. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of the Facultad de Filosofía y Letras of the University of Buenos Aires (research project UBACYT 20020160100123BA).

3.1.2. Materials

The critical stimuli were the same as those used for the previous experiment without the LB + OS condition. We excluded that condition because it presented differences in the acceptability rating, and the reading times could be affected by these differences. Thus, four lists of 55 dialogues, 15 critical, and 40 fillers were used. For this experiment, *the rest* phrase was included in the LB + S condition in order to compare the reading times for this segment. In this case, the fillers consisted of the 20 HAS stimuli from the previous experiment plus 20 more with the same characteristics (no HUS stimuli were used). Comprehension questions about the last dialog read were presented in 33% of the stimuli, only for fillers, and they were true or false questions that served to check if the participants were reading comprehensively.

3.1.3. Procedure

The task consisted of a self-paced reading with masked segments. At the beginning of each dialog, four red Xs appeared for 2000 ms. Then, they changed to four green plus signs that indicated that the participant could start reading. To move on to the next segment, the participant had to press the spacebar (cfr. Figure 2 for an example of the procedure).

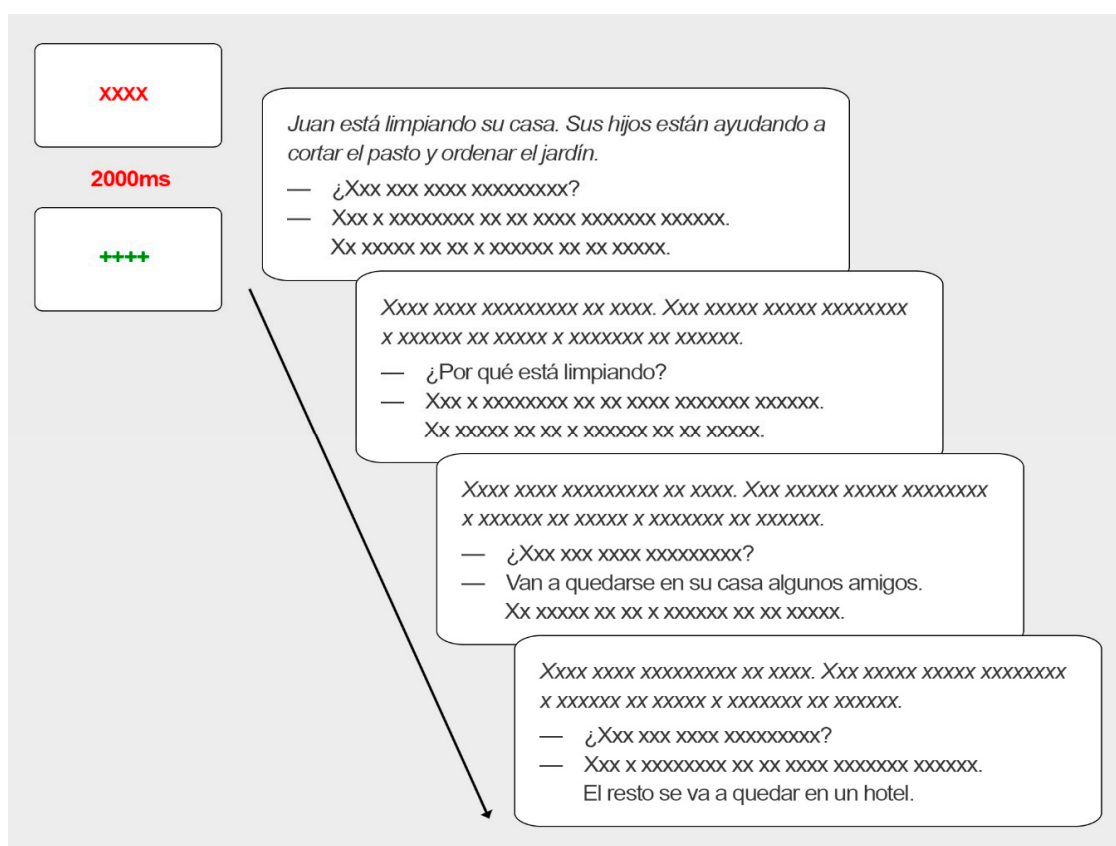


Figure 2. Procedure of experiment 2. Example of the screens that participants read during one item of the experiment.

The stimuli were fully randomized in blocks. The text was presented in a monospaced font and the context was presented in italics. An .html and JavaScript code was used for running the experiment (De Leeuw and Motz 2016). For the comprehension questions, the participants had to press the 'F' key if it was false and the 'J' key if it was true. The reading time for each segment was limited to 13,000 ms and there was no time limit to answer the questions. Before starting the activity, the participants had to read four practice dialogs (two HAS and two HUS in random order) with one comprehension question about the last one.

3.1.4. Statistical Analysis

Data were analyzed with R 3.3.3 with {lme4} library for linear mixed models (LMMs) (Bates et al. 2015a) and {multcomp} library for multiple comparisons with the Tukey method to obtain p -values (Hothorn et al. 2008). Reading times were log-transformed for normality. For the removal of outliers, we followed the model criticism approach suggested by Baayen and Milin (2010). We removed all data points below 200 ms and above 13,000 ms ($n = 56$, 0.3%), leaving a total of 3180 data points for each segment. Then, we fitted the LMMs for each segment and removed all data points with absolute standardized residuals exceeding 2.5 standard deviations. For each segment, less than 4% of the data was removed. For the models, we used the parsimonious model approach. For context, question, and quantifier segments, the models used context (UB vs. LB) as fixed effect, and subject, item, order, and Latin squared number as random intercepts. For the rest segment, the structure of random intercepts was the same, but the fixed effects included the context (UB vs. LB) and the quantifier (*algunos* 'some' vs. *solo algunos* 'only some'). The R code is available in Supplementary Material S2.

3.2. Results and Discussion

3.2.1. Accuracy

Participants answered more than 95% of the comprehension questions accurately. We regard this as evidence that participants were reading comprehensively.

3.2.2. Reading Times: Results Summary

In Table 3, we present the non-transformed reading times in milliseconds for each segment and the condition with standard deviation in parentheses. These values were calculated after the removal of the outliers. The comparisons made are described in detail in the next sections.

Table 3. Summary of results of the self-paced reading task. Reading times in ms. (SD in parentheses).

	Context	Question	Quantifier <i>Some</i>	<i>the Rest + Some</i>	<i>the Rest + only Some</i>
Upper Bound	2996 (1415)	1183 (597)	1405 (586)	1287 (716)	1252 (684)
Lower Bound	3007 (1487)	1187 (604)	1471 (676)	1397 (757)	

3.2.3. Context and Question

The results for context and question showed no significant differences, as expected (context: $SE = 0.01$, $z = -0.05$, $p = 0.95$; question: $SE = 0.005$, $z = -0.6$, $p = 0.5$). This suggests that both UB and LB contexts are equally processed and that there is no additional effort for any of the conditions that could cause a spillover effect over the next segment.

3.2.4. Quantifier Segment

Our results showed that the reading times for *some* in UB contexts are significantly lower than for LB contexts ($SE = 0.004$, $z = -2.52$, $p = 0.01$). These results differ from Breheny et al. (2006), who have found higher reading times for UB contexts, and from Politzer-Ahles and Fiorentino (2013), who have not found any significant differences.

3.2.5. The Rest Segment

For this segment, we obtained the same results as [Breheny et al. \(2006\)](#) and [Politzer-Ahles and Fiorentino \(2013\)](#). The reading times of UB + S and UB + OS showed no significant differences, and reading times for LB + S were significantly higher than for UB + S and UB + OS ($SE = 0.006$, $z = -1.10$, $p = 0.27$ for S vs. OS and $SE = 0.006$, $z = -5.63$, $p < 0.0001$ for UB vs. LB). As in previous experiments, our results suggest that in LB + S, the *only some* meaning of the quantifier is not the one derived. Therefore, reading times for *the rest* are higher because it is necessary to recover the reference for *the rest of the quantified noun*. Conversely, for UB and S/OS, the meaning selected for the quantifier is *some but not all* in both cases (for OS, the meaning is explicitly asserted), and thus the reading times for *the rest* are lower, as the reference is rapidly accessed. These results, combined with the ones from experiment 1, suggest that the UB contexts presented in our experiments are facilitating the inference, whereas the LB contexts are not.

4. General Discussion

In our study, we conducted two experiments that explored the processing of scalar implicatures with *algunos* (Spanish equivalent to ‘some’) in adjacency pair contexts. The novelty in our approach was twofold. On the one side, we tested the implicature in another language, Spanish. Although there is previous research that explores the topic in Spanish ([Vargas-Tokuda et al. 2009](#); [Miller et al. 2016](#)), it focused in acquisition in monolingual Spanish-speaking children and comprehension in L2 learners. To our knowledge, our study is the first in this language to approach implicature processing in adult native speakers and to support, with experimental evidence, the idea that the word *algunos* ‘some’ triggers the same kind of scalar implicature that it does in other languages like English, Greek, and French. Furthermore, using an acceptability judgment task, we tested if the contexts of our stimuli were constraining the interpretation of the quantifiers. Our results showed that dialogues that combined lower-bound contexts with the quantifier *only some* were less acceptable, suggesting that lower-bound contexts made less relevant the meaning of *some* as *only some* under that condition.

On the other side, we proposed a simplification of the stimuli in order to rule out other cues that can interact, to either facilitate or inhibit, with the inference process. The results of the self-paced reading task suggested that with an adjacency pair structure, the derivation of scalar implicatures is context sensitive, rapid, and effortless, whereas the retrieving of the logical meaning of *some* is effortful.

In general terms, our results are in line with the predictions of default models and support the idea that there is not a ‘literal first’ interpretation ([Breheny 2019](#)) of the quantifier *algunos* ‘some’ in Spanish.

Our findings could be explained in three different ways. [Degen and Tanenhaus \(2015\)](#) have proposed, in line with context-based approaches, a unique mechanism for the derivation of implicatures, sensitive to different cues that can make inferences appear context-based (effortful and slow) or default (effortless and rapid). This could explain our results in terms of the manipulation of the structure of our stimuli: adjacency pairs rule out the complexity of the reported speech structure and make the contextual cues that influence the processing of the dialog more readily available. The adjacency pair structure, along with the upper-bound context, supports the expectation of a certain answer (that is, the quantified element) and this expectation results in a lower reading time for the answer segment with *some*. In this sense, we argue that the adjacency pair structure that we use in our stimuli offered a more natural, conversational-like input to process, with a higher and more accessible contextual support. In line with this, [Huang and Snedeker \(2011\)](#) proposed a two-route derivation (bottom-up and top-down). Scalar implicatures could be derived by a top-down route when the information structure of the context causes a specific meaning to be expected, even before the appearance of the GCI trigger. In a less supportive context (LB), a bottom-up mechanism is necessary to arrive at the actual meaning of the trigger.

Another alternative is to consider scalar inferences as a case of lexical ambiguity. From this perspective, *some* would have two meanings specified in the lexical entry, and instead of the implicated meaning being derived by a global-inference mechanism, it would be a result of ambiguity resolution.

Seminal research in lexical ambiguity (Simpson 1994) has proposed two models for processing ambiguity, which differ in the role of the disambiguating context in the initial phase of access to meaning. These models may be linked to the implicature models proposed in the introduction. One model, the selective access model, states that context highly constrains the access process, and that only the appropriate meaning of an ambiguous word is accessed. This model is in line with the guided-by-context approach, although it will predict no differences in reading times for quantifiers (such as the study of Politzer-Ahles and Fiorentino 2013). For the other model, the autonomous access model, prior context does not affect the access phase because all of the meanings are accessed exhaustively. However, after the access phase, there is an integration phase where the appropriate meaning of the ambiguous word is selected. For a biased ambiguous word (a word that has one dominant meaning), the context will help in the selection between the options listed in the lexicon. The autonomous model can be considered to be in line with the default approach if it is assumed that the implicated meaning of scalar terms is the dominant one. If it is not the case (i.e., in lower-bound contexts), pragmatic processes would be involved in the cancellation of the dominant meaning and a further pragmatic derivation would be necessary to access the subordinate meaning. Nevertheless, it is not clear to what extent the implicated meaning is the dominant one. One possibility is that most frequent meanings are more dominant in terms of lexical access. The meaning of *some* could be biased towards the *only some* interpretations, if this is the most frequent use. This could be the case in Spanish where, as we have mentioned, the indefinite plural article (*unos/unas*) alternates with *algunos* 'some'. A possible objection to the dominance of the so-called derived meaning of *some* is the logical contradiction that arises if *at least one* is considered as a subordinated meaning. However, it could be the case that the derivation of strengthened meanings from scalar terms is not actually a restriction from a looser meaning to a more restrictive one, but a case of lexical broadening (Wilson and Cartson 2007), where the derivation starts at a restrictive meaning and extends its meaning to other loose uses. A recent study (Barbet and Thierry 2018) showed that the *not all* meaning of *some* played a role in a stroop-like task even when it was not necessary for the task itself, suggesting that the lexical item *some* could be ambiguous in the mental lexicon and that the *not all* interpretation is the preferred one in the absence of special circumstances.

A final explanation for our data could be modeled in terms of an extended sentence gestalt (SG) comprehension model (McClelland et al. 1989; St. John and McClelland 1999; Rabovsky et al. 2018). Under this explanation, which is related to the top-down/bottom-up approach, UB and LB contexts and the adjacency pair structure create a framework that acts as a representation of an event that is occurring and that includes expectations or predictions about what is coming next. For UB contexts, the expectation of encountering the meaning *only some* is higher than the expectations for *at least one*, whereas for LB contexts, it is the opposite.

Assuming that in Spanish, the dominant meaning of *some* is the restricted one (*some but not all*), the probability of encountering the word *algunos* 'some' is lower in LB contexts than in UB contexts, or than the probability of encountering the word *unos*. Therefore, a bigger effort is required to merge the word to the ongoing representation of the sentence gestalt (in a broader sense of what McClelland et al. (1989) meant by SG and that goes beyond sentences and acts over adjacency pairs as language units) when it appears, as broadening its meaning is necessary to make it compatible. This will be reflected in higher reading times for *some* in LB contexts. The same model can explain the higher reading times for *the rest* segment in LB contexts + *some*. In this case, *the rest* segment requires a bigger effort to be adjusted to the SG representation, as it is not expected and not in line with the SG representation, which treated *some* as *at least one and it could be all*.

In contrast to default explanations, this model does not claim that there is a backtracking or cancellation of meaning. Instead, there is only an adjustment of an SG representation that is being built online. In these terms, the cancellation of an implicature does not involve 'going back' and reanalyzing previous representations, but an effort to adjust the evidence provided by the new input to the current representation (and not the other way around).

This approach to implicature processing could also explain the theoretical distinction between PCIs and GCIs without appealing to two differentiated mechanisms and, what is more, contribute to a better understanding of how context modulates the relevance of newer inputs.

Under a model that combines relevance (Sperber and Wilson 1995) and a maximal incremental account of communication (Altmann and Mirkovic 2009), context yields not only assumptions (factual propositions), but also expectations about the following inputs (predicted information/structure). Thus, not only is the context merged into the processing after the new stimuli are presented, but it creates in advance a set of expected characteristics in the utterances that will follow. These expectations could be syntactic/structural (e.g., after a question an answer is expected) or semantic (the answer will be semantically related to the question, e.g., if we ask *what happened*, an unknown event must be described). When a new utterance appears, the processor assumes that this new input needs to fit into the expectations (communicative principle of relevance), and to do so, the process will follow the optimal path to meet those expectations (cognitive principle of relevance). This explains why GCIs are rapid and effortless when the context provides clear expectations. For example, in scalar implicatures, if what is expected is *only some* and the input is *some*, the predicted meaning of *some* is *only some*, but if the context does not create that expectation or creates expectations of a different kind, the meaning of *some* will need to be defined a posteriori, resulting in the extra effort reflected in the reading time latency of our experiment. Nevertheless, this proposal still gives rise to some unanswered questions, such as when these expectations are created and what aspects of the context provide the cues to create them, so further studies to analyze the role of prediction in the processing of conversational implicatures are needed.

To conclude, in our two experiments, we explored the processing of scalar implicatures in adjacency pair contexts with *some*. As in previous studies, our results showed that the access to the meaning of *some* as *only some* is context sensitive. However, our results suggested that the adjacency pair structure that we proposed helped to make that meaning more rapidly available.

Supplementary Materials: The following are available online at <https://doi.org/10.17605/OSF.IO/FW7Q8>. File S1: Stimuli used in experiment 1 and 2. File S2: R code and data for experiment 1 and 2.

Author Contributions: Conceptualization: R.L. Formal analysis: R.L., J.E.K. Writing original draft: R.L., V.J., J.E.K. Funding Acquisition: V.J.

Funding: This research was partially funded by UNIVERSIDAD DE BUENOS AIRES with the UBACYT project number 20020160100123BA, and CONICET with a PhD scholarship awarded to R.L.

Acknowledgments: We thank María Elina Sánchez, Julia Carden, and P. H. for their thoughtful reading and comments.

Conflicts of Interest: The authors declare no conflict of interest. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

- Altmann, Gerry T. M., and Jelena Mirkovic. 2009. Incrementality and prediction in human sentence processing. *Cognitive Science* 33: 583–609. [CrossRef] [PubMed]
- Baayen, R. Harald, and Petar Milin. 2010. Analyzing Reaction Times. *International Journal of Psychological Research* 3: 12–28. [CrossRef]
- Barbet, Cécile, and Guillaume Thierry. 2018. When some triggers a scalar inference out of the blue. An electrophysiological study of a Stroop-like conflict elicited by single words. *Cognition* 177: 58–68. [CrossRef] [PubMed]
- Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015a. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67: 1–48. [CrossRef]
- Bates, Douglas, Reinhold Kliegl, Shravan Vasishth, and R. Harald Baayen. 2015b. Parsimonious mixed models. *arXiv*.

- Bott, Lewis, Todd M. Bailey, and Daniel Grodner. 2012. Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language* 66: 123–42. [[CrossRef](#)]
- Breheny, Richard. 2019. Scalar implicatures. In *The Oxford Handbook of Experimental Semantics and Pragmatics*. Edited by C. Cummin and Napoleon Katsos. Oxford: Oxford University Press. [[CrossRef](#)]
- Breheny, Richard, Napoleon Katsos, and John Williams. 2006. Are generalized scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition* 100: 434–63. [[CrossRef](#)]
- Carston, Robyn. 1998. Informativeness, relevance and scalar implicature. In *Relevance Theory, Applications and Implications*. Edited by Robyn Carston and Seiji Uchida. Amsterdam: John Benjamins, pp. 179–238.
- Chemla, Emmanuel, and Lewis Bott. 2012. Processing: Free choice at no cost. In *Logic, Language and Meaning*. New York: Springer, pp. 143–49.
- Cowart, W. 1997. *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. Thousand Oaks: Sage Publications.
- De Leeuw, Joshua. R., and Benjamin A. Motz. 2016. Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behavior Research Methods* 48: 1–12. [[CrossRef](#)]
- De Neys, Wim, and Walter Schaeken. 2007. When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental Psychology* 54: 128–33. [[CrossRef](#)]
- Degen, Judith, and Michael K. Tanenhaus. 2015. Processing scalar implicature: A constraint-based approach. *Cognitive Science* 39: 667–710. [[CrossRef](#)]
- Dupuy, Ludivine, J. Van der Henst, and Anne Reboul. 2016. Context in generalized conversational implicatures: The case of some. *Frontiers in Psychology* 7: 381. [[CrossRef](#)] [[PubMed](#)]
- Enochson, Kelly, and Jennifer Culbertson. 2015. Collecting psycholinguistic response time data using Amazon mechanical Turk. *PLoS ONE* 10: e0116946. [[CrossRef](#)]
- Grice, H. Paul. 1975. Logic and conversation. In *Syntax and Semantics 3: Speech Acts*. Edited by Peter Cole and Jerry Morgan. New York: Academic Press, pp. 41–58.
- Grodner, Daniel J., Natalie M. Klein, Kathleen M. Carbary, and Michael K. Tanenhaus. 2010. “Some,” and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition* 116: 42–55. [[CrossRef](#)] [[PubMed](#)]
- Gutiérrez-Rexach, Javier. 2001. The semantics of Spanish plural existential determiners. *Probus* 13: 113–54. [[CrossRef](#)]
- Horn, Laurence R. 1984. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In *Meaning, Form, and Use in Context: Linguistic Applications*. Edited by Deborah Schiffrin. Washington, DC: Georgetown University Press, pp. 11–42.
- Hothorn, Torsten, Frank Bretz, and Peter Westfall. 2008. Simultaneous Inference in General Parametric Models. *Biometrical Journal* 50: 346–63. [[CrossRef](#)] [[PubMed](#)]
- Huang, yi Ting, and Jesse Snedeker. 2009. Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology* 58: 376–415. [[CrossRef](#)] [[PubMed](#)]
- Huang, yi Ting, and Jesse Snedeker. 2011. Logic and conversation revisited: Evidence for a division between semantic and pragmatic content in real-time language comprehension. *Language and Cognitive Processes* 26: 1161–72. [[CrossRef](#)]
- Levinson, Stephen. 1983. *Pragmatics*. Cambridge: Cambridge University Press.
- Levinson, Stephen. 2000. *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge: MIT Press.
- Marty, Paul, Emmanuel Chemla, and Benjamin Spector. 2013. Interpreting numeral and scalar items under memory load. *Lingua* 133: 152–63. [[CrossRef](#)]
- McClelland, Jay L., Mark St. John, and Roman Taraban. 1989. Sentence comprehension: A parallel distributed processing approach. *Language and Cognitive Processes* 4: 287–335. [[CrossRef](#)]
- Miller, David, David Giancaspro, Michael Iverson, Jason Rothman, and Roumyana Slabakova. 2016. Not just algunos, but indeed unos L2ers can acquire scalar implicatures in L2 Spanish. In *Language Acquisition Beyond Parameters. Studies in Honour of Juana M. Liceras*. Edited by Anahí Alba de la Fuente, Elena Valenzuela and Cristina Martínez Sanz. Amsterdam: John Benjamins, pp. 125–45. [[CrossRef](#)]

- Norman, Geoff. 2010. Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education* 15: 625. [CrossRef]
- Noveck, Ira, and Anne Reboul. 2008. Experimental Pragmatics: A Gricean turn in the study of language. *Trends in Cognitive Sciences* 12: 425–31. [CrossRef] [PubMed]
- Politzer-Ahles, Stephen, and Robert Fiorentino. 2013. The realization of scalar inferences: Context sensitivity without processing cost. *PLoS ONE* 8: e63943. [CrossRef] [PubMed]
- Politzer-Ahles, Stephen, and E. Matthew Husband. 2018. Eye movement evidence for context-sensitive derivation of scalar inferences. *Collabra: Psychology* 4: 3. [CrossRef]
- Rabovsky, Milena, Steven Hansen, and J. L. McClelland. 2018. Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nature Human Behaviour* 2: 693–705. [CrossRef]
- Schegloff, Emanuel A., and Harbey Sacks. 1973. Opening up closings. *Semiotica* 8: 289–327. [CrossRef]
- Simpson, Greg. 1994. Context and the processing of ambiguous words. In *Handbook of Psycholinguistics*. Edited by M. A. Gernsbacher. San Diego: Academic Press, pp. 359–74.
- Sperber, Dan, and Deirdre Wilson. 1995. *Relevance: Communication and Cognition*. Oxford: Blackwell and Cambridge.
- Sprouse, Jon. 2011. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods* 43: 155–67. [CrossRef]
- St. John, Mark, and James L. McClelland. 1999. Learning and Applying Contextual Constraints in Sentence Comprehension. *Artificial Intelligence* 46: 217–57. [CrossRef]
- Vargas-Tokuda, Marissa, John Grinstead, and Javier Gutiérrez-Rexach. 2009. Context and the Scalar Implicatures of Indefinites in Child Spanish. In *Hispanic Child Languages*. Edited by J. Grinstead. Amsterdam: John Benjamins, pp. 93–114. [CrossRef]
- Wilson, Deirdre, and Robyn Cartson. 2007. A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts. In *Pragmatics*. Edited by Noel Burton-Roberts. Basingstoke and New York: Palgrave Macmillan, pp. 230–59.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).