

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Bayesian Pronoun Interpretation in Mandarin Chinese

Permalink

<https://escholarship.org/uc/item/09q935qk>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 38(0)

Authors

Zhan, Meilin

Levy, Roger P.

Kehler, Andrew

Publication Date

2016

Peer reviewed

Bayesian Pronoun Interpretation in Mandarin Chinese

Meilin Zhan (mezhan@ucsd.edu)¹, Roger P. Levy (rplevy@mit.edu)^{1,2}, Andrew Kehler (akehler@ucsd.edu)¹

¹Department of Linguistics
University of California, San Diego
San Diego, CA 92093 USA

²Department of Brain & Cognitive Sciences
Massachusetts Institute of Technology
Cambridge, MA 02139 USA

Abstract

Kehler and Rohde (2013) proposed a Bayesian theory of pronoun interpretation where the influence of world knowledge emerges as effects on the prior and the influence of information structure as effects on the likelihood: $P(\text{referent}|\text{pronoun}) \propto P(\text{pronoun}|\text{referent})P(\text{referent})$. Here we present two experiments on Mandarin Chinese that allow us to test the generality of the theory for a language with different syntactic-semantic associations than English. Manipulations involving two different classes of implicit-causality verbs and passive vs. active voice confirmed key predictions of the Bayesian theory: effects of these manipulations on the prior and likelihood in production were consistently reflected in pronoun interpretation preferences. Quantitative analysis shows that the Bayesian model is the best fit for Mandarin compared to two competing analyses. These results lend both qualitative and quantitative support to a cross linguistically general Bayesian theory of pronoun interpretation.

Keywords: Bayesian modeling; pronoun interpretation; Mandarin Chinese

Introduction

Successful language understanding requires comprehenders to resolve uncertainty in language. One source of potential uncertainty emerges from pronouns (e.g., *he, she, they, it*) since pronouns carry little information and usually do not fully specify the intended referent semantically (e.g., *Jane invited Anne to her house and she had a great time*). Nevertheless, humans generally interpret pronouns rapidly and accurately. Previous work has proposed a variety of factors that influence how comprehenders resolve pronouns. One set of approaches has focused on grammatical factors such as subjecthood preferences (Crawley et al., 1990), first-mention preferences (Gernsbacher & Hargreaves, 1988), and parallel grammatical role preferences (Smyth, 1994). Others have focused on information structural features such as topicality. Centering Theory (Grosz et al., 1995), for instance, uses information-structural relationships within and between utterances and the grammatical roles of potential referents to guide pronoun interpretation. Another approach has argued for the role of world knowledge in referent assignment. The coherence-driven approach (Hobbs, 1979), for example, models pronoun interpretation as a side-effect of the inference processes that underpin general discourse processing.

Recent studies on pronoun interpretation (Kehler et al., 2008; Kehler & Rohde, 2013) integrate aspects of the information structural approach and coherence-driven approach by way of a probabilistic Bayesian framework. Kehler and Rohde (2013) proposed that comprehenders reverse-engineer speakers' intended referents in terms of Bayes' rule, as shown in (M1). The posterior $P(\text{referent}|\text{pronoun})$ represents the interpretation bias: upon hearing a pronoun, the probability of

the pronoun referring to a particular referent. On the other hand, the likelihood $P(\text{pronoun}|\text{referent})$ represents the production bias: the probability of the speaker using a pronoun to refer to an intended referent. The prior $P(\text{referent})$ denotes the NEXT-MENTION BIAS: the probability of a specific referent getting mentioned next in the context, independent of the form of referring expression used.

$$P(\text{referent}|\text{pronoun}) = \frac{P(\text{pronoun}|\text{referent})P(\text{referent})}{\sum_{\text{referent} \in \text{referents}} P(\text{pronoun}|\text{referent})P(\text{referent})} \quad (\text{M1})$$

Equation (M1) says that the relationship between pronoun interpretation and pronoun production follows Bayesian principles, without further specifying which factors affect each term in the numerator. Here we refer to this as the WEAK form of the Bayesian hypothesis. Kehler & Rohde (2013) further suggest that the factors conditioning the prior and the likelihood are different: the influence from world knowledge emerges as effects on the prior, and the influence from information structure as effects on the likelihood. We refer to this specification as the STRONG form of the Bayesian hypothesis. This Bayesian model successfully accounts for a range of English data. However, little work has investigated whether the model generalizes across different languages.

The Present Study

This paper has two objectives. The first is to test the generality of the Bayesian pronoun interpretation theory crosslinguistically, specifically in the case of Mandarin Chinese. Experiment 1 serves as a base-line test case, where we use a passage completion paradigm that allows us to tease apart the influences of world knowledge from those of grammatical and information structural factors. We provide both qualitative and quantitative model evaluations for the experimental data to test the Bayesian account.

The second objective is to further test the theory by way of a syntactic manipulation, specifically the voice used in a context sentence (passive or active). Here we follow up on two results found by Rohde & Kehler (2014) for English; to facilitate our descriptions, for the passive voice (e.g. *Jane was amazed by Sue*) we use the "first noun phrase" NP₁ to refer to the clause's *logical object* (the *syntactic subject* Jane), and the "second noun phrase" NP₂ to refer to the *logical subject* Sue. First, on the assumption that being the syntactic subject of a passive clause is a stronger indicator of topichood than being the syntactic subject of an active clause in English (because the speaker chose a syntactically marked con-

struction to get the logical object into the syntactic subject position), their theory predicts that the rate of pronominalization of the syntactic subject in the passive will be higher than that in the corresponding active case, a prediction that was confirmed by their study. Second, they also found an unexpected effect whereby the voice manipulation influenced next-mention biases. Specifically, Rohde et al. found that passivization increased the rate of next-mention for NP₂, the logical subject. This result was unexpected on the STRONG form of the Bayesian theory. However, Rohde et al. also found that pronoun interpretation reflected this effect of passivization on next-mention preference, as predicted by the WEAK form of the theory.

Here we examine whether similar effects are found for Mandarin Chinese. The syntactic-semantic properties of Chinese passives make these constructions a good test case. Passive voice in Mandarin Chinese is generally realized via the *bei* construction, with linear arrangement NP₁ *bei* NP₂ verb (e.g., Li & Thompson, 1981). In this construction, NP₁ is the logical object followed by the passive signaling word *bei*, which introduces the logical subject NP₂. The *bei* construction conveys the notion of AFFECTEDNESS (LaPolla, 1988); it describes an event in which the logical object is affected physically or psychologically in some way (Li & Thompson, 1981). We expect that the conveyance of AFFECTEDNESS may increase the probability of mentioning the logical object in the next sentence. If this expectation is borne out in the data, it affords the opportunity for an additional test of the WEAK Bayesian theory, which predicts that any effect of passivization on next-mention preferences should have a corresponding effect on pronoun interpretation preferences. We test these predictions in Experiment 2.

Experiment 1

Experiment 1 provides a first test case for investigating the generality of the Bayesian pronoun interpretation theory in Mandarin Chinese. We used a passage completion paradigm with a 2-by-2 design that allowed us to tease apart the influence of world knowledge-based inference that emerges from verb semantics, and the influence of information structural and grammatical factors by conditioning on the grammatical role of the referent in the analysis.

Methods

Participants We recruited 50 self-reported native Mandarin speakers over Witmart (a China-based online crowdsourcing platform). Each of them was paid \$4 for their participation in the experiment.

Materials and procedures Participants completed two-sentence passages, writing a second sentence after a transitive-verb context sentence with two like-gender animate arguments. We used a Prompt Type by Verb Type design. Prompt Type has two levels: The Free condition (1) included no material in the second sentence prompt, allowing us to estimate next-mention preferences $P(\text{referent})$ by analyzing

the first-mentioned referent in each condition, and the likelihood $P(\text{pronoun}|\text{referent})$ that a pronoun is produced to refer to that referent by analyzing the referential forms that participants used. The Pronoun condition (2) included an overt pronoun in the second sentence, allowing us to measure empirical pronoun interpretation preferences $P(\text{referent}|\text{pronoun})$.

The verbs in the first sentence were taken from one of two implicit causality (IC) classes (Garvey & Caramazza, 1974; Brown & Fish, 1983). The use of IC verbs allowed us to manipulate the prior, with IC-1 and IC-2 favoring the syntactic subject (NP₁) and the non-subject (NP₂) next-mentions respectively. A norming study (N=45) was conducted prior to Experiment 1 to ensure the verbs we selected have a clear bias towards re-mentioning either the subject or the non-subject. We selected sixteen subject-biased IC-1 verbs and twenty object-biased IC-2 verbs for the main experiments based on the results of the norming study.

- | | | |
|-----|---|-----------------|
| (1) | Meihui dadong-le Jieyi. ...
Meihui impressed Jieyi. ... | [IC-1, Free] |
| (2) | Meihui dadong-le Jieyi. Ta ...
Meihui impressed Jieyi. She ... | [IC-1, Pronoun] |
| (3) | Meihui jiegu-le Jieyi. ...
Meihui fired Jieyi. ... | [IC-2, Free] |
| (4) | Meihui jiegu-le Jieyi. Ta ...
Meihui fired Jieyi. She ... | [IC-2, Pronoun] |

Because IC verbs exhibit the most polarized effects on next-mention biases when the follow-on sentence provides an explanation of (i.e., a cause or reason for) the event described by the prompt sentence (Rohde, 2008), we instructed participants to limit their continuations to such explanations. Each participant completed 36 target items and 36 filler items with pseudo-randomization. Prompt Type varied within participants and within items; Verb Type varied only within participants. Each item was presented via the web interface on a separate page with a text box where the participants were instructed to write their continuations.

Coding The responses were coded by two trained native speakers, the first author and a UC San Diego graduate student who was blind to the hypothesis of the study. Each coder went through the responses independently to code the first-mentioned referent (or the assignment of the pronoun in the case of the Pronoun prompt condition) in each continuation as one of the five categories: *NP₁*, *NP₂*, *both*, *unclear*, and *other*. If the two coders did not agree on a reference, or there was not enough information available to identify the intended referent, the response was coded as *unclear*. For continuations in the Free prompt conditions, choice of referring expressions were coded as *name*, *overt pronoun*, *null pronoun*, and *other*.

Results and discussion

The analysis only included continuations for which the first-mentioned entity was NP₁ or NP₂; hence continuations were excluded if the referent was coded as *unclear* (3.8%), *both*

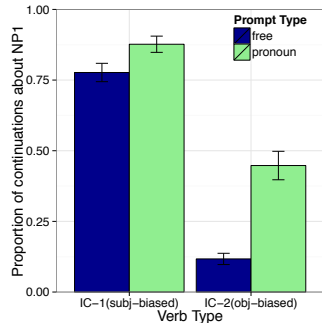


Figure 1: Proportion of continuations about the syntactic subject NP₁, by Verb Type and Prompt Type, in Experiment 1

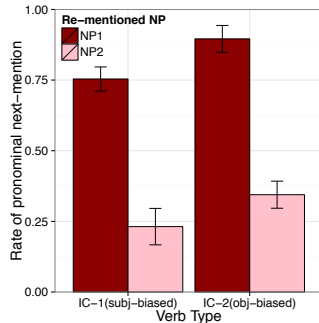


Figure 2: Rate of pronominalization by Verb Type and Re-mentioned NP in Experiment 1

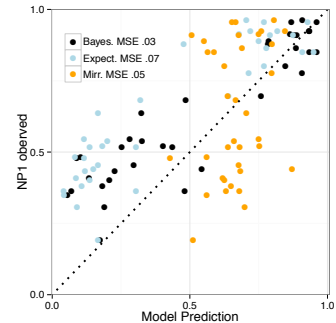


Figure 3: Quantitative model comparisons for Experiment 1

(0.8%), or *other* (2.7%). Continuations were also excluded if the choice of referring expression was other than a name, an overt pronoun, or a null pronoun (1.0%). After applying these restrictions, 1651 out of 1800 continuations in the dataset remained. All statistical analyses in this paper report results from mixed-effect logistic regression models with the maximal random-effect structure justified by the design (Barr et al., 2013), conducted using the `lme4` R package (Bates et al., 2015; R Core Team, 2015); we report significance levels based on the Wald z statistic. In cases where we encountered convergence failure with `lme4`, we report analyses carried out using the `MCMCglmm` R package (Hadfield, 2010), which we indicate by p_{MCMC} in reporting statistical significance. Error bars in all figures are standard errors over by-participant means.

The WEAK form of the Bayesian hypothesis predicts that pronoun interpretation biases and production biases follow Bayesian principles, without fully specifying what factors affect each component. The STRONG form of the hypothesis additionally predicts that semantic factors (Verb Type) affect only next-mention biases $P(\text{referent})$ and not pronoun production biases $P(\text{pronoun}|\text{referent})$; the primary factor affecting pronoun production biases is instead predicted to be the grammatical role of the referent (Re-mentioned NP). We test these predictions in the following analyses.

Next-mention biases Figure 1 shows the proportion of continuations about NP₁ in both Free prompt and Pronoun prompt conditions. We first evaluate the next-mention biases (dark blue bars) in the Free prompt data, which serves as the prior in the Bayesian model. Analyses showed a main effect of Verb Type ($p < 0.001$): subject-biased IC-1 items prompted significantly more continuations about NP₁ (77.7%) than object-biased IC-2 items (11.7%). This indicates that verb semantics had a strong effect on next-mention biases, an expected result on both the STRONG and WEAK forms of the Bayesian hypothesis.

Production biases Figure 2 shows the rate of pronominalization conditioning on Re-mentioned NP collapsing null and overt pronouns, which serves as the likelihood in the

Bayesian model. Analyses showed a large main effect of Re-mentioned NP ($\beta = 1.37$, $p < 0.001$), with NP₁ re-mentions much more likely to be pronominalized than NP₂ re-mentions. There was no significant interaction between Re-mentioned NP and Verb Type ($\beta = -0.053$, $p = 0.716$). Both results align with the predictions of both the STRONG and WEAK hypotheses, indicating a clear disassociation between next-mention biases and production biases. There was also a smaller, unanticipated main effect of Verb Type ($\beta = -0.43$, $p < 0.05$), with pronominalization rates higher in the IC-2 context than in the IC-1 context. The reasons for this effect remain unclear.

Interpretation biases Now we examine the interpretation biases (the *posterior*) and compare them to next-mention biases (the *prior*). As shown in Figure 1, there was a main effect of Verb Type ($\beta = -2.021$, $p < 0.001$), with IC-1 verbs eliciting more NP₁ mentions than IC-2 verbs. There was a main effect of Prompt Type ($\beta = 0.896$, $p < 0.001$), with the proportion of continuations about NP₁ in the Pronoun prompt condition higher than that in the Free prompt condition. Analyses also showed a interaction between Prompt Type and Verb Type ($\beta = -0.408$, $p < 0.001$), whereby the effect of Pronoun prompts was larger for IC-2 verbs than IC-1 verbs. Given next-mention and production biases, the main effects of Verb Type and Prompt Type are predicted by the Bayesian theory (the interaction’s presence was neither predicted nor precluded).

Quantitative model comparisons Following Rohde & Kehler (2014), we further evaluate the Bayesian account by comparing its predictions with those of two competing models. The first competing model is the Expectancy model, according to which the interpretation bias towards a referent is equal to the probability the referent gets mentioned next (Arnold, 2001). For this model, the predicted interpretation bias is estimated to be the prior next-mention bias. We express this model in Equation (M2) below, using the \leftarrow assignment operator to emphasize that this model does not follow normative probability theory.

$$P(\text{referent}|\text{pronoun}) \leftarrow P(\text{referent}) \quad (\text{M2})$$

The second competing model we call the Mirror model, according to which the interpretation bias towards an entity is proportional to the likelihood that a reference to that entity would have been pronominalized by the speaker. This model captures the intuition that pronoun interpretation and production are effectively mirror images of each other, as expressed by the assignment in (M3).

$$P(\text{referent}|\text{pronoun}) \leftarrow \frac{P(\text{pronoun}|\text{referent})}{\sum_{\text{referent} \in \text{referents}} P(\text{pronoun}|\text{referent})} \quad (\text{M3})$$

To determine the quantitative predictions of each model (M1)–(M3), we used the Free prompt data to estimate [condition+item]-specific prior and likelihood probabilities $P(\text{referent})$ and $P(\text{pronoun}|\text{referent})$, using add-one smoothing to avoid zero probabilities, and then compared these predictions to [condition+item]-specific human interpretation preferences from our Pronoun prompt data. Figure 3 plots observed NP₁ interpretation rates against item-specific predictions of the three models. The $x = y$ dotted line would be a perfect model fit. The Bayesian model had the least mean-squared error (0.03), indicating the Bayesian model is a better fit than either of the competing models in predicting pronoun interpretation. In comparison, the Mirror model dramatically underpredicts cross-item/condition variability in interpretation preference, because it lacks the influence of the world-knowledge-derived prior. The Expectancy model systematically underpredicts the rate of NP₁ pronoun interpretation, because it lacks the likelihood-derived bias toward NP₁ obtaining from the speaker’s choice of pronominal form.

Experiment 2

Experiment 2 further tests the generality of the Bayesian pronoun theory by pursuing a set of findings regarding the effect of voice on pronoun behavior identified for English by Rohde & Kehler (2014). Recall that Rohde and Kehler hypothesized that a difference in syntactic form — specifically active vs. passive voice — would have an effect on production biases. This expectation is based on the STRONG Bayesian hypothesis, which posits that the likelihood of a speaker pronominalizing a mention of a referent is based on the referent’s degree of topicality. Hypothesizing that being the subject of a passive clause is a stronger indicator of topichood than being the subject of an active clause, Rohde & Kehler (2014) thus predicted that the syntactic subject of a passive clause is more likely to be pronominalized than that of an active clause. The results of their study confirmed this prediction. However, the results also revealed a separate effect, unexpected under the STRONG hypothesis, whereby passivization not only affected the next-mention bias as well, but did so in the opposite direction that one would expect: Passivization increased the next-mention rate of the *logical* subject, that is, the entity mentioned from within a *by*- adjunct.

Here we evaluate the predictions of the models by employing a similar voice manipulation in Mandarin Chinese. The

STRONG hypothesis predicts that the difference in information structure between active and passive voice will affect production biases, and in turn interpretation biases. A finding similar to Rohde and Kehler’s whereby passivization increases the next-mention bias for the logical subject is not predicted by the STRONG hypothesis, however. The WEAK hypothesis makes no commitment to either of these predictions. Instead, the WEAK hypothesis simply predicts that any change in the pronoun production biases or the next-mention biases would affect interpretation biases, in accordance with Bayes’ rule.

Methods

Participants We recruited 71 self-reported native Mandarin speakers over Witmart. Each of them was paid \$4 for their participation in the experiment. None of them participated in Experiment 1.

Materials and procedures The same passage completion paradigm as Experiment 1 was employed. In addition to the Verb Type × Prompt Type design, we included a Voice (active vs. passive) manipulation. Participants were asked to write a natural continuation given the prompt (unlike Experiment 1, participants were not limited to writing explanations). The verbs were identical with the ones in Experiment 1 with one exception: We replaced “anwei” (*comfort*) with “xiaokan” (*look down upon*) since the former sounds unnatural in a passive clause. Example stimuli are shown in (5)–(8).

- (5) Meihui dadong-le Jieyi. (Ta) ... [IC-1, active]
Meihui impressed Jieyi. (She) ...
- (6) Jieyi bei Meihui dadong-le. (Ta) ... [IC-1, passive]
Jieyi was_by Meihui impressed. (She) ...
- (7) Meihui jiegou-le Jieyi. (Ta) ... [IC-2, active]
Meihui fired Jieyi. (She) ...
- (8) Jieyi bei Meihui jiegou-le. (Ta) ... [IC-2, passive]
Jieyi was_by Meihui fired. (She) ...

Each participant completed 36 target items and 36 filler items with pseudo-randomization. Prompt Type and Voice varied within participants and within items; Verb Type varied only within participants. Participants wrote passage continuations using the same web interface as in Experiment 1.

Coding The same judges coded the responses using the same criteria as those in Experiment 1.

Results and discussion

Out of 2556 continuations, 8.7% were excluded because the next-mentioned referent was coded as *unclear*, *both*, or *other*, or the choice of referring expressions was other than a name, an overt pronoun, or a null pronoun, leaving a total of 2334 continuations in the dataset.

Next-mention biases Next-mention results are depicted in Figure 4 in terms of the proportion of continuations about the *logical object*, which more clearly illustrates the effect of the

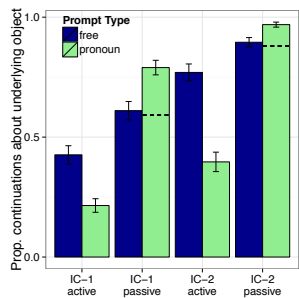


Figure 4: Prop. of continuations about the logical object, by Verb Type, Voice and Prompt Type, in Experiment 2

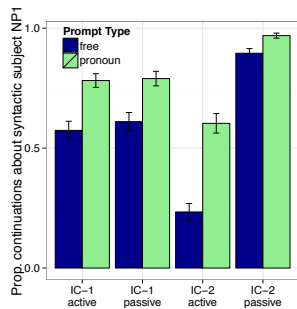


Figure 5: Prop. of continuations about the syntactic subject in Experiment 2 (representation of the data of Figure 4)

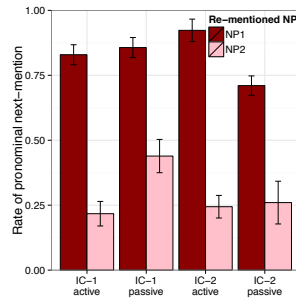


Figure 6: Rate of pronominalization, by Verb Type, Voice, and Re-mentioned NP, in Experiment 2

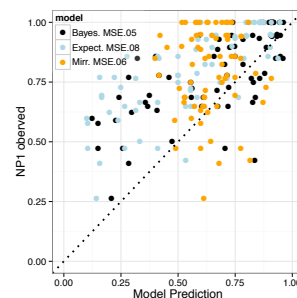


Figure 7: Quantitative model comparisons for Experiment 2

voice manipulation. The Free prompt condition (blue bars) showed a main effect of Verb Type ($\beta = -0.987$, $p < 0.001$), whereby IC-2 verbs elicited more continuations about the logical object than IC-1 verbs. Analyses also showed a main effect of Voice ($\beta = -0.477$, $p < 0.001$), whereby passivization increased continuations about the logical object. The effect of Verb Type is compatible with both the STRONG and the WEAK versions of the Bayesian hypothesis. However, the effect of Voice is only consistent with the WEAK version since the STRONG version predicts that information structure should only affect the likelihood.

Production biases Figure 6 presents the rate of pronominalization, collapsing both null and overt pronouns. Results showed a main effect of Re-mentioned NP ($p_{MCMC} < 0.001$), with NP₁ more likely to be pronominalized than NP₂ regardless of its semantic roles. There was no main effect of Voice ($p_{MCMC} = 0.38$) or Verb Type ($p_{MCMC} = 0.75$). The higher rate of pronominalization of NP₁ is compatible with both the STRONG and WEAK versions of the Bayesian hypothesis, as is the lack of effect of Verb Type. The lack of Voice effect is not compatible with the STRONG version, however, which predicts that the syntactic subject of a passive clause is more likely to be pronominalized than that of an active clause. These results instead suggest that grammatical role is the primary factor affecting pronoun production biases.

Interpretation biases To better illustrate the comparisons between next-mention biases (the prior) and interpretation biases (the posterior) in relation to NP₁ re-mentions, we graph the data from Figure 4 in terms of re-mentions of the syntactic subject in Figure 5. Analyses showed a main effect of Prompt Type ($p_{MCMC} < 0.001$), with Pronoun prompts being associated with more NP₁ re-mentions than Free prompts. Analyses also showed a main effect of Voice ($p_{MCMC} < 0.001$), with passive voice eliciting more NP₁ re-mentions than active voice, as well as a marginal effect of Verb Type ($p_{MCMC} = 0.063$), with IC-1 verbs eliciting numerically more NP₁ re-mentions collapsing Voice and Prompt Type. The Bayesian hypothesis predicts both the observed effect of Prompt Type,

due to the likelihood term in Equation (M1), and the close relationship between the NP₁ re-mention rate in the Pronoun and Free prompt types across Verb Type and Voice conditions, due to the effect of the prior.

We have shown in Figure 4 (blue bars) that passivization increased next-mention rate for the logical object NP₂. The Bayesian hypothesis predicts that any effect of passivization on next-mention biases should have a corresponding effect on pronoun interpretation biases. The dashed lines in Figure 4 illustrate the NP₁ re-mention rate that would be predicted for passives under an alternative model that is identical to M1 except that the normatively correct prior term $P(\text{referent} = \text{NP}_1 | \text{passive})$ is replaced with the active-voice prior term $P(\text{referent} = \text{NP}_1 | \text{active})$. This alternative model underpredicts the observed rate of NP₂ re-mention, showing that the passive voice’s effect of increased next-mention prior bias toward the logical object indeed manifests in pronoun interpretation preferences.

Quantitative model comparisons We again used the Free prompt data to estimate the interpretation biases of overt pronouns for each item, and then compared the predictions to the actual interpretation data measured in the Pronoun prompt condition. Figure 7 plots observed NP₁ interpretation rates against item-specific predictions of the three models. The $x = y$ dotted line would be perfect model fit. As also seen in Experiment 1, the Mirror model underpredicts cross-item variability in interpretation preferences, and the Expectancy model systematically underestimates the likelihood-derived preference toward NP₁. Once again, the Bayesian model had the least mean-squared error (0.05) of the three models.

General Discussion

The results from Experiment 1 showed that verb semantics had a strong effect on next-mention biases, where IC-1 verbs elicited more mentions of the subject than IC-2 verbs. The results also showed that the grammatical roles of potential referents had a significant impact on the likelihood of producing a pronoun given a particular referent, with re-mentions of

the subject of the previous sentence (NP₁) more likely to be pronominalized than re-mentions of the non-subject (NP₂). Importantly, this held even in the object-biased IC-2 conditions. This and the lack of interaction between Verb Type and Re-mentioned NP in Figure 2 reveal a disassociation between the factors that determine what entities get mentioned next versus those that determine whether those mentions get pronominalized. When interpreting an ambiguous pronoun, the comprehender takes into account both the prior probability of a referent being mentioned next in the context and the likelihood that the speaker would choose a pronoun to mention that referent. As illustrated in Figure 1, the presence of an overt pronoun boosted NP₁ mentions in both IC-1 and IC-2 contexts compared to the Free prompt condition. This indicates the effect of the likelihood (the probability of NP₁ being realized pronominally) on pronoun interpretation when the prior next-mention probability is kept constant.

The results from Experiment 2 show that passivization increases the rate of next-mention of the logical object in Mandarin Chinese, reflecting the affectedness of the logical object in the passive *bei* construction. Results again showed that occupying the subject position increases the likelihood of pronominalizing the re-mention of an entity. On the other hand, we found no consistent evidence showing that re-mentions of passive subjects are more often pronominalized in Mandarin than re-mentions of active subjects, unlike what Rohde & Kehler (2014) found in English.

Our results indicate that pronoun interpretation biases do not straightforwardly mirror expectations about next-mention, nor do they simply mirror pronoun production biases. Rather, pronoun interpretation biases reflect the joint influence of next-mention biases and production biases, where the comprehender reasons about the states of the world as well as the speaker's linguistic choices in conveying those states. Comparisons between the full Bayesian model and two reduced variants showed that the full Bayesian model best predicted human behavior in pronoun interpretation, lending quantitative support for the Bayesian theory. These findings are broadly consistent with previous studies in English (Kehler & Rohde, 2013; Rohde & Kehler, 2014), but also show that cross-linguistic variation in effects of grammatical voice on production preferences is mirrored in interpretation. Overall these results lend both qualitative and quantitative support to a cross-linguistically general Bayesian theory of pronoun interpretation.

Acknowledgments

We gratefully acknowledge valuable feedback from members of the UCSD Computational Psycholinguistics Laboratory, and support from NSF grant IIS-0953870 and BCS-1456081, and NIH grant 1R01HD065829 to RPL.

References

Arnold, J. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes*, 21(2), 137-162.

- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi: 10.18637/jss.v067.i01
- Brown, R., & Fish, D. (1983). The psychological causality implicit in language. *Cognition*, 14(3), 237-273.
- Crawley, R. A., Stevenson, R. J., & Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research*, 19(4), 245-264.
- Garvey, C., & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry*, 459-464.
- Gernsbacher, M. A., & Hargreaves, D. J. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language*, 27(6), 699-717.
- Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2), 203-225.
- Hadfield, J. (2010). MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*, 33(2), 1-22.
- Hobbs, J. R. (1979). Coherence and coreference. *Cognitive Science*, 3, 67-90.
- Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics*, 25(1), 1-44.
- Kehler, A., & Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics*, 39(1-2), 1-37.
- LaPolla, R. J. (1988). Topicalization and the question of lexical passives in Chinese. In M. Chan & T. Ernst (Eds.), *Proceedings of the Third Annual Conference on Chinese Linguistics* (pp. 170-188). Indiana University Linguistics Club, Bloomington, IN.
- Li, C. N., & Thompson, S. A. (1981). *Mandarin chinese: A functional reference grammar*. Univ of California Press.
- R Core Team. (2015). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <http://www.R-project.org/>
- Rohde, H. (2008). *Coherence-driven effects in sentence and discourse processing*. Unpublished doctoral dissertation, University of California, San Diego.
- Rohde, H., & Kehler, A. (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition and Neuroscience*, 29(8), 912-927.
- Smyth, R. (1994). Grammatical determinants of ambiguous pronoun resolution. *Journal of Psycholinguistic Research*, 23(3), 197-229.