

## Exploring Scalar Implicature variation Across Different Mandarin Existential Quantifiers

**Abstract:** This study examines scalar implicature (SI) variation among four existential quantifiers in Mandarin—*you*, *youde*, *youxie*, and *yixie*—using a binary sentence-picture judgment task. Despite sharing the same existential meaning, the expressions yielded divergent SI rates: *youde* and *youxie* triggered significantly more implicatures than *you* and *yixie*. Two sources of variation are explored: partitivity and semantic similarity. Structural analysis identifies *youde* as an overt partitive and *youxie* as forming a covert partitive via discourse linkage, whereas *you* and *yixie* lack such structure. A BERT-based semantic similarity analysis further indicates that *youxie* and *youde* are more similar to the universal quantifier *suoyou* (“all”), potentially increasing alternative accessibility and supporting implicature computation. These findings demonstrate that variation in SI strength can emerge from internal lexical and structural properties, even within expressions of the same scalar meaning. Implications for Mandarin SI research highlight the need to account for quantifier form and its pragmatic effects.

### 1. Introduction

Scalar implicature (SI) has long been a central topic in pragmatics research. It refers to the phenomenon where the use of a weaker term gives rise to the inference that a stronger, more informative term does not apply. For example, the sentence “Some children like apples” is commonly understood to imply that “not all children like apples” (Horn, 1984). This kind of inference plays a crucial role in how listeners enrich utterances beyond their literal meanings. In recent years, researchers have become increasingly interested in the variability of SI strength, that is, how likely different scalar expressions are to trigger an implicature, and how this likelihood is shaped by linguistic form and context (Degen, 2015; van Tiel et al., 2016). While much of this work has focused on English, less is known about how SI behaves in other languages, especially in Mandarin Chinese, which features a variety of existential quantifiers that may encode subtle differences in meaning and use.

The present study examines whether 4 different commonly used existential quantifier expressions in Mandarin—*you*, *yixie*, *youde*, and *youxie* (all broadly equivalent to the English *some*, rather than expressions like *few*, *a couple*, or *most*)—differ in the extent to which they give rise to scalar implicatures. To do so, I conducted an online binary choice task in which participants judged whether sentences containing different existential quantifiers appropriately described accompanying images. The results reveal clear variation: *youxie* and *youde* consistently elicited higher rates of SI compared to *you* and *yixie*.

For this pattern, two possible sources of variation were considered. The first involves partitivity: whether the quantifier signals that it refers to a subset of a contextually relevant set. In Mandarin, a closer examination of the internal structure

and semantics of the existential quantifiers reveals that quantifiers encoding overt or covert partitivity are associated with higher SI rates, while those lacking such cues consistently yield lower rates. This correspondence between partitivity and SI variation aligns with earlier observations in English and supports the view that partitive structure plays a key role in facilitating SIs. The second concerns semantic similarity, specifically, how closely each existential quantifier is related in meaning to the universal quantifier *suoyou* (“all”), which serves as the stronger alternative in the scalar relation. A semantic similarity analysis based on contextualized embeddings (BERT) suggests that *youxie* and *youde* are more similar to *suoyou*, possibly increasing the accessibility of the stronger alternative and thereby encouraging implicature computation. Together, these findings show that even subtle differences in quantifier form can shape the strength of pragmatic inferences.

## 2. Previous Work

A growing body of experimental evidence has emerged to question the long-standing assumption that scalar implicatures are uniform in strength and arise automatically across contexts (Gazdar, 1979; Grice, 1975; Horn, 1984; Levinson, 2000). Recent studies have demonstrated that SI strength can vary systematically depending on a range of linguistic, contextual, and cognitive factors.

### 2.1 Experimental Evidence and Contributing Factors of SI Variation

Degen (2015) used large-scale corpus data and web-based judgment experiments to challenge the theoretical view that scalar implicatures from *some* to *not all* form a homogeneous, context-independent type of implicature that qualifies as a Generalized Conversational Implicature (Grice, 1975), as well as the Default Model assumption that these implicatures are computed without additional cognitive effort (Horn, 1984; Levinson, 2000). Specifically, she found that SI strength increases when *some* is used in the partitive form (i.e., *some of*), when it is prosodically prominent, and when the embedded noun phrase is more discourse-accessible (for example, previously mentioned, in subject position, or modified). Extending this work, Degen and Tanenhaus (2015) used eye-tracking and acceptability judgment experiments to show that partitive *some of* (referred to as *summa*) generally led to faster rejection of total set interpretations and stronger implicatures, particularly for intermediate quantities. However, this effect was weakened or delayed when more specific alternatives like exact number expressions were made salient in the context, suggesting that partitivity interacts dynamically with the availability of lexical alternatives. Together, these findings suggest that scalar implicatures are more variable and sensitive to lexical and contextual factors than previously assumed under the Homogeneity Assumption.

The SI variability in English is not confined to within the *some*–*all* scale. Van Tiel et al. (2016) conducted a comprehensive investigation of scalar implicature diversity across forty-three English expressions and analyzed potential predictors of variation. They found that properties influencing the distinctness between the weak term (e.g., *some*) and the strong term (e.g., *all*) made a significant contribution to scalar

inference rates. Specifically, the semantic distance, i.e., the difference in strength between the weak and strong term, showed a positive correlation with the likelihood of triggering an implicature; and scalar expressions on bounded scales, where the stronger term denotes an endpoint, were more likely to give rise to implicatures. Van Tiel et al. also considered the accessibility of alternatives (i.e. whether the alternatives of the scalar term is likely to be mentally activated or retrieved) as a possible influencing factor. Despite the fact that their analysis did not find significant effects for availability, subsequent studies (McNally, 2017; Westera & Boleda, 2020) have questioned the validity of their methodology and the robustness of the results. More recently, Hu et al. (2023) used neural language models to approximate human predictive distributions. They found that scalar implicature rates can be accounted for by the expectedness of the stronger alternative in context, further confirming that the accessibility of alternatives exerts a reliable influence on SI variation..

Beyond English, cross-linguistic studies have identified meaningful differences in how scalar implicatures are computed, shaped by lexical properties and usage patterns in each language. Stateva et al. (2019) investigated interpretations of the quantifier *some* and its equivalents in English, French, German, and Slovenian through questionnaire ratings and a picture-selection task, and found that English speakers accepted *some* in contexts with up to 80% inclusion, whereas French speakers applied much stricter bounds to *quelques*, suggesting that different languages may associate quantifiers with different perceived semantic distances between scalar terms, which in turn affects the strength of implicatures. Dionne and Coppock (2021) compared interpretations of *finger* in English and Spanish using production and comprehension tasks, and observed that English speakers frequently produced *thumb* as a distinct term and inferred that *finger* excludes it, while Spanish speakers used *dedo* to refer to all digits and did not derive the same inference, indicating that implicature strength is closely tied to the frequency and contextual salience of lexical alternatives.

Overall, there is now substantial experimental evidence demonstrating that SI variation is a robust phenomenon. It arises not only within a single scale in a given language, but also across different scales within the same language, as well as across equivalent scales in different languages. A wide range of factors have been identified as potential contributors to this variation, including lexical properties such as partitivity, the perceived distinctness between scalar pairs, the accessibility of alternatives, and cross-linguistic differences in the frequency with which particular expressions are used.

## 2.2 Mandarin Chinese SI: Limited Exploration on Variation

Research on SIs in the Sino-Tibetan language family remains limited, particularly with respect to variation. Empirical studies on Mandarin SI emerged relatively late and have primarily focused on the developmental timeline of its acquisition in children. For instance, Su (2013) investigated whether the computation of SIs in child Mandarin is influenced by the monotonicity of linguistic contexts. Through three experiments involving the scalar term *huozhe* (“or”) in different syntactic environments, Su examined the derivation of scalar implicatures involving the <or, and>

scale, where *or* can implicate *not and*—a distinction known as inclusive *or* (allows both) versus exclusive *or* (allows only one). Su found that while Mandarin-speaking adults exhibited sensitivity to downward-entailing contexts, Mandarin-speaking children did not. Children persistently interpreted *huozhe* as inclusive “or”, regardless of context, unless explicitly prompted with alternatives, suggesting immature or less spontaneous implicature computation. Similarly, Zhao et al. (2021) tested children’s ability to compute scalar implicatures triggered by *youxie* (“some”) and found that while other pragmatic inferences were accessible from age 4, consistent SI interpretation did not emerge until age 6. Wang (2023) further explored processing constraints by testing 4- to 7-year-old Mandarin- and English-speaking children. Using a covered-box task and a digit span task, Wang found that children’s SI performance correlated significantly with working memory capacity, providing support for a processing-based account of delayed SI acquisition. Zhang and Wu (2023), using mouse-tracking techniques, showed that participants differed in their reasoning procedures when interpreting the same scalar term *yixie* (“some”), reflecting stable individual differences in inferential strategies. However, despite confirming the presence of scalar implicature in Mandarin, these studies have primarily focused on the developmental trajectory and processing mechanisms of SI. Consequently, exploration of variation in Mandarin SI strength remains largely underdeveloped.

### 3. Research Question and Significance

Interestingly, Mandarin offers multiple expressions for the same existential meaning, especially in the case of quantifiers. Common forms include *you*<sup>1</sup>, *you (yi)xie*, *youde* (Tsai, 2003), and *yixie*, exceeding the two corresponding expressions in English— “some” and “some of”. However, prior experimental research has largely overlooked this lexical richness. Most studies have only selected one of these forms as test material. For example, Zhao et al. (2021) only used *youxie* as the target quantifier in their SI experiments; although *you* also appeared in their materials, it was used in its non-scalar verbal meaning (“have”) rather than as an existential quantifier. Similarly, many other researchers such as Wang (2023) and Zhang & Wu (2023), tested only *yixie* in their experimental materials.

When it comes to the topic of SI variation, Wang (2023) has observed cross-linguistic variation: under the same experimental settings, Mandarin SIs were found to be more robustly computed than English SIs (95% vs. 78%). However, prior research has rarely investigated or compared the potential SI variation across different

---

<sup>1</sup> The term *you* in Mandarin Chinese fulfills five distinct grammatical functions: possessive, locative, presentational, perfective, and assertive (Cheng, 1978; Huang, 1988; Tsai, 2003). The present study focuses exclusively on the *presentational* use of *you*, that is, its function as an existential quantifier. It is important to note that this usage should not be conflated with the *possessive you* meaning ‘have’, which functions as a lexical verb.

Mandarin scalar expressions. This leaves open a theoretically and empirically important question: can scalar implicature variation in Mandarin also arise across synonymous expressions of a single scalar point, as has been observed in English? Addressing this question is important for several reasons. It contributes to a more comprehensive understanding of SI variability beyond Indo-European languages. More importantly, it highlights a previously underexplored source of variation in Mandarin SI research—the distinction and impact among different expressions of the same scale. This calls for greater attention to how such differences may affect both the design of experimental materials and the interpretation of results. The existence of multiple expressions for “some” in Mandarin is not merely a lexical detail, it may reflect a meaningful source of variation in how scalar inferences are derived. Therefore, the present study asks:

a. Is there SI variation among different existential quantifier expressions in Mandarin?

b. If such variation exists, what are the specific patterns?

c. And what are the main factors contributing to these patterns? E.g., are they influenced by lexical properties such as partitivity, as in the case of English *some* vs. *some of*? Or, similar to differences across English scalar pairs, do these Mandarin expressions differ in perceived distinctness or in the accessibility of alternatives, thus giving rise to different SI patterns?

By identifying the SI patterns associated with different existential expressions and examining possible explanatory factors, this study provides new insights into the SI variation in Mandarin.

#### 4. SI Experiment

To address the above questions, this study first conducts an experiment to test SI strength, examining whether different existential quantifiers in Mandarin trigger SIs of varying strength. The method employed is a binary choice task, adapted primarily from Tiel and Schaeken (2017), in which participants judge sentence–picture pairs by pressing the button which represents “good” or “bad” to determine whether a sentence appropriately describes a picture.

##### 4.1 Participants

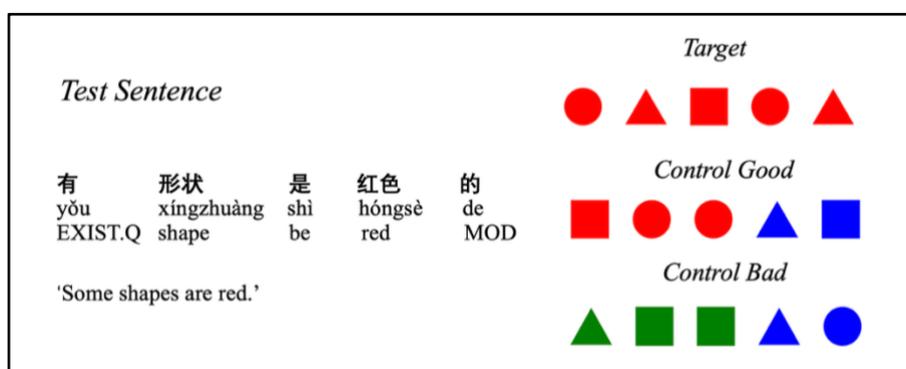
The experiment was hosted on the PCIBex platform. Forty-five participants ( $M_{age} = 29$ ; range = 19–60; 20 females) were recruited via e-flyers and compensated with a payment in RMB equivalent to 5 USD at the current exchange rate for completing the 20-minute task. They were unaware of the purpose of the experiment and had not participated in any prior related studies. Before the experiment began, they were asked to complete a brief questionnaire indicating their gender, age, native language, childhood location, and length of residence in China. All participants were native speakers of Mandarin who grew up in China. The study was approved by the Northwestern University IRB (STU00223481).

##### 4.2 Material

The experimental materials were adapted from the scalar inference trials in Tiel and Schaeken (2017) and Ronai and Xiang (2021). Each critical item consisted of one

test sentence paired with three picture conditions: a target (underinformative) condition, a control good condition, and a control bad condition. For example, as illustrated in Figure 1, when the test sentence was “Some shapes are red,” the target condition showed a picture where all the shapes were red. This pairing was semantically true but pragmatically false. If a participant judged this as “not good,” it indicated that they computed a scalar implicature, interpreting “some” as “some but not all.” In the good control condition, the sentence matched both semantic and pragmatic expectations—for instance, only some but not all of the shapes were red. In the bad control condition, the sentence contradicted both semantic and pragmatic meaning—such as when none of the shapes were red.

Figure 1. Example of Critical Item with Three Conditions



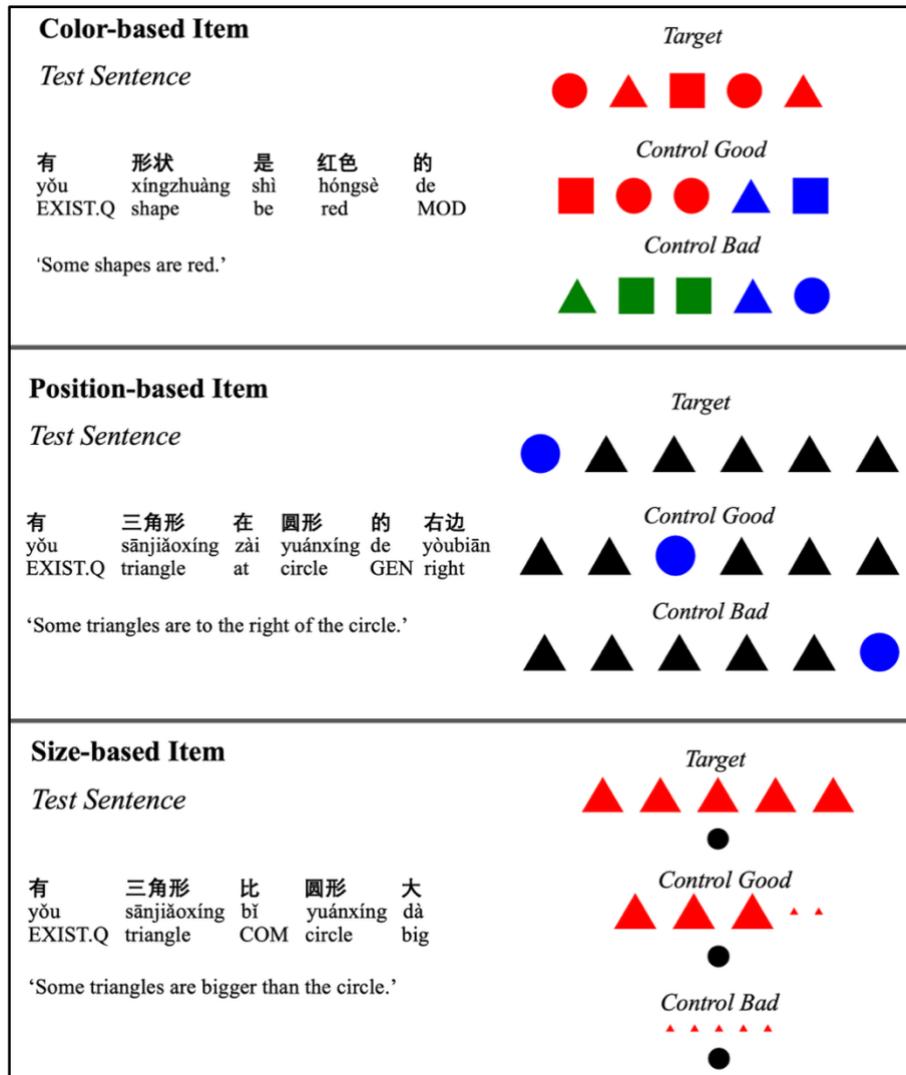
The tested quantifiers included the four most common Mandarin existential expressions: *you*, *you (yi) xie*, *youde* (as noted in Tsai, 2003), along with *yixie*, which is frequently used in prior Mandarin SI experiments. To ensure consistency and simplicity, *you (yi) xie* was tested in its abbreviated form *youxie* throughout the experiment. Since *you* and *youde* cannot appear in postverbal object positions (Tsai, 2003), all test sentences were structured with the quantifier in preverbal subject position, following a *quantifier-NP + VP* format, to control for potential syntactic position effects.<sup>2</sup>

The main experiment employed a within-subjects design, in which all participants encountered all four quantifier expressions.<sup>3</sup> To reduce the risk of fixed response strategies due to item repetition, especially given that this study focused exclusively on existential quantifiers (unlike Tiel and Schaecken (2017) or Ronai and Xiang (2021), who tested a broader range of implicature types), the critical items were diversified across three types: color-based, position-based, and size-based sentence-picture pairs (see Figure 2 for examples).

<sup>2</sup> The test sentence in Figure 1 is an example from the *you* condition.

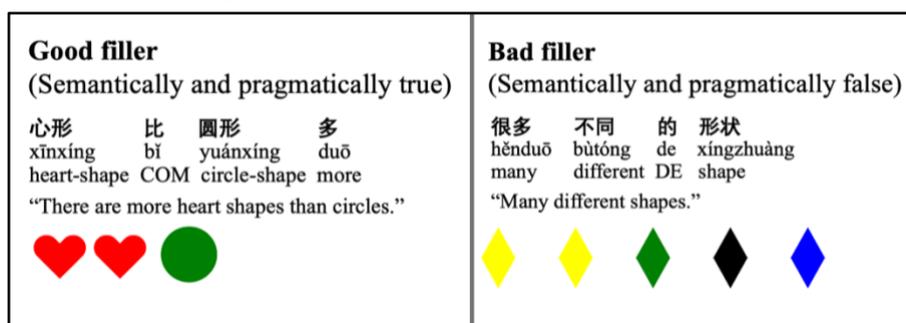
<sup>3</sup> A pilot experiment with a between-subjects design was conducted beforehand, in which each participant was exposed to only one quantifier. The results (reported in Appendix A) suggested that participants' response patterns were strongly influenced by individual biases, motivating a within-subjects design for the main study in order to allow direct comparison across quantifiers within participants.

Figure2. Three Types (Color/Position/Size-based) of Critical Items



In total, each participant saw 72 critical items, evenly distributed across the four existential quantifiers (*you*, *youxie*, *youde*, *yixie*) and the three item types (color, position, size). The three conditions for each item were assigned following a Latin square design, so that each participant saw only one condition per item. Additionally, 48 filler trials (40%) were included. These filler trials did not contain any test existential quantifiers and had fixed semantic & pragmatic truth values, as illustrated in Figure 3.

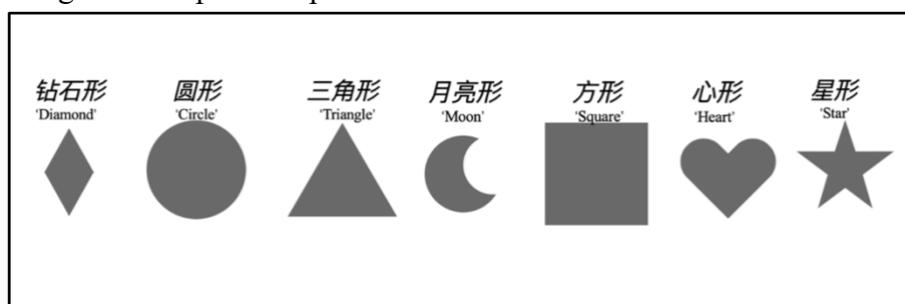
Figure3. Example of Filler Items



### 4.3 Procedure

Before the main trials, participants were presented with a brief instruction in Mandarin. The instruction explained that in each item, they would see a picture along with a sentence describing the picture. The pictures could involve various colors, sizes, and seven types of shapes: diamond, circle, triangle, moon, square, heart, and star (with corresponding image examples shown in Figure 4; English labels are provided here for illustration but were not shown in the actual materials). Participants were instructed to judge whether the sentence accurately described the picture by clicking the corresponding button labeled “恰当” (“good”) or “不恰当” (“not good”) using a mouse or touchscreen. The instructions emphasized: “We are interested in your first response, so please do not think too long before answering.”

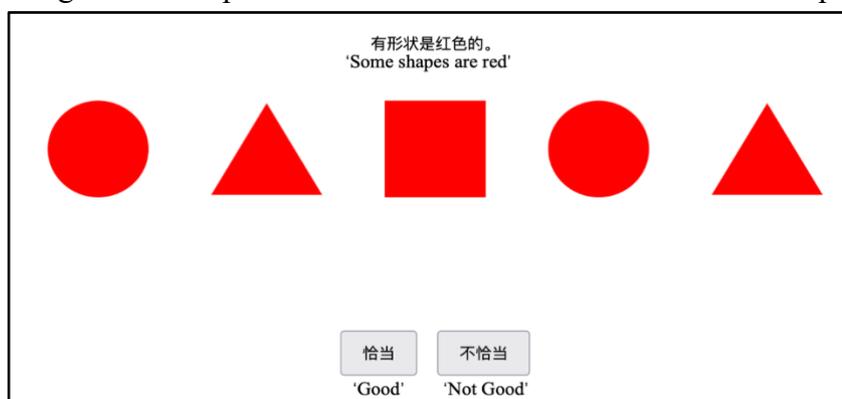
Figure4. Shape Examples Shown in the Instruction Screen



After reading the instructions, participants completed three practice trials, which were not included in the final data analysis. They then proceeded to the main task.

An example trial is shown in Figure 5. In each trial, the sentence and picture appeared simultaneously on the screen. Participants responded by clicking either the “good” or “not good” button. Once a response was made, the experiment automatically advanced to the next trial. This continued until all 120 trials were completed.

Figure5. Example Test Trial: Sentence–Picture Pair with Response Buttons



## 4.4 Results

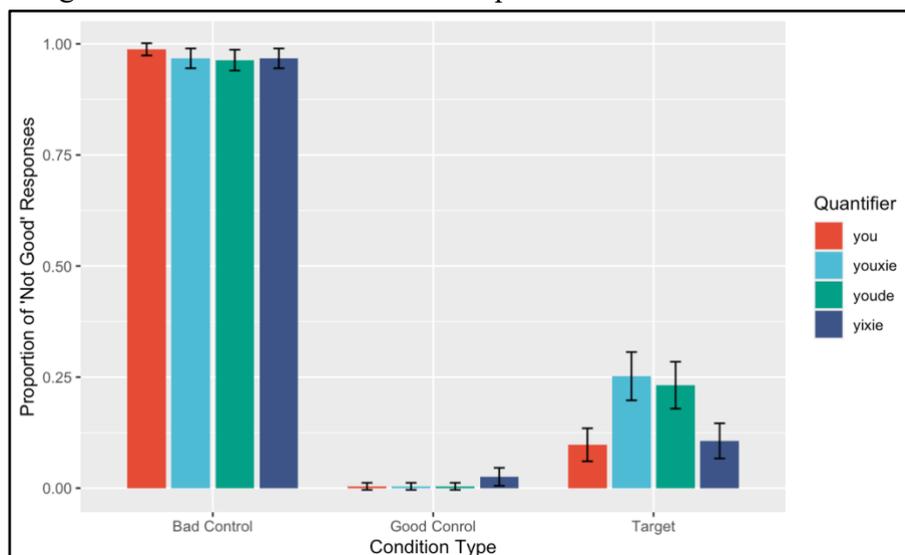
### 4.4.1 Data Treatment

Four participants were excluded for making errors on more than 20% of the filler and control items. Additionally, one participant was removed due to anomalous response patterns in the *yixie* trials, consistently selecting “not good” for all good control items. Data from the remaining 40 valid participants were included in the final analysis.

### 4.4.2 Main Results

Figure 6 plots the mean SI rates for the four quantifiers. The y-axis represents the proportion of trials where participants judged the sentence as “not good,” indicating that they computed a scalar implicature in target conditions (by interpreting “some” as “some but not all.”) The x-axis shows, from left to right, the bad condition, the good condition, and the target condition. Different colors represent the four quantifiers, and error bars indicate bootstrapped 95% confidence intervals.

Figure 6. Mean SI rates across four quantifiers in each condition.



As shown, greater variation appeared in the target condition. For the statistical analysis, a mixed-effects logistic regression model was fit using the *lme4* package in R, predicting binary responses (0 = good, 1 = not good) from quantifier type in the target condition only. The quantifier type variable included four levels (*you*, *youxie*, *youde*, *yixie*), with *you* set as the baseline. Initially, a model with random intercepts for both participants and items was tested but resulted in a singular fit warning. The final model included quantifier type as a fixed effect and a random intercept for participants only. Pairwise comparisons were conducted using the *emmeans* package in (Lenth, 2022), comparing all quantifier types. The results revealed a significant effect of quantifier type on scalar implicature rates. Specifically, *you* elicited significantly lower SI rates than both *youxie* ( $\beta = -2.22$ ,  $SE = 0.38$ ,  $t = -5.83$ ,  $p < .001$ ) and *youde* ( $\beta = -1.97$ ,  $SE = 0.38$ ,  $t = -5.22$ ,  $p < .001$ ). Both *youxie* and *youde* produced significantly higher SI rates

than *yixie* (*youxie* vs. *yixie*:  $\beta = 2.07$ ,  $SE = 0.38$ ,  $t = 5.52$ ,  $p < .001$ ; *youde* vs. *yixie*:  $\beta = 1.82$ ,  $SE = 0.37$ ,  $t = 4.90$ ,  $p < .001$ ). No significant differences were observed between *you* and *yixie* ( $\beta = -0.15$ ,  $SE = 0.40$ ,  $t = -0.38$ ,  $p = .982$ ), nor between *youxie* and *youde* ( $\beta = 0.25$ ,  $SE = 0.32$ ,  $t = 0.79$ ,  $p = .859$ ).

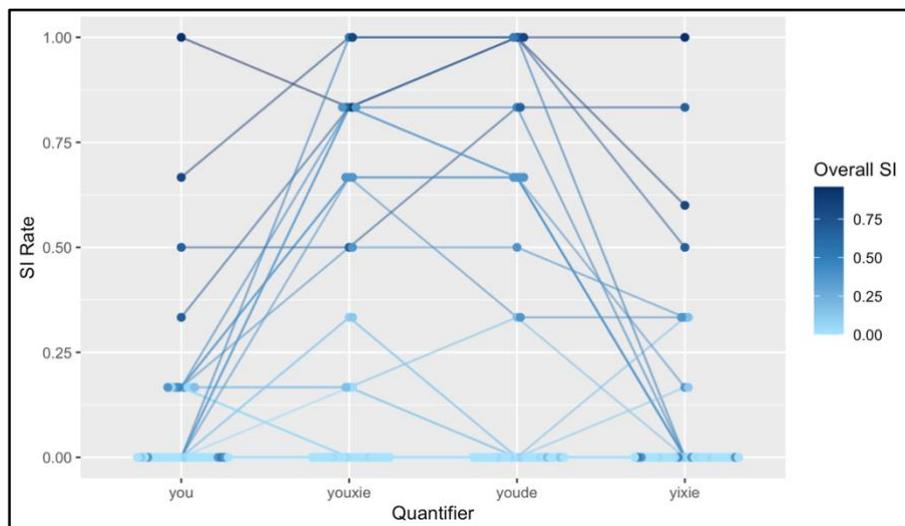
A separate mixed-effects logistic regression with random intercepts for both participants and items was conducted to compare the three types of critical items (color-based, position-based, size-based). The model predicted binary responses from sentence type, with random intercepts for both participants and items. Sentence type was treatment-coded with color-based items as the reference level. Pairwise comparisons revealed no significant differences in SI rates across sentence types (Color vs. position:  $\beta = 0.70$ ,  $SE = 0.40$ ,  $t = 1.76$ ,  $p = .183$ ; color vs. size:  $\beta = 0.27$ ,  $SE = 0.39$ ,  $t = 0.69$ ,  $p = .767$ ; position vs. size:  $\beta = -0.43$ ,  $SE = 0.40$ ,  $t = -1.09$ ,  $p = .521$ .)

These results further confirm the pattern observed in Figure 6: *youxie* and *youde* consistently elicited higher SI rates than *you* and *yixie*.

#### 4.4.3 Individual SI Rates by Quantifier

To further examine participant-level variation, I visualized each participant's SI rates across the four quantifiers. Lighter lines represent participants with lower overall SI rates, while darker lines represent those with higher SI rates. Figure 7 displays this plot, where each line connects a participant's SI rates for the four quantifiers: *you*, *youxie*, *youde*, and *yixie*. The y-axis shows the proportion of target trials where participants gave a pragmatic response.

Figure 7. Line plot of each participant's SI rate across four quantifiers.



The plot reveals two main trends. First, a large group of participants clusters near the bottom with very light-colored lines, indicating consistently low SI rates across quantifiers. This suggests that in my experimental setup, that is, in Mandarin without additional contextual cues, a considerable proportion of participants did not compute scalar implicatures. Second, among participants with higher overall SI rates (darker

lines), nearly all show a consistent middle-high pattern: SI rates are higher for *youxie* and *youde* and lower for *you* and *yixie*. This closely mirrors the aggregate results and indicates that among participants who do compute scalar implicatures, this specific pattern of SI variation across quantifiers is robustly observed.

## 5. General Discussion

### 5.1 Summary

This study employed an online binary choice task to examine whether different existential quantifier expressions in Mandarin elicit variation in scalar implicature (SI) computation when no additional contextual cues are provided. The experimental results confirmed two key points: a. There is clear SI variation among different existential quantifiers in Mandarin; b. The specific pattern shows that for the majority of participants who computed SIs, *youxie* and *youde* consistently elicited higher SI rates than *you* and *yixie*. What remains to be addressed is a crucial question: c. What are the main factors contributing to this observed pattern?

The following analysis turns to two frequently discussed factors which have been independently identified as contributing to scalar implicature strength in prior experimental and theoretical work. Both factors relate directly to the intrinsic properties of scalar expressions themselves and serve as possible explanations for the observed variation: (1) differences in the lexical property of partitivity (as in the case of English *some* vs. *some of*), and (2) variation in the semantic relationship between scalar terms, such as their relative distinctness or the accessibility of stronger alternatives during interpretation (as in the SI variation across multiple different scalar pairs in English).

### 5.2 Partitivity

#### 5.2.1 *Youde* as Overt Partitivity

According to predictions from partitivity-based accounts, expressions with partitive cues are more likely to trigger scalar implicatures (Degen, 2015). Among the four Mandarin existential quantifiers investigated in this study, *youde* is the only one whose partitive nature has been systematically demonstrated. Following the structural analysis proposed by Tsai (2003), we adopt the structure for *youde* as:

S1. [*you*[(part)[*de*[NP]]DP]<sub>NP</sub>]<sub>DP</sub>

This mirrors the structure of the English *some of* (Tsai, 2003, 2004; Tang, 2004):

S2. [*some*[(part)[*of*[DP]]PP]<sub>NP</sub>]<sub>DP</sub>

which has been analyzed as an overt partitive construction (Falco & Zamparelli, 2019). For example:

E1. 有的      学生      通过      了      考试。  
*Yǒude      xuéshēng      tōngguò      le      kǎoshì*

Some of students pass PERF exam  
“Some of the students passed the exam.”

Even in the absence of discourse information or additional contextual cues, *youde* clearly conveys that the described set (‘students who passed’) is a subset of a larger set of students. In other words, *youde* carries an inherent partitive interpretation, directly encoding that the subject is part of a previously known or contextually accessible whole. This aligns well with the SI results observed in our experiment: *youde* consistently elicited significantly higher SI rates.

But what about the other three existential quantifiers? How do they pattern with respect to partitivity? The following section examines this question, drawing on recent insights from Falco and Zamparelli (2024), whose study offers a framework for understanding how certain noun phrases may implicitly encode partitivity even in the absence of overt morphosyntactic cues.

5.2.2 Discourse-Linked DPs as Covert Partitives (Falco & Zamparelli, 2019, 2024)

The concept of *covert partitivity* was introduced by Falco and Zamparelli (2019) as a contrast to traditional *overt partitivity*. Specifically, consider the sentence “Ten students took the exam; three students passed.” Although no overt partitive marker appears in the second clause, the noun phrase *three students* is interpreted as a subset of the previously mentioned *ten students*. This is because *three students* is a *discourse-linked* (D-linked) DP: it is interpreted in relation to a previously introduced set. In this context, D-linked DPs refer to entities that are part of a discourse-salient whole. In contrast, in a sentence such as “Ten students took the exam; three students were exempt,” the second occurrence of *three students* is not necessarily interpreted as a subset of the original *ten students*. Here, *three students* is a non-D-linked DP: its referent is not constrained by the previous discourse context.

In the classic semantic approach, the distinction between D-linked and non-D-linked readings is attributed to contextual domain restriction: a mechanism whereby quantificational expressions are interpreted over contextually salient sets (von Stechow, 1994; Stanley & Szabó, 2000). On this view, sentences like *three students passed following ten students took the exam* are interpreted as referring to a subset of the ten students, not because of any underlying syntactic structure, but due to pragmatic enrichment based on discourse context. However, according to the structural analysis proposed in Falco and Zamparelli (2024), D-linked and non-D-linked DPs are not merely interpreted differently; they also differ structurally. D-linked DPs are analyzed as involving an implicit partitive structure (e.g., *Ten students...Three of the students*), whereas non-D-linked DPs have a simpler nominal structure without partitivity (e.g., *Ten students...Three students*).

This structural distinction is supported by several diagnostics, such as the definiteness effect (e.g., ungrammaticality of definites in existential there-sentences), and the unacceptability of partitives formed with conjoined antecedents, both of which

also apply to covert D-linked constituents but not to non-D-linked cases. Importantly, the cross-linguistic validity of their account strengthens its theoretical appeal: Italian and English show parallel constraints on D-linked readings, but only a structural explanation can capture phenomena such as obligatory clitic realization (*ne*), sensitivity to antecedent structure, and proportional expression behavior. While the semantic approach offers a simpler explanation based on context-driven domain restriction (von Stechow, 1994; Stanley & Szabó, 2000), it cannot account for the syntactic dependencies observed in covert D-linked constructions. For these reasons, this paper adopts Falco and Zamparelli’s (2024) covert partitivity account, which offers a more robust explanation of D-linked readings. This framework serves as the basis for the following analysis of the remaining three Mandarin existential quantifiers in this study.

### 5.2.3 *You* & *Yixie*: No Inherent Partitivity

*You* and *Yixie* themselves do not exhibit overt partitivity. While some analyses have proposed a partitive interpretation for *you*, the present study argues that its apparent partitivity arises only when *you* forms part of a discourse-linked(D-linked) DP, resulting in an implicit partitive structure.

Specifically, Lena (2021) proposes that the [*yǒu* + *rén* + *VP*] pattern allows two distinct readings. The first is a non-partitive presentational reading, which introduces a new, specific referent into the discourse and typically co-occurs with eventive predicates (e.g., motion or action verbs), as shown in the following example:

E2. 有 人 来了  
*Yǒu rén lái*  
 EXIST.Q person come-PERF  
 “There is/are a person/people coming.”

(Tsai, 2003)

In this example, *you* serves solely to introduce new information (“person/people coming”) into the discourse, without any presupposition of a larger set or partitivity.

The second interpretation is the partitive reading. Lena provides the following example:

E3. 有 人 就 讲 哈尔滨 那儿 冷 得 很。  
*Yǒu rén jiù jiǎng Hā'ěrbīn nàr lěng de hěn.*  
 EXIST.Q person then say Harbin there cold COMP much  
 “Some people say that it is very cold in Harbin.”

In this Beijing Mandarin example, the presence of *jiu* signals that the sentence refers to old or previously discussed information in the discourse. For example, it may follow a prior utterance like “No one said Harbin was cold,” where *jiu* marks a contrast or rebuttal to the prior claim. Lena argues that when *you* does not introduce a new referent, it implies partitivity.

However, from the perspective of *D-linked DPs as Covert Partitives* (Falco & Zamparelli, 2024), the source of the partitive meaning in E3 is not *you* itself, but rather the underlying structure of *youren*, which functions as a D-linked constituent. Its deep structure resembles an implicit *youde* structure. In other words, because *jiu* marks the subject as part of the prior discourse (i.e., old information), *youren* here is a D-linked constituent, which can be reanalyzed structurally as *youde*ren.

This suggests that the partitive interpretation in E3 may be consistent with an implicit *youde* structure, not from *you* alone.

Similarly, *yixie* does not carry inherent partitivity. The phrase *yixie + NP* is syntactically a simple nominal structure, where *yixie* functions as a collective operator (Tsai, 2003), indicating that the NP it modifies should be interpreted as a collective subject

E4. 一些 小朋友 在 跳 绳  
*Yixie xiǎopéngyǒu zài tiào shēng*  
 EXIST.Q children PROG jump rope  
 “Some children are jumping rope.”

In this case, *yixie* merely selects a group of children as a whole to serve as the subject of the event, without implying partitivity. However, once the NP becomes discourse-linked, as in E5, an implicit partitive interpretation emerges:

E5. 一班 的 小朋友 在 上 体育课  
*yībān de xiǎopéngyǒu zài shàng tǐyùkè*  
 ClassOne DE children PROG attend P.E. class  
 ‘The children in Class One are having P.E. class.’

一些 小朋友 在 跳 绳  
*Yixie xiǎopéngyǒu zài tiào shēng*  
 EXIST.Q children PROG jump rope  
 “Some children are jumping rope.”

Here, the second mention of *xiaopengyou* (“children”) is clearly D-linked to the initial referent “children in Class One,” thus interpreted as part of the original group. The partitive reading does not arise from *yixie* itself, but rather from its status as a D-linked DP.

Given that the D-linked DPs are involving in an implicit partitive structure, I further hypothesize that this partitive interpretation derives from a covert structure akin to *youde*, though without surface *de*. The actual structure of such a phrase could be represented as:

S3. [[*qizhong* ‘among them’]<sub>pp</sub>[(part)[*de*[*yixie*[*xiaopengyou*]<sup>D-linked</sup>]<sub>NP</sub>]<sub>DP</sub>]<sub>NP</sub>]<sub>NP</sub>

Here, *qizhong* (‘among them’) is a common overt expression in Mandarin discourse that can be optionally omitted. It represents the superset expression, indicating that the referent is drawn from the previously mentioned group, in this case, “children in Class One”. The subsequent implicit partitive structure mirrors the internal configuration of *youde*, where a semantic ‘(part)’ operator introduces an NP that contains a DP complement headed by an elided *de* as its D-head.

#### 5.2.4 *You(yi)xie*: Obligatory Binding to D-linked constituents and the Emergence of Covert Partitivity

The final existential quantifier, *you(yi)xie*, presents an interesting case. Although it does not exhibit overt partitive morphology, I argue that it obligatorily binds to a D-linked constituent, thereby yielding a covert partitive interpretation. Existing research has shown that *you* contributes specificity to the *youxie*-NP, such that the NP is interpreted as definite or anchored (Milsark, 1974; Tsai, 2003). In other words, the determiner *you* ensures that the following *yixie*-NP constitutes a specific entity. This implies that, even in the absence of an explicitly mentioned antecedent, a *youxie*-NP is necessarily associated with a particular set. This set may have been explicitly introduced in the preceding discourse (as in Example 5), or may be pragmatically salient in the common ground shared by speaker and hearer (e.g., when both interlocutors are jointly attending to a known group).

Consequently, the *yixie*-NP commanded by *you* is always D-linked, and thus inherently partitive—even without additional discourse cues. For instance:

E6. 有些 小朋友 在 跳 绳  
*Yǒuxie xiǎopéngyǒu zài tiào shēng*  
 EXIST.Q children PROG jump rope  
 “Some (of the) children are jumping rope.”

In contrast to E4 and E5, even in the absence of explicit discourse such as “children in Class One,” *youxie xiaopengyou* (“some (of the) children”) in E6 can naturally be interpreted as referring to a subset of a contextually salient set. In such cases, while *youxie* may not be inherently partitive in the syntactic sense, it frequently gives rise to a partitive reading due to its association with a specific, contextually salient subset, thus effectively forcing a D-linked interpretation.

Similar to the structure proposed in S3, the deep structure here should also be analyzed as an implicit partitive structure. Specifically, the relative position of *de*, *you*, and *yixie* in this structure can be inferred by drawing on syntactic evidence:

E7. \*其中 有 的 一些 小朋友  
*qizhong yǒu de yixie xiǎopéngyǒu*

among.them EXIST DE some child  
*Ungrammatical*

E8. 其中 的 有 一些 小朋友  
*qízhōng de yǒu yìxiē xiǎopéngyǒu*  
 among.them DE EXIST some child  
 “Some of the children among them.”

Given that in overt expressions, E7 is ungrammatical while E8 is acceptable, I propose that *youxie xiaopengyou* (“some children”) has the following underlying structure:

S4. [[~~*qízhōng*~~]<sub>pp</sub>[(part)[~~*de*~~[*you*[*yixie*[*xiaopengyou*]<sup>D-linked</sup>]<sub>NP</sub>]<sub>DP</sub>]<sub>DP</sub>]<sub>NP</sub>]<sub>NP</sub>

This suggests that the expression *you(yi)xie* does not contain a partitive structure internally; however, because it obligatorily associates with a D-linked constituent, it is consistently embedded within a covert partitive configuration.

#### 5.2.5 Summary of Partitivity

In summary, among the four existential quantifier expressions in Mandarin, we find a clear correspondence between the degree of partitivity and the observed scalar implicature (SI) strength:

*You* and *yixie* do not inherently encode partitivity. Any partitive interpretation must be licensed via prior discourse.

*Youde* is an overt partitive structure, directly parallel to *some of* in English, and consistently carries a partitive meaning.

*You(yi)xie* inherently links to D-linked constituent, forming a covert partitive structure, even without explicit discourse cues.

This partitivity-based analysis aligns closely with the SI variation patterns observed in the experiment: quantifiers with or involved in (overt or covert) partitive structures—*youde* and *youxie*—elicited significantly higher SI rates than those without—*you* and *yixie*. These results suggest that partitivity provides an explanation for SI strength variation across existential quantifiers in Mandarin.

### 5.3 Semantic Similarity

As noted previously, in addition to lexical cues of partitivity, this study also examines whether the observed variation in SI rates across Mandarin existential quantifiers resembles patterns found in different English scalar pairs. Specifically, this involves exploring whether different quantifier expressions vary in the degree of semantic similarity with their stronger scalar alternative *all*. While all four existential expressions in question share the same stronger alternative (the universal quantifier *suoyou*<sup>4</sup> “all”), it remains an open question whether their relation to *suoyou* varies.

<sup>4</sup> I selected *suoyou* as the universal quantifier because, semantically, it most closely parallels *all* in the scalar contrast with *some*, primarily highlighting quantity. While I also considered the expression *quanbu*, it more

Prior work has identified two key dimensions by which the relation between a scalar term and its alternative may influence SI computation: (1) distinctness, where greater perceived contrast between a weaker and stronger term leads to more robust SI derivation (Horn, 1972; van Tiel et al., 2016); and (2) alternative accessibility, where the stronger term is more likely to be considered or retrieved in context, thereby increasing the likelihood of implicature (van Tiel et al., 2016; Hu et al., 2023). A factor that potentially affects both distinctness and accessibility—though its role remains debated—is semantic similarity, defined as the perceived closeness in meaning between a weaker scalar term (e.g., *some*) and its stronger alternative (e.g., *all*).

Van Tiel et al. (2016) hypothesized that higher semantic similarity increases the accessibility of scalar alternatives, reasoning that semantically similar terms are more easily retrieved or expected. On this basis, they predicted a positive correlation between semantic similarity and SI strength. To test this, they measured static semantic similarity using Latent Semantic Analysis (LSA), alongside SI rate, for 43 English scalar expressions. However, they found no significant relationship between semantic similarity and SI rates.

This null result was critically addressed by McNally (2017), who argued that LSA’s context-independent nature fails to capture fine-grained, sense-specific meanings. For instance, while *warm* and *hot* may co-occur in descriptions of temperature, they do not necessarily share meaning when describing personalities or popularity. McNally contended that such context-insensitivity likely obscured semantic distinctions crucial for implicature generation.

Building on this critique, Westera and Boleda (2020) employed context-sensitive word embeddings (ELMo) to re-evaluate the role of semantic similarity in SI computation. Contrary to van Tiel et al.’s hypothesis, they proposed that semantic similarity should be better treated not as a predictor of alternative accessibility, but rather as a negative correlate of distinctness: the more semantically similar two expressions are, the less distinct they appear, and the less likely a scalar implicature is to be generated, because without a clear contrast, the listener has no reason to assume the speaker deliberately chose the weaker term over the stronger one. Their findings showed a negative correlation between semantic similarity and SI strength across the same 43 scalar expressions analyzed by van Tiel et al., although the effect accounted for only 6% of variance.

Following this line of inquiry, I calculated semantic similarity scores between each existential quantifier in Mandarin and the stronger scalar alternative *suoyou* (“all”), using BERT-based contextual embeddings. While Westera and Boleda found ELMo to be well-suited for English, BERT offers comparable context-sensitive representations and is more appropriate for Mandarin, because it is trained on large Chinese corpora

---

typically conveys a sense of wholeness or event-level completeness rather than quantitative scope. For completeness, I conducted an additional analysis averaging the semantic similarity scores derived from both *suoyou* and *quanbu*, but this yielded no substantial change in the results: *youde* and *youxie* still exhibited the highest similarity scores among the four existential quantifiers.

and uses character-based tokenization, which better matches the structure of Mandarin. To ensure make the semantic similarity estimates more representative of the actual linguistic contexts used in the experiment, I constructed three sentence templates corresponding to the three sentence types used in my experiment (color-based, position-based, and size-based), and computed similarity scores within each context. These were then averaged to yield a composite similarity score for each quantifier, grounded in the kinds of environments participants encountered during the task (see Appendix B for full code)<sup>5</sup>.

Based on Westera and Boleda’s study, I initially hypothesized a negative correlation between semantic similarity and SI rate, that is, greater similarity should imply reduced distinctness and therefore a lower likelihood of SI. However, the results revealed the opposite pattern. The actual similarity scores are reported in Table 1.

Table 1. Semantic similarity scores between 4 existential quantifiers and *suoyou* (“all”)

Existential Quantifier	<i>you</i>	<i>youxie</i>	<i>youde</i>	<i>yixie</i>
Semantic Similarity Score	0.6821	0.6960	0.7003	0.6750

*Youxie* and *youde* were found to be more semantically similar to *suoyou* (“all”) than the other two quantifiers, and notably, they also triggered more scalar implicatures. That is, the present findings point to an interpretation of SI variation across different Mandarin expressions that encode the same scalar meaning that runs counter to the prediction based on Westera and Boleda’s account. A positive relationship was observed between semantic similarity to the universal alternative *suoyou* and SI rates. This seems support the view that increased semantic similarity enhances the accessibility of stronger alternatives, thereby facilitating scalar implicature computation.

The idea that alternative accessibility facilitates SI derivation is further supported by experimental evidence from Hu et al. (2023), who used neural language models to derive weighted surprisal values of stronger alternatives as a proxy for their expectedness. Their results showed that pragmatic inferences are shaped by context-driven expectations, with stronger alternatives that are more expected (i.e., associated with lower surprisal) being more likely to trigger scalar implicatures. The current findings extend this account by demonstrating that, even in the absence of various contextual support (with all measures taken in minimal and comparable contexts), semantic similarity can positively influence the activation of stronger alternatives in Mandarin, and thereby exert a positive effect on scalar implicature computation.

---

<sup>5</sup> AI Disclosure: Due to the author’s limited programming experience, the BERT-based semantic similarity analysis code was developed with the assistance of ChatGPT (OpenAI, GPT-4o), which was used to refine Python scripts. This assistance was limited to code implementation and did not contribute to the conceptual development of the study or the interpretation of the semantic similarity results.

In sum, what can be clearly observed is that Mandarin existential quantifiers vary in their semantic similarity to *suoyou*, and this variation appears to correlate positively with the scalar implicature strength they elicit.

## 6. Conclusion and Future Directions

This study investigated whether different existential quantifier expressions in Mandarin Chinese (*you*, *youxie*, *youde*, *yixie*) give rise to systematic variation in scalar implicature (SI) strength. Using an online binary choice task, I tested native Mandarin speakers' pragmatic interpretations of underinformative sentences featuring each of the four quantifiers. The results revealed variation in SI rates across the quantifiers: *youxie* and *youde* triggered significantly higher SI rates than *you* and *yixie*. This pattern was observed both in aggregate results and at the individual participant level, where nearly all SI-generating participants showed the same quantifier ranking.

To understand the source of this variation, two possible explanatory factors were explored. First, the role of partitivity was examined. *Youde* is identified as an overt partitive expression in Mandarin, structurally analogous to English *some of*, while *youxie*—though not overtly partitive—tends to bind a D-linked constituent, thereby inducing a covert partitive structure. In contrast, *you* and *yixie* lack inherent or obligatory partitivity, unless supported by prior discourse. This structural distinction offers a compelling explanation for the observed SI variation: quantifiers that encode (covert or overt) partitivity facilitate scalar enrichment by clearly indicating a subset relation with an accessible superset, thereby strengthening the contrast with the universal alternative. Second, semantic similarity between each existential quantifier and the Mandarin universal quantifier *suoyou* (“all”) was assessed using contextualized BERT-based embeddings. Interestingly, the results revealed a positive correlation between semantic similarity and SI strength: *youxie* and *youde* were more semantically similar to *suoyou* and also triggered higher SI rates. This contrasts with previous results in English scalar expressions (Westera & Boleda, 2020), which reported a negative correlation between semantic similarity and SI strength and interpreted semantic similarity as reducing distinctness. That is, in this case of Mandarin existential expressions, the findings appear to lend greater support to the view that semantic similarity increases the perceived availability of the stronger alternative (*suoyou*), as originally hypothesized by van Tiel et al. (2016), thereby facilitating scalar implicature computation through enhanced alternative accessibility, as demonstrated by Hu et al. (2023).

Nevertheless, this study has certain limitations. The semantic similarity measure was derived from BERT embeddings, which capture distributional semantics but may not fully reflect human judgments of conceptual similarity or scalar relatedness. Future work should include experimental tasks to more directly measure perceived distinctness or similarity. Additionally, while this study separately examined partitivity and semantic similarity, the interaction between structural cues and contextual accessibility of alternatives remains underexplored. For instance, Degen & Tanenhaus (2015)

showed that the presence of contextually salient, more specific alternatives can override the effect of partitivity—a possibility worth investigating in Mandarin as well. Future research could thus test how explicit alternatives or modified QUDs affect the strength of SI triggered by different quantifiers, and whether partitive structures influence this interaction.

The cross-linguistic variation in SI strength also remains puzzling. In this study, Mandarin adults showed relatively low SI rates—around 10% for *you* and *yixie*, and about 25% for *youxie* and *youde*. In contrast, English-speaking participants in similar experiments showed much higher rates—for example, 42% in van Tiel and Schaeken (2016), and more than 45% in Ronai & Xiang (2021). Interestingly, this contrasts with findings from Wang (2022), who reported higher SI rates among Mandarin speakers than English speakers. Future work could further explore the factors contributing to cross-linguistic differences in SI computation.

In sum, this study contributes to our understanding of scalar implicature variability in Mandarin by demonstrating that even among different lexical expressions of the same scalar meaning (existential quantification), systematic variation can emerge. Both partitive structure and semantic similarity to stronger alternatives help explain the observed variation in SI computation, offering novel insights into the pragmatic behavior of quantifiers in Mandarin. These findings highlight the importance of future research on scalar implicature in Mandarin paying closer attention to the choice of scalar expressions and their potential effects, as different forms may themselves shape implicature computation. Prior studies have often relied on a single quantifier: *youxie* (Zhao et al., 2021) or *yixie* (Wang, 2023; Zhang & Wu, 2023), without systematically comparing alternatives. A more careful consideration of lexical variation is therefore crucial for improving our understanding of scalar inference in Mandarin.

## **Abbreviations**

COM = comparative marker

COMP = complementizer

DE = attributive marker *de*

D-linked = discourse-linked

EXIST.Q = existential quantifier

GEN = genitive/relational marker

MOD = modifier marker

PERF = perfect tense marker

PROG = present progressive tense marker

QUD = questions under discussion

SI = scalar implicature

## References

- Cheng, R. L. (1978). Tense interpretation of four Taiwanese modal verbs. In R. Cheng, Y.-C. Li, & T.-C. Tang (Eds.), *Proceedings of Symposium on Chinese Linguistics, 1977 Linguistic Institute of the Linguistic Society of America* (pp. 243–266).
- Degen, J. (2015). Investigating the distribution of *some (but not all)* implicatures using corpora and web-based methods. *Semantics and Pragmatics*, 8(11), 1–55.
- Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive Science*, 39(4), 667–710.
- Dionne, D., & Coppock, E. (2021). Tattoos as a window onto cross-linguistic differences in scalar implicature. In *Proceedings of ELM 1: Experiments in Linguistic Meaning* (pp. 147–158). Linguistic Society of America.
- Falco, M., & Zamparelli, R. (2019). Partitives and partitivity. *Glossa: A Journal of General Linguistics*, 4(1), 1–34.
- Falco, M., & Zamparelli, R. (2024). Discourse-linked DPs as covert partitives: Ellipsis and *pro*-form strategies in Italian and English. *Isogloss: Open Journal of Romance Linguistics*, 10(3), 1–33.
- Gazdar, G. (1979). *Pragmatics: Implicature, presupposition, and logical form*. New York, NY: Academic Press.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). Academic Press.
- Horn, L. R. (1972). *On the Semantic Properties of Logical Operators in English*. PhD Thesis, University of California.
- Horn, L. R. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (Ed.), *Meaning, form, and use in context* (pp. 11–42). Georgetown University Press.
- Hu, J., Levy, R., Degen, J., & Schuster, S. (2023). Expectations over unspoken alternatives predict pragmatic inferences. *Transactions of the Association for Computational Linguistics*, 11, 885–901.
- Huang, C.-T. J. (1988). Shuo *shi* he *you* [On “Be” and “Have” in Chinese]. *The Bulletin of the Institute of History and Philology, Academia Sinica*, 59, 43–64.
- Kursat, L., & Degen, J. (2020). Probability and processing speed of scalar inferences is context-dependent. In *Proceedings of the 42nd Annual Conference of the Cognitive Science Society* (pp. 1236–1242).
- Lena, L. (2021, July). *The semantics and pragmatics of you ren ‘there is someone’ in Chinese*. Paper presented at the 34th Paris Meetings on East Asian Linguistics (JLAO34), CRLAO, Paris, France.

- Lenth, R. V. (2022). *emmeans: Estimated marginal means, aka least-squares means* (R package version 1.8.1).
- Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT Press.
- McNally, L. (2017). Scalar alternatives and scalar inference involving adjectives: A comment on van Tiel et al. (2016). In J. Ostrove, R. Blust, & M. Kim (Eds.), *Asking the right questions: Essays in honor of Sandra Chung* (pp. 17–28). University of California eScholarship.
- Milsark, G. (1974). *Existential sentences in English* (Doctoral dissertation, Massachusetts Institute of Technology).
- Ronai, E., & Xiang, M. (2021). Pragmatic inferences are QUD-sensitive: An experimental study. *Journal of Linguistics*, 57(4), 841–870.
- Stanley, J., & Szabó, Z. G. (2000). On quantifier domain restriction. *Mind and Language*, 15(2–3), 219–261.
- Stateva, P., Stepanov, A., Déprez, V., Dupuy, L. E., & Reboul, A. C. (2019). Cross-linguistic variation in the meaning of quantifiers: Implications for pragmatic enrichment. *Frontiers in Psychology*, 10, 957.
- Su, Y. E. (2013). Scalar implicatures and downward entailment in child Mandarin. *Journal of East Asian Linguistics*, 22(2), 167–187.
- Tang, S.-W. (2004, December 11–12). *Some notes on the syntax of Chinese partitives and indefinites*. Paper presented at the Linguistic Society of Hong Kong Annual Research Forum, Chinese University of Hong Kong.
- Tsai, W.-T. D. (2003). Three types of existential quantification in Chinese. In A. Li & A. Simpson (Eds.), *Form, interpretation and functional structure: Perspectives from Asian languages* (pp. 161–179). London: Curzon/Routledge.
- Tsai, W.-T. D. (2004). 談「有人」「有的人」和「有些人」 [On *you ren*, *you de ren*, and *you xie ren*]. *漢語學報* (Journal of Chinese Linguistics), (2), 16–25.
- Van Tiel, B., & Schaeken, W. (2017). Processing conversational implicatures: Alternatives and counterfactual reasoning. *Cognitive Science*, 41(Suppl. 5), 1119–1154.
- Van Tiel, B., Van Miltenburg, E., Zevakhina, N., & Geurts, B. (2016). Scalar diversity. *Journal of Semantics*, 33(1), 137–175.
- von Fintel, K. (1994). *Restrictions on quantifier domains* (Doctoral dissertation). University of Massachusetts, Amherst.
- Wang, S. (2023). Effects of processing limits on computing scalar implicatures: Evidence from child English and child Mandarin. *Language Learning and Development*, 19(3), 345–368.

- Westera, M., & Boleda, G. (2020). A closer look at scalar diversity using contextualized semantic similarity. *Proceedings of Sinn Und Bedeutung*, 24(2), 439–454.
- Zhao, S., Ren, J., Frank, M. C., & Zhou, P. (2021). The development of quantity implicatures in Mandarin-speaking children. *Language Learning and Development*, 17(4), 343–365.
- Zhang, J., & Wu, Y. (2023). “Yixie” is likely not created equal for everyone: Distinct reasoning procedures in interpreting the same scalar quantifier. *Lingua*, 288, 103523.

## Appendix A

A binary sentence–picture judgment task was conducted using a between-subjects design to examine whether five common Mandarin existential quantifier expressions (*you*, *youde*, *youxie*, *yixie*, *youyixie*) would lead to different scalar implicature (SI) rates. Notably, although *youxie* and *youyixie* are morphologically equivalent, they were tested separately to allow for a finer-grained comparison across groups in the between-subjects design. In the main experiment (within-subjects), *youyixie* was removed to reduce participant burden.

### Participants.

A total of 150 native Mandarin speakers were recruited via e-flyers and randomly assigned to one of five quantifier groups (30 participants each). Each participant was exposed to only one quantifier throughout the experiment. A lottery system selected one participant from each group to receive a \$40 USD prize (or equivalent in RMB) for completing the approximately 5-minute task.

### Materials.

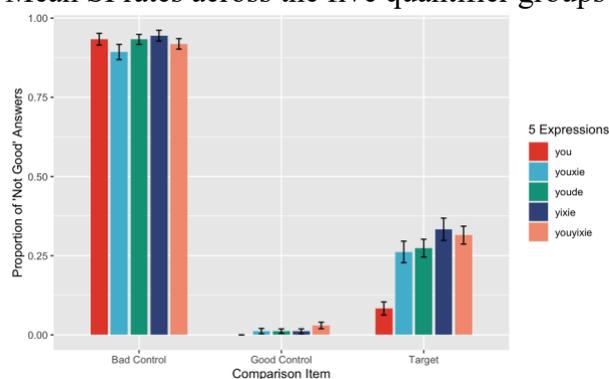
Each participant saw 18 critical items (same as Figure 1 in the main text, Latin-square counterbalanced across three conditions) and 12 fillers (same as Figure 3 in the main text). The only difference across groups was the existential quantifier used in the critical items.

### Procedure.

The procedure was identical to the main experiment (see Section 4.3).

### Results.

Mean SI rates across the five quantifier groups are shown in the plot below.



Two generalized linear mixed-effects models (GLMMs) were fitted. The first model included only a random intercept for item:

```
gml1 <- glmer(Answer_coded ~ Group +  
              (1 | Item),  
              data = some_data_filtered,  
              family = binomial,  
              control = glmerControl(optimizer = "bobyqa"))  
emm_ex1 <- emmeans(gml1, ~ Group)
```

```
pairs(emm_ex1)
```

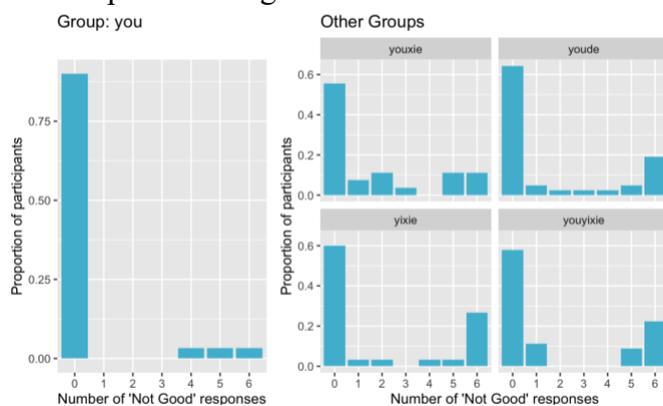
Pairwise comparisons showed significantly lower SI rates for you compared to all other quantifiers (all  $p < .001$ ).

The second model added random intercepts for both participants and items:

```
gml2 <- glmer(Answer_coded ~ Group +  
              (1 | ParticipantID+Item) ,  
              data = some_data_filtered,  
              family = binomial,  
              control = glmerControl(optimizer = "bobyqa"))  
emm_ex2 <- emmeans(gml2, ~ Group)  
pairs(emm_ex2)
```

With this model, no pairwise comparisons reached significance ( $p > .87$  for all contrasts).

The figure below displays the distribution of pragmatic responses (0–6) in the target condition across the five groups. Each participant completed six target trials. A value of 0 indicates no pragmatic responses; 6 indicates all responses were pragmatic. As shown, responses were strongly polarized, particularly in the *you* group, where most participants gave no pragmatic responses at all. That is, many participants responded uniformly across all target items, consistently giving either semantic or pragmatic responses. This pattern was likely due to the repetitive structure of the critical items, which all followed the same *quantifier + shape + are + color* format. This issue was addressed in the main experiment by diversifying the critical items to include color-based, position-based, and size-based sentence–picture pairs, reducing the likelihood of fixed response strategies.



## Appendix B

### BERT-Based Semantic Similarity Code

```
import statistics
from transformers import BertTokenizer, BertModel
import torch
import torch.nn.functional as F

# Load the Chinese BERT model
tokenizer = BertTokenizer.from_pretrained('hfl/chinese-roberta-wwm-ext')
model = BertModel.from_pretrained('hfl/chinese-roberta-wwm-ext')

# List of some quantifiers (in Chinese)
quantifiers = ['有', '有些', '有的', '一些']

# List of all quantifiers (in Chinese)
target_words = ['所有']

# Mapping Chinese quantifiers to pinyin for output
quantifier_pinyin = {
    '有': 'you',
    '有些': 'youxie',
    '有的': 'youde',
    '一些': 'yixie'
}

# Sentence templates: color, size, position
templates = {
    "color": "{} 圆形是红色的。",
    "size": "{} 圆形比三角形大。",
    "position": "{} 圆形在三角形的左边。"
}

# Helper function: extract word or subword embeddings and average them
def get_embedding(hidden_state, tokens, word):
    indices = [i for i, token in enumerate(tokens) if token == word]
    if indices:
        return hidden_state[0, indices[0]]
    else:
        sub_tokens = list(word)
        sub_indices = [i for i, token in enumerate(tokens) if token in sub_tokens]
```

```

        if not sub_indices:
            raise ValueError(f'Cannot find {word} or its subparts in tokens.')
        return hidden_state[0, sub_indices].mean(dim=0)

# Store total scores for averaging
total_scores = {q: [] for q in quantifiers}

# Loop over each sentence type
for template_name, template in templates.items():
    # Construct sentences
    sentences = {}
    for q in quantifiers:
        sentences[q] = template.format(q)
    for t in target_words:
        sentences[t] = template.format(t)

    # Tokenize sentences
    inputs = {k: tokenizer(v, return_tensors='pt') for k, v in sentences.items()}

    # Get hidden states
    with torch.no_grad():
        outputs = {k: model(**v).last_hidden_state for k, v in inputs.items()}

    # Compute average similarity
    for q in quantifiers:
        tokens_q = tokenizer.convert_ids_to_tokens(inputs[q]['input_ids'][0])
        emb_q = get_embedding(outputs[q], tokens_q, q)

        sims = []
        for t in target_words:
            tokens_t = tokenizer.convert_ids_to_tokens(inputs[t]['input_ids'][0])
            emb_t = get_embedding(outputs[t], tokens_t, t)
            sim = F.cosine_similarity(emb_q, emb_t, dim=0).item()
            sims.append(sim)

        avg_sim = sum(sims) / len(sims)
        total_scores[q].append(avg_sim)

# Print final averaged results with standard deviation
print("\n==== Average semantic similarity across all sentence types ==== \n")
for q in quantifiers:
    scores = total_scores[q]

```

```
avg = sum(scores) / len(scores)
std = statistics.stdev(scores)
print(f' {quantifier_pinyin[q]:<6} avg similarity to 'all': {avg:.4f}, std: {std:.4f}')
```