The Influence of Commonsense Knowledge and Discourse Context on Pronoun Resolution

Authors: Wu Yan, Gao Yuefei, Zhao Simin, Wang Suiping, Wang Suiping

Date: 2018-11-20T00:00:00+00:00

Abstract

This study investigates the role and temporal dynamics of readers' commonsense knowledge and discourse context in Chinese pronoun comprehension. Experiment 1 first examined whether occupational gender stereotypes, as a type of commonsense information, could exert early influence on pronoun resolution during Chinese reading comprehension. With two experimental conditions (consistent vs. conflicting, e.g., guard-he; guard-she), results revealed gender conflict effects triggered by occupational gender stereotypes in gaze duration, regression time, and total reading time. Building upon Experiment 1, Experiment 2 added a second pronoun (e.g., guard-he-he; guard-she-she) to investigate whether updated textual context could override the influence of commonsense knowledge on pronoun processing. The results demonstrated that discourse context could indeed override commonsense information, exerting early influence on pronoun processing. However, commonsense knowledge of occupational gender stereotypes continued to play a role at later stages of pronoun processing. Given that the context created by pronouns alone is relatively implicit, Experiment 3 employed more explicit ways to define the gender of occupational terms, such as "male actor," "father," etc., before presenting a pronoun whose gender was always consistent with the preceding gender description but conflicted with occupational gender stereotypes. This also yielded two conditions (consistent vs. conflicting, e.g., nanny-wife-she; guard-wife-she). Results showed that only the updated contextual information influenced pronoun processing, with occupational gender stereotypes no longer exerting any effect. This indicates that in Chinese, a language with high context-dependency, context can override commonsense knowledge to exert early influence on pronoun processing. However, the persistence of contextual effects is modulated by the explicitness of gender information in the discourse context.

Full Text

The Effects of World Knowledge and Discourse Context on Pronoun Resolution

WU Yan[1], GAO Yuefei[1], ZHAO Simin[1], WANG Suiping[2]
([1] School of Psychology, Northeast Normal University, Changchun 130024, China)
([2] School of Psychology, South China Normal University, Guangzhou 510631, China)

Abstract

This study investigates the role and time course of readers' world knowledge and discourse context in Chinese pronoun comprehension. Experiment 1 first examined whether occupational gender stereotype, as a type of world knowledge, could exert early influence on pronoun resolution. Two conditions were created: congruent (e.g., guard-he) and incongruent (e.g., guard-she). Results revealed gender conflict effects triggered by occupational gender stereotypes in gaze duration, second reading time, and total reading time. Building upon Experiment 1, Experiment 2 added another pronoun (e.g., guard-he-he; guard-she-she) to examine whether updated information—that is, textual context—could override the effect of world knowledge on pronoun processing. The findings demonstrated that discourse context could indeed override world knowledge and influence pronoun processing at an early stage. However, this world knowledge of occupational gender stereotypes continued to exert influence during later stages of pronoun processing. Given that the context created by pronouns is relatively implicit, Experiment 3 employed more explicit ways to define the gender of occupational terms, such as "male actor" or "father," before presenting a pronoun whose gender always matched the prior gender description but conflicted with the occupational gender stereotype. This again created congruent and incongruent conditions (e.g., nanny-wife-she; guard-wife-she). The results showed that only the updated contextual information influenced pronoun processing, with occupational gender stereotypes no longer playing a role. These findings indicate that in Chinese, a language with high context dependency, discourse context can override world knowledge to exert early effects on pronoun processing. However, the persistence of contextual effects is influenced by the explicitness of gender information in the discourse context.

Keywords: pronoun; discourse context; world knowledge; occupational gender stereotype
Classification: B824

Introduction

Pronouns are words with specific functions and a special status in language, and pronoun processing constitutes an indispensable component of language comprehension. Determining the referent of a pronoun is crucial for maintaining coherence at both the sentence and discourse levels. Consequently, the investigation of pronoun resolution processes has long been a concern for psycholinguists (Garnham, 2001, a review).

As substitutes for nouns referring to people or things, pronouns carry limited semantic information (Esaulova, Reali, & Stockhausen, 2014). However, pronouns typically exhibit distinctions in gender (e.g., "他" [he] and "她" [she] in Chinese) and number (singular vs. plural). Therefore, numerous previous studies investigating pronoun processing have focused on these gender and number features, examining the pronoun processing mechanism and its influencing factors by creating mismatches between the pronoun and the antecedent' s gender, gender stereotype, or number (Canal, Garnham, & Oakhill, 2015; Duffy & Keir, 2004; Esaulova et al., 2014; Kennison & Trofe, 2003; Qiu, Chen, & Wang, 2012; Qiu, Swaab, Chen, & Wang, 2012; Xu, Jiang, & Zhou, 2013). For instance, previous research has shown that both number information and biological gender of antecedents (e.g., father-she) can affect pronoun processing (Canal et al., 2015; Xu et al., 2013). Additionally, gender stereotypes associated with antecedents (e.g., guard-she) also influence pronoun resolution processes (Duffy & Keir, 2004; Esaulova et al., 2014; Qiu et al., 2012).

However, it is important to note that violations of number or biological gender differ fundamentally from violations of gender stereotypes in pronoun resolution. The former are irreconcilable and unintegrable, whereas the latter can be coordinated and integrated. Gender stereotypes represent world knowledge acquired through everyday experience and are not necessarily accurate. For example, although a guard is most likely male, this is not necessarily always the case—a guard could also be female. Therefore, when readers encounter the sentence "The guard hurriedly stopped the incoming taxi because she needed to check the documents," they may experience processing difficulty, but they can infer from the pronoun "she" that the guard is female and integrate this information with what they have read previously.

In this case, the first occurrence of the pronoun "she" in fact constitutes a form of context that reveals the true gender of the character it refers to in the text. In a longer discourse segment, readers will likely encounter the pronoun "she"a second time. For instance, after reading the sentence "The guard hurriedly stopped the incoming taxi because she needed to check the documents," suppose readers then encounter the following segment: "This is the station' s regulation; this is also her job, and all visitors must comply." In this segment, readers encounter a second pronoun "she." At this point, two opposing factors will influence pronoun processing. Since the pronoun still refers to the guard, the gender stereotype that guards are male may continue to cause processing difficulty. Simultaneously,

the corrective information provided by the first pronoun—that the occupational gender stereotype was wrong—may facilitate processing of the second pronoun. How, then, will readers process the pronoun when these two types of information coexist?

This question in fact concerns how different types of semantic knowledge—such as semantic information provided by the text and readers' relevant world knowledge —influence the processing of current information.

Previous research has offered different theoretical perspectives on whether updated information can replace pre-existing knowledge in readers' minds to produce faster and more sustained effects on current word processing. Two-stage processing models propose that pronoun processing occurs in two distinct stages. While updated information can indeed influence current word processing, it only does so during the later stage of pronoun processing. For example, Kintsch' s Construction-Integration Model (1988) posits that comprehension of current information can be divided into construction and integration stages. During the early construction stage, readers' pre-existing concepts are first activated and participate in processing the current target, while discourse contextual information only comes into play during the integration stage. Garrod and Terras' s Bonding-Resolution Model (2000) suggests that the early stage of reading comprehension is an automatic, low-level process. During this stage, information semantically related to the current input is activated through low-level matching mechanisms. The later stage is a high-level process in which readers evaluate the matches formed during the early stage based on discourse contextual information.

In contrast to two-stage theories, an alternative perspective holds that word processing in reading comprehension can be completed in a single stage, with context exerting immediate effects on current word processing just like low-level information such as world knowledge and semantics. For example, Hess, Foss, and Carroll' s (1995) Lexical Reinterpretation Model proposes that readers reinterpret word meanings based on sentence or discourse context. To ensure coherence of sentence or discourse information, only the reinterpreted information influences current word processing.

Building upon these theoretical debates, several studies have investigated this issue (Duffy & Keir, 2004; Cook & Myers, 2004; Garrod & Terras, 2000; Nieuwland & Van Berkum, 2006b). However, evidence from pronoun processing research remains limited. A representative study in this area comes from Duffy and Keir (2004), who used eye-tracking technology to examine the roles of context and world knowledge in reflexive pronoun processing. Eye-tracking methodology enables real-time monitoring of readers' reading processes, and temporal measures such as "fixation duration" can provide effective evidence for different processing stages (Rayner & Pollatsek, 1989). For instance, first fixation duration and gaze duration are important indicators that typically reflect early lexical processing (Kliegl, Grabner, Rolfs, & Engbert, 2004). First fixation duration refers to the duration of the initial fixation on an interest area during

first-pass reading, while gaze duration, also known as first-pass reading time, refers to the total duration from the first fixation until the fixation point first leaves the current interest area. Second reading time, which refers to the sum of all fixation durations when readers regress to the current interest area, is an important indicator of later lexical processing stages (Inhoff & Liu, 1998).

Duffy and Keir (2004) measured the effects of world knowledge and context using gaze duration and second reading time as indices of early and late lexical processing stages. They presented participants with materials such as the following while recording their eye movements:

The electrician taught herself/himself a lot.

The results showed that when the above material was presented alone, readers exhibited longer gaze durations and second reading times for "herself" in the incongruent condition. However, when preceded by information such as "The electrician is a cautious woman," the difference between congruent and incongruent conditions in readers' processing of "herself" was no longer significant for either gaze duration or second reading time. This indicates that information provided by the text completely overrode world knowledge, producing immediate and sustained effects on pronoun processing—findings that provide evidence for the Lexical Reinterpretation Model proposed by Hess et al. (1995).

Nevertheless, beyond this study, no other research has been found that combines world knowledge and context to investigate their joint effects on pronoun processing. There are studies that examine the effect of world knowledge alone, such as the aforementioned research on antecedent gender stereotypes (Duffy & Keir, 2004; Esaulova et al., 2014; Kennison & Trofe, 2003; Qiu et al., 2012), and studies that investigate context alone (Chen, Cheung, Tang, & Wong, 2000; Nieuwland & Van Berkum, 2006a). However, research integrating both factors remains scarce. Evidently, evidence from pronoun processing studies on how words are processed under the combined influence of context and world knowledge is very limited.

However, we cannot simply equate non-pronoun processing with pronoun processing, as pronoun processing differs from that of other words and even other types of anaphoric expressions (e.g., "this person") (Crocker, Pickering, & Clifton, 2000; Esaulova et al., 2014). Esaulova et al. (2014) argue that because pronouns carry limited semantic information, readers must immediately identify the pronoun's referent to maintain coherence at the sentence and discourse levels. Other words, by contrast, contain rich semantic information and can often be understood without immediate reference to prior context (Esaulova et al., 2014, p. 798). This leads to differential roles of antecedent and discourse contextual information in pronoun processing versus other word types. In Esaulova et al.' s study, the researchers found that gender information carried by antecedents affected pronouns earlier than it affected other types of anaphoric expressions.

Moreover, research in this area lacks evidence from Chinese studies. Compared to English, Chinese possesses many unique characteristics. Written Chinese

is characterized as highly context-dependent (Chen et al., 2000; Wang, Chen, Yang, & Mo, 2008), manifested in the fact that the meaning of individual Chinese characters is often ambiguous. In many cases, the meaning of a single character must be determined based on context. Even pronouns, which are relatively straightforward in meaning, such as "他" (he), may function as singular third-person pronouns or as morphemes in other words (e.g., "其他"[other], "他乡" [foreign land], "排他" [exclusive]), where "他" does not refer to any specific entity and thus does not require readers to resolve its reference. By contrast, in most Western languages such as English, words are separated by fixed spaces and have relatively transparent meanings that generally do not change with surrounding words. How might this characteristic of Chinese affect the role of context in pronoun processing? One possibility is that because contextual information is frequently needed to determine word meanings, contextual effects may be faster or stronger. Another possibility, however, is that because Chinese characters and words are often semantically ambiguous, more processing resources must be allocated to determining word meanings, thereby delaying the utilization of contextual information. Overall, although many researchers believe that semantic factors play a more important role than grammatical structure in successfully comprehending written Chinese, regrettably, few studies have systematically examined how different types of semantic factors, such as context and world knowledge, operate.

Based on these considerations, we employed the sensitive eye-tracking technique to conduct three experiments investigating the effects and time course of context and world knowledge on pronoun processing in Chinese. Experiment 1 first examined whether occupational gender stereotype, as a type of world knowledge, could exert early influence on pronoun resolution in Chinese reading comprehension. The materials in Experiment 1 included a pronoun that created congruent and incongruent versions based on the relationship between the pronoun' s gender and the occupational gender stereotype (e.g., guard-he, guard-she).

Experiment 1

This experiment investigated the effect of occupational gender stereotypes on pronoun processing. Based on previous research findings, we expected that gender stereotypes would influence pronoun processing and exert effects during early stages of pronoun processing.

Method

Participants Following references (Anderson, Kelley, & Maxwell, 2017; Taylor & Muller, 1996), we set the a priori effect size to 0.8 and used G*Power 3.1 software (http://www.gpower.hhu.de/) to calculate the planned sample size of 24 participants. We actually recruited 25 participants, all of whom had normal or corrected-to-normal vision, were native Chinese speakers, and had no reading

disabilities. Three participants whose question-answering accuracy was below 75% were excluded, leaving 22 valid participants (11 male, 11 female).

Design and Materials We employed a single-factor within-subjects design with consistency as the independent variable, which had two levels: pronouns that were congruent or incongruent with the gender stereotype of the occupational term. First, we conducted a rating assessment of gender stereotypicality for occupational terms to identify those with clear gender stereotypes for use in constructing experimental materials, ensuring that occupational terms and pronouns indeed formed congruent or incongruent relationships. Thirty-two university students who did not participate in the formal experiment rated the gender stereotypicality of 80 occupational terms on a 7-point scale. For half of the participants, 1 indicated "very likely female" and 7 indicated "very likely male," while for the other half the scale was reversed. For data analysis, we standardized the scale so that 1 represented "very likely male" and 7 represented "very likely female." Based on the mean ratings, we selected 20 occupations with strong male stereotypes ($M = 2.13$, $SD = 0.40$) and 20 with strong female stereotypes ($M = 5.58$, $SD = 0.40$). An independent samples t-test revealed a significant difference between the two groups, $t(38) = 23.99$, $p < 0.001$.

We constructed experimental materials using the selected gender-stereotyped occupations. Each discourse consisted of five sentences: the first sentence introduced the occupational term, followed by two transitional sentences, then a target sentence containing the target pronoun, and finally a concluding sentence. The experimental materials were divided into congruent and incongruent versions based on the relationship between the pronoun' s gender cue and the occupational gender stereotype. For example:

Congruent condition: The magician had an important performance tonight but had a terrible headache, so he walked backstage to rest. But he found the backstage area chaotic with no place to rest.

Incongruent condition: The magician had an important performance tonight but had a terrible headache, so he walked backstage to rest. But she found the backstage area chaotic with no place to rest.

We used a Latin square design to balance sentence topics across the two experimental conditions, creating two sets of experimental materials. Each set contained 40 experimental items, with each participant receiving only one condition for each discourse.

Procedure Participants were randomly assigned to one of the two experimental sets and were tested individually. The experiment took approximately 40 minutes to complete. We used an EyeLink II eye-tracker developed by SR-Research with a sampling rate of 500 Hz. Participants' eye fixations and movements were recorded via two miniature infrared cameras mounted on a headband, and we recorded data only from the right eye. The distance between the

monitor screen and participants' eyes was approximately 75 cm, with a screen refresh rate of 150 Hz. All materials were presented in 34-point Kai font, with each Chinese character measuring 0.95 cm × 0.95 cm. Adjacent characters were separated by 0.25 cm, resulting in a visual angle of approximately 0.73° per character.

Calibration was required before each experimental session to ensure accurate recording of eye movement trajectories. Each calibration procedure included calibration, validation, and drift correction. During calibration, nine calibration points (white dots) appeared sequentially at random locations in the center or periphery of the screen. Participants were instructed to fixate on each point until it disappeared. Following calibration, validation was performed using the same nine-point procedure. If validation was successful, drift correction was conducted, during which a single calibration point appeared randomly in the center or periphery of the screen, and participants were required to fixate on it until it disappeared. Only after successful calibration could participants proceed to the formal experiment.

The 66 discourses were presented in 22 groups of three. The specific experimental procedure is illustrated in Figure 1 [Figure 1: see original paper]. Before each group, a prompt "Group X" was displayed. The experimental materials were then presented, with each discourse displayed on a single screen in multiple lines, with the first line indented by two spaces. All materials were presented as white text on a black background. After each group, a comprehension question requiring a "yes" or "no" response was presented to encourage careful reading. Participants controlled their own reading pace and pressed any button on the response pad to advance to the next screen. Before presenting prompts, discourses, or questions, the computer performed drift correction to ensure eye-tracking accuracy. Specifically, a calibration point appeared in the upper left corner of the screen, and the next screen only appeared when participants' fixation coincided with this point. Three practice trials were administered before the experiment to familiarize participants with the procedure.

Results and Analysis

Fixations shorter than 60 ms or longer than 600 ms were excluded from analysis as they are considered not to reflect normal reading processes (Rayner, 1998; Angele, Slattery, Yang, Kliegl, & Rayner, 2008; Qiu et al., 2012).

Chinese singular personal pronouns consist of a single character and have relatively high word frequency. Consequently, we found a very high skipping probability for pronouns in this experiment, with first-pass skipping rates exceeding 70%. This is consistent with previous Chinese studies (Qiu et al., 2012) and similar to the pattern observed for English singular personal pronouns (Ehrlich & Rayner, 1983). Since statistical analyses are based on fixated characters, high skipping rates cause substantial data loss and reduce result reliability. To obtain more valid data, researchers often combine the pronoun with surrounding

linguistic stimuli into a single interest area for analysis (Van Gompel & Majid, 2004; Qiu et al., 2012). The analysis regions included the pronoun region and the post-pronoun region. The pronoun region comprised the pronoun and the two characters before and after it (with approximately 10 pronouns preceded by punctuation marks), while the post-pronoun region consisted of the three characters immediately following the pronoun region. The post-pronoun region was primarily examined for potential delayed or spillover effects.

The measures used in data analysis included first fixation duration, gaze duration, second reading time, and total reading time. The first three measures were introduced earlier. Total fixation duration refers to the sum of all fixation durations within an interest area and is sensitive to slower and longer cognitive processing (Kliegl et al., 2004). Based on analyses of these four measures, we can obtain information about early, late, and overall pronoun processing. In the following analyses, t1 refers to t-values from participant-based analyses, while t2 refers to t-values from item-based analyses, and 95%CI represents the 95% confidence interval. Means and standard deviations for each interest area are presented in Table 1 .

Table 1 Means and Standard Deviations (ms) of Eye Movement Measures in the Pronoun Region and Post-Pronoun Region

Region	Condition	First Fixation	Gaze Duration	Second Reading Time	Total Reading Time
Pronoun	Congruent	225 (37)	217 (22)	123 (104)	381 (137)
Pronoun	Incongruent	228 (30)	224 (26)	136 (99)	365 (110)
Post-Pronoun	Congruent	296 (59)	277 (63)	99 (70)	340 (104)
Post-Pronoun	Incongruent	262 (45)	265 (56)	121 (110)	347 (108)

Note: Standard deviations are shown in parentheses.

Analysis of the pronoun region data revealed no significant difference in first fixation duration between the two conditions, t1(21) = 1.52, p = 0.43, d = 0.33; t2(39) = 1.72, p = 0.093, d = 0.27. However, compared to the congruent condition, the incongruent condition showed longer gaze duration, t1(21) = 2.25, p = 0.036, d = 0.50; t2(39) = 2.06, p = 0.046, d = 0.33; 95% CI = [1.4, 36]; longer second reading time, t1(21) = 2.02, p = 0.056, d = 0.43; t2(39) = 1.95, p = 0.058, d = 0.31; 95% CI = [-0.7, 48]; and longer total reading time, t1(21) = 3.08, p = 0.006, d = 0.66; t2(39) = 2.93, p = 0.006, d = 0.47; 95% CI = [13, 67]. This conflict effect, however, did not extend to the post-pronoun region. Analysis of the post-pronoun region revealed no significant differences between the two conditions in first fixation duration, gaze duration, second reading time, or total reading time (all ps > 0.2). According to Cohen' s (1988) criteria, the effect sizes for gaze duration and total reading time were medium to large,

further confirming the influence of occupational gender stereotypes on pronoun processing.

These results indicate that during Chinese discourse reading, readers can quickly associate gender information carried by occupational terms with pronouns, and conflicting world knowledge thus impedes pronoun comprehension. This finding is consistent with Duffy and Keir's (2004) results on English pronoun processing but differs slightly from Esaulova et al.' s (2014) findings on German pronoun processing, where occupational gender stereotypes influenced later processing measures such as total reading time. This suggests that the time course of antecedent gender stereotype effects varies across languages. The present findings build upon previous research by confirming that pronoun resolution can proceed very rapidly even in a highly context-dependent language like Chinese. How, then, does corrected information influence pronoun processing when the text contains information that updates this world knowledge of gender stereotypes? Experiment 2 addresses this question.

Experiment 2

The materials in Experiment 2 contained two pronouns that were always consistent with each other and formed congruent and incongruent conditions with the gender stereotype of the antecedent occupational term. Therefore, for the second pronoun, two sources of information would influence its processing. First, since the pronoun' s antecedent remained the occupational term, the occupational gender stereotype would continue to influence pronoun processing. Second, the gender information provided by the first pronoun would also affect pronoun processing. Thus, if the contextual information provided by the first pronoun could update readers' pre-existing gender stereotypes about the occupational term and influence pronoun processing, then readers should not experience difficulty processing the second pronoun in the incongruent condition. However, if the occupational gender stereotype continued to influence pronoun processing despite contextual cues, reading times for the second pronoun would still be longer in the incongruent condition compared to the congruent condition.

Method

Participants As before, the planned sample size was 24 participants. However, considering that post-hoc statistical power might not reach the desired level, we actually recruited 32 participants (16 male, 16 female) with a mean age of 20 years. All participants had normal or corrected-to-normal vision, were native Chinese speakers, and had no reading disabilities.

Design and Materials We employed a single-factor within-subjects design with consistency as the independent variable, which had two levels: pronouns that were congruent or incongruent with the occupational gender stereotype.

The experimental materials were similar to those in Experiment 1, also consisting of five clauses. Unlike Experiment 1, the first target sentence appeared immediately after the opening sentence, containing the first pronoun that was either congruent or incongruent with the occupational gender stereotype. This was followed by a transitional sentence, then a second target sentence containing the second pronoun, which was always consistent with the first pronoun. The discourse concluded with a final sentence. Moreover, the position of the second pronoun in Experiment 2 was approximately the same as the pronoun position in Experiment 1. Example sentences:

Congruent condition: The magician was stunned by the sudden incident—this was something he had not expected. The performance was ruined, which was really unfair to him; he should never have hired such an assistant.

Incongruent condition: The magician was stunned by the sudden incident—this was something she had not expected. The performance was ruined, which was really unfair to her; she should never have hired such an assistant.

We used a Latin square design to balance sentence topics across the two experimental conditions, creating two sets of experimental materials. Each set contained 40 experimental items, with each participant receiving only one condition for each topic. Additionally, we included 47 filler items with a structure similar to the experimental materials. Each set thus contained 87 discourses, presented in groups of three for a total of 29 groups.

Procedure The procedure was identical to Experiment 1.

Results and Analysis

Participants' question-answering accuracy was above 81%, indicating that they read the materials carefully. As before, fixations shorter than 60 ms or longer than 600 ms were excluded from analysis. Data analysis was also conducted on interest areas, including T1 (the first pronoun and the two characters before and after it, with approximately 10 pronouns preceded by commas), T1+1 (consisting of the three characters following the T1 region), T2 (the second pronoun and the two characters before and after it), and T2+1 (consisting of the three characters following the T2 region).

The experimental results are shown in Table 2 . When readers first encountered a pronoun that violated the occupational gender stereotype of its antecedent, a conflict effect again emerged. Specifically, in the T1 region, the incongruent condition showed longer gaze duration, $t1(31) = 2.31$, $p = 0.028$, $d = 0.41$; $t2(39) = 2.46$, $p = 0.018$, $d = 0.40$; 95% CI = [2.5, 42]; longer second reading time, $t1(31) = 1.96$, $p = 0.059$, $d = 0.35$; $t2(39) = 1.75$, $p = 0.089$, $d = 0.27$; 95% CI = [-1.1, 53]; and longer total reading time, $t1(31) = 3.64$, $p = 0.001$, $d = 0.64$; $t2(39) = 2.77$, $p = 0.009$, $d = 0.43$; 95% CI = [22, 78]. According to Cohen' s (1988) criteria for d values, both gaze duration and total reading time showed medium to large effect sizes, indicating reliable differences

between conditions. However, this conflict effect did not extend to the T1+1 region, where no significant differences between conditions were found for first fixation duration, gaze duration, second reading time, or total reading time (all ps > 0.1). These results are consistent with Experiment 1, demonstrating that occupational gender stereotypes exerted early and sustained effects on pronoun processing.

However, when readers encountered the pronoun again (in the T2 region), first fixation duration was actually shorter in the incongruent condition, t1(31) = 2.70, p = 0.007, d = 0.46; t2(39) = 2.82, p = 0.011, d = 0.47; 95% CI = [3.0, 22]. No difference in gaze duration was observed between conditions. However, second reading time was longer in the incongruent condition than in the congruent condition, t1(31) = 2.27, p = 0.031, d = 0.41; t2(39) = 1.88, p = 0.068, d = 0.30; 95% CI = [1.9, 37]. Additionally, in the T2+1 region, the incongruent condition showed longer gaze duration compared to the congruent condition, t1(31) = 3.22, p = 0.003, d = 0.75; t2(39) = 2.87, p = 0.007, d = 0.44; 95% CI = [9.1, 40]; and longer total reading time, t1(31) = 2.62, p = 0.013, d = 0.46; t2(39) = 2.74, p = 0.009, d = 0.43; 95% CI = [9.2, 73].

These results suggest that the contextual information created by the first pronoun exerted early influence on processing of the second pronoun, leading to shorter first-pass reading times and gaze durations in the incongruent condition —indeed, even shorter than in the congruent condition. However, as processing progressed, world knowledge in the form of occupational gender stereotypes continued to influence pronoun processing during later stages, as evidenced by longer second reading times in the T2 region and longer gaze duration and total reading time in the T2+1 region in the incongruent condition.

Thus, although the position of the second pronoun in Experiment 2 was essentially the same as the pronoun position in Experiment 1, the presence of the first pronoun produced different results. In Experiment 1, occupational gender stereotypes exerted early and sustained effects on pronoun processing. In Experiment 2, the effect of occupational gender stereotypes was delayed, emerging during later lexical processing stages or more extended cognitive processing. This occurred because readers attended to the gender information provided by the first pronoun, reprocessed the occupational gender stereotype, and consequently facilitated early processing of the second pronoun. Unexpectedly, however, first fixation duration—an indicator of early processing—was actually shorter in the incongruent condition. We believe this result is related to expectations generated during reading. Because the two pronouns in the text were always consistent, when readers processed the first pronoun, the conflict between its gender information and the occupational name made them realize that their pre-existing knowledge about the occupational gender stereotype might be inappropriate in the current text, leading them to maintain an expectation regarding the occupational gender stereotype. In the congruent condition, the non-conflicting nature of the first pronoun did not create such an expectation. Consequently, when encountering the second pronoun, first fixation duration—reflecting initial

processing—was shorter in the incongruent condition. However, this expectation effect quickly dissipated, as evidenced by the absence of differences in gaze duration between conditions.

Nevertheless, it is noteworthy that occupational gender stereotypes could still exert influence during repeated or more extended cognitive processing. The likely reason is that a single pronoun cannot completely revise readers' entrenched beliefs about antecedent gender stereotypes, allowing the occupational gender stereotype to continue playing a role in subsequent processing. Thus, in Experiment 2, readers' processing of the gender information provided by the first pronoun was rather convoluted: they initially formed strong expectations, then accepted the gender provided in the text, but during more extended cognitive processing, this conflict resurfaced as the occupational gender stereotype continued to exert influence. When the text reveals the antecedent' s gender more explicitly through terms such as "male actor" or "father," will readers still experience this conflict during extended processing? Can occupational gender stereotypes still affect pronoun processing? These are the questions addressed in Experiment 3.

Experiment 3

In Experiment 3, we provided explicit and clear descriptions of the occupational term's gender, such as "male actor"or "father,"before presenting a pronoun. The pronoun' s gender always matched the prior gender description but conflicted with the occupational gender stereotype. If the contextual information provided by explicit gender descriptions in the text could completely update readers' entrenched gender stereotypes about occupational terms and influence pronoun processing, then readers should experience no difficulty processing the second pronoun in the incongruent condition, and no differences in second reading time or total reading time—measures reflecting later processing—should emerge between conditions. Conversely, if occupational gender stereotypes persisted, second reading time or total reading time would be longer in the incongruent condition.

Method

Participants Twenty-four university students (12 male, 12 female) participated in the experiment. All had normal or corrected-to-normal vision, were native Chinese speakers, and had no reading disabilities.

Design and Materials The experimental design was identical to Experiment 2. In the congruent condition, the pronoun was consistent with both the occupational gender stereotype and the gender description. In the incongruent condition, the pronoun was inconsistent with the occupational gender stereotype but consistent with the gender description. The materials were similar to those

in Experiment 2. The distance between the gender description word and the occupational term was similar to that between the first pronoun and the occupational term in Experiment 2, and the distance between the pronoun and the occupational term was similar to that between the second pronoun and the occupational term in Experiment 2. Additionally, the total number of characters was essentially the same across the two experiments. Example sentences:

Congruent condition: The magician performed exceptionally well in the art troupe and received the title of outstanding male actor, for which the troupe commended him. This kept him excited for days, and he mentioned it to everyone he met.

Incongruent condition: The dancer performed exceptionally well in the art troupe and received the title of outstanding male actor, for which the troupe commended him. This kept him excited for days, and he mentioned it to everyone he met.

The two occupational terms used to create congruent and incongruent conditions with pronouns did not differ significantly in number of characters (3.35 vs. 3.35), average stroke count per character (7.97 vs. 8.24), word frequency (2.57 vs. 2.39), or average character frequency (5.19 vs. 5.22; all ps > 0.30). To confirm that the contextual explicitness provided by pronouns in Experiment 2 and gender terms in Experiment 3 indeed differed, we recruited an additional 28 participants to evaluate the gender of occupational terms in materials from both experiments. Sample materials were as follows:

Experiment 3: The magician/dancer performed exceptionally well in the art troupe and received the title of outstanding male actor.

The evaluation materials were presented on a computer screen. Participants were required to identify the protagonist' s gender in the text and respond by pressing the F (male) or J (female) key. Response times and accuracy were recorded. The four conditions were balanced using a Latin square design to create four versions, with each participant randomly assigned to one version.

For item-based results, we conducted two-way repeated measures ANOVAs with consistency and gender description type as independent variables and accuracy and response time as dependent variables. For accuracy, we found a significant interaction between gender description type and consistency, $F(1, 39) = 4.15$, $p = 0.04$, $^2 = .096$. Simple effects analysis revealed significant differences between congruent and incongruent conditions under both gender description types (ps < 0.001). However, the difference in accuracy between the two conditions was larger when gender was described by pronouns (0.96 vs. 0.82) than when described by gender terms (0.95 vs. 0.86). This was because participants' response accuracy in the incongruent condition improved substantially when occupational gender was described using gender terms. These results demonstrate that gender descriptions indeed enhanced the explicitness of contextual information for gender specification.

Procedure The procedure was identical to Experiment 1.

Results and Analysis

Participants' question-answering accuracy was above 80%, indicating that they read the materials carefully. The analysis regions included T1 (the gender description term), T1+1 (the three characters following the gender description term), T2 (the pronoun and the two characters before and after it), and T2+1 (the three characters following the pronoun region). Fixations shorter than 60 ms or longer than 600 ms were excluded from analysis. The results are presented in Table 3 .

Table 3 Means and Standard Deviations (ms) of Eye Movement Measures in T1, T1+1, T2, and T2+1 Regions

Region	Condition	First Fixation	Gaze Duration	Second Reading Time	Total Reading Time
T1	Congruent	211 (28)	187 (29)	173 (94)	430 (117)
T1	Incongruent	198 (33)	186 (35)	133 (87)	359 (105)
T1+1	Congruent	288 (58)	262 (67)	154 (121)	384 (169)
T1+1	Incongruent	256 (53)	246 (59)	129 (91)	349 (183)
T2	Congruent	212 (30)	253 (58)	79 (50)	297 (73)
T2	Incongruent	206 (19)	238 (69)	77 (54)	276 (84)
T2+1	Congruent	191 (29)	216 (46)	91 (78)	289 (96)
T2+1	Incongruent	194 (23)	215 (37)	73 (68)	269 (89)

Note: Standard deviations are shown in parentheses.

Analysis of the T1 region data revealed longer first fixation duration in the incongruent condition, significant in the participant-based analysis, $t1(23) = 2.29$, $p = 0.023$, $d = 0.47$; $t2(39) = 1.54$, $p = 0.132$, $d = 0.24$; 95% CI = [1.3, 25]; longer gaze duration, $t1(23) = 2.94$, $p = 0.007$, $d = 0.59$; $t2(39) = 2.10$, $p = 0.042$, $d = 0.33$; 95% CI = [9.6, 55]; longer second reading time, $t1(23) = 3.68$, $p = 0.001$, $d = 0.75$; $t2(39) = 2.02$, $p = 0.049$, $d = 0.32$; 95% CI = [17, 62]; and longer total reading time, $t1(23) = 5.39$, $p < 0.001$, $d = 1.11$; $t2(39) = 3.45$, $p = 0.001$, $d = 0.55$; 95% CI = [44, 98].

The results indicate that readers experienced processing difficulty when encountering gender descriptions inconsistent with occupational gender stereotypes. However, readers quickly accepted this gender information and immediately updated their pre-existing occupational gender stereotypes. Consequently, when encountering pronouns inconsistent with occupational gender stereotypes, no reading difficulty emerged. This was evidenced by the absence of differences in all reading time measures between conditions in both the pronoun region and the post-pronoun region. Furthermore, comparing the results of Experiments 2 and 3 reveals that the explicitness of context indeed influenced the role

of context in pronoun processing. When contextual information was relatively implicit (as in Experiment 2), context produced immediate but non-sustained effects. However, when contextual information was explicit (as in Experiment 3), context produced immediate and sustained effects.

General Discussion

The present study investigated the effects and time course of discourse context and world knowledge on pronoun processing in Chinese. Experiment 1 primarily examined whether world knowledge in the form of occupational gender stereotypes could exert early influence on pronoun processing. Building upon this, Experiments 2 and 3 added information to the text to update occupational gender stereotypes, examining the effects of this updated information on pronoun processing.

First, the results of Experiments 1 and 2 both confirmed the effect of occupational gender stereotypes on pronoun processing, as evidenced by longer gaze duration, second reading time, and total reading time for pronouns that were inconsistent with the antecedent' s occupational gender stereotype. Additionally, this study found that the distance between antecedent and pronoun did not interfere with the influence of occupational gender stereotypes on pronoun resolution. Although the distance between the occupational term and the first pronoun differed between Experiments 1 and 2, this world knowledge of gender stereotypes could be quickly activated and participate in current pronoun processing. This finding differs from Qiu et al. (2012), who found that distance could modulate the effect of antecedent gender stereotypes on pronoun processing: when antecedent and pronoun were close, gender stereotypes exerted immediate effects; when they were far apart, gender stereotypes only influenced later stages of pronoun processing. This discrepancy may be attributable to differences in antecedent accessibility across discourse contexts. In Qiu et al.' s study, antecedents were names with certain gender stereotypes, whereas in the present study, antecedents were occupational terms with gender stereotypes. When using names as antecedents, materials can be constructed arbitrarily. However, when using occupational terms as antecedents, to maintain natural and coherent discourse content, the texts primarily revolved around the industries associated with each occupation, as shown in most materials in the present study. This resulted in higher accessibility of the antecedent occupational terms. Antecedent accessibility is precisely a critical factor influencing whether pronouns can be fully resolved. For example, according to Love and McKoon (2011), pronouns are not always fully resolved but may remain in a partially resolved state, with full resolution depending on both antecedent accessibility and the establishment of coherent text representations. Based on this view, in the present study, due to the antecedent type and material characteristics, antecedent accessibility was relatively high. Without interference from other information, the antecedent could be immediately activated when the pronoun appeared.

However, when contextual information updates the occupational gender stereotype, the time course of world knowledge effects on pronoun processing changes, and this change is related to the salience of gender information in the context. In Experiment 2, when pronouns were used to update occupational gender stereotypes, the updated information exerted early effects: the difference in gaze duration between incongruent and congruent conditions was no longer significant, and the effect of occupational gender stereotypes was delayed, manifesting as longer second reading time and total reading time in the incongruent condition. In Experiment 3, when gender descriptions were used to update occupational gender stereotypes, occupational gender stereotypes no longer influenced pronoun processing. These results are difficult to accommodate within existing theoretical frameworks.

For example, according to Hess' s Lexical Reinterpretation Model (Hess et al., 1995), once world knowledge in readers' minds is updated, only the updated contextual information influences pronoun processing, and world knowledge no longer affects pronoun processing. Although this theory can adequately explain the results of Experiment 3, where context completely overrode world knowledge and exerted immediate and sustained effects on pronouns, it cannot account for the results of Experiment 2, where context only partially overrode world knowledge, producing immediate but non-sustained effects on pronoun processing. Furthermore, the present results cannot be explained by Kintsch' s (1998) Construction-Integration Model or Garrod and Terras' s (2000) Bonding-Resolution Model. According to these models, pronoun processing occurs in two stages, with world knowledge influencing early stages and context only affecting later stages of pronoun processing. However, in both Experiments 2 and 3, we found contextual effects on gaze duration, an indicator of early processing. Therefore, it is certain that context can exert influence during early stages of pronoun processing.

Thus, whether pronoun processing under the influence of world knowledge and context occurs in stages or is completed in a single step appears to depend on the explicitness of the context employed. When contextual expression is relatively implicit, as with pronouns in Experiment 2, results suggest two-stage lexical processing. When contextual expression is explicit, as with gender descriptions in Experiment 3, strong contextual effects emerge, leading to single-step lexical processing. Additionally, evidence from non-pronoun word processing by Nieuwland and Van Berkum (2006b) also confirms that contextual explicitness is a factor influencing contextual effects. These researchers created a fairy tale atmosphere to render originally implausible scenarios plausible. Through repeated exposure, a context was formed that led readers to accept the fairy tale world. Results showed that when readers first encountered words inconsistent with world knowledge, a semantic violation effect (N400) emerged. When the violating word appeared for the third time, no N400 effect was found for the violating noun, but the preceding verb still showed a violation effect. However, when the word appeared for the fifth time in the discourse, neither the world knowledge-violating noun nor its preceding verb produced any effect. This seems

to suggest that when readers heard the world knowledge-violating noun for the third time, the fairy tale context that had been formed was not yet explicit enough to completely override the effect of world knowledge on word processing. Only when the sentential context became sufficiently explicit—by the fifth occurrence—did it exert full influence. These results indicate that contextual explicitness appears to be an important mediating variable affecting the influence of context on word processing. The theoretical divide between two-stage processing theories (Kintsch' s Construction-Integration Model and Garrod and Terras' s Bonding-Resolution Model) and one-stage processing theories (Hess' s Lexical Reinterpretation Model) may stem from differences in contextual explicitness employed by researchers.

Additionally, it is noteworthy that although the overall time course of contextual effects differed between the two experiments, both demonstrated early mechanisms of contextual influence. This was reflected in Experiment 2 by shorter first fixation duration in the incongruent condition for the second pronoun region and non-significant differences in gaze duration between conditions, and in Experiment 3 by non-significant differences in both first fixation duration and gaze duration between incongruent and congruent conditions in the pronoun region. These results are consistent with recent findings using eye-tracking and ERP techniques (Cook & Myers, 2004; Duffy & Keir, 2004; Nieuwland & Van Berkum, 2006b; Hald, Steenbeek-Planting, & Hagoort, 2007). They suggest that in Chinese reading comprehension, although word segmentation is difficult and can cause comprehension problems, this very difficulty makes Chinese word comprehension more context-dependent, leading to faster contextual effects on pronoun processing. These findings align with research on the time course of Chinese reading comprehension (Wang et al., 2008) and provide evidence supporting the immediate role of context in Chinese reading comprehension.

Finally, it is worth noting that the present study employed a paradigm in which context updated world knowledge, yielding findings that context can exert immediate influence on pronoun processing. In this design, context represents correct information explicitly stated in the materials, whereas world knowledge represents pre-existing but incorrect information in readers' minds. Compared to incorrect world knowledge, correct contextual information is stronger. Therefore, it is understandable that stronger contextual information can produce immediate effects on pronoun processing when contrasted with incorrect world knowledge. Whether context can still exert early influence on pronoun processing when both context and world knowledge are incorrect and of equal strength remains a question worthy of further investigation.

In summary, the present findings demonstrate that in Chinese, a language with high context dependency, context can override world knowledge to exert early effects on pronoun processing, consistent with previous research on the time course of contextual effects in Chinese processing (Wang et al., 2008). Furthermore, the explicitness of context can influence the persistence of contextual effects. The theoretical divide between two-stage and one-stage theories of word

processing in discourse may be attributable to differences in contextual explicitness employed across studies.

References

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: a method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science, 28*(11), 1547−1562.

Angele, B., Slattery, T. J., Yang, J., Kliegl, R., & Rayner, K. (2008). Parafoveal processing in reading: Manipulating n+1 and n+2 previews simultaneously. *Visual Cognition, 16*(6), 697−707.

Canal, P., Garnham, A., & Oakhill, J. (2015). Beyond gender stereotypes in language comprehension: self sex-role descriptions affect the brain' s potentials associated with agreement processing. *Frontiers in Psychology, 6*(9),

Chen, H. C., Cheung, H., Tang, S. L., & Wong, Y. T. (2000). Effects of antecedent order and semantic context on Chinese pronoun resolution. *Memory & Cognition, 28*(3), 427−438.

Cohen, J. (1988). *The statistical power analysis for the behavioral sciences*. Journal of the American Statistical Association, 2nd(334).

Cook, A. E., & Myers, J. L. (2004). Processing discourse roles in scripted narratives: the influences of context and world knowledge. *Journal of Memory & Language, 50*(3), 268−288.

Crocker, M. W., Pickering, M., & Clifton, C. (2000). *Architectures and mechanisms for language processing*. Linguistic Inquiry, 17(4), 623−662.

Duffy, S. A., & Keir, J. A. (2004). Violating stereotypes: eye movements and comprehension processes when text conflicts with world knowledge. *Memory & Cognition, 32*(4), 551−559.

Ehrlich, K., & Rayner, K. (1983). Pronoun assignment and semantic integration during reading: Eye movements and immediacy of processing. *Journal of Verbal Learning and Verbal Behavior, 22*(1), 75-87.

Esaulova, Y., Reali, C., & Stockhausen, L. V. (2014). Influences of grammatical and stereotypical gender during reading: eye movements in pronominal and noun phrase anaphor resolution. *Language Cognition & Neuroscience, 29*(7), 781−803.

Garnham, A. (Ed). (2001). *Mental models and the interpretation of anaphora*. Philadelphia: Psychology Press/Taylor & Francis.

Garrod, S., & Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: bonding and resolution. *Journal of Memory & Language, 42*(4), 526−544.

Hald, L. A., Steenbeek-Planting, E. G., & Hagoort, P. (2007). The interaction of discourse context and world knowledge in online sentence comprehension. Evidence from the N400. *Brain Research, 1146*, 210−218.

Hess, D. J., Foss, D. J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology General, 124*(1), 62−82.

Inhoff, A. W., & Liu, W. (1998). The perceptual span and oculomotor activity during the reading of Chinese sentences. *Journal of Experimental Psychology: Human Perception & Performance, 24*(1), 20−34.

Kennison, S. M., & Trofe, J. L. (2003). Comprehending pronouns: a role for word-specific gender stereotype information. *Journal of Psycholinguistic Research, 32*(3), 355−378.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review, 95*(2), 163−182.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology, 16*(1−2), 262−284.

Love, J., & McKoon, G. (2011). Rules of engagement: Incomplete and complete pronoun resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(4), 874−887.

Nieuwland, M. S., & Van Berkum, J. J. A. (2006a). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. *Brain Research, 1118*(1), 155−167.

Nieuwland, M. S., & Van Berkum, J. J. A. (2006b). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience, 18*(7), 1098−1111.

Qiu, L., Swaab, T. Y., Chen, H. C., & Wang, S. (2012). The role of gender information in pronoun resolution: Evidence from Chinese. *Plos One, 7*(5), e36156.

Qiu, L. J., Wang, S. P., & Chen, H. C. (2012). Pronoun processing during language comprehension: the effects of distance and gender stereotype. *Acta Psychologica Sinica, 44*(10), 1279−1288. [邱丽景, 王穗苹, 陈烜之. (2012). 阅读理解中的代词加工：先行词的距离与性别刻板印象的作用. 心理学报, 44(10), 1279−1288.]

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice-Hall.

Taylor, D. J., & Muller, K. E. (1996). Bias in linear model power and sample size calculations due to estimating noncentrality. *Communications in Statistics:*

Theory and Methods, 25(7), 1595-1610.

Van Gompel, R. P. G., & Majid, A. (2004). Antecedent frequency effects during the processing of pronouns. *Cognition, 90*(3), 255-264.

Wang, S., Chen, H.C., Yang, J., & Mo, L. (2008). Immediacy of integration in discourse comprehension: Evidence from Chinese readers' eye movements. *Language and Cognitive Processes, 23*(2), 241—257.

Xu, X., Jiang, X., & Zhou, X. (2013). Processing biological gender and number information during Chinese pronoun resolution: ERP evidence for functional differentiation. *Brain &Cognition, 81*(2), 223—236.

The Effects of Discourse Context and World Knowledge on Pronoun Resolution

WU Yan[1]; GAO Yuefei[1]; ZHAO Simin[1]; WANG Suiping[2]
([1]School of Psychology, Northeast Normal University, Changchun 130024, China)
([2]School of Psychology, South China Normal University, Guangzhou 510631, China)

Abstract

Pronoun resolution can play a vital role in narrative comprehension. Understanding nature of pronoun resolution can help us to learn more about the cognitive processes underlying comprehension. Studies have shown that comprehension processes will be interrupted when a pronoun mismatches its prior context or the gender stereotype of its antecedent. This indicates that discourse context and world knowledge about gender stereotype can play an important role in pronoun resolution. Recently, researchers tried to combine these two factors together and to examine which factor is crucial to the pronoun resolution. The most controversial issue is that whether the discourse context could override the world knowledge which was told to be wrong by the passage, and exert earlier influence on the pronoun resolution. Therefore, the present study examined the effects of context and world knowledge as well as its time course on pronoun resolution with eye tracking measures.

In the Experiment 1, participants were asked to read the discourse with a personal pronoun congruent or incongruent with the gender stereotype of its antecedent, an occupation name. The results revealed that reading times (including gaze, second reading time and total reading time) increased when the gender of the pronoun mismatched with the gender stereotype of its antecedent.

In the Experiment 2, another personal pronoun indicating the gender of the antecedent would be inserted into the discourse as the prior context to update the readers' gender stereotype of the occupation name. Therefore, readers would

meet two identical personal pronouns while reading the passage. The first pronoun provided the updated gender information for the second pronoun. Again, the results of the first pronoun indicated that the gender stereotype of occupation could influence pronoun processing immediately. As for the second pronoun, the complicated results showed discourse context had an early influence on resolution of pronouns, but with the processing went on, the gender stereotype of occupation continued to influence integration. However, when the first pronoun was changed into an obvious gender description in Experiment 3, the discourse context was found not only to exert an earlier effect but the effect would be continued as the only factor to influence the pronoun resolution.

The current results clearly suggest that both gender stereotype and discourse context can affect the comprehension of Chinese pronouns. However, when the discourse context updates the gender stereotype of the antecedents, the updating information can override the world knowledge information to exert an earlier effect on pronoun resolution. But whether the effects will be continued depend on the strength of the discourse context. These findings provide evidence for the interactive model of sentence comprehension.

Key words: pronoun; discourse context; world knowledge; occupation gender stereotype

Note: Figure translations are in progress. See original paper for figures.