Resolving Conflicts Between Syntax and Plausibility in Sentence Comprehension

Glenda Andrews[1,2], Jessica E. Ogden[1], and Graeme S. Halford[3]

[1] School of Applied Psychology, Griffith University, Gold Coast, Australia

[2] Menzies Institute of Health Queensland, Griffith University, Gold Coast, Australia

[3] School of Applied Psychology, Griffith University, Mt Gravatt, Australia

KEYWORDS

syntax-plausibility conflict, relative clause sentence comprehension, working memory, fluid intelligence

ABSTRACT

Comprehension of plausible and implausible object- and subject-relative clause sentences with and without prepositional phrases was examined. Undergraduates read each sentence then evaluated a statement as consistent or inconsistent with the sentence. Higher acceptance of consistent than inconsistent statements indicated reliance on *syntactic analysis*. Higher acceptance of plausible than implausible statements reflected reliance on *semantic plausibility*. There was greater reliance on semantic plausibility and lesser reliance on syntactic analysis for more complex object-relatives and sentences with prepositional phrases than for less complex subject-relatives and sentences without prepositional phrases. Comprehension accuracy and confidence were lower when syntactic analysis and semantic plausibility yielded conflicting interpretations. The conflict effect on comprehension was significant for complex sentences but not for less complex sentences. Working memory capacity predicted resolution of the *syntax-plausibility conflict* in more and less complex items only when sentences and statements were presented sequentially. Fluid intelligence predicted resolution of the conflict in more and less complex items under sequential and simultaneous presentation. Domain-general processes appear to be involved in resolving syntax-plausibility conflicts in sentence comprehension.

INTRODUCTION

Many cognitive and linguistic tasks require us to integrate different types of information. For example, deductive reasoning involves considering the logical form of the problem in the presence of content that might support or conflict with conclusions based on logical form. Another example is sentence comprehension. To interpret a sentence, information conveyed by its syntactic structure and its semantic content must be integrated (Traxler, 2011, 2014). In some cases, syntactic and semantic information are mutually supportive. In other cases, they lead to conflicting interpretations of the sentence. How these conflicting interpretations are reconciled is not yet fully understood. In the current research, we used an off-line procedure adapted from reasoning research to examine the effect of *syntax-plausibility conflict* on comprehension of relative clause sentences, and the extent to which verbal working memory and fluid intelligence are involved in resolving the conflict.

Relative clause sentences vary in *syntactic complexity*. There is consensus that object-extracted relatives, such as *Sentence 1*, are more complex than simple subject relatives, such as *Sentence 2*, and subject-extracted relatives, such as *Sentence 3*. *Sentences 1* and *3* are similar in structure in that the relative clause is embedded in the main clause, whereas in *Sentence 2* the relative clause follows the main clause. *Sentences 1* and *2* express virtually the same meaning in that *woman* is the patient of *helped* and the agent of *sang*, while *man* is the agent of *helped*. *Sentence 3* has a different meaning in that *man* is the agent of both *helped* and *sang* while *woman* is the patient of *helped*.

The woman that the man helped sang well.   (1)

Corresponding author: Glenda Andrews, School of Applied Psychology, Griffith University, Gold Coast, 4222, Australia. Phone: +61756788613. E-mail: g.andrews@griffith.edu.au

The man helped the woman that sang well. (2)

The man that helped the woman sang well. (3)

The semantic content of *Sentences 1*, *2*, and *3* is neutral in the sense that our real-world knowledge tells us that both *man* and *woman* are potential agents of the verbs *helped* and *sang*. In such cases comprehension relies heavily on syntactic processing because the semantic content provides no clues regarding who helped whom and who sang.

The effect of *semantic plausibility* can be illustrated by comparing the object-relative *Sentence 4* and the subject-relative *Sentences 5* and *6* to *Sentences 1*, *2*, and *3*, respectively.

The comedian that the millionaire funded left town. (4)

The millionaire funded the comedian that left town. (5)

The millionaire that funded the comedian left town. (6)

Unlike *Sentences 1*, *2*, and *3*, the semantic content of *Sentences 4*, *5*, and *6* is not neutral. Rather, our real-world knowledge tells us that *millionaire* is a more plausible agent of *funded* than *comedian* is.

Much previous research has shown that when semantic content is neutral, object-relative sentences are more difficult to understand than subject-relative sentences (Caplan & Waters, 1999). This difference in difficulty has been shown using many dependent variables, including word-by-word reading times (Caplan, DeDe, Waters, Michaud, & Tripodis, 2011; King & Just, 1991), eye-movements during reading (Traxler, Williams, Blozis, & Morris, 2005), true/false comprehension questions (Gordon, Hendrick, & Levine, 2002; King & Just, 1991), and comprehension questions assessing understanding of thematic roles (Andrews, Birney, & Halford, 2006; Andrews et al., 2013; Andrews & Todd, 2008; Hansell et al., 2015). These differences in difficulty are attributed to the greater syntactic complexity of object-relative than subject-relative sentences (Frazier, 1985; Gibson, 2000). Syntactic complexity effects are in turn attributed to memory limitations in some accounts (e.g., Carpenter, Miyake, & Just, 1994; Gibson, 2000) and to the role of the parser's expectations about the upcoming structure in other accounts (e.g., Hale, 2001; Levy, 2008; Levy, Bicknell, Slattery, & Rayner, 2009). Based on eye-movement evidence, Staub (2010) concluded that both violation of expectations and memory-related factors contribute to the greater difficulty of English object-extracted relatives (see *Sentence 1*) compared to subject-extracted relatives (see *Sentence 3*). Levy, Fedorenko, and Gibson (2013) reached a similar conclusion based on their experiments examining on-line comprehension of Russian relative clauses.

Regarding semantic plausibility, there is evidence that sentence processing and comprehension are easier when the semantic content supports the syntactic analysis (King & Just, 1991; Traxler et al., 2005). Thus, *Sentences 4*, *5*, and *6* should be easier to understand than *Sentences 1*, *2*, and *3*, respectively. Semantically plausible sentences are easier than semantically implausible sentences, especially when sentences are syntactically complex (Ferreira, 2003).

Interpretations of the effects of syntactic complexity and semantic plausibility on relative clause comprehension often make reference to verbal working memory (e.g., Carpenter et al., 1994; Gibson, 2000). Working memory capacity has been conceptualized as a finite resource that can be flexibly allocated to maintenance and computational or integration functions. Complex span tests, such as reading span, require simultaneous storage and processing and are based on this conceptualisation. In the reading span task (Daneman & Carpenter, 1980), participants must retain a number of sentence-final words while reading and comprehending sentences.

Comprehension of object-relatives is thought to impose a higher working memory load than subject-relatives. Subject-relatives can be structured and interpreted incrementally, but this is not the case for object-relatives. One implication is that in object-relatives, the first noun in *Sentence 1*, *woman*, must be held in an active state until the verb, *helped*, is encountered later in the sentence. This imposes a maintenance load on working memory. Another implication of non-incremental interpretation relates to thematic role assignments. The verbs in object-relatives occur in close proximity toward the end of the sentence, so assigning nouns to the thematic roles of the verbs (e.g., agent, patient) is concentrated in this region of the sentence. This imposes an integration cost (Gibson, 2000) or computational load (Carpenter et al., 1994) on working memory, which is higher for object-relatives than for subject-relatives.

King and Just (1991, Experiment 1) contrasted semantically neutral object-relatives and subject-relatives similar to *Sentences 1* and *3*, respectively, using a methodology in which sentences were presented visually, one word at a time. On-line reading times were recorded and end-of-sentence comprehension probes were presented. Participants were classified into high- and low-span groups based on their reading spans. High-span participants had better comprehension overall than low-span participants, but the difference was greater for object-relatives than for subject-relatives. The comprehension data and the low-span participants' longer reading times at the verbs were interpreted as showing that object-relatives imposed a higher demand on verbal working memory than subject-relatives (King & Just, 1991). In Experiment 2, semantic cues were provided in the relative clause, the main clause, in both clauses, or in neither clause. The presence of semantic cues in the relative clause improved low-span participants' comprehension of the relative clauses, and semantic cues in the main clause improved their comprehension of the main clauses. Low-span participants were able to use semantic information locally but were less able to use it to constrain their responses when the other clause was probed. High-span participants' comprehension was more accurate than low-span participants', especially for the more difficult relative clauses, and it was less dependent on the presence of semantic cues. King and Just concluded that working memory capacity plays an important role in comprehending syntactically complex object-relative sentences and in using semantic information to comprehend complex sentences.

Other findings are not entirely consistent with King and Just's (1991) conclusions. In an eye-movement study, Traxler et al. (2005, Experiment 1) found no evidence that working memory capacity moderated the difference in reading times between object- and subject-relatives, similar to *Sentences 1* and *3*, respectively. In Experiment 2, helpful semantic information in the form of inanimate sentential subjects greatly reduced the magnitude of object-subject penalty on relative clause reading times. The influence of animacy on the

object-subject penalty was moderated by working memory capacity. In sentences with animate sentential subjects, the object-subject penalty was significant in all span groups. Use of inanimate sentential subjects eliminated the object-subject penalty for high-span participants. For mid-span and low-span participants, the object-subject penalty was reduced but not eliminated. Traxler et al. (2005) concluded that higher span participants were more sensitive to semantic cues than lower span participants. Higher span participants benefited more from helpful semantic cues when assigning nouns to the thematic roles of the verbs. Thus, Traxler et al.'s (2005) findings suggest that working memory capacity plays a role in using semantic information to process complex sentences, but it does not play a role in syntactic processing.

Andrews et al. (2006) reported three experiments in which comprehension of semantically neutral object- and subject-relatives similar to *Sentences 1* and *2*, respectively, was examined. Their multiple regression analyses showed that performance on working memory tasks including reading span accounted for variance in comprehension of object-relatives, but not when comprehension of subject-relatives was controlled by entering it as a predictor in the analyses. This is not consistent with King and Just's (1991) conclusion that working memory plays a role in syntactic processing. Andrews et al. (2006) found that comprehension of object-relatives relied more heavily on a domain-general capacity to process complex relational information than on verbal working memory capacity. The effect of semantic plausibility was not examined in this study.

The effects of semantic plausibility and syntactic complexity on comprehension have been examined using different types of sentences. Ferreira (2003, Experiment 1) presented active and passive sentences that were either plausible or implausible. After each sentence, comprehension was assessed by having participants identify either the agent (the "do-er") or patient (the "acted-on") of the verb. There was no significant effect of plausibility on comprehension of simple active sentences. However, on the more complex passives, comprehension was significantly more accurate for plausible than implausible sentences. A similar pattern was observed in Experiment 3, in which subject-cleft (e.g., "It was the man that bit the dog") and object-cleft (e.g., "It was the dog that the man bit") sentences were contrasted. There was no effect of plausibility on comprehension of subject-cleft sentences. However, on the more complex object-clefts, comprehension was significantly more accurate for plausible than implausible sentences.

Ferreira's (2003) findings are consistent with the view that we do not always conduct a full syntactic analysis of the sentence we are attempting to comprehend. A full syntactic analysis would involve applying morpho-syntactic constraints on thematic role assignment. This would have allowed Ferreira's participants to determine "who did what to whom" and to respond correctly to the do-er and acted-on questions following plausible and implausible sentences. Ferreira argued that we use shallower forms of heuristic processing in combination with syntactic algorithms to interpret sentences. Heuristics, such as semantic plausibility, yield a "good enough" rather than a precise linguistic representation of the sentence meaning (Ferreira, 2003).

Gibson, Bergen, and Piantadosi (2013) also observed a greater reliance on semantic plausibility for passive than active sentences. Gibson et al. interpreted their findings for this and four other pairs of constructions in terms of the *noisy channel hypothesis*. The noisy channel approach assumes that the input to language comprehension is noisy due to environmental factors and errors made by producers and perceivers. Perceivers compute the most likely interpretation given the cues available to them. When implausible sentences are presented, perceivers might infer that the producer intended a different, more plausible meaning. This is more likely when the implausible sentence differs minimally from its plausible alternative. Their findings were consistent with predictions, although some challenges for noisy channel models remain, as noted by Traxler (2014).

Our focus in the current research is on situations in which syntactic processing and semantic plausibility yield conflicting interpretations of sentences and how the conflict is resolved. Ye and Zhou (2009) reported an fMRI study in which they examined the role of domain-general executive control in resolving the conflict between interpretations based on syntactic analysis and semantic plausibility. While in the scanner, participants read active and passive Chinese sentences that were plausible or implausible. Following each sentence, a probe statement was presented and participants judged whether it was consistent or inconsistent with the preceding sentence. They also completed the Stroop Color-Word task and the flanker task. These are non-linguistic tasks that require executive control of a perceptual or a response conflict. There was substantial overlap in brain regions associated with conflicting interpretations of the sentences and those associated with congruency effects in the Stroop and flanker tasks. A network of frontal-parietal structures was implicated. The findings indicated that domain-general mechanisms are recruited to deal with conflicts during sentence comprehension.

The issue of how syntactic and semantic systems interact during sentence comprehension is analogous to how logical (analytic) and heuristic (e.g., belief-based) processes interact during reasoning. Analytical processing in reasoning can be likened to syntactic processing in sentence comprehension because both involve interpreting the underlying structure. Belief-based processing in reasoning can be likened to semantic processing in sentence comprehension in that both involve a reliance on existing semantic knowledge. Analytic and belief-based processing have been examined in several types of deductive reasoning problems, including categorical syllogisms (Evans & Curtis-Holmes, 2005; Stupple & Ball, 2008), relational reasoning (Banks, 2013; Roberts & Sykes, 2003), and transitive inference (Andrews, 2010; Andrews & Hewitt-Stubbs, 2015; Andrews & Mihelic, 2014; Banks & Hope, 2014).

The *conclusion evaluation technique* is typically employed. Premises (e.g., "The man is bigger than the elephant" and "The elephant is bigger than the horse") and a conclusion (e.g., "Therefore, the man is bigger than the horse") are presented either simultaneously or sequentially. Participants accept the conclusion if they judge that it follows logically from the premises and reject it otherwise. The four problem types are: valid believable, valid unbelievable (as in the example), invalid

believable, and invalid unbelievable. The conclusion follows logically from the premises in valid problems, but not in invalid problems. The conclusion is consistent with real-world knowledge in believable problems, but not in unbelievable problems. Analytic processing and belief-based processes yield the same conclusion in valid believable and invalid unbelievable problems, which are called *no-conflict* problems. Analytic and belief-based processes yield different responses in valid unbelievable and invalid believable problems, which are called *conflict* problems. Reasoning is more accurate on non-conflict than conflict problems. Domain-general capacities, including fluid intelligence (Andrews, 2010; Andrews & Hewitt-Stubbs, 2015; Andrews & Mihelic, 2014; Newstead, Handley, Harley, Wright, & Farrelly, 2004; Stanovich & West, 2008) and relational processing (Andrews, 2010), have been shown to predict analytic responding in the presence of conflicting semantic content.

Current Research

The conclusion evaluation methodology (described above) was adapted to examine syntactic and semantic processing in comprehension of relative clause sentences. Each item consisted of a sentence, followed by a statement. Participants read each sentence, then evaluated whether the statement was consistent with the preceding sentence. They also rated their level of confidence in each decision.

Syntactic processing and semantic processing were inferred from acceptance rates for four item types: consistent plausible (CP), consistent implausible (CImp), inconsistent plausible (IncP), and inconsistent implausible (IncImp). The statements were consistent with a correct syntactic analysis of the sentences in the consistent items, but not in the inconsistent items. The statements were consistent with real-world knowledge in the plausible items, but not in the implausible items. Thus, in the consistent plausible and the inconsistent implausible items, there was no conflict between interpretations based on syntactic analysis and semantic processing. The statements for consistent plausible items could be accepted on the grounds of either syntactic consistency or semantic plausibility. The statements for inconsistent implausible items could be rejected on the grounds of either syntactic inconsistency or semantic implausibility. By contrast, in the consistent implausible and inconsistent plausible items, there was a conflict between interpretations based on syntactic processing and semantic plausibility. In the consistent implausible items, a correct syntactic analysis implied acceptance of the statement, whereas the semantically implausible content implied rejection. In the inconsistent plausible items, a correct syntactic analysis implied rejection of the statement, whereas the semantically plausible content implied acceptance. In this procedure, use of syntactic processing is indicated by significantly higher acceptance of consistent than inconsistent statements. Use of semantic processing is indicated by significantly higher acceptance of plausible than implausible statements.

In reasoning research, reliance on belief-based processing is sometimes greater on problems with invalid than valid conclusions (Evans & Curtis-Holmes, 2005). Presenting the four item types and using the statement evaluation procedure allowed us to examine whether reli-

ance on semantic processing is greater for inconsistent than consistent items.

It was expected that reliance on semantic processing would increase with syntactic complexity. This was examined by contrasting items that included object-relative and subject-relative sentences. We contrasted object-relative sentences, such as *Sentence 1*, with subject-relative sentences, such as *Sentence 2*, because these sentence forms express the same noun-verb relations. An additional structural variation was that the object- and subject-relative sentences either included or did not include a prepositional phrase, as described in the Method section.

Based on findings from reasoning research (Andrews, 2010) and Ferreira's (2003) findings for active versus passive and subject- versus object-cleft sentences, we expected greater reliance on semantic processing and less reliance on syntactic processing when complexity was higher (sentences with object-relative clauses and/or prepositional phrases) than when complexity was lower (sentences with subject-relative clauses and/or no prepositional phrases).

To the extent that participants rely on both semantic plausibility and syntactic analysis, they should experience conflict when these processes yield different responses. Conflict effects on comprehension and confidence were examined. Comprehension accuracy (derived from acceptance rates) and confidence were expected to be higher for no-conflict items than for conflict items. Conflict effects were expected to be greater for more complex sentences.

Potential links between syntactic processing in conflict items and two measures of cognitive ability, namely, *fluid intelligence* (Culture Fair Test, Cattell & Cattell, 1973) and verbal working memory capacity (reading span, Daneman & Carpenter, 1980; Daneman & Merikle, 1996), were examined. Fluid or *analytic* intelligence is the ability to reason and solve problems involving novel information without extensive reliance on declarative knowledge. Based on previous research in reasoning, in which fluid intelligence predicted analytic processing in conflict problems (Andrews, 2010; Andrews & Hewitt-Stubbs, 2015; Andrews & Mihelic, 2014), and Ye and Zhou's (2009) finding that a network of frontal-parietal structures underpins resolution of conflicts in sentence comprehension and domain-general executive tasks, we predicted that fluid intelligence would account for significant variance in accuracy of conflict items after controlling for comprehension of no-conflict items. This hypothesis was examined in items containing subject-relative sentences, object-relative sentences, sentences with prepositional phrases, and sentences without prepositional phrases.

On the basis of previous research suggesting that working memory plays a role in syntactic (King & Just, 1991) and semantic processing (King & Just, 1991; Traxler et al., 2005), it seemed plausible that working memory might be required to resolve the conflict between responses based on syntactic analysis and semantic plausibility. If so, then reading span should account for variance in comprehension accuracy of conflict items after controlling for no-conflict items. The shared and unique contributions of fluid intelligence and working memory were examined. We conducted two experiments investigating these research questions.

EXPERIMENT 1

Method

PARTICIPANTS

The participants were 98 first-year undergraduate students (70 females, 28 males) at the Gold Coast campus of Griffith University, Australia, who participated in return for course credit. Their ages ranged from 17 to 49 years (M_{age} = 22.95, SD = 7.28). All participants reported their first language as English.

MATERIALS AND PROCEDURES

Participants attended a single session that lasted approximately 1.25 hrs. They completed the sentence evaluation task, reading span, and the Culture Fair Test of Intelligence (Cattell & Cattell, 1973). These tasks were administered in the same order to groups of up to six participants. The research protocols for Experiments 1 and 2 were approved by the institutional ethics committee, and students gave written informed consent before participating.

In the sentence evaluation task, there were two 32-item sets (set A and set B). Each item consisted of a sentence and a statement. The sentences were object-relative and subject-relative sentences. Fifty percent of the object- and subject-relatives were of the forms shown in Sentences 4 and 5, respectively, and repeated here:

The comedian that the millionaire funded left town.   (4)

The millionaire funded the comedian that left town.   (5)

The remaining sentences were similar, except that they included a prepositional phrase. Examples of object-relatives and subject-relatives with prepositional phrases are shown in Sentences 7 and 8, respectively:

The art dealer that the detective warned the buyer about sold fakes.   (7)

The detective warned the buyer about the art dealer that sold fakes.   (8)

Semantic content and sentence form were balanced across sets A and B, such that the semantic content of object-relatives in set A matched the semantic content of subject-relatives in set B, whereas the semantic content of subject-relatives in set A matched semantic content of object-relatives in set B. Fifty percent of the sentences were semantically plausible and the remainder were implausible.

A statement was generated for each sentence. Each statement was either consistent or inconsistent with the correct syntactic analysis of the sentence. This resulted in the aforementioned four item types: CP, CImp, IncP, and IncImp. As noted above, the CImp and IncP items embodied a syntax-plausibility conflict, but there was no such conflict in the CP and the IncImp items. Within each item type, two items had object-relative sentences with prepositional phrases, two items had object-relative sentences with no prepositional phrase, two items had subject-relative sentences with prepositional phrases, and two items had subject relative sentences with no prepositional phrases. The statements used with object-relatives in set A were used with subject-relatives in set B, and vice versa. Statements referring to the relative clause and the main clause of the sentence and expressed using active and passive constructions occurred with equal frequency across item types and sentence forms. By varying the form of the sentences and statements as described, we sought to encourage respondents to comprehend entire sentences (rather than focussing on a specific clause) and to ensure that they were less able to anticipate the precise form

TABLE 1.
Examples of Object- and Subject-Relative Sentences and Probe Statements

Form	Sentences	Probe Statement	Item Type
Object	The banker that the teacher instructed enjoyed the meal	The teacher instructed the banker	CP
Subject	The teacher instructed the banker that enjoyed the meal		
Object	The art dealer that the detective warned the buyer about sold fakes	The buyer was warned about the art dealer	
Subject	The detective warned the buyer about the art dealer that sold fakes		
Object	The student that the professor despised graded the essay	The essay was graded by the student	CImp
Subject	The professor despised the student that graded the essay.		
Object	The waitress that the customer criticised the chef to owned the restaurant.	The waitress owned the restaurant.	
Subject	The customer criticised the chef to the waitress that owned the restaurant.		
Object	The doctor that the lawyer visited drafted the contract	The lawyer drafted the contract	IncP
Subject	The lawyer visited the doctor that drafted the contract		
Object	The father that the girl transferred the funds to travelled overseas.	The father transferred the funds.	
Subject	The girl transferred the funds to the father that travelled overseas.		
Object	The detective that the barber respected investigated the crime	The barber investigated the crime	IncImp
Subject	The barber respected the detective that investigated the crime		
Object	The governor that the lobbyist sent the businessman to signed the legislation	The legislation was signed by the lobbyist	
Subject	The lobbyist sent the businessman to the governor that signed the legislation.		

of the sentences and statements. Examples of the sentences and statements are shown in Table 1.

In a pilot study, a separate group of fourteen undergraduate students with a mean age of 20.54 years (SD = 2.11) rated the plausibility of the 32 statements presented without the sentences. Plausibility was rated on a 7-point scale, from 1 (*highly implausible*) to 7 (*highly plausible*). Single sample t tests confirmed that the mean plausibility rating of statements that were intended to be plausible (M = 5.97, SD = 0.50) was significantly higher than the neutral midpoint (4 on the plausibility scale), $t(13)$ = 14.62, p < .001. The mean plausibility rating of statements that were intended to be implausible (M = 3.01, SD = 0.65) was significantly lower than the neutral midpoint, $t(13)$ = −5.64, p < .001.

In the study proper, items were presented on PowerPoint slides projected onto a screen at the front of the room. Each sentence was displayed for 15 s, followed by the statement for 5 s, then a *respond now* slide for 3 s, during which participants made their responses. Participants read each sentence, then the statement. Then they indicated whether the statement was consistent or inconsistent with the preceding sentence by writing a cross in the *accept* column or *reject* column on the response form. Participants also rated their confidence in their decisions using a 3-point scale (0 = *low*, 1 = *medium*, and 2 = *high*).

Four practice items with neutral semantic content were presented to ensure that students understood the procedure. In two practice items, the sentences were object-relatives, one with a consistent statement, the other with an inconsistent statement. In two practice items, the sentences were subject-relatives, one with a consistent statement, the other with an inconsistent statement. Feedback was provided after each practice item. Participants were encouraged to read the sentences for comprehension while they were visible on the screen.

Following the sentence evaluation task, working memory was assessed using a version of the reading span task (Daneman & Carpenter, 1980) that was based on Andrews et al. (2006) and Turner and Engle (1989). The stimuli were 46 sentences, 11 to 17 words in length. The final words were one-syllable, high frequency nouns. Half the sentences were made nonsensical by reversing the order of the last four to six pre-terminal words. For example, "The possum took the apple from the sill and then disappeared into the night" became "The possum took the apple from the sill the into disappeared then and night." Sentences were assigned to three sets at each set size (2, 3, 4, and 5) and two practice sets at set size 2. The sentences were presented using PowerPoint software. The slide presentation times were controlled by the software. The response sheets had spaces for *make sense* judgements on one side, and *recall of final words* on the reverse side. Two practice sets were administered to ensure that participants understood the procedure. Sentences were printed in Times Roman lettering (font size 32) and each was presented on a separate slide for 8 s. After each sentence, a slide with the words "Make sense?" was displayed for 3 s, then the next sentence appeared. Participants were instructed to read each sentence, then record their make sense judgements by circling "Yes" or "No" on the response sheet. When all sentences in the set had been presented, a

slide with the word "Recall" was displayed, accompanied by a hammer sound, and participants attempted to write the final words of the sentences in the set on the response sheet. The time allowed for recalling the final words was 5, 6, 7, and 8 s for sets of 2, 3, 4, and 5 sentences, respectively. Set size began at 2 and increased by 1 after every third set of sentences, up to a maximum set size of 5. Reading span scores were calculated by subtracting the number of make sense judgement errors from the number of final words correctly recalled (Maximum of 42).

Fluid intelligence was assessed using the Culture Fair Intelligence Test, Scale 2, Form A (Cattell & Cattell, 1973). The four subtests (Series, Classifications, Matrices, and Conditions) were administered using the standard instructions and time limits. The items require participants to induce the relations between the stimuli (geometric shapes or figures) and to make inferences based on these relations. Correct responses for the four subtests were summed and the total score (Maximum of 46) was used in the analyses.

ANALYTIC APPROACH

A recent trend in psycholinguistic research is to report mixed linear models analyses, which treat items as well as participants as random factors. Consistent with this trend, we attempted to test six models (corresponding to the three dependent variables in each of the present Experiments 1 and 2). The contributions of the random factors to overall variance were very small according to BIC in Bayesian estimation and proved to be unstable with maximum likelihood models failing to converge. We concluded that mixed effects models were not necessary for our data. Therefore, in both experiments, we used analyses of variance (ANOVA) to analyse acceptance rates, accuracy rates, and confidence rates. These analyses were conducted using SPSS version 22.

Results

Preliminary analyses revealed no significant effects or interactions involving gender, so this variable was not included in analyses for Experiment 1.

ACCEPTANCE

To the extent that evaluations of the statements were based on participants' syntactic analysis of the sentences, acceptance rates should be higher for consistent than inconsistent statements. Our main prediction was that this consistency effect would be stronger for the less complex sentences than for more complex sentences. To the extent that evaluations were based on semantic processing, acceptance rates should be higher for plausible than implausible statements. Our main prediction was that the plausibility effect would be stronger for the more complex sentences than for the less complex sentences.

A mixed ANOVA with acceptance rates as the dependent variable was conducted. The within-subject variables were Consistency (2), Plausibility (2), Form (2), and Prepositional Phrase (2). Set (2) was a between-subjects variable.

There were significant effects of consistency, $F(1, 96)$ = 5291.53, p < .001, partial η^2 = .982, plausibility, $F(1, 96)$ = 15.31, p < .001, partial η^2 = .138, form, $F(1, 96)$ = 9.86, p = .002, partial η^2 = .093, prepositional

phrase, $F(1, 96) = 13.18$, $p < .001$, partial $\eta^2 = .121$, and set, $F(1, 96) = 4.14$, $p = .045$, partial $\eta^2 = .041$. Acceptance rates were higher for consistent than inconsistent statements, for plausible than implausible statements, for statements that referred to subject- than to object-relative sentences, for statements that referred to sentences with rather than without prepositional phrases, and for set B than A.

The consistency effect interacted with the two structural variables. As predicted, there was a significant Form × Consistency interaction, $F(1, 96) = 66.96$, $p < .001$, partial $\eta^2 = .411$. The means suggest a stronger consistency effect for subject-relatives (consistent: $M = .971$, $SE = .006$; inconsistent: $M = .024$, $SE = .007$) than object-relative sentences (consistent: $M = .869$, $SE = .013$; inconsistent: $M = .077$, $SE = .010$). The significant Prepositional Phrase × Consistency interaction, $F(1, 96) = 38.67$, $p < .001$, partial $\eta^2 = .287$, indicated a stronger consistency effect for sentences without prepositional phrases (consistent: $M = .964$, $SE = .008$; inconsistent: $M = .034$, $SE = .008$) than for those with prepositional phrases (consistent: $M = .875$, $SE = .012$; inconsistent: $M = .066$, $SE = .009$).

These two-way interactions were modified by a significant Prepositional Phrase × Form × Consistency interaction, $F(1, 96) = 24.08$, $p < .001$, partial $\eta^2 = .201$, which is shown in Figure 1. For each sentence form, the effect of consistency was highly significant, $F(1, 96) = 1,818.62$, $p < .001$, partial $\eta^2 = .949$ (for object-relatives), and $F(1, 96) = 7,568.69$, $p < .001$, partial $\eta^2 = .987$ (for subject-relatives). The effect of prepositional phrase was significant for object-relative items, $F(1, 96) = 11.18$, $p = .001$, partial $\eta^2 = .103$, but not for subject-relative items ($p = .105$). The Prepositional Phrase × Consistency interaction was significant for object-relatives, $F(1, 96) = 36.52$, $p < .001$, partial $\eta^2 = .274$, and for subject relatives, $F(1, 96) = 4.381$, $p = .039$, partial $\eta^2 = .043$. Follow-up tests showed that the effect of prepositional phrase was significant for consistent object-relative items, $F(1, 97) = 34.64$, $p < .001$, partial $\eta^2 = .263$, for inconsistent object-relative items, $F(1, 97) = 11.85$, $p = .001$, partial $\eta^2 = .109$, and for consistent subject-relative items, $F(1, 97) = 5.50$, $p = .021$, partial $\eta^2 = .054$, but not for inconsistent subject-relative items ($p = .78$).
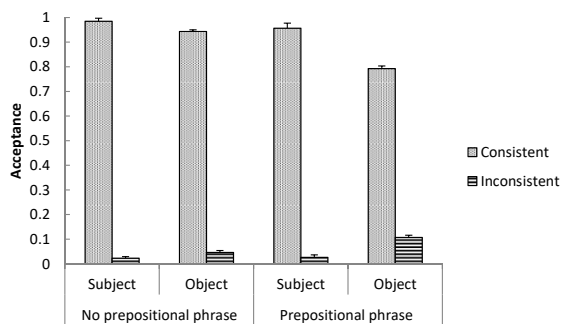
To further understand the three-way interaction, the effects of consistency were examined separately for object- and subject-relatives with and without prepositional phrases. The consistency effects were highly significant for subject-relative without prepositional phrases, $F(1, 97) = 7521.28$, $p < .001$, partial $\eta^2 = .987$, subject-relative with prepositional phrases, $F(1, 97) = 3929.11$, $p < .001$, partial $\eta^2 = .976$, object-relative without prepositional phrases, $F(1, 97) = 2468.92$, $p < .001$, partial $\eta^2 = .962$, and object-relative with prepositional phrases, $F(1, 97) = 481.91$, $p < .001$, partial $\eta^2 = .832$. Comparison of the effect sizes and inspection of Figure 1 indicate that although participants' discrimination between consistent and inconsistent statements was very accurate overall, responses were least accurate for object-relative sentences with prepositional phrases. The pattern is consistent with a lesser reliance on syntactic processing as sentence structure becomes more complex.

Plausibility also interacted with the two structural variables. As predicted, the main effect of plausibility was modified by a significant Plausibility × Form interaction, $F(1, 96) = 14.28$, $p < .001$, partial $\eta^2 = .129$, which is shown in Figure 2. The plausibility effect was significant for object-relatives, $F(1, 97) = 21.12$, $p < .001$, partial $\eta^2 = .179$, but not for subject-relatives, $F(1, 97) = 1.28$, $p = .26$.

The significant Plausibility × Prepositional Phrase interaction, $F(1, 96) = 13.09$, $p < .001$, partial $\eta^2 = .120$, was due to a significant effect of plausibility when prepositions were present, $F(1, 97) = 21.42$, $p < .001$, partial $\eta^2 = .181$ (implausible: $M = .440$, $SE = .009$; plausible: $M = .501$, $SE = .009$), but not when prepositional phrases were absent, $F(1, 97) = 1.43$, $p = .24$ (implausible: $M = .494$, $SE = .007$; plausible: $M = .505$, $SE = .007$). The Prepositional Phrase × Form × Plausibility interaction was not significant. These patterns are consistent with a greater reliance on semantic plausibility when the sentence structure is complex either because the sentences are object-relatives or because they included prepositional phrases.

There were three interactions involving the set variable. The significant two-way interactions of Set × Form, $F(1, 96) = 4.05$, $p = .047$, partial $\eta^2 = .04$, and Set × Prepositional Phrase, $F(1, 96) = 6.25$, $p = .014$, partial $\eta^2 = .061$, were modified by the significant three-way interaction
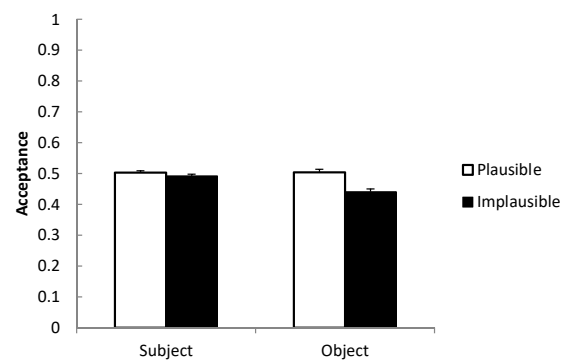

FIGURE 1.

The significant Prepositional Phrase × Form × Consistency interaction on acceptance rates in Experiment 1. Error bars represent standard errors.

FIGURE 2.

The significant Plausibility × Form interaction on acceptance rates in Experiment 1. Error bars represent standard errors.

of Set × Prepositional Phrase × Form, $F(1, 96) = 16.21$ $p < .001$, partial $\eta^2 = .144$. For participants who received set A, there were significant effects of prepositional phrase, $F(1, 48) = 18.98$, $p < .001$, partial $\eta^2 = .283$, and form, $F(1, 48) = 12.07$, $p < .001$, partial $\eta^2 = .201$, and a significant Prepositional Phrase × Form interaction, $F(1, 48) = 21.61$, $p < .001$, partial $\eta^2 = .310$. Acceptance of statements referring to object-relative sentences with prepositional phrases ($M = .411$, $SE = .015$) was significantly lower than for statements referring to object-relatives without prepositional phrases ($M = .495$, $SE = .009$), $F(1, 48) = 27.87$, $p < .001$, partial $\eta^2 = .367$. For subject–relatives, this effect was not significant (prepositional phrases present: $M = .495$, $SE = .009$; prepositional phrases absent: $M = .500$, $SE = .007$). These effects were not significant for participants who received set B. For set B, the means and standard errors were .490 and .015, respectively, for object-relatives with prepositional phrases, .487 and .011, respectively, for object-relatives without prepositional phrases; .487 and .009, respectively, for subject–relatives with prepositional phrases, and .508 and .007, respectively, for subject–relatives without prepositional phrases. Acceptance rates were significantly higher in set B than set A for object-relatives with prepositional phrases, $F(1, 96) = 14.33$, $p < .001$, partial $\eta^2 = .130$, but not for object-relatives without prepositional phrases, nor for subject-relatives with or without prepositional phrases, ($F < 1.0$ in each case).

The non-significant Consistency × Plausibility interaction showed that the plausibility effect did not differ for consistent versus inconsistent items. Reliance on semantic processing was not significantly greater when participants evaluated inconsistent versus consistent statements. There were no other significant interactions for acceptance rates.

CONFLICT EFFECT IN COMPREHENSION ACCURACY

Acceptance rates were converted to comprehension accuracy and expressed as percentages. Acceptance responses were correct for the consistent items and incorrect for inconsistent items. Rejection responses were correct for inconsistent items and incorrect for consistent items. As noted previously, syntax-semantic conflict was present for the CImp and IncP items and absent for CP and IncImp items. Accuracy scores were computed separately for object- and subject-relatives with and without prepositional phrases. In line with acceptance rates, our main prediction was that accuracy would be lower for conflict items than for no-conflict items, especially for the more complex sentences.

A $2 \times 2 \times 2 \times 2$ (Conflict × Form × Prepositional Phrase × Set) mixed ANOVA yielded significant effects of conflict, $F(1, 96) = 15.31$, $p < .001$, partial $\eta^2 = .138$, form, $F(1, 96) = 66.96$, $p < .001$, partial $\eta^2 = .411$, and prepositional phrase, $F(1, 96) = 38.67$, $p < .001$, partial $\eta^2 = .287$. Comprehension was more accurate for no-conflict than conflict sentences, for subject- than object-relatives, and for sentences without than with prepositional phrases.

There was a significant Conflict × Form interaction, $F(1, 96) = 14.28$, $p < .001$, partial $\eta^2 = .129$, which is shown in Figure 3. For object-relatives, accuracy was significantly higher for no-conflict than for conflict items, $F(1, 97) = 21.12$, $p < .001$, partial $\eta^2 = .179$. For subject-relatives, accuracy for no-conflict and for conflict items did not differ significantly, $p = .26$. There was a significant Conflict ×
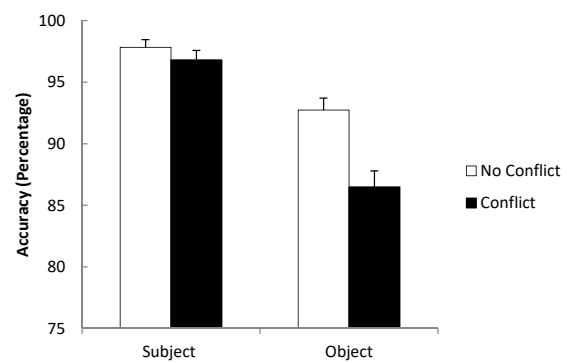

FIGURE 3.

The significant Conflict × Form interaction on comprehension accuracy in Experiment 1. Error bars represent standard errors.

Prepositional Phrase interaction, $F(1, 96) = 13.09$, $p < .001$, partial $\eta^2 = .120$. For items with prepositional phrases, accuracy was significantly higher for no-conflict ($M = 93.50$, $SE = 0.876$) than for conflict items ($M = 87.37$, $SE = 1.29$), $F(1, 97) = 21.42$, $p < .001$, partial $\eta^2 = .181$. For items without prepositional phrases, accuracy for no-conflict ($M = 97.07$, $SE = 0.74$) and for conflict items ($M = 95.92$, $SE = 0.85$) did not differ significantly, $p = .24$.
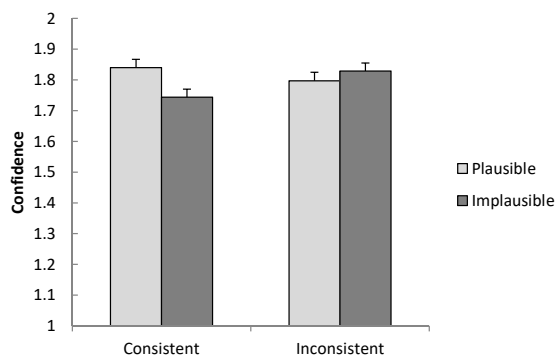
The significant Prepositional Phrase × Form interaction, $F(1, 96) = 24.08$, $p < .001$, partial $\eta^2 = .201$, indicated that inclusion of a prepositional phrase significantly reduced accuracy in object-relatives, $F(1, 97) = 36.52$, $p < .001$, partial $\eta^2 = .274$ (no prepositional phrase: $M = 94.90$, $SE = 0.91$; prepositional phrase: $M = 84.31$, $SE = 1.57$), and to a lesser, but still significant, extent in subject relatives, $F(1, 97) = 4.38$, $p = .039$, partial $\eta^2 = .043$ (no preposition phrase: $M = 98.09$, $SE = 0.56$; prepositional phrase: $M = 96.56$, $SE = 0.73$). There were no other significant effects or interactions for accuracy.
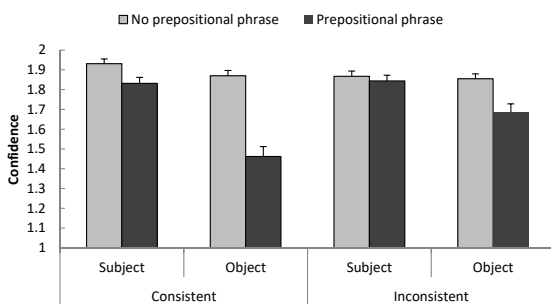
CONFIDENCE

Our main prediction was that confidence would be lower in item types in which a syntax-plausibility conflict was present (CImp and IncP) than in no-conflict (CP and IncImp) items. The impact of structural variables (form, prepositional phrase) was also examined.

A $2 \times 2 \times 2 \times 2 \times 2$ (Consistency × Plausibility × Form × Prepositional Phrase × Set) mixed ANOVA with Consistency, Plausibility, Form, and Prepositional Phrase as within-subject variables and Set as a between-subjects variable yielded significant main effects of consistency, $F(1, 96) = 5.34$, $p = .023$, partial $\eta^2 = .053$, form, $F(1, 96) = 62.23$, $p < .001$, partial $\eta^2 = .393$, and prepositional phrase, $F(1, 96) = 60.784$, $p < .001$, partial $\eta^2 = .388$. Confidence was higher for inconsistent than consistent items, for subject- than object-relative items, and higher for items without prepositional phrases than with them.

The significant Consistency × Plausibility interaction, $F(1, 96) = 14.95$, $p < .001$, partial $\eta^2 = .135$, is shown in Figure 4. For consistent statements, confidence was significantly higher for plausible than implausible items, $F(1, 96) = 11.31$, $p = .001$, partial $\eta^2 = .105$. For inconsistent statements, the plausibility effect approached significance, $F(1$,

FIGURE 4.

The significant Consistency × Plausibility interaction on confidence in Experiment 1. Error bars represent standard errors.

FIGURE 5.

The significant Consistency × Form × Prepositional Phrase interaction on confidence in Experiment 1. Error bars represent standard errors.
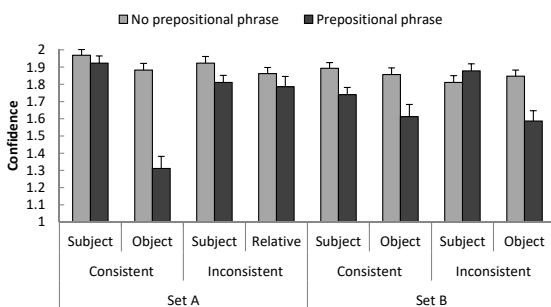
FIGURE 6.

The significant Consistency × Form × Prepositional Phrase × Set interaction on confidence in Experiment 1. Error bars represent standard errors.

96) = 3.71, p = .057, partial η^2 = .037. There was a trend for confidence to be higher for the inconsistent than consistent statements. The disordinal nature of this interaction and the higher confidence for CP and IncImp items than for CImp and IncP items suggests that respondents detected the syntax-plausibility conflict when it was present.

The significant Consistency × Form interaction, $F(1, 96)$ = 18.96, $p < .001$, partial η^2 = .165, the significant Consistency × Prepositional Phrase interaction, $F(1, 96)$ = 24.24, $p < .001$, partial η^2 = .202, and the significant Form × Prepositional Phrase interaction, $F(1, 96)$ = 51.33, $p < .001$, partial η^2 = .348, were in turn modified by a Consistency × Form × Prepositional Phrase interaction, $F(1, 96)$ = 9.21, p = .003, partial η^2 = .087, which is shown in Figure 5. The interactions reflected the differential effect of prepositional phrase for consistent and inconsistent object- and subject-relatives. The presence of prepositional phrases significantly reduced confidence for consistent object-relatives, $F(1, 97)$ = 15.44, $p < .001$, partial η^2 = .137, inconsistent object-relatives, $F(1, 97)$ = 66.20, $p < .001$, partial η^2 = .406, and consistent subject-relatives, $F(1, 97)$ = 24.84, $p < .001$, partial η^2 = .204, but not for inconsistent subject-relatives (p = .45).

There were also interactions involving the set variable. The significant Form × Set interaction, $F(1, 96)$ = 5.79, p = .018, partial η^2 = .057, and the significant Consistency × Set interaction, $F(1, 96)$ = 4.05, p = .047, partial η^2 = .040, were further modified by a significant Consistency × Form × Set interaction, $F(1, 96)$ = 34.71, $p < .001$, partial η^2 = .266. In turn, these two- and three-way interactions were modified by a significant Consistency × Form × Prepositional Phrase × Set interaction, $F(1, 96)$ = 54.69, $p < .001$, partial η^2 = .363, which is shown in Figure 6. The Consistency × Form × Prepositional Phrase interaction was significant in set A, $F(1, 48)$ = 47.98, $p < .001$, partial η^2 = .500, and in set B, $F(1, 48)$ = 10.98, p = .002, partial η^2 = .186. The interaction effect was somewhat stronger for set A than set B.

RESOLVING SYNTAX-PLAUSIBILITY CONFLICTS IN OBJECT-RELATIVES AND SUBJECT-RELATIVES

Multiple regression analyses were used to examine the nature of the cognitive load imposed by resolving the syntax-plausibility conflict. In the first analysis, the criterion was comprehension of object-relative conflict items, and the predictors were comprehension of object-relative no-conflict items, working memory (M = 34.21, SD = 6.53), and fluid intelligence (M = 36.18, SD = 4.09). The set variable was significant in the analyses reported above, so it was included as a control variable. The predictors together accounted for 29.6% variance in comprehension of object-relative conflict items, $F(4, 93)$ = 9.76, $p < .001$. Working memory (13.5%) accounted for significant unique variance, but the other predictors did not. Thus, people with higher working memory capacity were better able to resolve the syntax-plausibility conflict in object-relatives, after controlling for fluid intelligence, set, and comprehension of object-relative items in which there was no such conflict. When the analysis was re-conducted after excluding reading span, total variance was reduced to 16.1%. Fluid intelligence (4.84%) and object-relative no-conflict items (6.4%) each accounted for unique variance.

In the second analysis, the criterion was comprehension of subject-relative conflict items, and the predictors were subject-relative no-conflict items, working memory, fluid intelligence, and set. The predictors together accounted for 23.1% variance in comprehension of subject-relative conflict items, $F(4, 93)$ = 6.97, $p < .001$. Fluid intel-

TABLE 2.

Zero-Order Correlations and Standard Multiple Regression of No-Conflict Item Accuracy, Working Memory, Fluid Intelligence, and Set on Conflict Item Accuracy for Items With Object-Relative (Upper) and Subject-Relative (Lower) Sentences in Experiment 1

	1.	2.	3.	4.	B	β	sr^2(unique)	p
1. Object conflict								
2. Object no-conflict	.30				.161	.132	.015	.162
3. Working Memory	.51	.36			.849	.443	.135	< .001
4. Fluid intelligence	.27	.18	.49		.095	.031	.001	.757
5. Set	.15	.01	-.00	-.02	3.83	.154	.024	.081
	1.	2.	3.	4.	B	β	sr^2(unique)	p
1. Subject conflict								
2. Subject no-conflict	.20				.067	.055	.003	.581
3. Working Memory	.26	.36			.043	.036	.001	.737
4. Fluid Intelligence	.45	.27	.49		.787	.419	.132	< .001
5. Set	-.15	-.12	-.00	-.02	-2.069	-.135	.018	.143

Notes. N = 98

ligence (13.18%) accounted for significant unique variance, but the other predictors did not. Thus, people with higher fluid intelligence were better able to resolve the syntax-plausibility conflict in subject-relative items after controlling for working memory, set, and comprehension of subject-relative items in which there was no such conflict. These analyses are summarized in Table 2. When fluid intelligence was excluded, total variance was reduced to 9.9% and reading span (4.37%) accounted for unique variance.

RESOLVING SYNTAX-PLAUSIBILITY CONFLICTS IN PRESENCE/ABSENCE OF PREPOSITIONAL PHRASES

These multiple regression analyses were repeated after grouping the items in terms of whether they included prepositional phrases. In the first analysis, the criterion was comprehension of conflict items with prepositional phrases and the predictors were no-conflict items with prepositional phrases, working memory, fluid intelligence, and set. The predictors together accounted for 24.6% variance in comprehension of conflict items with prepositional phrases, $F(4, 93) = 7.57$, $p < .001$. Working memory (5.86%) accounted for significant unique variance, but the other predictors did not. Thus, people with higher working memory capacity were better able to resolve the syntax-plausibility conflict in sentences with prepositional phrases after controlling for fluid intelligence, set, and comprehension of comparable sentences in which there was no such conflict. When the analysis was re-conducted after excluding reading span, total variance was reduced to 18.7%. Fluid intelligence (9.24%) and no-conflict items with prepositional phrases (5.66%) each accounted for unique variance.

When the criterion was comprehension of conflict items without prepositional phrases and the predictors were no-conflict items without prepositional phrases, working memory, fluid intelligence, and set, 18.3% variance in comprehension of conflict items without prepositional phrases was accounted for, $F(4, 93) = 5.22$, $p = .001$. Fluid intelligence (3.96%) and no-conflict items without prepositional phrases (3.06%) each accounted for unique variance, but working memory did not. These analyses are summarized in Table 3. When fluid intelligence

was excluded, total variance accounted for was reduced to 14.4%. Working memory (6.45%) and no-conflict items without prepositional phrases (3.76%) each accounted for unique variance.

Discussion

The acceptance rates show that participants used both semantic plausibility and syntactic analysis to interpret the sentences. However, they relied more on semantic plausibility and less on syntactic analysis when syntactic structure was more complex either because the sentences were object-relatives rather than subject-relatives, or because they contained prepositional phrases, or both. The findings are consistent with the interpretation that we rely on syntactic and non-syntactic information (e.g., semantic plausibility) to interpret sentences (Ferreira, 2003; Gibson et al., 2013; Traxler, 2014).

A cautionary note relates to the finding of higher acceptance rates in set B than set A for statements referring to the structurally complex object-relative sentences with prepositional phrases. The reason for this is unclear. One potential explanation is that the semantic content of the items was responsible. If so, higher acceptance rates for subject-relatives with prepositional phrases might be expected in set A than set B because the same semantic content was employed for object-relatives in set B as for subject-relatives in set A. However, this was not the case. The difference might stem from some other characteristics of the items or from differences between the participants who received set A versus set B. This issue will be addressed further in the General Discussion section.

When acceptance rates were expressed in terms of comprehension accuracy and the item types were characterized in terms of the presence/absence of a syntax-plausibility conflict, conflict items were more difficult than no-conflict items, but only for object-relative items and items that included prepositional phrases. For the less complex subject-relative items and items without prepositional phrases in which syntactic processing was easier, the presence of conflicting semantic information did not have a significant impact on accuracy.

TABLE 3.

Zero-Order Correlations and Standard Multiple Regression of No-Conflict Item Accuracy, Working Memory, Fluid Intelligence and Set on Conflict Item Accuracy for Items With Prepositional Phrases (Upper) and Without Prepositional Phrases (Lower) in Experiment 1

	1.	2.	3.	4.	B	β	sr^2(unique)	p
1. Conflict_Prep[a]								
2. No-conflict_Prep[b]	.30				.190	.129	.013	.207
3. Working Memory	.45	.30			.596	.305	.058	.009
4. Fluid intelligence	.36	.19	.49		.574	.184	.026	.078
5. Set	-.05	.07	-.00	-.02	-.141	-.056	.003	.540
	1.	2.	3.	4.	B	β	sr^2(unique)	p
1. Conflict_No-prep[c]								
2. No-conflict_No-prep[d]	.27				.210	.185	.031	.065
3. Working Memory	.33	.29			.205	.160	.019	.149
4. Fluid Intelligence	.35	.22	.49		.468	.228	.040	.037
5. Set	.03	-.12	-.00	-.02	.967	.058	.003	.541

Notes. N = 98. [a] Conflict_Prep = Accuracy on conflict items with prepositional phrases; [b] No-conflict_Prep = Accuracy on no-conflict items with prepositional phrases; [c] Conflict_No-Prep = Accuracy on conflict items without prepositional phrases; [d] No-Conflict_No-Prep = Accuracy on no-conflict items without prepositional phrases.

The higher confidence ratings for no-conflict (CP and IncImp) items than for conflict (CImp and IncP) items suggest that participants noticed the syntax-plausibility conflict, but the lower accuracy for conflict than no-conflict items shows that the conflict was not always resolved in favour of the syntactically correct interpretation.

Fluid intelligence has been linked to the ability to resolve conflict between analytic and heuristic processing in reasoning tasks (Andrews, 2010; Andrews & Hewitt-Stubbs, 2015; Andrews & Mihelic, 2014; Stanovich & West, 2008). In the current study, participants with higher fluid intelligence were better able to resolve the syntax-plausibility conflict in the less complex subject-relative items and items with no prepositional phrases. Participants with higher working memory capacity were better able to resolve this conflict in object-relative items and in items with prepositional phrases.

Reading span and fluid intelligence were significantly correlated, and they accounted for shared variance in comprehension, as shown by the follow-up analyses. Fluid intelligence accounted for variance in object-relative conflict items and conflict items with prepositional phrases when reading span was excluded as a predictor. Reading span accounted for variance in subject-relative conflict items and conflict items without prepositional phrases when fluid intelligence was excluded. However, reading span accounted for more variance in conflict resolution in the more complex items than fluid intelligence did.

The regression analyses suggest that resolving syntax-plausibility conflicts in complex sentences imposes an increased demand on working memory. As outlined in the Introduction, the non-incremental way in which object-relatives must be interpreted means that they impose higher demands than comparable subject-relative sentences on storage and/or integration functions of working memory. Similarly, the extra length of sentences with prepositional phrases would increase their working memory demand as compared to sentences without these phrases. The need to resolve the syntax-plausibility conflict might load

working memory further, such that the total load exceeds the working memory capacity of many individuals. This would explain why reading span accounted for variance in the more complex conflict items after controlling for the corresponding non-conflict items.

However, sequential presentation of the sentences and statements might also have imposed an additional load on working memory. On each item, the sentence was presented first, then the statement was presented. The sentence and statement were not available simultaneously. The need to retain the sentence while reading and comprehending the statement might load working memory further. This possibility was explored in Experiment 2.

EXPERIMENT 2

The purpose of Experiment 2 was to better understand the finding in Experiment 1 that verbal working memory capacity (reading span) accounted for variance in performance on the structurally complex conflict items independently of the other predictors. To do this, we minimized one source of the working memory load. Whereas sequential presentation of the sentences and statements was employed in Experiment 1, simultaneous presentation was employed in Experiment 2. The purpose of this change was to eliminate retention of the sentence as a source of working memory load. All other aspects of the stimuli and procedure were identical to Experiment 1. If the working memory load imposed by the object-relative conflict items and conflict items with prepositional phrases was due to resolving the syntax-plausibility conflict, reading span should be a significant predictor in the multiple regression analyses, as it was in Experiment 1. However, if the variance accounted for by reading span was due to retention of the sentence while the statement was processed, reading span would be less likely to emerge as a significant predictor in the regression analyses.

Method

PARTICIPANTS

The participants were 75 first-year undergraduate students (57 females, 18 males) at the Gold Coast campus of Griffith University, Australia, who participated in return for course credit. Their ages ranged from 17 to 55 years (M_{age} = 22.53, SD = 7.37). All participants reported their first language as English.

MATERIALS AND PROCEDURE

In the sentence evaluation task, the sentences and statements were the same as in Experiment 1. However, the following changes were made to the presentation. Each sentence and its probe statement were printed on the same slide. The sentence was displayed alone in the upper half of the slide for 10 s. Then the statement appeared in the lower half of the slide. The statement and the sentence were displayed simultaneously for a further 10 s. This was followed by a respond now slide for 3 s, during which participants made their responses. Reading span and fluid intelligence were assessed in the same manner as in Experiment 1.

Results and discussion

Preliminary analyses revealed no significant effects of set and no interactions involving set on any dependent variable, therefore set was dropped from all analyses.

ACCEPTANCE

A 2 × 2 × 2 × 2 (Consistency × Plausibility × Form × Prepositional Phrase) ANOVA with Consistency, Plausibility, Form, and Prepositional Phrase as within-subject variables yielded significant main effects of consistency, $F(1, 74)$ = 4,236.13, p < .001, partial η^2 = .983, plausibility, $F(1, 74)$ = 8.86, p = .004, partial η^2 = .107, and prepositional phrase, $F(1, 74)$ = 27.21, p < .001, partial η^2 = .269. The effect of form approached significance, $F(1, 74)$ = 2.93, p = .091, partial η^2 = .038. Acceptance rates were higher for consistent than inconsistent statements, for plausible than implausible statements, for statements that referred to subject- than to object-relative sentences, and for statements that referred to sentences with prepositional phrases than without them.

As in Experiment 1, the consistency effect interacted with two structural variables. There was a significant Form × Consistency interaction, $F(1, 74)$ = 47.68, p < .001, partial η^2 = .392. The means suggest a stronger consistency effect for subject-relatives (consistent: M = .973, SE = .007; inconsistent: M = .005, SE = .003) than object-relative sentences (consistent: M = .880, SE = .016; inconsistent: M = .065, SE = .014). The significant Prepositional Phrase × Consistency interaction, $F(1, 74)$ = 40.03, p < .001, partial η^2 = .35, indicated a stronger consistency effect for sentences without prepositional phrases (consistent: M = .978, SE = .006; inconsistent: M = .025, SE = .008) than for those with prepositional phrases (consistent: M = .875, SE = .016; inconsistent: M = .045, SE = .009).

These two-way interactions and the significant Prepositional Phrase × Form interaction, $F(1, 74)$ = 7.62, p = .007, partial η^2 = .093, were

modified by a significant Prepositional Phrase × Form × Consistency interaction, $F(1, 74)$ = 21.167, p < .001, partial η^2 = .222, which is shown in Figure 7. For both sentence forms, the effect of consistency was highly significant, $F(1, 74)$ = 1175.43, p < .001, partial η^2 = .941 (for object-relatives), and $F(1, 74)$ = 16,499.01, p < .001, partial η^2 = .996 (for subject-relatives). The effect of prepositional phrase was significant for object-relative items, $F(1, 74)$ = 20.64, p < .001, partial η^2 = .218, and for subject-relative items, $F(1, 74)$ = 5.11, p = .027, partial η^2 = .065. The Prepositional Phrase × Consistency interaction was significant for object-relatives, $F(1, 74)$ = 34.55, p < .001, partial η^2 = .318, and also for subject-relatives, $F(1, 74)$ = 4.01, p = .049, partial η^2 = .051. Further tests showed that, as in Experiment 1, the effect of prepositional phrase was significant for consistent object-relative items, $F(1, 74)$ = 39.27, p < .001, partial η^2 = .347, for inconsistent object-relative items, $F(1, 74)$ = 5.80, p = .019, partial η^2 = .073, and for consistent subject-relative items, $F(1, 74)$ = 5.29, p = .024, partial η^2 = .067, but not for inconsistent subject-relative items (p = .18).

The effects of consistency were also examined separately for object-relatives and subject-relatives with and without prepositional phrases. The consistency effects were highly significant for subject-relative without prepositional phrases, $F(1, 74)$ = 10,148.28, p < .001, partial η^2 = .993, subject-relative with prepositional phrases, $F(1, 74)$ = 6,028.79, p < .001, partial η^2 = .988, object-relative without prepositional phrases, $F(1, 74)$ = 2230.14, p < .001, partial η^2 = .968, and object-relative with prepositional phrases, $F(1, 74)$ = 349.89, p < .001, partial η^2 = .825. Comparison of the effect sizes and Figure 7 are consistent with a lesser reliance on syntactic processing as sentence structure becomes more complex.

The significant Plausibility × Form interaction, $F(1, 74)$ = 7.06, p = .010, partial η^2 = .087, shown in Figure 8, was due to a significant plausibility effect for object-relatives, $F(1, 74)$ = 9.07, p = .004, partial η^2 = .109, but not for subject-relatives, $F(1, 74)$ = 1.48, p = .228, partial η^2 = .02. As in Experiment 1, semantic plausibility affected evaluation of statements in the more complex object-relative sentences, but not in the less complex subject-relative sentences. The Plausibility × Prepositional Phrase interaction on acceptance rates was not significant in Experiment 2.
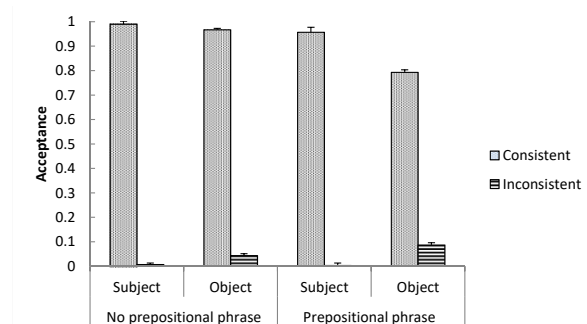

FIGURE 7.

The significant Prepositional Phrase × Form × Consistency interaction on acceptance rates in Experiment 2. Error bars represent standard errors.
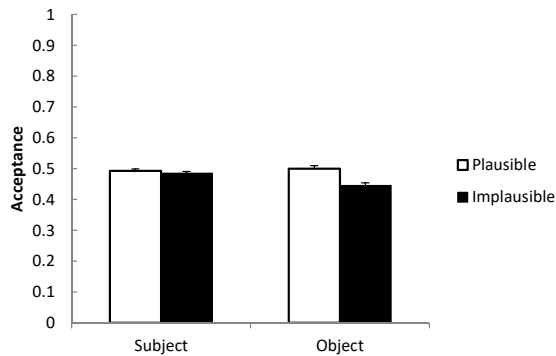
FIGURE 8.

The significant Plausibility × Form interaction on acceptance rates in Experiment 2. Error bars represent standard errors.

CONFLICT EFFECTS IN COMPREHENSION ACCURACY

Acceptance rates were converted to accuracy. A $2 \times 2 \times 2$ (Conflict × Form × Prepositional Phrase) ANOVA yielded significant effects of conflict, $F(1, 74) = 8.86$, $p = .004$, partial $\eta^2 = .107$, form, $F(1, 74) = 47.68$, $p < .001$, partial $\eta^2 = .392$, and prepositional phrase, $F(1, 74) = 40.03$, $p < .001$, partial $\eta^2 = .351$, and a significant Conflict × Form interaction, $F(1, 74) = 7.06$, $p = .01$, partial $\eta^2 = .087$, which is shown in Figure 9. For object-relatives, accuracy was significantly higher for no-conflict than for conflict items, $F(1, 74) = 9.07$, $p = .004$, partial $\eta^2 = .109$. For subject-relatives, accuracy for no-conflict and for conflict items did not differ significantly, $p = .228$. As in Experiment 1, the effect of syntax-plausibility conflict was evident only on the more complex object-relative items. The Conflict × Prepositional Phrase interaction was not significant in Experiment 2.

There was also a significant Form × Prepositional Phrase interaction, $F(1, 74) = 21.17$, $p < .001$, partial $\eta^2 = .222$. While the effect of prepositional phrase on comprehension was significant for subject-relatives, $F(1, 74) = 4.01$, $p = .049$, partial $\eta^2 = .051$ (no prepositional phrase: $M = 99.17$, $SE = 0.43$; prepositional phrase: $M = 97.67$, $SE =$
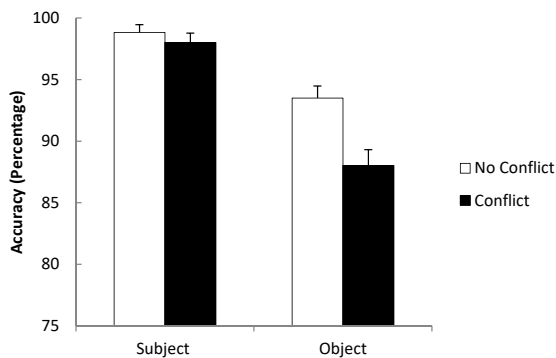
0.61), and for object relatives, $F(1, 74) = 34.55$, $p < .001$, partial $\eta^2 = .318$ (no prepositional phrase: $M = 96.17$, $SE = 0.98$; prepositional phrase: $M = 85.33$, $SE = 1.89$), the effect was larger for object-relatives.

CONFIDENCE

Preliminary analyses revealed a significant effect of gender, $F(1, 73) = 4.10$, $p = .047$, partial $\eta^2 = .053$. Males ($M = 1.90$, $SE = .06$) reported higher confidence than females ($M = 1.76$, $SE = .03$). Gender did not interact with any other variables, so the analyses excluding gender are reported.

A $2 \times 2 \times 2 \times 2$ (Consistency × Plausibility × Form × Prepositional Phrase) ANOVA, with Consistency, Plausibility, Form and Prepositional Phrase as within-subject variables, yielded significant main effects of consistency, $F(1, 74) = 17.63$, $p = .001$, partial $\eta^2 = .192$, form, $F(1, 74) = 80.49$, $p < .001$, partial $\eta^2 = .521$, and prepositional phrase, $F(1, 74) = 86.51$, $p < .001$, partial $\eta^2 = .539$. Confidence was higher for subject- than object-relative items, for inconsistent than consistent statements, and for items without rather than with prepositional phrases.

The significant Consistency × Plausibility interaction, $F(1, 74) = 9.43$, $p = .003$, partial $\eta^2 = .113$, is shown in Figure 10. For consistent items, confidence was higher for plausible than implausible statements, $F(1, 74) = 6.45$, $p = .013$, partial $\eta^2 = .08$. For inconsistent items, confidence was higher for implausible than plausible statements, $F(1, 74) = 6.71$, $p = .012$, partial $\eta^2 = .083$. As in Experiment 1, confidence was sensitive to the syntax-plausibility conflict.

The significant Consistency × Form interaction, $F(1, 74) = 7.50$, $p = .008$, partial $\eta^2 = .092$, the significant Consistency × Prepositional Phrase interaction, $F(1, 74) = 21.42$, $p < .001$, partial $\eta^2 = .224$, and the significant Form × Prepositional Phrase interaction, $F(1, 74) = 70.10$, $p < .001$, partial $\eta^2 = .486$, were in turn modified by a Consistency × Form × Prepositional Phrase interaction, $F(1, 74) = 7.07$, $p = .010$, partial $\eta^2 = .087$, which is shown in Figure 11. The interaction reflected the differential effect of prepositional phrase for consistent and inconsistent object- and subject-relatives. As in Experiment 1, the presence of prepositional phrases significantly reduced confidence for consistent object-relatives, $F(1, 74) = 82.56$, $p < .001$, partial $\eta^2 = .527$, inconsistent

FIGURE 9.

The significant Conflict × Form interaction on comprehension in Experiment 2. Error bars represent standard errors.
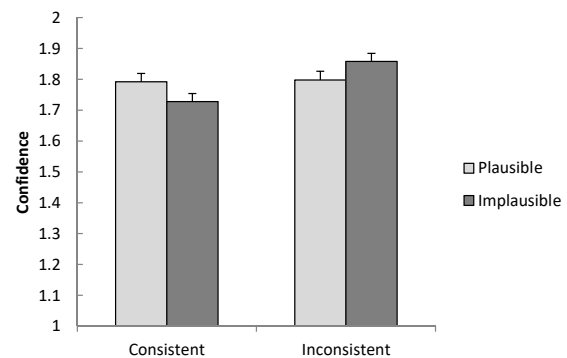

FIGURE 10.

The significant Consistency × Plausibility interaction on confidence in Experiment 2. Error bars represent standard errors.
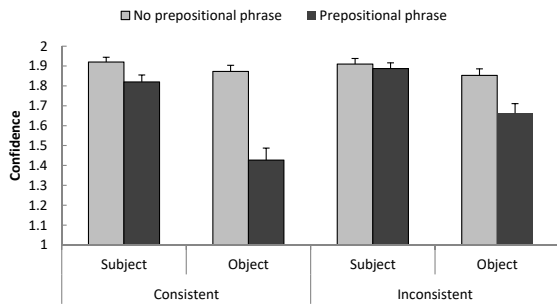
FIGURE 11.

The significant Consistency × Form × Prepositional Phrase interaction on confidence in Experiment 2. Error bars represent standard errors.

object-relatives $F(1, 74) = 38.31$, $p < .001$, partial $\eta^2 = .341$, and consistent subject-relatives, $F(1, 74) = 10.57$, $p = .002$, partial $\eta^2 = .125$, but not for inconsistent subject-relatives ($p = .40$).

RESOLVING SYNTAX-PLAUSIBILITY CONFLICTS IN OBJECT-RELATIVES AND SUBJECT-RELATIVES

In the first multiple regression analysis, the criterion was comprehension of object-relative conflict items and the predictors were comprehension of object-relative no-conflict items, working memory ($M = 35.23$, $SD = 5.10$), and fluid intelligence ($M = 35.84$, $SD = 4.28$). The predictors together accounted for 17.2% variance in comprehension of object-relative conflict items, $F(3, 71) = 4.91$, $p = .004$. Fluid intelligence (5.8%) accounted for significant unique variance, but the other predictors did not.

In the second analysis, the criterion was comprehension of subject-relative conflict items and the predictors were comprehension of subject-relative no-conflict items, working memory, and fluid intelligence. The predictors together accounted for 14.2% variance in comprehen-

sion of subject-relative conflict items, $F(3, 71) = 3.91$, $p = .012$. Fluid intelligence (11.02%) accounted for significant unique variance, but the other predictors did not. These analyses are summarised in Table 4. Contrary to Experiment 1, reading span was not a significant predictor in these analyses, and this remained the case when fluid intelligence was excluded as a predictor.

RESOLVING SYNTAX-PLAUSIBILITY CONFLICTS IN PRESENCE/ABSENCE OF PREPOSITIONAL PHRASES

These multiple regression analyses were repeated for items with and without prepositional phrases. In the first analysis, the criterion was comprehension of conflict items with prepositional phrases and the predictors were no-conflict items with prepositional phrases, working memory, and fluid intelligence. The predictors together accounted for 21.1% variance in comprehension of conflict items with prepositional phrases, $F(3, 71) = 6.31$, $p = .001$. Fluid intelligence (6.50%) and no-conflict items (5.66%) each accounted for significant unique variance, but working memory did not. Thus, people with higher fluid intelligence were better able to resolve the syntax-plausibility conflict in sentences with prepositional phrases after controlling for working memory and comprehension of comparable sentences in which there was no such conflict. When the analysis was re-conducted after excluding fluid intelligence, total variance accounted for was reduced to 14.6%. Comprehension of no-conflict items with prepositional phrases (11.36%) accounted for unique variance, but working memory did not.

When the criterion was comprehension of conflict items without prepositional phrases and the predictors were no-conflict items without prepositional phrases, working memory, and fluid intelligence, the predictors together accounted for 14.0% variance in comprehension of conflict items without prepositional phrases, $F(3, 71) = 3.84$, $p = .013$. Fluid intelligence (6.76%) accounted for unique variance, but working memory did not. When the analysis was re-conducted after excluding fluid intelligence, the regression equation was no longer significant ($p = .068$).

TABLE 4.

Zero-Order Correlations and Standard Multiple Regression of No-Conflict Item Accuracy, Working Memory, and Fluid Intelligence on Conflict Item Accuracy for Items With Object-Relative (Upper) and Subject-Relative (Lower) Sentences in Experiment 2

	1.	2.	3.	B	β	sr^2(unique)	p
1. Object Conflict							
2. Object No-conflict	.30			.269	.152	.018	.217
3. Working Memory	.20	.18		.002	.072	.005	.534
4. Fluid intelligence	.39	.47	.36	.011	.288	.058	.029
	1.	2.	3	B	β	sr^2(unique)	p
1. Subject Conflict							
2. Subject No-conflict	.10			.042	.031	.001	.787
3. Working Memory	.14	-.04		.011	.011	.001	.926
4. Fluid Intelligence	.38	.19	.36	.431	.365	.110	.004

Notes. $N = 75$.

Zero-order Correlations and Standard Multiple Regression of No-conflict Item Accuracy, Working Memory, and Fluid Intelligence on Conflict Item Accuracy for Items with Prepositional Phrases (Upper) and without Prepositional Phrases (Lower) in Experiment 2

	1.	2.	3.	B	β	sr^2(unique)	p
1. Conflict_Prep[a]							
2. No-conflict_Prep[b]	.36			.393	.253	.057	.027
3. Working Memory	.18	.12		.114	.044	.002	.700
4. Fluid intelligence	.39	.34	.36	.898	.289	.065	.018
	1.	2.	3	B	β	sr^2(unique)	p
1. Conflict_No-Prep[c]							
2. No-conflict_No-Prep[d]	.20			.059	.032	.000	.804
3. Working Memory	.20	.09		.127	.083	.006	.488
4. Fluid Intelligence	.37	.49	.36	.586	.320	.068	.021

Notes. $N = 75$. [a] Conflict_Prep = accuracy on conflict items with prepositional phrases; [b] No-conflict_Prep = accuracy on no-conflict items with prepositional phrases; [c] Conflict_No-Prep = accuracy on conflict items without prepositional phrases; [d] No-Conflict_No-Prep = accuracy on no-conflict items without prepositional phrases.

GENERAL DISCUSSION

The acceptance-rate findings in Experiments 1 and 2 showed that participants relied on semantic plausibility as well as on syntactic analysis to interpret the sentences. In both experiments, there was greater reliance on semantic plausibility and a lesser reliance on syntactic analysis when the sentences being read and evaluated were more complex than when the sentences were syntactically simpler. The findings are consistent with those of Ferreira (2003), who used different sentence forms, and with interpretations that include a role for non-syntactic factors as well as syntactic algorithms in sentence interpretation (Gibson et al., 2013; Hale, 2001; Levy, 2008; Traxler, 2014). As sentence complexity increases, we rely more on non-syntactic factors, such as semantic plausibility, which yield what Ferreira (2003) described as a good enough interpretation of sentence meaning.

Acceptance rates for the most complex object-relatives with prepositional phrases were higher in set B than set A in Experiment 1 but not Experiment 2. This finding might be due to sample or item characteristics. For example, the participants in Experiment 1 who received set B might have had a stronger tendency to accept (rather than reject) consistent and inconsistent statements when they were uncertain of the correct response on these complex sentences. Alternately, it might reflect differences between the items in the two sets that impacted evaluations when sequential presentation was employed in Experiment 1, but not when simultaneous presentation was used in Experiment 2.

That rated confidence was sensitive to the syntax-plausibility conflict suggests that readers noticed the conflict when it occurred. However, the accuracy data showed that the conflict was not always resolved in favour of syntax. The regression analyses show that resolving the conflict in favour of syntax imposes a demand on cognitive resources in the form of working memory and/or fluid intelligence.

Experiment 1 suggested the involvement of working memory in resolving the syntax-plausibility conflict, especially when sentence complexity was high. However, the working memory demand in Experiment 1 appears to have stemmed primarily from the need to hold the sentence in mind while reading the statement rather than from resolving the syntax-plausibility conflict per se. When the need to retain the sentence in working memory was minimized by presenting each sentence and its probe statement simultaneously in Experiment 2, the magnitude of the conflict effect on object-relative items was reduced from .179 (Experiment 1) to .109 (Experiment 2) and working memory capacity, which was a strong predictor of comprehension of object-relative conflict items and conflict items with prepositional phrases in Experiment 1, was no longer a significant predictor. Reading span did not account for variance in comprehension of conflict items, even when fluid intelligence was excluded from the analyses. The regression analyses in Experiments 1 and 2 suggest that reading span accounts for the maintenance demands imposed during sentence comprehension to a greater extent than it accounts for computation/integration costs. Resolving the conflict between responses based on syntactic structure and semantic plausibility would appear to impose an integration cost as participants attempt to combine, integrate, and reconcile different types of information.

By contrast, fluid intelligence accounted for significant unique variance in resolving the syntax-plausibility conflict in both experiments. That the predictive ability of fluid intelligence was somewhat independent of sentence form (being significant for both object- and subject-relative conflict items and for conflict items with and without prepositional phrases) and of procedural demands (being significant with sequential and simultaneous presentation) suggests that it is involved specifically in conflict resolution rather than in some other aspect of our task. This interpretation is consistent with research demonstrating the involvement of fluid intelligence in resolving conflicts between analytic and heuristic processing in deductive reasoning (Andrews, 2010; Andrews & Hewitt-Stubbs, 2015; Andrews & Mihelic, 2014; Stanovich & West, 2008) and with the fMRI research (Ye & Zhou, 2009) supporting the involvement of domain-general processes in resolving the syntax-plausibility conflict.

An interesting finding was that fluid intelligence accounted for more unique variance in less complex subject-relative conflict items than the more complex object-relative conflict and no-conflict items. While fluid intelligence captures the computational load associated

with conflict resolution, it might not capture the higher maintenance loads imposed by object-relatives as compared to subject-relatives. Further investigation would be needed to confirm this explanation. However, comparison of the correlations of fluid intelligence with conflict and no-conflict object relatives in Experiments 1 and 2 provides some support. These correlations were somewhat stronger in Experiment 2, when maintenance load in the sentence evaluation task was reduced, than in Experiment 1, when it was higher.

The amounts of variance in conflict items accounted for by fluid intelligence might seem modest. However, it should be recognised that variance shared with reading span and with no-conflict items was controlled for in these analyses. Moreover, the non-verbal nature of the Culture Fair Test (Cattell & Cattell, 1973) means that the variance is unlikely to reflect common surface characteristics of the test items and the sentence evaluation task. Associations between reading span and sentence comprehension are open to such interpretations because both tasks involve reading and comprehending sentences.

Conflict effects in the subject-relative items and the items without prepositional phrases did not reach significance in our sample of undergraduates. However, it is possible that significant conflict effects in these less complex items would be observed in samples of aging adults (Andrews & Todd, 2008; Caplan et al., 2011; DeDe, 2015), brain injured adults (Andrews et al., 2013, 2014; Novick, Kan, Trueswell, & Thompson-Schill, 2009), or in young children (Andrews & Halford, 2002). These populations experience greater difficulty with sentence comprehension and other language tasks than young unimpaired adults do.

Our findings suggest that resolving syntax-plausibility conflicts in sentence comprehension involves integrating syntactic and semantic information. One mechanism by which this integration might be achieved is through alignment of mental representations of syntactic structure and semantic knowledge in a manner akin to analogical mapping. According to this view, the experience of conflict occurs when a misalignment between semantic knowledge and syntactic structure is detected. Structural alignment is a key process in analogical and other types of reasoning (Halford, Andrews, & Wilson, 2015; Halford, Wilson, Andrews, & Phillips, 2014; Holyoak, 2012; Smith & Gentner, 2012), and it imposes a cognitive load that increases with structural complexity (Cho, Holyoak, & Cannon, 2007). Participants with the cognitive capacity to make the alignments will be less likely to exhibit plausibility effects in sentence comprehension and also belief bias in reasoning. The role of structural alignment in plausibility effects in sentence comprehension and in belief bias in reasoning is speculative at this stage and it requires further investigation.

REFERENCES

Andrews, G. (2010). Belief-based and analytic processing in transitive inference depends on premise integration difficulty. *Memory & Cognition, 38*, 928-940. doi: 10.3758/MC.38.7.928

Andrews, G., Birney, D. P., & Halford, G. S. (2006). Relational processing and working memory capacity in comprehension of relative clause sentences. *Memory & Cognition, 34*, 1325-1340. doi: 110.3758/BF03193275

Andrews, G., & Halford, G. S. (2002). A cognitive complexity metric applied to cognitive development. *Cognitive Psychology, 45*, 153-219. doi: 110.1016/S0010-0285(02)00002-6

Andrews, G., Halford, G. S., Shum, D. K. H., Maujean, A., Chappell, M., & Birney, D. P (2013). Relational processing following stroke. *Brain & Cognition, 81*, 44-51. doi: 110.1016/j.bandc.2012.09.003

Andrews, G., Halford, G. S., Shum, D. K., Maujean, A., Chappell, M., & Birney, D. P. (2014). Verbal learning and memory following stroke. *Brain Injury, 28*, 442-447. doi: 110.3109/02699052.2014.888758

Andrews, G., & Hewitt-Stubbs, G. (2015). Further exploration of the belief bias in transitive inference. In A. M. Columbus (Ed.), *Advances in Psychology Research* (Vol. 102, pp. 117-129). New York, NY: Nova Science Publishers Inc.

Andrews, G., & Mihelic, M. (2014). Belief-based and analytic processing in transitive inference: Further evidence for the importance of premise integration. *Journal of Cognitive Psychology, 26*, 588-596. doi: 110.1080/20445911.2014.909434

Andrews, G., & Todd, J. M. (2008). Two sources of age-related decline in comprehension of complex relative clause sentences. In N. B. Johansen (Ed.), *New research on short-term memory* (pp. 93-123). New York, NY: NovaScience Publishers.

Banks, A. P. (2013). The influence of activation level on belief bias in relational reasoning. *Cognitive Science, 37*, 544-577. doi: 110.1111/cogs.12017

Banks, A. P., & Hope, C. (2014). Heuristic and analytic processes in reasoning: An event-related potential study of belief bias. *Psychophysiology, 51*, 290-297. doi: 110.1111/psyp.12169

Caplan, D., DeDe, G., Waters, G., Michaud, J., & Tripodis, Y. (2011). Effects of age, speed of processing and working memory on comprehension of sentences with relative clauses. *Psychology and Aging, 26*, 439-450. doi: 10.1037/a0021837

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences, 22*, 77-94.

Carpenter, P. A., Miyake, A., & Just, M. A. (1994). Working memory constraints in comprehension: Evidence from individual differences, aphasia, and aging. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp.1075-1122). San Diego, CA: Academic Press.

Cattell, R. B., & Cattell, A. K. S. (1973). *Culture Fair Test of "g": Scale 2*. Champaign, IL: Institute for Personality and Ability Testing.

Cho, S., Holyoak, K. J., & Cannon, T. (2007). Analogical reasoning in working memory: Resources shared among relational integration, interference resolution, and maintenance. *Memory & Cognition, 35*, 1445-1455. doi: 110.3758/BF03193614

Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 405-438. doi: 110.1016/S0022-5371(80)90312-6

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin and Review, 3*, 422-433. doi: 110.3758/BF03214546

DeDe, G. (2015). Effects of animacy on processing relative clauses in older and younger adults. *Quarterly Journal of Experimental Psychology, 68*, 487-498. doi: 110.1080/17470218.2014.956766

Evans, J. St. B. T., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning, 11*, 305-416. doi: 110.1080/13546780542000005

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47*, 164-203. doi: 110.1016/S0010-0285(03)00005-7

Frazier, L. (1985). Syntactic complexity. In D. Dowty, L. Kartunnen, & A. Zwicky (Eds.), *Natural language parsing* (pp. 129-189). Cambridge, MA.: Cambridge University Press.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In Y. Miyashita, A. Marantz, & W. O'Neil, (Eds.), *Image, language, brain* (pp. 95-126). Cambridge, MA: MIT Press.

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences of the United States of America, 110*, 8051-8056. doi: 110.1073/pnas.1216438110

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science, 13*, 425-430. doi: 110.1111/1467-9280.00475

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics and Language Technologies* (pp. 1-8). Association for Computational Linguistics. doi: 110.3115/1073336.1073357

Halford, G. S., Andrews, G., & Wilson, W. H. (2015). Relational processing in reasoning: The role of working memory. In A. Feeney, & V. A. Thompson (Eds.), *Reasoning as memory* (pp. 34-52). Hove, UK: Psychology Press.

Halford, G. S., Wilson, W. H., Andrews, G., & Phillips, S. (2014). *Categorizing cognition: Toward conceptual coherence in the foundations of psychology*. Cambridge, MA: The MIT Press.

Hansell, N. K. , Halford, G. S., Andrews, G., Shum, D. H. K., Harris, S.E., Davies, G., . . . Wright, M. J. (2015). Genetic basis of a cognitive complexity metric. *PLoS ONE, 10*: e0123886. doi:10.1371/journal.pone.0123886

Holyoak, K. J. (2012). Analogy and relational reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning*. New York, NY: Oxford University Press. doi:10.1093/oxfordhb/9780199734689.013.0013

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30*, 580-602. doi: 110.1016/0749-596X(91)90027-H

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*, 1126-1177. doi: 110.1016/j.cognition.2007.05.006

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 21086-21090. doi: 110.1073/pnas.0907664106

Levy, R., Fedorenko, E., & Gibson, E. (2013). The syntactic complexity of Russian relative clauses. *Journal of Memory & Language, 69*, 46-495. doi: 110.1016/j.jml.2012.10.005

Newstead, S. E., Handley, S. J., Harley, C., Wright, H., & Farrelly, D. (2004). Individual differences in deductive reasoning. *Quarterly Journal of Experimental Psychology, 57*A, 33-60. doi: 110.1080/02724980343000116

Novick, J. M., Kan, I. P., Trueswell, J. C., & Thompson-Schill, S. L. (2009). A case for conflict across multiple domains: Memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive Neuropsychology, 26*, 527- 567. doi: 110.1080/02643290903519367

Roberts, M. J., & Sykes, E. D. A. (2003). Belief bias and relational reasoning. *The Quarterly Journal of Experimental Psychology, 56*A, 131-154. doi: 110.1080/02724980244000233

Smith, L. A., & Gentner, D. (2012). Using spatial analogy to facilitate graph learning. In C. Stachniss, K. Schill, & D. Uttal (Eds.), *Spatial cognition* (pp. 196-209). Berlin, Germany: Springer-Verlag.

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive abilities. *Journal of Personality and Social Psychology, 94*, 672-695. doi: 110.1037/0022-3514.94.4.672

Staub, A. (2010). Eye movements and processing difficulty in object-relative clauses. *Cognition, 116*, 71-86. doi: 110.1016/j.cognition.2010.04.002

Stupple, E. J. N., & Ball, L. J. (2008). Belief-logic conflict resolution in syllogistic reasoning: Inspection time evidence for a parallel process model. *Thinking & Reasoning, 14*, 168-181. doi: 110.1080/13546780701739782

Traxler, M. J. (2011). Parsing. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*, 353-364. doi: 110.1002/wcs.112

Traxler, M. J. (2014). Trends in syntactic parsing: Anticipation, Bayesian estimation, and good-enough parsing. *Trends in Cognitive Sciences, 18*, 605-611. doi: 110.1016/j.tics.2014.08.001

Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language, 53*, 204-224. doi: 110.1016/j.jml.2005.02.010

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*, 127-154. doi: 110.1016/0749-596X(89)90040-5

Ye, Z., & Zhou, X. (2009). Conflict control during sentence comprehension: fMRI evidence. *NeuroImage, 48*, 280-290. doi: 10.1016/j.neuroimage.2009.06.032