

Resumptive pronouns form on-line dependencies with fillers in English: Evidence from the Maze task

Chung-hye Han, Simon Fraser University and Ewha Womans University, chunghye@sfu.ca

Trevor Block, Simon Fraser University, tblock@sfu.ca

Holly Gendron, Simon Fraser University, gendronholly@gmail.com

Dennis Storoshenko, University of Calgary, dstorosh@ucalgary.ca

Jesse Weir, University of Calgary, jesse.weir1@ucalgary.ca

Sara Williamson, Simon Fraser University, slw28@sfu.ca

Keir Moulton, University of Toronto, keir.moulton@utoronto.ca

We investigated the processing of resumptive pronouns in English relative clauses. We asked whether resumptive pronouns in complex relative clauses formed from islands and non-islands establish dependencies with head nouns in incremental processing. Based on the findings from two reading time experiments using the Maze task with the gender-mismatch paradigm, as well as an off-line comprehension study, we conclude that (i) a dependency is formed with the head noun immediately upon encountering the resumptive pronoun; (ii) dependency formation is easier across islands than across non-islands; (iii) dependency formation privileges the head noun over intervening noun phrases; and (iv) these dependencies are maintained in off-line comprehension. The data suggest that although resumptive pronouns are judged to be unacceptable in English, comprehenders automatically and immediately integrate them using routines similar to filler-gap dependencies. We discuss the implications of our findings for the question of whether resumptive pronouns facilitate filler dependency formation in comparison to gaps, a possible reason why dependencies would be easier to establish across islands than across non-islands, and how the use of resumptive pronouns may benefit listeners as well as speakers.



1. Introduction

Resumptive pronouns (RPs) in English are often produced in long or island-violating *wh*-dependency structures, as confirmed by many corpus and production studies (Ferreira & Swets, 2005; Morgan & Wagers, 2018; Prince, 1990, 1995). For instance, (1a) is an example of a naturally occurring RP in a verb complement clause within a relative clause, collected by Anthony Kroch; and (1b) is an example of an RP contained in a *wh*-complement clause within a relative clause produced in a lab setting.

- (1) a. He's the kind of guy that I think **he** respects the job we did. (Prince, 1995, p. 224)
 b. This is a donkey that I don't know where **it** lives. (Ferreira & Swets, 2005, p. 270)

According to the theoretical syntax literature, RPs aid *wh*-dependency formation in otherwise hard-to-form dependency structures or island-violating structures (Cann et al., 2005; Erteschik-Shir, 1992; Kayne, 1981; Kroch, 1981; Prince, 1990; Ross, 1967). But in acceptability judgment studies, RPs have been shown to (i) have uniformly low acceptability across different types of island-violating *wh*-dependencies, and (ii) decrease the acceptability of *wh*-dependencies across verb complement clauses (Alexopoulou & Keller, 2007; Clemens et al., 2012; Ferreira & Swets, 2005; Han et al., 2012; Heestand et al., 2011; Keffala & Goodall, 2011; Meltzer-Asscher, 2021). However, in self-paced reading studies, RPs have been shown to lead to faster reading times compared to gaps after the RP/gap region (Dickey, 1996; Hammerly, 2022; Hofmeister & Norcliffe, 2013; Morgan et al., 2020).

There are two competing interpretations of the fast reading times of RPs. Under one view, RPs facilitate otherwise difficult-to-form dependencies with fillers in on-line processing. The facilitation approach predicts that RPs enhance comprehension, compared to gaps (Dickey, 1996; Hammerly, 2022; Hofmeister & Norcliffe, 2013). Under the other view, RPs are read fast because they are shallowly processed and not fully integrated. Under this shallow processing approach, RPs should not enhance comprehension (Morgan & Wagers, 2018; Morgan et al., 2020).

In order to address the matter, we first need to establish whether RPs indeed form dependencies with fillers in incremental processing, i.e., at the point readers encounter the pronoun. In this article, using a Maze task with the gender-mismatch paradigm, we present data showing that RPs do very quickly form a dependency with fillers in relative clauses formed into both non-islands (verb complement clauses) and islands (adjunct or *wh*-complement clauses). We also show that dependency formation privileges the head noun over intervening noun phrases, suggesting a mechanism similar, if not identical to, a standard *wh*-filler/gap dependency. Moreover, the data show that filler-RP dependencies are formed more easily across islands than across non-islands, indicating that the processor is sensitive to structural distinctions regarding clause islandhood when integrating RPs. We then present data from a comprehension task showing that sentences containing RPs exhibit enhanced comprehension compared to sentences containing gaps.

1.1 Previous studies

Previous studies have asked whether RPs form dependencies with the head noun filler by looking at how RPs are interpreted. If RPs form *wh*-dependencies with fillers, they should also be interpreted as fillers, and, thus, should facilitate comprehension, as argued by Hammerly (2022), based on a combination of reading and comprehension studies. Morgan et al. (2020), using different types of configurations, provide evidence that RPs do not facilitate comprehension, even though they confer reading-time advantages.

Morgan et al. (2020) used *it*-cleft sentences, such as in (2), to compare reading times of RPs and gaps in both non-islands and islands in *that*-clauses. After each sentence was read, participants were asked the question *Who did what to whom?*, with four choices given as possible answers, as in (3). The TARGET response interpreted the gap/RP as the clefted phrase, and the LOCAL response interpreted the gap/RP as the matrix subject of the *that*-clause (the *intervenor*).

- (2)
- a. It is **Miss Piggy** that **Miss Cat** reported that Mr. Dog poked ___ /**her** with a pencil.
NONISLAND
 - b. It is **Miss Piggy** that **Miss Cat** understood why Mr. Dog poked ___ /**her** with a pencil.
WEAK ISLAND
 - c. It is **Miss Piggy** that **Miss Cat** exercised while Mr. Dog poked ___ /**her** with a pencil.
STRONG ISLAND
- (3) Who did what to whom?
- a. Mr. Dog poked Miss Piggy with a pencil. TARGET
 - b. Mr. Dog poked Miss Cat with a pencil. LOCAL
 - c. Mr. Dog poked Miss Rabbit with a pencil. DANGLE
 - d. Miss Cat poked Miss Piggy with a pencil. BONKERS

Morgan et al. (2020) found that in both non-islands and islands, in the spillover region after the gap/RP, RPs were read faster than gaps. However, in comprehension, RPs elicited fewer TARGET responses and more LOCAL responses than gaps in both non-islands and islands, and RPs in islands had the lowest proportion of TARGET responses. They therefore conclude that there is no evidence that RPs aid in comprehension.

Hammerly (2022) used test sentences with an RP or a gap in a relative clause formed from a non-island (4a) or an island (4b). The RPs were gender-congruent with the fillers (*the spy*), but not with the matrix subjects of the relative clauses (*Beth*). Comprehension questions asked for the interpretation of the gap/RP, with four choices given as possible answers, as in (5).

- (4)
- a. Mary trained **the spy** who **Beth** announced that the agency had recruited ___ /**him** over the summer for the program.
NONISLAND
 - b. Mary trained **the spy** who **Beth** announced which agency had recruited ___ /**him** over the summer for the program.
ISLAND

- | | | |
|-----|------------------------------------|-----------|
| (5) | Who was recruited over the summer? | |
| | a. The spy | FILLER |
| | b. Beth | NONLOCAL |
| | c. The agency | LOCAL |
| | d. I don't know | CONFUSION |

In the self-paced reading task, Hammerly (2022) found that in the spillover region right after the gap/RP, RPs were read slower than gaps in the NONISLAND condition, but faster than gaps in the ISLAND condition. In the comprehension question task, Hammerly reports that numerically, in the ISLAND condition, the sentences with RPs showed a slightly higher proportion of FILLER responses and a lower proportion of CONFUSION responses compared to the sentences with gaps. Taking together the findings that the island-violating structures containing RPs were read faster than gaps and the numerical tendency that the RPs in islands led to more FILLER responses than gaps, Hammerly concludes that RPs enhance the comprehension of otherwise difficult-to-form dependencies.

This conclusion is not fully supported by the data, though, as the observed numerical tendency regarding the proportion of FILLER responses to the comprehension questions does not reach statistical significance. Hammerly notes that the observed numerical tendency might not have reached statistical significance because the FILLER response rates across the board were high and close to ceiling. However, another reason why the FILLER responses were so high in sentences with RPs might have been due to the fact that between the FILLER responses (*the spy*) and the NONLOCAL responses (*Beth*), only the FILLER responses were gender-congruent with the RPs (*him*). Given this possibility, we cannot be sure whether there is a general comprehension advantage for RPs; in these cases, the gender matching alone could be driving the FILLER responses.

In addition to a self-paced reading study, Morgan et al. (2020) present a visual-world eye tracking study as a more direct way of testing the processing profile of RPs. Using *it*-cleft sentences similar to (2), they tested whether RPs in islands and non-islands form dependencies with the clefted phrase (the *target*). They found that RPs reduced looks to the referent corresponding to the clefted phrase (*Miss Piggy*) in comparison to gaps. Moreover, whether in islands or non-islands, in the RP region, looks to the referent corresponding to the clefted phrase (*Miss Piggy*) and looks to the referent corresponding to the subject of the *that*-clause (*Miss Cat*) showed similar frequencies. Morgan et al. (2020) take this to mean that RPs may be fully ambiguous between these two potential referents in on-line processing. They argue that since RPs are ungrammatical, the processing of RPs does not reliably involve dependency formation with fillers, and RPs end up confusing the comprehender. Therefore, when forced to select a referent, comprehenders approach chance in selecting between the discourse entities corresponding to the two potential antecedents.

As noted by Morgan and colleagues, in both self-paced reading and visual-world eye tracking, participants are not necessarily fully integrating the RPs in sentences. Furthermore, the self-paced reading task cannot tell us whether the RP is forming a dependency with the filler or something else in real time. In moving to studies using a Maze task, we test the processing of RPs in a more cognitively demanding task, requiring potentially deeper processing of the sentence (Han et al., 2025; Sloggett et al., 2020). Furthermore, our trial items, described below, differ from those used in these prior studies in crucial ways that allow us to more effectively detect quick dependency formation.

1.2 The present studies

In our studies, we tested whether RPs form a dependency with the head noun in incremental processing and asked how quickly this dependency is formed, using a gender-mismatch paradigm (Garnham, 2001; Sturt, 2003; van Gompel & Liversedge, 2003). We were also interested in whether RPs discriminate between potential NP antecedents and true fillers. To address these questions, we investigated RPs in relative clauses using the Maze task (Boyce et al., 2020; Forster et al., 2009; Freedman & Forster, 1985).

Our test sentences are existential *there was* constructions followed by a noun, referred to as the *head noun*, modified by a *wh*-relative clause. The expected gap position in the relative clause is either in an island (an adjunct island introduced by *because* or a *wh*-complement clause), or in a non-island verb complement clause from which extraction should be grammatical. Instead of gaps, these positions are filled with RPs that (mis)match with the gender of the head noun, as in (6) (used in Experiment 1) and (7) (used in Experiment 2). The matrix subject of the relative clause serves as an intervenor (which we will refer to as the *intervening noun*) between the head noun and the RP. In the stimuli used in Experiment 1, as in (6), the intervening noun is gender-neutral, making it compatible with any gendered RP. Further, the intervening noun is closer to the RP than the head noun and, as such, it is a potential sentence-internal antecedent for the RP in incremental processing. The local subject of the embedded clause containing the RP is always the first person pronoun *I*, which should not be considered as a potential antecedent for the third person RP. With this, there are only two possible antecedents to consider for the RP: the head noun and the intervening noun.¹

- (6) a. There was a **policeman** who a **civilian** was glad because I acknowledged **him/her** after the arrest. ISLAND
- b. There was a **policeman** who a **civilian** had said that I acknowledged **him/her** after the arrest. NONISLAND

¹ Strictly speaking, under the most widespread analysis of these relative clauses, the filler associated with the gap position filled by the RP is the *wh*-phrase. This is, in turn, associated with the head noun (Carlson, 1977; Sauerland, 2000). Setting aside the exact semantic mechanism for this association, we refer to the head noun itself as the filler for discursive ease.

- (7) a. There was a **watchman** who a **policewoman** was glad because I acknowledged **him/her** after the arrest. ISLAND
- b. There was a **watchman** who a **policewoman** had said that I acknowledged **him/her** after the arrest. NONISLAND

In the stimuli used in Experiment 2, both the head noun and the intervening noun are gendered and manipulated to match or mismatch in gender with the RP, as can be seen in (7). Following assumptions in previous studies that used the gender-mismatch paradigm in pronoun processing, we interpret processing difficulty at gender-mismatched RPs as evidence that the processor attempts to establish a dependency (Cunnings et al., 2015; Garnham, 2001; Kazanina et al., 2007; Sturt, 2003; van Gompel & Liversedge, 2003). Thus, if RPs form on-line dependencies with head nouns and not with intervening nouns, we predict processing difficulty – that is, gender-mismatch effects – when RPs mismatch in gender with the head noun, but not when they mismatch with the intervening noun.

The stimuli we tested differ in important ways from Morgan et al. (2020) and Hammerly (2022). In the Morgan et al. (2020) stimuli, there was a third noun phrase, the local subject of the embedded clause (*Mr. Dog* in (2)), whose referent was in a similar semantic domain as the other characters: all three were animals. This local subject should be ruled out as a potential antecedent of the RP through Principle B of the Binding Theory (Chomsky, 1981); in their items, it was also always gender-contrasted, which should act to obviate a local reading when an RP is present. However, its presence still added complexity, with a third character to keep track of. As noted in Meltzer-Asscher (2021), similarity-based interference can hinder the retrieval of the filler, and the addition of another possible antecedent may have added to the complexity of the experiment. Similarly, in Hammerly’s stimuli, the local subject introduced another referent to consider: it was sometimes an organization (*the union, the agency*), but sometimes a profession noun (*the coach, the teacher*) with a clearly human referent. In our stimuli, rather than introducing a new referent as the subject of the lowest clause, and relying on gender or some other semantic/grammatical factor to steer participants away from considering it as a possible antecedent for the RP, we use a first person pronoun *I* in this position. We believe it should be impossible for any participant to consider forming a connection between the RP and *I* in incremental processing.

Another issue raised by Meltzer-Asscher (2021) is the role that gender agreement can play in the interpretation of RPs. She notes that “[i]f more than one noun phrase in the sentence agrees with the RP, interference may arise” (p. 190). This is exactly the situation in the Morgan et al. (2020) items, where both the filler and the intervenor were gender-matched with the RP. In Hammerly’s stimuli, we find the opposite: between the head nouns and the intervening nouns, only the head nouns were systematically gender-congruent with the RPs, as noted above. Hence, participants could use this information to arrive at the filler choice instead of a non-filler

choice, as the RP's gender feature made it uniquely disambiguating. Rather than having gender mismatches deter participants from forming a connection between the RP and a non-filler phrase, or introducing ambiguity through complete gender-matching, our stimuli combine gendered and ungendered nouns, manipulating (mis)matches, to induce gender mismatch effects at a single position in the sentence, creating a clear diagnostic for immediate dependency formation.

Our on-line experiments also use a different task than those discussed above. To guard against shallow processing as a possible strategy and to promote the integration of RPs, Experiments 1 and 2 were deployed using the Maze task (Boyce et al., 2020; Forster et al., 2009; Freedman & Forster, 1985). In the Maze task, the reader must make a choice about which of two words can result in a grammatical continuation of the sentence. This task requires integration of each word into the previous material and, thus, should promote assigning an interpretation to the RP in incremental processing. Similar reasoning about the Maze task discouraging shallow processing is suggested in Han et al. (2025) and Sloggett et al. (2020).

To preview the results: in both Experiments 1 and 2, we found gender-mismatch effects at the RP region when it was mismatched in gender with the head noun. Furthermore, in Experiment 2, no gender-mismatch effects were found at the RP region when it was mismatched with the intervening noun. Moreover, in both experiments, the RPs in islands were read much faster than the ones in non-islands. These results suggest that RPs form on-line dependencies with head nouns in complex relative clauses. Also, as readers were encouraged to integrate each incoming word, the faster reading time of RPs in island relative clauses is unlikely to be due to shallow processing.

Experiment 3 was a comprehension question experiment probing the off-line interpretation of RPs versus gaps in relative clauses formed from both islands and non-islands. Here, both the head noun and the intervening noun were gender-matched with the RP, so as not to bias the participants' responses towards a particular interpretation. We found that, across the board, RPs were interpreted as head nouns, not as intervening nouns, and that in relative clauses formed from islands, the rate at which RPs were interpreted as head nouns was even higher than the rate at which gaps were interpreted as head nouns. These results, taken together with the results from Experiments 1 and 2, suggest that RPs quickly form a dependency with the filler, and this dependency is especially facilitated in islands, resulting in enhanced comprehension.

2. Experiment 1: Gender-mismatch effects of the head noun

Experiment 1 tested whether RPs in relative clauses form on-line dependencies with head nouns, using the gender-mismatch paradigm in a Maze task.

2.1 Participants

100 native English speakers (50 female, 50 male) were recruited through Prolific Academic and completed the experiment on PClbex Farm (Zehr & Schwarz, 2018). Participants ranged in age

from 19 to 65, with a mean age of 36.7. All participants gave their informed consent through PCIBex Farm prior to the experiment and were compensated £3.00 upon completion.

2.2 Task, design, and materials

The version of the Maze task that we used is the G-Maze, where participants read through a sentence by making a forced-choice between a word that grammatically continues the sentence and a distractor that is a real word, but is not a possible grammatical continuation (Boyce et al., 2020; Forster et al., 2009).² The time a participant takes to make their choice is the reaction time (RT). If a participant chooses the distractor, they can then make a different selection, following Boyce et al. (2020).

The Maze task, and G-Maze specifically, has been shown to replicate results in both self-paced reading and eye-tracking (Forster et al., 2009; Witzel et al., 2012). Moreover, the Maze task has been shown to localize effects better than self-paced reading (Boyce et al., 2020; Forster et al., 2009): reading time effects tend to appear at the target word rather than in a spillover region, as is common in self-paced reading. An additional issue with self-paced reading is that it allows participants to move through a sentence without engaging deeply, potentially leading to shallow processing (Christianson et al., 2001; Ferreira & Patson, 2007; Ferreira et al., 2002). The Maze task, by contrast, requires incremental processing, as it forces readers to determine which word integrates grammatically with the preceding string. As noted above, we chose the Maze task to decrease the likelihood that participants would shallowly process RPs, as suggested by Morgan et al. (2018, 2020).

Experiment 1 had a total of 48 items (24 test items, 24 fillers). The test items were constructed in a 2×2 factorial design, illustrated in the sample trial set in (8), repeated from (6). The distractors associated with (8) are given in (9). The distractors in all Maze task experiments reported here were automatically generated using A(uto)-maze, developed by Boyce et al. (2020).³

- (8)
- a. $\text{There}_{/0} \text{ was}_{/1} \text{ a}_{/2} \text{ policeman}_{/3} \text{ who}_{/4} \text{ a}_{/5} \text{ civilian}_{/6} \text{ was}_{/7} \text{ glad}_{/8} \text{ because}_{/9} \text{ I}_{/10} \text{ acknowledged}_{/11} \text{ him}_{/12} \text{ after}_{/13} \text{ the}_{/14} \text{ arrest}_{/15}$. ISLAND-MATCH
 - b. $\text{There}_{/0} \text{ was}_{/1} \text{ a}_{/2} \text{ policeman}_{/3} \text{ who}_{/4} \text{ a}_{/5} \text{ civilian}_{/6} \text{ was}_{/7} \text{ glad}_{/8} \text{ because}_{/9} \text{ I}_{/10} \text{ acknowledged}_{/11} \text{ her}_{/12} \text{ after}_{/13} \text{ the}_{/14} \text{ arrest}_{/15}$. ISLAND-MISMATCH
 - c. $\text{There}_{/0} \text{ was}_{/1} \text{ a}_{/2} \text{ policeman}_{/3} \text{ who}_{/4} \text{ a}_{/5} \text{ civilian}_{/6} \text{ had}_{/7} \text{ said}_{/8} \text{ that}_{/9} \text{ I}_{/10} \text{ acknowledged}_{/11} \text{ him}_{/12} \text{ after}_{/13} \text{ the}_{/14} \text{ arrest}_{/15}$. NONISLAND-MATCH
 - d. $\text{There}_{/0} \text{ was}_{/1} \text{ a}_{/2} \text{ policeman}_{/3} \text{ who}_{/4} \text{ a}_{/5} \text{ civilian}_{/6} \text{ had}_{/7} \text{ said}_{/8} \text{ that}_{/9} \text{ I}_{/10} \text{ acknowledged}_{/11} \text{ her}_{/12} \text{ after}_{/13} \text{ the}_{/14} \text{ arrest}_{/15}$. NONISLAND-MISMATCH

² In contrast, the L-Maze uses nonce word distractors.

³ A(uto)-maze automates the generation of distractors using natural language processing techniques. For all sentences, the distractor for the first word is x-x-x. For each of the subsequent words, a distractor word is generated that matches in length and approximate frequency, but which has a low probability given the preceding context, according to a language model.

- (9) /₀ x-x-x /₁ pre /₂ cent /₃ theorists /₄ laws /₅ yes /₆ estimate /₇ yeah /₈ lake /₉ senate /₁₀ fund /₁₁
 cancellation /₁₂ grow /₁₃ click /₁₄ west /₁₅ expand.

As described above, all test trial items begin with an existential construction, followed by a noun modified by a complex *wh*-relative clause. Inside this relative clause, an RP appears instead of a gap. The first variable is the Gender of the RP, with two levels, MATCH and MISMATCH. In the MATCH condition, the gender of the pronoun matches that of the gendered noun which is the head of the relative clause. In the MISMATCH condition, the pronoun does not match (i.e., *her* with a masculine noun, or *him* with a feminine noun). The gender of the head noun was counterbalanced so that in the MATCH condition, the RP was *him* on half of the trials, and *her* on the other half. The same counterbalancing was applied in the MISMATCH condition. The second variable, Clause, has two levels, ISLAND and NONISLAND, manipulating the internal structure of the complex relative clause. In the ISLAND condition, a filler-gap dependency would be ungrammatical, as the *wh*-filler would have been extracted out of an adjunct *because*-clause or a *wh*-complement clause inside the larger relative clause. Conversely, in the NONISLAND condition, a filler-gap dependency would be grammatical, with extraction out of a verb complement clause beginning with the complementizer *that*.

In all conditions, the subject of the most deeply embedded clause is the first person pronoun *I*, which should not be considered as a potential antecedent for the RP. With this, participants only have two possible sentence-internal antecedents to consider for the RP: the gendered head noun (*policeman* in (8)), or the intervening subject of the relative clause (*civilian* in (8)). Crucially, the intervening subject is always an ungendered noun. While the RP could refer to this more local antecedent in on-line processing, it would not be expected to result in a clash.

The test items were divided into 16 single-word regions. Regions 2–3 introduce the head noun, and Regions 5–6, the intervening noun. Region 12, the region of interest, contains the RP. We will refer to this region as the *RP region*.

At the end of each trial, a *yes-no* comprehension question was presented to check for attention to the overall sentence content. Although all test sentences are ungrammatical due to the presence of RPs, with some containing relative clauses formed across islands and some containing RPs with mismatched fillers, we assigned each question a “correct” answer based on the identification of the subject of the relative clause. For the examples in (8), the correct answers to the comprehension questions *Was a civilian glad?* or *Did a civilian say something?* would be *yes*, while the answers to *Was a policeman glad?* or *Did a policeman say something?* would be *no*. There was an equal number of questions with correct *yes* and *no* responses across the whole stimulus set. We anticipate that the accuracy of responses may be low, due to the ungrammatical nature of these stimuli. None of the comprehension questions involved interpretation of the RP.

The 24 filler items were drawn from an unrelated experiment. None of the fillers contained sentences with relative clauses or with RPs.

2.3 Procedure

The items for Experiment 1 (24 test items, 24 fillers) were distributed across four lists in a Latin Square design. Items appeared in a randomized order.

Experiment 1 was published on PCIbex Farm (Zehr & Schwarz, 2018). Each trial began with the first word of the target sentence presented on the left of the screen, and the distractor *x-x-x* on the right. For subsequent words, the left-to-right order of the target and distractor was randomized. Participants were instructed to select the word that continued the sentence by pressing a key ('e' for the word on the left, 'i' for the word on the right). An incorrect selection caused the message *Incorrect. Try again!* to appear on-screen, and the participant could make a new selection. A correct selection advanced to the next target-distractor word pair.⁴ Following each trial, participants were presented with a *yes-no* comprehension question. Upon selecting an answer, they were given feedback as to whether the chosen answer was correct or not, and then instructed to press any key to continue to the next trial.

Four practice items served to familiarize participants with the Maze task before the start of the experiment. At the end of the experiment, participants were asked to complete a brief demographic survey.

2.4 Predictions

We used the gender-mismatch paradigm (Garnham, 2001; Sturt, 2003; van Gompel & Liversedge, 2003) to test whether the RP forms an on-line dependency with the head noun. If the RP forms a dependency with the head noun, we expect gender-mismatch effects exhibited by processing difficulty at the RP region. That is, as a dependency is formed, a gendered RP would clash with a gender-mismatched head noun (as opposed to a gender-matched head noun). This would result in longer reading times at the RP region in sentences with a gender-mismatched head noun than with a gender-matched head noun.

By contrast, if the RP does not form a dependency with the head noun, and instead connects with the intervening noun (which is closer in distance to the RP), then we do not expect any gender-mismatch effects. The intervening nouns in our test sentences are all gender-neutral: they should not clash with the RP, and so there should be no difference in reading times at the RP region, regardless of the gender of the head noun.

Moreover, if the way an RP forms a dependency depends on whether it occurs inside an island clause or not, then we should find reading time differences or gender-mismatch effect differences

⁴ As pointed out by a reviewer, in reality, since all test sentences were ungrammatical, participants were making a selection that would result in a less unacceptable option. However, the ungrammatical nature of the stimuli did not have any effect on the Maze task, as selection times in Maze tasks reflect relative integration difficulty. As discussed in 2.5, the target word selection rate was overwhelmingly high across conditions in all regions.

in the RP region between the two types of clause. For instance, if dependency formation is easier across one type of clause than across the other, then we expect to see reading time differences in the RP region between the two clause types. Further, if an RP forms a dependency with the head noun in islands, but not in non-islands, we only expect longer reading times at the RP region in sentences with a gender-mismatched head noun than with a gender-matched head noun *in islands*; no reading time differences are expected in non-islands, regardless of the gender specification of the head noun.

2.5 Results

For analysis, we removed data from three participants whose mean comprehension question response scores on test sentences were below 40%. We also removed data from two participants whose target word selection rates on their initial try were below 75%.

The grand mean comprehension question response score on test sentences was 65%. The mean proportions of correct responses for the comprehension questions are reported by condition in **Table 1**. The comprehension questions tested participants' attention to the overall sentence content. As anticipated, the results show somewhat low comprehension across conditions, but, importantly, the manipulated factors had no effect on comprehension.⁵

Table 1: Proportion of correct responses (SE), Experiment 1.

	MATCH	MISMATCH
NONISLAND	.63 (.005)	.66 (.005)
ISLAND	.66 (.005)	.65 (.005)

The comprehension question response scores were analyzed using a generalized mixed-effects model with fixed factors of Clause and Gender, and a random-effects structure consisting of random intercepts for participants and items, as well as random slopes for Clause and Gender for both participants and items. For all analyses reported in this article, we fit the maximal random-effects structure that converged without producing a singular fit (Barr et al., 2013; Bates et al., 2018) in R (R Development Core Team, 2020). The lme4 package was used to fit the model (Bates, 2005), and the lmerTest package was used to obtain *p*-values (Kuznetsova et al., 2014). Both predictors were sum coded: the NONISLAND level as 1 and the ISLAND level as -1 for

⁵ The distribution of mean comprehension question response scores was as follows, after the removal of five participants with low comprehension question response scores or low target word selection rates: 10% of participants scored 40–50%, 30% scored 50–60%, 23% scored 60–70%, 29% scored 70–80%, 6% scored 80–90%, and 2% scored 90–100%.

Clause, and the MATCH level as 1 and the MISMATCH level as -1 for Gender. We found no effect of Clause, Gender, or interaction between the two, as can be seen in **Table 2**.

Table 2: Fixed effects, comprehension question responses, Experiment 1.

	Estimate	SE	z value	Pr(> z)
(Intercept)	1.44431	0.45333	3.186	0.00144 **
Clause1	0.11256	0.12149	0.926	0.35420
Gender1	-0.04977	0.14557	-0.342	0.73242
Clause1:Gender1	-0.05181	0.12354	-0.419	0.67494
Formula in R: CorrectAns ~ Clause*Gender + (1 + Clause*Gender Participant) + (1 + Clause*Gender Item), family = "binomial"				

Significance levels: *** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .1$.

The low comprehension question response scores were possibly due to the fact that the test sentences were not grammatical. Recall that they contained relative clauses formed across islands, and also RPs in both island and non-island relative clauses, which are intrusive in English (Sells, 1984). Participants were paying attention and engaged with the task, as evidenced by the fact that the mean comprehension question response score of filler sentences was high (88%). The filler sentences were all grammatical. The target word selection rate in the Maze task was also high.

Overall, participants selected the target word on their initial try 97% of the time. The mean proportions of target word selection by condition from Region 10 to Region 15 are summarized in **Table 3**. Region 10 is two regions before the region of interest, the RP region (Region 12), and Region 15 is the last region. The table shows that the target word selection rate was high across the board.

Table 3: Proportions of target word selection, Experiment 1.

Region		10	11	12	13	14	15
NONISLAND	MATCH	0.96	0.98	0.97	0.98	0.96	0.95
	MISMATCH	0.96	0.98	0.96	0.97	0.97	0.97
ISLAND	MATCH	0.98	0.98	0.97	0.97	0.98	0.99
	MISMATCH	0.98	0.96	0.99	0.97	0.98	0.99

In order to exclude extreme outliers from the analysis, the raw reading times (RTs) of a region that were below 50 ms and above 4000 ms were removed. This resulted in removing 0.4% of

the observations from the data. In addition, we removed RTs for which the target word was not selected initially from the analysis. This further removed 3% of the observations from the data.

The graph in **Figure 1** summarizes mean RTs by condition from Region 10 to Region 15. It represents RTs for all trimmed data, regardless of whether the comprehension question was answered correctly. Trials with incorrectly answered comprehension questions were retained, because the comprehension questions probed global understanding of the test sentences and not word-by-word processing. As such, an incorrect answer to a comprehension question does not imply that the critical region was not processed. Further, given the low overall comprehension question response scores, removing all trials with incorrect comprehension question answers would lead to a large decrease in the sample size, which would negatively affect the distribution of the data.

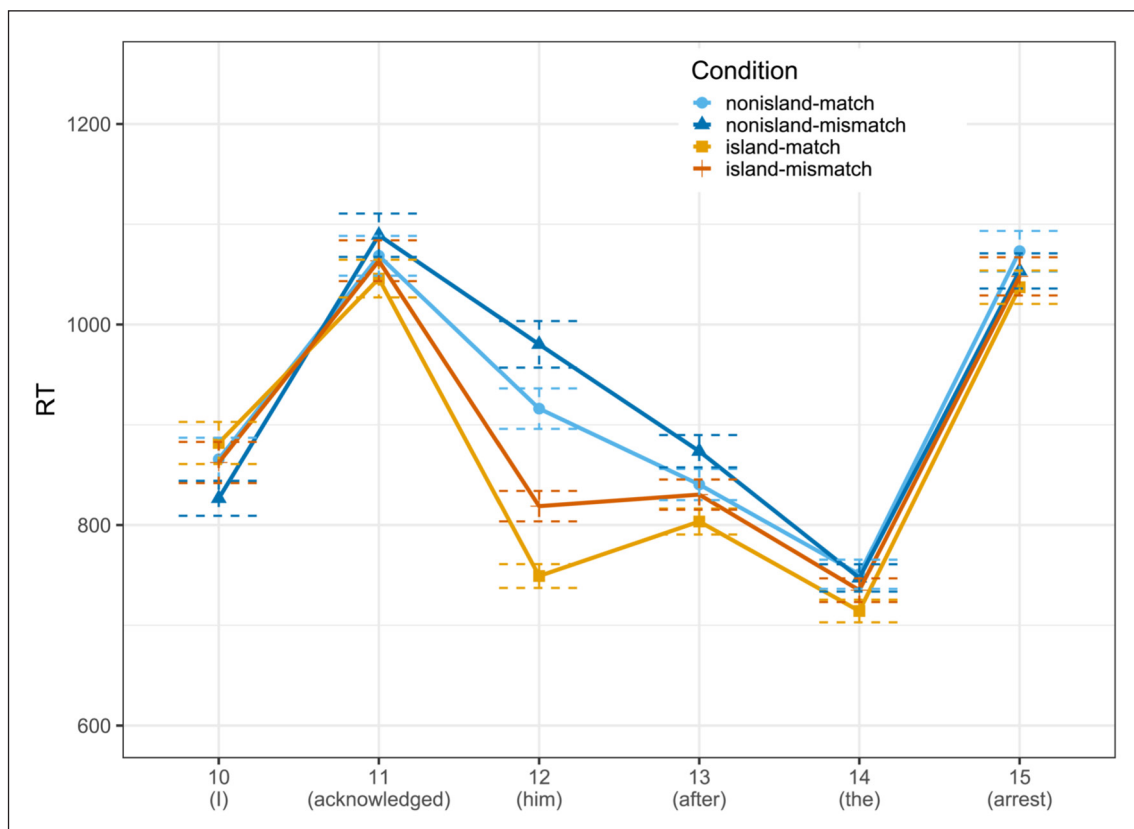


Figure 1: Mean RTs by condition with standard errors, Experiment 1.

We analyzed log reading times ($\log(\text{RTs})$) of the region of interest, the RP region (Region 12), with linear mixed-effects models using random-effects structures with by-participant random intercepts and random slopes for Clause, and by-item random intercepts and random slopes for

Clause and Gender. Distributions of raw RTs are typically right-skewed, and so log(RTs) were used to better approximate normality, as in Boyce et al. (2020). The model was fit to the log(RTs) with the fixed factors of Clause, Gender, and the log(RT) of the previous region (logPrevRT_c). The log(RT) of the previous region (Region 11) was included as a predictor, to control for any spillover effect that the previous region (Region 11) may have on the region of interest (Region 12). Both Clause and Gender were sum coded, with the NONISLAND level as 1 and the ISLAND level as -1 for Clause, and the MATCH level as 1 and the MISMATCH level as -1 for Gender. The log(RT) of the previous region was centered.

The results of the analysis are summarized in **Table 4**. The analysis showed effects of Clause and Gender. Overall, the non-island sentences showed longer log(RTs) than the island sentences in the RP region. Moreover, across the board, the sentences showed gender-mismatch effects: the RPs that mismatched in gender with the head noun had longer log(RTs) than the RPs that matched in gender with the head noun.

Table 4: Fixed effects, Region 12, Experiment 1.

	Estimate	SE	df	t value	Pr(> t)
(Intercept)	6.633	0.01868	88.78	355.161	< 0.0001 ***
Clause1	0.0703	0.00977	32.03	7.193	< 0.0001 ***
Gender1	-0.0298	0.00847	22.41	-3.515	0.002 **
logPrevRT_c	0.0856	0.00845	1770	10.138	< 0.0001 ***
Clause1:Gender1	0.00578	0.00697	1937	0.830	0.4068
Formula in R: $\log(\text{RT}) \sim \text{Clause} * \text{Gender} + \text{logPrevRT_c} + (1 + \text{Clause} \text{Participant}) + (1 + \text{Clause} + \text{Gender} \text{Item})$					

Significance levels: *** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .1$.

Planned comparisons between the log(RTs) of the MATCH and the MISMATCH sentences in the two Clause conditions confirmed that both the non-island sentences and the island sentences exhibit gender-mismatch effects. In each sentence type, the gender-mismatched RPs had significantly longer log(RTs) than the gender-matched RPs (NONISLAND: Estimate = -0.024, SE = 0.011, $t = -2.089$, $p = 0.037$; ISLAND: Estimate = -0.035, SE = 0.008, $t = -4.225$, $p < 0.0001$).⁶

⁶ A reviewer noted that there has been some evidence that log-transforming RTs can lead to an increase in Type 1 errors (Liceralde & Gordon, 2018). We reanalyzed the data with the raw RTs, and this showed the same pattern of results.

2.6 Discussion

The results show gender-mismatch effects for RPs in both island and non-island relative clauses. Further, the RPs in non-island clauses had longer reading times than the ones in island clauses, regardless of the gender of the head noun. The two results, taken together, suggest that the processor immediately attempts to form a dependency at the RP with the head noun in both islands and non-islands. Moreover, this happens even though there is a closer intervening noun. Finally, we also find that dependency formation is easier in islands than in non-islands.

Experiment 1 manipulated the gender of the head noun to (mis)match with the gender of the RP but did not manipulate the gender of the intervening noun. As such, while the results tell us that RPs form dependencies with head nouns, we cannot rule out the possibility that they can also form dependencies with intervening nouns. This is addressed in Experiment 2.

3. Experiment 2: Gender-mismatch effects of the head noun and the intervening noun

Experiment 2 tested whether RPs in relative clauses form on-line dependencies with head nouns or with intervening nouns.

3.1 Participants

120 native English speakers (71 female, 48 male, 1 prefer not to say) were recruited through Prolific Academic and completed the experiment on PCIBex Farm (Zehr & Schwarz, 2018). Participants ranged in age from 18 to 82, with a mean age of 37.5. All participants gave their informed consent through PCIBex Farm prior to the experiment and were compensated £3.00 upon completion. No participants had participated in Experiment 1.

3.2 Task, design, and materials

The task for Experiment 2 was the same as in Experiment 1, as was the general form of the trial items. However, the items for Experiment 2 follow a 3×2 factorial design, as in (10). The Match condition has three levels: MATCH, INTERVENOR.MISMATCH and HEAD.MISMATCH. For all conditions, both the head noun and the intervening noun are gendered. In the MATCH condition, both the head noun and the intervening noun match the gender of the RP. In the INTERVENOR.MISMATCH condition, the head noun matches the gender of the RP, while the intervening noun does not. This is reversed in the HEAD.MISMATCH condition. As in Experiment 1, this manipulation was achieved by changing the gender of the RP for the two MISMATCH conditions. In the MATCH condition, the gender of the intervening noun was changed to match the head noun. The genders of the head and the intervening nouns were counterbalanced. The second variable, Clause, was constructed as in Experiment 1, with the levels ISLAND and NONISLAND. Test items were divided into 16 regions, where the RP was contained in Region 12. Distractors for (10) are given in (11).

- (10) a. /₀ There /₁ was /₂ a /₃ watchman /₄ who /₅ a /₆ policewoman /₇ was /₈ glad /₉ because /₁₀ I /₁₁ acknowledged /₁₂ him /₁₃ after /₁₄ the /₁₅ arrest. ISLAND-INTERVENOR.MISMATCH
- b. /₀ There /₁ was /₂ a /₃ watchman /₄ who /₅ a /₆ policewoman /₇ was /₈ glad /₉ because /₁₀ I /₁₁ acknowledged /₁₂ her /₁₃ after /₁₄ the /₁₅ arrest. ISLAND-HEAD.MISMATCH
- c. /₀ There /₁ was /₂ a /₃ watchman /₄ who /₅ a /₆ policeman /₇ was /₈ glad /₉ because /₁₀ I /₁₁ acknowledged /₁₂ him /₁₃ after /₁₄ the /₁₅ arrest. ISLAND-MATCH
- d. /₀ There /₁ was /₂ a /₃ watchman /₄ who /₅ a /₆ policewoman /₇ had /₈ said /₉ that /₁₀ I /₁₁ acknowledged /₁₂ him /₁₃ after /₁₄ the /₁₅ arrest. NONISLAND-INTERVENOR.MISMATCH
- e. /₀ There /₁ was /₂ a /₃ watchman /₄ who /₅ a /₆ policewoman /₇ had /₈ said /₉ that /₁₀ I /₁₁ acknowledged /₁₂ her /₁₃ after /₁₄ the /₁₅ arrest. NONISLAND-HEAD.MISMATCH
- f. /₀ There /₁ was /₂ a /₃ watchman /₄ who /₅ a /₆ policeman /₇ had /₈ said /₉ that /₁₀ I /₁₁ acknowledged /₁₂ him /₁₃ after /₁₄ the /₁₅ arrest. NONISLAND-MATCH
- (11) /₀ x-x-x /₁ anti /₂ knew /₃ seabirds /₄ sad /₅ sun /₆ antibodies /₇ pre /₈ yeah /₉ digital /₁₀ vs /₁₁ mathematical /₁₂ mid /₁₃ click /₁₄ grow /₁₅ funded.

Comprehension questions once again checked for attention to the overall sentence content by testing the participants' ability to accurately identify the subject of the relative clause. The questions associated with the sample set of items in (10) would be: *Was a policewoman glad?*, *Was a policeman glad?*, *Did a policewoman say something?*, and *Did a policeman say something?*

An equal number of fillers appeared in each list. None of the fillers contained sentences with relative clauses or with RPs.

3.3 Procedure

The procedure for Experiment 2 was identical to that described for Experiment 1, except that the items (24 trial items, 24 fillers) were distributed across six lists in a Latin Square design. The items appeared in a randomized order.

As in Experiment 1, four practice items served to familiarize participants with the Maze task. At the end of the experiment, participants were asked to complete a brief demographic survey.

3.4 Predictions

If the RP forms an on-line dependency with the head noun, we expect gender-mismatch effects at the RP region when the head noun mismatches in gender and the intervening noun matches in gender with the RP. That is, we expect a longer reading time at the RP region in sentences with a gender-mismatched head noun and a gender-matched intervening noun than in sentences where both the head noun and the intervening noun match in gender with the RP.

If the RP forms a dependency with the intervening noun, we expect gender-mismatch effects at the RP region when the intervening noun mismatches in gender and the head noun matches in

gender with the RP. That is, we expect a longer reading time at the RP region in sentences with a gender-mismatched intervening noun and a gender-matched head noun than in sentences with a gender-matched head noun and intervening noun.

Moreover, if the way an RP forms a dependency depends on whether it occurs inside an island clause or not, we should find reading time differences or gender-mismatch effect differences in the RP region between the two types of clause. For instance, if an RP forms a dependency more easily across one type of clause than across the other, we expect to see reading time differences in the RP region between the two clause types. Further, if an RP forms a dependency with the head noun in islands, but not in non-islands, then only in islands do we expect to see gender-mismatch effects at the RP region, when the head noun mismatches, and the intervening noun matches, in gender with the RP.

3.5 Results

As in Experiment 1, we removed data from two participants whose mean comprehension question response scores on test sentences were below 40%, and two participants whose target word selection rates were below 75%.

The grand mean comprehension question response score on test sentences was 65% after the removal. The mean proportions of correct responses for the comprehension questions are reported in **Table 5**. Comprehension was somewhat low across all conditions, as in Experiment 1, due to the ungrammatical nature of the stimuli. In addition, on visual inspection, the response scores in the MATCH condition were lower than the ones in the MISMATCH conditions.⁷

Table 5: Proportion of correct responses (SE), Experiment 2.

	MATCH	INTERVENOR.MISMATCH	HEAD.MISMATCH
NONISLAND	.61 (.006)	.66 (.006)	.67 (.005)
ISLAND	.60 (.006)	.67 (.006)	.71 (.006)

A generalized mixed-effects model with fixed factors of Clause and Gender was fitted to the comprehension question response scores. The random-effects structure included by-participant random intercepts and random slopes for Clause and Gender, and by-item random intercepts and random slopes for Gender. Clause was sum coded, with the NONISLAND level coded as 1, and

⁷ The distribution of mean comprehension question response scores was as follows, after the removal of four participants with low comprehension question response scores or low target word selection rates: 13% of participants scored 40–50%, 24% scored 50–60%, 24% scored 60–70%, 27% scored 70–80%, 9% scored 80–90%, and 3% scored 90–100%.

the ISLAND level as -1 . Gender was simple coded, such that INTERVENOR.MISMATCH (Gender2) and HEAD.MISMATCH (Gender3) were each compared to MATCH, the reference level.⁸ As shown in **Table 6**, we found an effect of Gender, such that the response scores in the MATCH condition were significantly lower than the ones in the HEAD.MISMATCH condition.

Table 6: Fixed effects, comprehension question responses, Experiment 2.

	Estimate	SE	z value	Pr(> t)
(Intercept)	1.43096	0.42162	3.394	<0.001 ***
Clause1	-0.11045	0.07705	-1.433	0.151746
Gender2	0.37992	0.31890	1.191	0.233525
Gender3	0.88237	0.36101	2.444	0.014519 *
Clause1:Gender2	-0.03156	0.03922	-0.805	0.420959
Clause1:Gender3	-0.01601	0.04132	-0.388	0.698357
Formula in R: CorrectAns ~ Clause*Gender + (1 + Clause + Gender Participant) + (1 + Gender Item), family = "binomial"				

Significance levels: *** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .1$.

In a separate analysis, we fit a model with a fixed factor of Gender only. This predictor was forward-difference coded, such that MATCH was compared to INTERVENOR.MISMATCH (Gender1), and INTERVENOR.MISMATCH was compared to HEAD.MISMATCH (Gender2).⁹ This analysis yielded no effect, as shown in **Table 7**.

Table 7: Fixed effects, Gender only, comprehension question responses, Experiment 2.

	Estimate	SE	z value	Pr(> t)
(Intercept)	1.313	0.397	3.305	<0.001 ***
Gender1	-0.323	0.298	-1.083	0.279
Gender2	-0.434	0.304	-1.426	0.154
Formula in R: CorrectAns ~ Gender + (1 + Gender Participant) + (1 + Gender Item), family = "binomial"				

Significance levels: *** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .1$.

⁸ Gender2 was coded 2/3 for INTERVENOR.MISMATCH and $-1/3$ for HEAD.MISMATCH and MATCH. Gender3 was coded 2/3 for HEAD.MISMATCH and $-1/3$ for INTERVENOR.MISMATCH and MATCH.

⁹ Gender1 was coded 2/3 for MATCH and $-1/3$ for INTERVENOR.MISMATCH and HEAD.MISMATCH. Gender2 was coded 1/3 for MATCH and INTERVENOR.MISMATCH, and $-2/3$ for HEAD.MISMATCH.

According to these results, while the sentences in the two mismatch conditions had similar comprehension response scores, the sentences in the MATCH condition had reduced comprehension response scores in comparison to the ones in the HEAD.MISMATCH condition. But this reduction in comprehension did not have an effect on the target selection rate or the reading time in the Maze task.

Overall, participants selected the target word on their initial try 97% of the time. The mean proportions of target word selection by condition from Region 10 to Region 15 are summarized in **Table 8**. Just as in Experiment 1, the target word selection rate was high across the board.

Table 8: Proportions of target word selection, Experiment 2.

Region		10	11	12	13	14	15
NONISLAND	MATCH	0.96	0.98	0.99	0.97	0.97	0.98
	INTERVENOR. MISMATCH	0.96	0.98	0.99	0.97	0.97	0.98
	HEAD.MISMATCH	0.97	0.98	0.98	0.97	0.97	0.98
ISLAND	MATCH	0.98	0.98	0.99	0.97	0.97	0.98
	INTERVENOR. MISMATCH	0.97	0.98	0.99	0.98	0.97	0.99
	HEAD.MISMATCH	0.99	0.98	0.98	0.98	0.98	0.97

As in Experiment 1, raw RTs of a region that were below 50 ms and above 4000 ms were removed from the analysis. This resulted in the removal of 0.5% of the observations from the data. Also, as in Experiment 1, we removed RTs with incorrect target word selection from the analysis. This further removed 2.6% of the observations from the data.

Mean RTs by condition from Region 10 to Region 15 are summarized in the graphs in **Figure 2**. As in Experiment 1, the graphs represent RTs for all trimmed data, regardless of which answer was selected for the comprehension question. The graph on the left compares the RTs from the INTERVENOR.MISMATCH condition and the MATCH condition, while the graph on the right compares the RTs from the HEAD.MISMATCH condition and the MATCH condition.

We analyzed the $\log(\text{RTs})$ of the RP region (Region 12) using linear mixed-effects models with random-effects structures that consisted of random intercepts for participants and items, and random slopes for Clause for participants. As in Experiment 1, fixed factors included Clause, Gender, and the $\log(\text{RT})$ of the previous region ($\log\text{PrevRT}_c$), which was centered. Clause was sum coded, with the NONISLAND level coded as 1, and the ISLAND level as -1 . Gender was simple coded, such that INTERVENOR.MISMATCH (Gender2) and HEAD.MISMATCH (Gender3) were each compared to MATCH, the reference level.

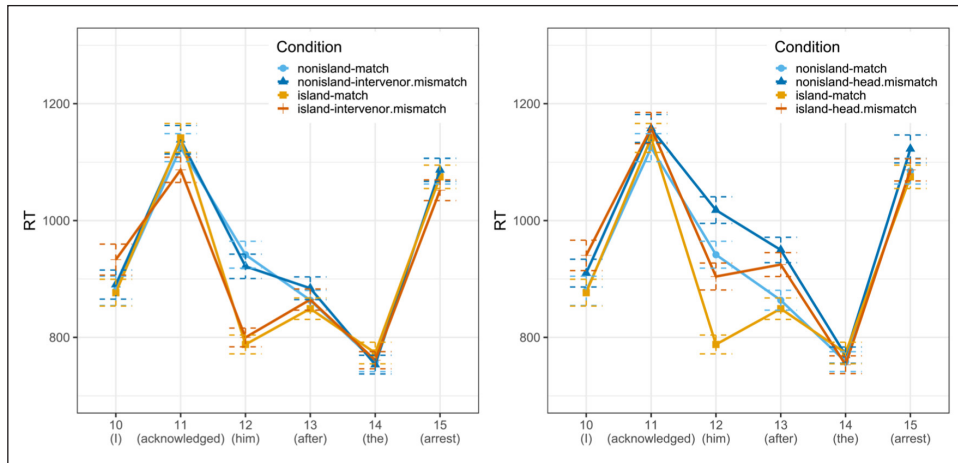


Figure 2: Mean RTs by condition with standard errors, Experiment 2.

In the RP region, we found significant effects of Clause and Gender, as shown in **Table 9**. Overall, the non-island sentences showed longer log(RTs) than the island sentences, as in Experiment 1. Also, the sentences across the board showed gender-mismatch effects with respect to the head noun (Gender3). That is, the RPs that mismatched in gender with the head noun but matched in gender with the intervening noun had longer log(RTs) than the RPs that matched in gender with both the head noun and the intervening noun. However, no gender-mismatch effects were found with respect to the intervening noun (Gender2): the RPs that mismatched in gender with the intervening noun but matched in gender with the head noun showed similar log(RTs) to the RPs that matched in gender with both the head noun and the intervening noun.

Table 9: Fixed effects, Region 12, Experiment 2.

	Estimate	SE	df	t value	Pr(> t)
(Intercept)	6.659	0.02123	90.35	313.74	< 0.0001 ***
Clause1	0.06212	0.007709	113.5	8.058	< 0.0001 ***
Gender2	0.005174	0.01516	2440	0.341	0.733
Gender3	0.08922	0.01521	2442	5.866	< 0.0001 ***
logPrevRT_c	0.09520	0.007896	2586	12.056	< 0.0001 ***
Clause1:Gender2	-0.02162	0.01518	2443	-1.425	0.154
Clause1:Gender3	-0.01274	0.01521	2444	-0.838	0.402
Formula in R: $\log(\text{RT}) \sim \text{Clause} * \text{Gender} + \log\text{PrevRT_c} + (1 + \text{Clause} \text{Participant}) + (1 \text{Item})$					

Significance levels: *** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .1$.

We conducted planned comparisons among the log(RTs) of the different levels in Gender in the two Clause conditions. Gender was forward-difference coded, so that the MATCH sentences were compared to the INTERVENOR.MISMATCH sentences (Gender1), and the INTERVENOR.MISMATCH sentences were compared to the HEAD.MISMATCH sentences (Gender2) in each Clause condition. In both the NONISLAND and the ISLAND conditions, the MATCH and the INTERVENOR.MISMATCH sentences had similar log(RTs), but the HEAD.MISMATCH sentences had significantly longer log(RTs) than the INTERVENOR.MISMATCH sentences. These results are shown in **Table 10** for sentences with non-island relative clauses and **Table 11** for sentences with island relative clauses. Thus, both the non-island relative clauses and the island relative clauses exhibit gender-mismatch effects with respect to the head noun, but not the intervening noun.¹⁰

Table 10: Fixed effects, sentences with non-island relative clauses, Experiment 2.

	Estimate	SE	df	t value	Pr(> t)
(Intercept)	6.71439	0.02463	82.73801	272.655	<0.0001 ***
Gender1	0.01696	0.02298	1206.14412	0.738	0.461
Gender2	-0.09152	0.02304	1206.14659	-3.972	<0.0001 ***
logPrevRT_c	0.10889	0.01179	1316.47652	9.234	<0.0001 ***
Formula in R: log(RT) ~ Gender + logPrevRT_c + (1 Participant) + (1 Item)					

Significance levels: *** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .1$.

Table 11: Fixed effects, sentences with island relative clauses, Experiment 2.

	Estimate	SE	df	t value	Pr(> t)
(Intercept)	6.59613	0.02063	78.67872	319.707	<.0001 ***
Gender1	-0.02661	0.01994	1207.58145	-1.334	0.182367
Gender2	-0.07438	0.01998	1208.17035	-3.723	0.000206 ***
logPrevRT_c	0.09747	0.01027	1289.41555	9.491	<.0001 ***
Formula in R: log(RT) ~ Gender + logPrevRT_c + (1 Participant) + (1 Item)					

Significance levels: *** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .1$.

¹⁰ Reanalyzing the data with the raw RTs yielded the same pattern of results.

3.6 Discussion

Just as in Experiment 1, the results in Experiment 2 exhibit gender-mismatch effects with respect to the head noun in both non-island and island relative clauses. Further, in Experiment 2, we found no evidence of gender-mismatch effects with respect to the intervening noun. These results show that RPs immediately form dependencies with head nouns, but not with intervening nouns, in both types of clause.

Moreover, the RPs in non-islands exhibited a processing difficulty in comparison to the ones in islands, regardless of whether they matched in gender with the head nouns or with the intervening nouns. This result is also consistent with the finding in Experiment 1, and indicates that RPs form dependencies with the head nouns in islands more easily than with the head nouns in non-islands.

4. Experiment 3: Comprehension of resumptive pronouns and gaps

In Experiments 1 and 2, we found evidence that RPs in relative clauses quickly form dependencies primarily with head nouns, and that dependency formation is especially facilitated across islands. Experiment 3 tested whether the dependency formation found in incremental processing is maintained in off-line comprehension and whether the on-line effect results in enhanced comprehension of RPs in comparison to gaps.

4.1 Participants

48 native English speakers (30 female, 18 male) were recruited through Prolific Academic and completed the experiment on PCIBex Farm (Zehr & Schwarz, 2018). Participants ranged in age from 18 to 71, with a mean age of 36.7. All participants gave their informed consent through PCIBex Farm prior to the experiment and were compensated £2.50 upon completion. No participants had participated in either of the previous experiments.

4.2 Task, design, and materials

This experiment used a forced choice sentence comprehension task. Participants were presented with complete relative clause sentences similar to those in the previous experiments. After reading each sentence, participants were asked a comprehension question that tested their understanding of the direct object of the verb within the relative clause, which was either a gap or an RP. This experiment had a 2×2 factorial design with a total of 48 items (24 test items, 24 fillers). A sample test trial item set is given in (12).

- (12) a. There was a watchman who a policeman was glad because I acknowledged ___ after the arrest. ISLAND-GAP

- b. There was a watchman who a policeman was glad because I acknowledged him after the arrest. ISLAND-RP
- c. There was a watchman who a policeman had said that I acknowledged __ after the arrest. NONISLAND-GAP
- d. There was a watchman who a policeman had said that I acknowledged him after the arrest. NONISLAND-RP

As in Experiments 1 and 2, the variable Clause has two levels: ISLAND and NONISLAND. In this experiment, the variable Object also has two levels: GAP, in which the object of the most deeply embedded clause is the gap of the relative clause, and RP, where the object position is occupied by an RP that is gender-matched to the head noun (*watchman* in (12)) as well as the intervening noun (*policeman* in (12)). The genders of the head and intervening nouns were counterbalanced.

The comprehension questions asked participants to select between four possible referents for the object of the verb in the relative clause. The options were the head noun (coded HEAD), the subject of the relative clause (INTERVENOR), the pronoun *I* (INCORRECT), or some other person (UNHERALDED). An example of the comprehension questions associated with (12) is shown in (13). The comprehension questions were matched to the factive/presumptive nature of the sentences (i.e., ISLAND sentences were accompanied by comprehension questions such as (13a), and NONISLAND sentences were accompanied by questions such as (13b)).

- (13) a. Who was acknowledged? (question for (12a,b))
- b. Who was said to be acknowledged? (question for (12c,d))
- | | |
|------------------------|------------|
| (i) watchman | HEAD |
| (ii) policeman | INTERVENOR |
| (iii) I | INCORRECT |
| (iv) some other person | UNHERALDED |

The 24 filler items were grouped into three different categories with eight items in each category: CATEGORICAL, BIAS, and NEUTRAL. All filler sentences were complex sentences with a matrix subject and a matrix object instantiated as referential noun phrases (e.g., *the child, the old man*). This was followed by a verb complement clause containing a pronoun (e.g., *him, he*) and a third referential noun phrase (e.g., *the street, the procedure*). An example set of sentences is shown in (14). The comprehension questions asked for the referent of the pronoun. The options were the matrix subject, the matrix object, the third noun phrase, or some other person. In CATEGORICAL items (14a), the pronoun categorically referred to either the matrix subject or the matrix object. In BIAS items (14b), the pronoun preferentially referred to either the matrix subject or the matrix object. In NEUTRAL items (14c), the pronoun could equally refer to the matrix subject or the matrix object.

- (14) a. The child asked the old man to assist him across the street.
Who should be assisted? CATEGORICAL
- b. The doctor said to the patient that he can perform the procedure next week.
Who can perform the procedure? BIAS
- c. The bus driver warned the passenger that he would not get to the station on time.
Who would not get to the station on time? NEUTRAL

The distribution of correct answers varied between the matrix subject and the matrix object, to ensure that participants did not develop a strategy for choosing the first NP.

4.3 Procedure

The items for Experiment 3 (24 trial items, 24 fillers) were distributed across four lists in a Latin Square design, and appeared in a randomized order.

Experiment 3 was published on PCIBex Farm (Zehr & Schwarz, 2018). Each trial was presented as a single screen: the target sentence, the comprehension question, and four possible answer options appeared on subsequent lines. The order of the answer options was randomized. Participants were instructed to click on the most appropriate answer to the comprehension question. After a brief delay – during which the participant’s selection was highlighted – the next trial was shown. No feedback was given on the participant’s answer selection.

Four practice items served to familiarize participants with the forced choice comprehension task before the start of the experiment. At the end of the experiment, participants were asked to complete a brief demographic survey.

4.4 Predictions

If the dependency formation of RPs of relative clauses in incremental processing is maintained in comprehension, then the selection rate for the head noun should be much higher than any of the other options.

Moreover, if RPs help comprehension in comparison to gaps, the selection rate for the head noun should be higher in sentences with RPs than in sentences with gaps. On the other hand, if RPs hinder comprehension, then the selection rate for the head noun should be lower in sentences with RPs than in sentences with gaps.

4.5 Results

The proportions of the four types of answer selections are summarized in **Figure 3**. As can be seen in **Figure 3**, the proportion of head noun selection is overwhelmingly higher than any other selection across all conditions.

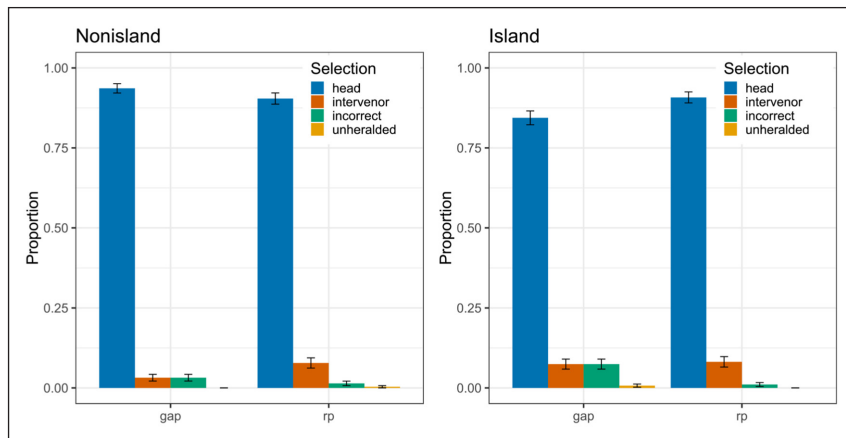


Figure 3: Proportions of answer choices by condition, Experiment 3.

For the purposes of statistical analysis, we used binary values (0 or 1) to code whether or not a participant selected a head noun answer to a comprehension question: a score of 1 was assigned if a head noun answer was chosen, and a score of 0 was assigned if any other answer was chosen. We analyzed the head noun answer selection rates by means of a generalized mixed-effects model (logistic/binomial regression model), using a random-effects structure with random intercepts for participants and items, and random slopes for Object for items. Fixed factors were Clause and Object.

We found a significant effect of Clause and an interaction between Clause and Object, as can be seen in **Table 12**. We conducted planned comparisons between the head noun answer selection rates of the sentences with gaps and the ones with RPs in the two Clause conditions. According to the planned comparisons, the interaction was due to the fact that in the ISLAND condition, the sentences with RPs had significantly higher head noun selection rates than the ones with gaps (Estimate = -0.38 , $SE = 0.15$, $z = -2.51$, $p = 0.01$), whereas in the NONISLAND condition, the sentences with gaps had marginally higher head noun selection rates than the ones with RPs (Estimate = 0.38 , $SE = 0.22$, $z = 1.76$, $p = 0.08$).

Table 12: Fixed effects, Experiment 3.

	Estimate	SE	z value	Pr(> z)
(Intercept)	3.3453	0.3644	9.181	<0.0001 ***
Clause1	-0.3682	0.1282	-2.872	0.00409 **
Object1	-0.1326	0.1702	-0.779	0.43594
Clause1:Object1	-0.3568	0.1276	-2.796	0.00518 **
Formula in R: Selection ~ Clause*Object + (1 Participant) + (1 + Object Item), family="binomial"				

Significance levels: *** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .1$.

4.6 Discussion

The results of the comprehension questions show that there is no difficulty in comprehending RPs as the head nouns in both island and non-island relative clauses. This shows that the on-line effect of RPs forming dependencies with head nouns carries over to off-line comprehension. In addition, we found no evidence that RPs hinder comprehension. In non-islands, RPs were interpreted as the head nouns at a similar rate as gaps, which was already quite high. Notably, in islands, RPs were interpreted as the head nouns at a higher rate than gaps, reaching the same rate as in non-islands. Thus, instead of hindering comprehension, RPs enhance it.

The reliability of our results is supported by the selection rates of the filler sentences patterning as expected, as summarized in **Figure 4**.

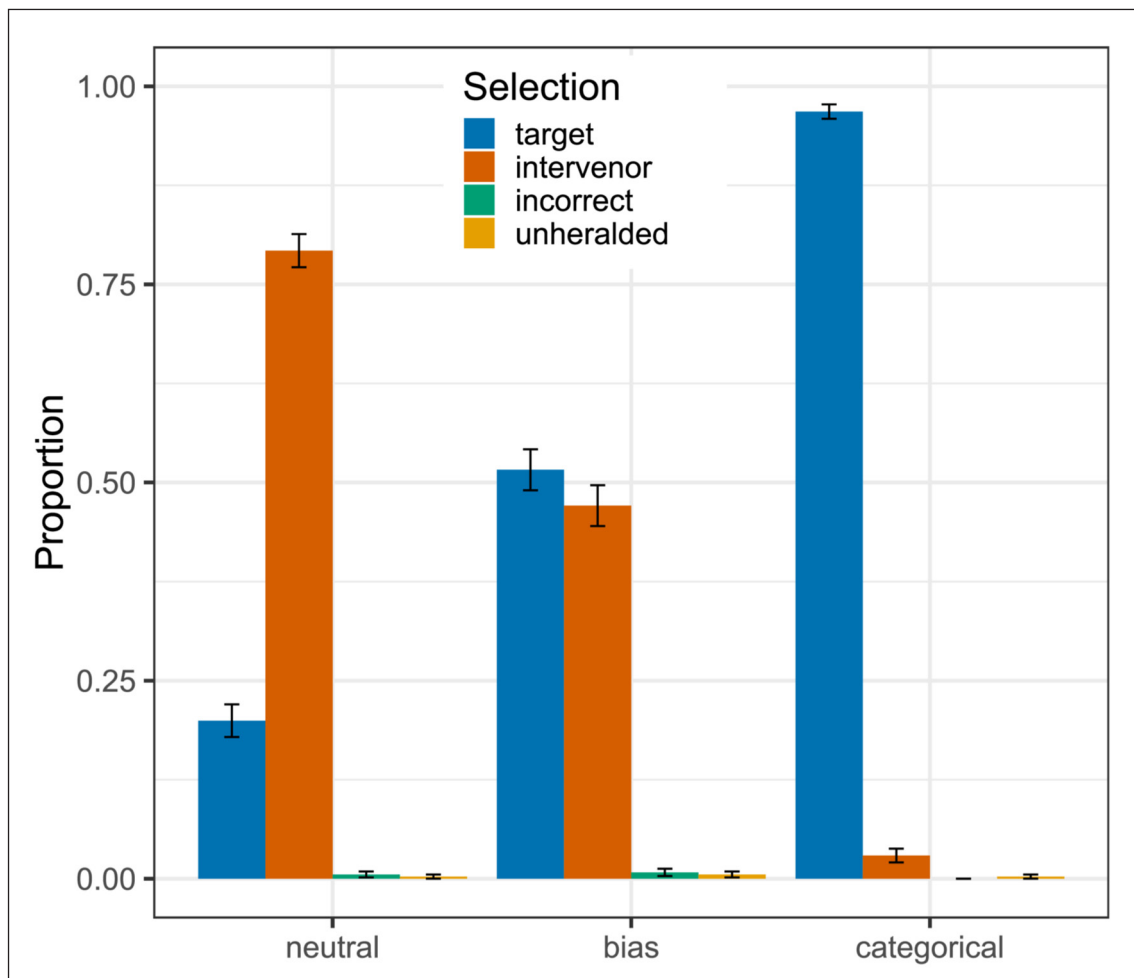


Figure 4: Proportions of answer choices by condition in filler sentences, Experiment 3.

For the sentences in the CATEGORICAL condition, between the matrix subject and the matrix object, we coded the noun phrase that is the correct antecedent of the pronoun as the TARGET and the noun phrase that is not as the INTERVENOR. A noun phrase that is neither the matrix subject nor the matrix object was coded as INCORRECT. For instance, in (14a), *the child* was coded as the TARGET, *the old man* was coded as the INTERVENOR, and *the street* was coded as INCORRECT. As can be seen in **Figure 4**, our participants almost always chose the correct answer, the TARGET, in the CATEGORICAL condition.

For the sentences in the BIAS and the NEUTRAL conditions, we coded the matrix subject as the TARGET, the matrix object as the INTERVENOR, and the third noun as INCORRECT. For instance, in (14b) and (14c), *the doctor* and *the bus driver* were coded as the TARGET, *the patient* and *the passenger* were coded as the INTERVENOR, and *the procedure* and *the station* were coded as INCORRECT. In the BIAS condition, half of the sentences were biased towards a matrix subject interpretation and the other half were biased towards a matrix object interpretation. As expected, selections between the TARGET and the INTERVENOR are split, reaching close to 50% each. In the NEUTRAL condition, the INTERVENOR was chosen 80% of the time, while the TARGET was chosen only 20% of the time. That is, in most cases, between the matrix subject and the matrix object, the participants chose the matrix object as the antecedent, which is closer in distance to the pronoun. This is consistent with reports in the literature that pronoun interpretation is generally subject to a recency condition: pronouns tend to associate with more recently mentioned referents (Ariel, 1990; Clark & Sengul, 1979; Cunnings et al., 2014).

Our findings show that RPs have no issues referring to head nouns, despite having closer intervening nouns as potential antecedents. This indicates that the recency effect is not as strong a factor in the interpretation of RPs as in the interpretation of referential pronouns, and suggests that the way RPs are interpreted should not be subsumed under how referential pronouns are interpreted. Instead, RP interpretation involves dependency formation with the filler.

5. General discussion

Our goal in this article was to investigate whether RPs in complex relative clauses formed from islands and non-islands establish dependencies with head nouns in incremental processing. In Experiments 1 and 2, in Maze tasks using the gender-mismatch paradigm, we found that RPs indeed form dependencies with head nouns across islands and non-islands, and they do not do so with intervening nouns that are closer to them in distance than the head nouns. Further, overall, RPs in islands were read faster than RPs in non-islands. This difference in reading times is unlikely to be due to shallow processing of RPs in islands, as readers were encouraged to integrate each incoming word through the use of the Maze task. Instead, the reading time difference suggests that dependency formation is easier – and, hence, facilitated – across islands in comparison to across non-islands.

The on-line effect of RPs forming dependencies with head nouns carries over to off-line comprehension. In Experiment 3, we found that comprehenders have no trouble interpreting RPs as head nouns. In non-islands, RPs were interpreted as head nouns at a similar rate as gaps. In islands, the rate at which RPs were interpreted as head nouns was higher than gaps, reaching the rates seen in non-islands. We thus found no evidence that RPs hinder sentence comprehension in comparison to gaps; on the contrary, our data suggest that they enhance comprehension.

Recall that the comprehension question response scores in Experiments 1 and 2 were low. Those comprehension questions were a check on overall sentence content and did not involve interpretation of the RPs. As speculated earlier, the low comprehension question response scores may be due to the fact that none of the test sentences were grammatical. However, our findings show that RP interpretation was not affected: comprehenders quickly connected RPs with fillers in on-line processing and understood RPs to be referring to the fillers in off-line comprehension.

The results of our studies contrast with those in Morgan et al. (2020). Their findings suggest that RPs do not reliably form dependencies with fillers, and that comprehenders have trouble understanding them to be referring to fillers. Perhaps the differences in stimuli and tasks contributed to obtaining results that paint different pictures. Test sentences in Morgan et al. (2020) had three animal characters as potential antecedents, while our test sentences had two potential antecedents, as the clause-mate subjects of the RPs were first person singular pronouns. In terms of task, in a visual-world eye tracking paradigm, participants are passively hearing test sentences, while in a Maze task, they are actively selecting words to form a possible sentence. Therefore, participants in a Maze task might be more encouraged to integrate incoming words, including RPs. Further, the format of the comprehension question task was different between Morgan et al.'s studies and our Experiment 3. In Morgan and colleagues' studies, after the test sentence was presented, a new screen appeared containing only a comprehension question and response options. In our Experiment 3, both the test sentence and the comprehension question with response options were presented on a single screen. This format may have made the task easier for our participants. Regardless, our results show that when comprehenders are encouraged to integrate RPs, they connect them with fillers in incremental processing, and they understand them to be referring to fillers. This connection is made even though RPs are judged to be unacceptable, as shown in many acceptability judgment studies (Alexopoulou & Keller, 2007; Clemens et al., 2012; Ferreira & Swets, 2005; Han et al., 2012; Heestand et al., 2011; Keffala & Goodall, 2011). According to Morgan and colleagues, the unacceptable status of RPs leads comprehenders to give up on integrating them. Our results, however, indicate that instead of giving up, the human sentence processor is capable of integrating and using RPs to form dependency structures with fillers.

5.1 Implications for facilitation of dependency formation

In our studies, we did not directly compare the on-line processing of RPs and gaps. Thus, although our results provide evidence that RPs form dependencies with fillers, we do not have direct evidence that RPs *facilitate* dependency formation in comparison to gaps. Nevertheless, in the off-line data, we saw that in islands, RPs enhance comprehension (as opposed to gaps). It was precisely in this context that our on-line studies found filler-RP dependency formation to be easiest. Moreover, it was also in this context that Hammerly (2022) found that in the spillover region right after the gap/RP, the reading time in the RP condition was faster than in the gap condition. Taking these points together, the findings in our study are consistent with the conclusion that RPs facilitate dependency formation with fillers, at least in islands.

That said, the comprehension rate of RPs in non-islands was still high – as high as that of gaps. As such, although we have no evidence that RPs in non-islands facilitate dependency formation in comparison to gaps, we see that the increased difficulty of processing RPs in this context does not lead to a reduction in comprehension (as opposed to gaps).

5.2 Implications for active search in islands

Our studies confirm previous findings that RP processing is easier in islands than in non-islands (Hammerly, 2022; Morgan et al., 2018). As Hammerly (2022) points out, the slow reading time at the RP in non-islands is most likely a filled-gap effect (Crain & Fodor, 1985; Stowe, 1986). Filled-gap effects arise because, according to the active filler strategy (Frazier & Clifton, 1989), the *wh*-filler associated with the head noun engages in a search for a gap at the first possible location. When this gap is filled, the parser's expectation is dashed, resulting in slow reading times at the filled phrase.

Of primary interest, however, are the effects in island sentences. The gender-mismatch effect at the RP shows that readers retrieve the filler immediately upon encountering the RP. This dependency, moreover, is maintained: end-of-sentence comprehension shows that readers interpret the RP as the filler. Things could have been different. Given that the RP is within an island, the filler should not be active (Phillips, 2006; Stowe, 1986; Traxler & Pickering, 1996). We might have expected the processor to “wait” to exit the island to attempt to resolve the *wh*-dependency. Further, the RP could take the intervenor as a grammatical antecedent. That's not what we found. One possibility is that the filler may be active inside the island, contrary to the general consensus. Another possibility is that the head noun filler is retrieved not because it is an active *wh*-filler, but because as a head noun it is highly topical and prominent inside the relative clause (Kuno, 1976; van Valin & Wilkins, 1996). Pronouns gravitate to topical prominent antecedents. These factors could help explain why the RP links to the head noun rather than to the intervenor. Following this line of thinking, the dependency formed between the filler and

the RP is not a *wh*-dependency *per se*, but an antecedent-pronoun dependency. Or there may be a hybrid approach, where the surprising activity of the head in the island is the joint effect of the properties of *wh*-fillers, which carry, perhaps, a baseline level of activity even in islands, and the properties of being a prominent antecedent. Put together, this lets the head noun filler beat out the intervenor for retrieval at the RP. These speculations require further elaboration within an explicit model of activity in sentence processing.

5.3 Implications for comprehension and production benefits

Our findings also have implications for the relationship between the production and comprehension of RPs. The production of RPs has been described as paradoxical (Morgan et al., 2020): why would speakers produce something that is unacceptable? The most widely-accepted explanation traces back to Asudeh (2004, 2011), with the claim that the production of an RP in an island or a non-island allows the speaker to maintain a locally grammatical structure, even if the sentence is globally ungrammatical. This can simply be a result of poor sentence planning, as described in, e.g., Kroch (1981), taking the view that the speaker's use of an RP is "selfish", and that any benefit to the comprehender is more coincidental than intentional. The findings in Morgan et al. (2020) align with this view: the suggestion is that the presence of an RP ends up confusing comprehenders, as they are tasked with processing an ungrammatical string.

An alternate view is that, in addition to being a production device benefiting the speaker, the RP is produced to help comprehension. It is certainly the case that RPs are shown to have faster reading times than gaps in experimental data (Hofmeister & Norcliffe, 2013; Morgan et al., 2020), and Beltrama and Xiang (2016) report participants rating sentences with RPs as more comprehensible than ones without. But these measures alone do not isolate the issue of whether the presence of an RP facilitates the connection with the filler during sentence processing.

Using the Maze task, in combination with a gender-mismatch paradigm, we have demonstrated that readers form filler-RP dependencies immediately upon encountering the RP. Furthermore, in off-line comprehension, we found a significant increase in identifications of the head noun in islands when an RP was present (vs. a gap). Thus, we see that even though RPs may not improve sentence acceptability, making the ungrammatical dependency in islands overt (in the form of an RP) does enhance comprehension. By showing that readers in our study are actively connecting the RP to the correct antecedent, we can see that the potential benefit on the comprehension side is not just a coincidence: the language processor is integrating the RP as part of a dependency. So, our results suggest that while there may be a local benefit for the speaker producing the RP, this is not the only benefit, and global comprehension for the listener also improves.

6. Conclusion

In the experiments reported in this article, we investigated whether English RPs in complex relative clauses formed from islands and non-islands establish dependencies with head nouns in incremental processing. Based on the findings from the Maze task and comprehension data, we argue that (i) RPs form dependency structures with head nouns quickly in on-line processing; (ii) this dependency formation is easier across islands than non-islands; (iii) the dependency formation privileges the head noun over other potential antecedents that are closer in distance to the RP; and (iv) the dependencies are maintained in off-line comprehension.

Data accessibility statement

The code and data supporting the conclusions of this article are available on OSF: <https://osf.io/r36q5/>. Stimuli used in the experiments are provided in the Supplemental Material file available at the link above.

Ethics and consent

The studies presented in this article were reviewed and approved by Simon Fraser University, Office of Research Ethics (30001133).

Acknowledgements

We are extremely indebted to the three anonymous reviewers for their insightful comments, which were crucial in improving this article. This work was supported in part by Social Sciences and Humanities Research Council (SSHRC) Insight Grant 435-2022-0475 to Chung-hye Han.

Competing interests

The authors have no competing interests to declare.

Author contributions

C-hH conceptualized the study. C-hH, KM, and DS designed the experiments. C-hH, KM, DS, TB, HG, JW, and SW constructed the stimuli and wrote the manuscript. TB, HG, JW, and SW programmed the experiments and collected data. C-hH performed statistical analysis. SW copy-edited the manuscript. All authors contributed to the article and approved the submitted version.

ORCID IDs

Trevor Block: <https://orcid.org/0009-0001-3531-8073>

Dennis Storoshenko: <https://orcid.org/0000-0002-4038-4796>

Jesse Weir: <https://orcid.org/0009-0003-6201-0589>

References

- Alexopoulou, T., & Keller, F. (2007). Locality, cyclicity, and resumption: At the interface between the grammar and the human sentence processor. *Language*, 83(1), 110–160. <https://doi.org/10.1353/lan.2007.0001>
- Ariel, M. (1990). *Accessing noun-phrase antecedents*. Routledge.
- Asudeh, A. (2004). *Resumption as resource management* [Doctoral dissertation]. Stanford University.
- Asudeh, A. (2011). Local grammaticality in syntactic production. In E. M. Bender & J. E. Arnold (Eds.), *Language from a cognitive perspective* (pp. 51–79). CSLI Publications.

- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D. (2005). Fitting linear mixed models in R: Using the lme4 package. *R News: The Newsletter of the R Project*, 5, 27–30.
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2018). *Parsimonious mixed models*. <https://arxiv.org/abs/1506.04967>
- Beltrama, A., & Xiang, M. (2016). Unacceptable but comprehensible: The facilitation effect of resumptive pronouns. *Glossa*, 1(1), 1–24. <https://doi.org/10.5334/gjgl.24>
- Boyce, V., Futrell, R., & Levy, R. P. (2020). Maze made easy: Better and easier measurement of incremental processing difficulty. *Journal of Memory and Language*, 111, Article 104082. <https://doi.org/10.1016/j.jml.2019.104082>
- Cann, R., Kaplan, T., & Kempson, R. (2005). Data at the grammar-pragmatics interface: The case of resumptive pronouns in English. *Lingua*, 115, 1551–1577. <https://doi.org/10.1016/j.lingua.2004.07.004>
- Carlson, G. N. (1977). Amount relatives. *Language*, 53(3), 520–542. <https://doi.org/10.2307/413175>
- Chomsky, N. (1981). *Lectures on government and binding*. Foris Publications.
- Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, 42(4), 368–407. <https://doi.org/10.1006/cogp.2001.0752>
- Clark, H., & Sengul, C. (1979). In search of referents for nouns and pronouns. *Memory & Cognition*, 7(1), 35–41. <https://doi.org/10.3758/BF03196932>
- Clemens, L. E., Morgan, A., Polinsky, M., & Xiang, M. (2012, March 14–16). *Listening to resumptives: An auditory experiment* [Conference poster]. CUNY 2012 Conference on Human Sentence Processing, New York, NY, United States.
- Crain, S., & Fodor, J. D. (1985). How can grammars help parsers? In D. Dowty, L. Karttunen, & A. M. Zwicky (Eds.), *Natural language parsing: Psycholinguistic, computational, and theoretical perspectives* (pp. 94–128). Cambridge University Press. <https://doi.org/10.1017/CBO9780511597855.004>
- Cunnings, I., Patterson, C., & Felser, C. (2014). Variable binding and coreference in sentence comprehension: Evidence from eye movements. *Journal of Memory and Language*, 71(1), 39–56. <https://doi.org/10.1016/j.jml.2013.10.001>
- Cunnings, I., Patterson, C., & Felser, C. (2015). Structural constraints on pronoun binding and coreference: Evidence from eye movements during reading. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00840>
- Dickey, M. W. (1996). Constraints on the sentence processor and the distribution of resumptive pronouns. In M. W. Dickey & S. Tunstall (Eds.), *Linguistics in the laboratory* (pp. 157–192, Vol. 19). GLSA.
- Erteschik-Shir, N. (1992). Resumptive pronouns in islands. In H. Goodluck & M. Rochemont (Eds.), *Island constraints: Theory, acquisition and processing* (pp. 89–109, Vol. 15). Kluwer Academic Publishers. https://doi.org/10.1007/978-94-017-1980-3_4

- Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11(1), 11–15. <https://doi.org/10.1111/1467-8721.00158>
- Ferreira, F., & Patson, N. D. (2007). The ‘good enough’ approach to language comprehension. *Language and Linguistics Compass*, 1(1–2), 71–83. <https://doi.org/10.1111/j.1749-818X.2007.00007.x>
- Ferreira, F., & Swets, B. (2005). The production and comprehension of resumptive pronouns in relative clause “island” contexts. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 263–278). Lawrence Erlbaum.
- Forster, K. I., Guerrero, C., & Elliot, L. (2009). The Maze task: Measuring forced incremental sentence processing time. *Behavior Research Methods*, 41, 163–171. <https://doi.org/10.3758/BRM.41.1.163>
- Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, 4(2), 93–126. <https://doi.org/10.1080/01690968908406359>
- Freedman, S. E., & Forster, K. I. (1985). The psychological status of overgenerated sentences. *Cognition*, 19(2), 101–131. [https://doi.org/10.1016/0010-0277\(85\)90015-0](https://doi.org/10.1016/0010-0277(85)90015-0)
- Garnham, A. (2001). *Mental models and the interpretation of anaphora*. Psychology Press.
- Hammerly, C. (2022). The pronoun which comprehenders who process it in islands derive a benefit. *Linguistic Inquiry*, 53, 823–835. https://doi.org/10.1162/ling_a_00422
- Han, C., Block, T., Gendron, H., Grant, M., Nederveen, S., Storoshenko, D., Weir, J., Williamson, S., & Moulton, K. (2025). Underspecified they becomes specified early in sentence processing. *Language, Cognition and Neuroscience*, 40(4), 508–526. <https://doi.org/10.1080/23273798.2024.2443975>
- Han, C., Elouazizi, N., Galeano, C., Görgülü, E., Hedberg, N., Hinnell, J., Jeffrey, M., Kim, K., & Kirby, S. (2012). Processing strategies and resumptive pronouns in English. In N. Arnett & R. Bennett (Eds.), *Proceedings of the 30th West Coast Conference on Formal Linguistics* (pp. 153–161). Cascadilla Proceedings Project.
- Heestand, D., Xiang, M., & Polinsky, M. (2011). Resumption still does not rescue islands. *Linguistic Inquiry*, 42, 138–152. https://doi.org/10.1162/LING_a_00032
- Hofmeister, P., & Norcliffe, E. (2013). Does resumption facilitate sentence comprehension? In P. Hofmeister & E. Norcliffe (Eds.), *The core and the periphery: Data-driven perspectives on syntax inspired by Ivan A. Sag* (pp. 225–246). CSLI Publications.
- Kayne, R. (1981). “ECP” extensions. *Linguistic Inquiry*, 12, 93–134.
- Kazanina, N., Lau, E. F., Lieberman, M., Yoshida, M., & Phillips, C. (2007). The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language*, 56, 384–409. <https://doi.org/10.1016/j.jml.2006.09.003>
- Keffala, B., & Goodall, G. (2011, March 24–26). *Do resumptive pronouns ever rescue illicit gaps in English?* [Conference poster]. CUNY 2011 Conference on Human Sentence Processing, Stanford, CA, United States.

- Kroch, A. (1981). On the role of resumptive pronouns in amnestying island constraint violations. In R. A. Hendrick, C. S. Masek, & M. F. Miller (Eds.), *Papers from the 17th Regional Meeting, Chicago Linguistic Society* (pp. 125–135). University of Chicago, Chicago Linguistic Society.
- Kuno, S. (1976). Subject, theme, and the speaker's empathy: A reexamination of relativization phenomena. In C. N. Li (Ed.), *Subject and topic* (pp. 417–444). Academic Press. <https://doi.org/10.4324/9781003766667-13>
- Kuznetsova, A., Brockho, P. B., & Christensen, R. H. B. (2014). *lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package)*. <http://CRAN.R-project.org/package=lmerTest>
- Licalalde, V. R. T., & Gordon, P. C. (2018). *Consequences of power transforms as a statistical solution in linear mixed-effects models of chronometric data*. PsyArXiv. <https://doi.org/10.31234/osf.io/f73mh>
- Meltzer-Asscher, A. (2021). Resumptive pronouns in language comprehension and production. *Annual Review of Linguistics*, 7(1), 177–194. <https://doi.org/10.1146/annurev-linguistics-031320-012726>
- Morgan, A. M., von der Malsburg, T., Ferreira, V. S., & Wittenberg, E. (2018, March 15–17). *This is the structure that we wonder why anyone produces it: Resumptive pronouns in English hinder sentence comprehension* [Conference presentation]. CUNY 2018 Conference on Human Sentence Processing, Davis, CA, United States.
- Morgan, A. M., von der Malsburg, T., Ferreira, V. S., & Wittenberg, E. (2020). Shared syntax between comprehension and production: Multi-paradigm evidence that resumptive pronouns hinder comprehension. *Cognition*, 205, 1–21. <https://doi.org/10.1016/j.cognition.2020.104417>
- Morgan, A. M., & Wagers, M. W. (2018). English resumptive pronouns are more common where gaps are less acceptable. *Linguistic Inquiry*, 49(4), 861–876. https://doi.org/10.1162/ling_a_00293
- Phillips, C. (2006). The real-time status of island phenomena. *Language*, 82, 795–823. <https://doi.org/10.1353/lan.2006.0217>
- Prince, E. (1990). Syntax and discourse: A look at resumptive pronouns. In K. Hall, J.-P. Koenig, M. Meacham, S. Reinman, & L. A. Sutton (Eds.), *Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society: Parasession on the legacy of Grice* (pp. 482–497). University of California, Berkeley Linguistics Society. <https://doi.org/10.3765/bls.v16i0.1719>
- Prince, E. (1995). On KIND-sentences, resumptive pronouns, and relative clauses. In G. Guy, J. Baugh, & D. Schiffrin (Eds.), *Towards a social science of language – A festschrift for William Labov* (pp. 223–235). Cambridge University Press. <https://doi.org/10.1075/cilt.128.16pri>
- R Development Core Team. (2020). *R: A language and environment for statistical computing*.
- Ross, J. R. (1967). *Constraints on variables in syntax* [Doctoral dissertation]. MIT.
- Sauerland, U. (2000). Two structures for English relative clauses. In M. Saito, Y. Abe, H. Aoyagi, M. Arimoto, K. Murasugi, & T. Suzuki (Eds.), *Proceedings of the Nanzan GLOW* (pp. 351–366). Nanzan University.

- Sells, P. (1984). *Syntax and semantics of resumptive pronouns* [Doctoral dissertation]. University of Massachusetts at Amherst.
- Sloggett, S., Van Handel, N., Sasaki, K., Duff, J., Rich, S., Orth, W., Anand, P., & Rysling, A. (2020, March 19–21). “Ambiguous” isn’t “underspecified”: Evidence from the Maze task [Conference presentation]. CUNY 2020 Conference on Human Sentence Processing, Amherst, MA, United States.
- Stowe, L. A. (1986). Parsing wh-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, 1(3), 227–245. <https://doi.org/10.1080/01690968608407062>
- Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48(3), 542–562. [https://doi.org/10.1016/S0749-596X\(02\)00536-3](https://doi.org/10.1016/S0749-596X(02)00536-3)
- Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, 35(3), 454–475. <https://doi.org/10.1006/jmla.1996.0025>
- van Gompel, R., & Liversedge, S. P. (2003). The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 1, 128–139. <https://doi.org/10.1037//0278-7393.29.1.128>
- van Valin, R. D., & Wilkins, D. P. (1996). The case for ‘effector’: Case roles, agents, and agency revisited. In M. Shibatani (Ed.), *Grammatical constructions: Their form and meaning* (pp. 289–322). Clarendon Press. <https://doi.org/10.1093/oso/9780198235392.003.0011>
- Witzel, N., Witzel, J., & Forster, K. (2012). Comparisons of online reading paradigms: Eye tracking, moving-window, and Maze. *Journal of Psycholinguistic Research*, 41, 105–128. <https://doi.org/10.1007/s10936-011-9179-x>
- Zehr, J., & Schwarz, F. (2018). *PennController for internet based experiments (IBEX)*. <https://doi.org/10.17605/OSF.IO/MD832>

