ZEITSCHRIFT FÜR SPRACHWISSENSCHAFT
Journal of the Linguistic Society of Germany

Volume 45, Issue 1, Article 3

https://doi.org/10.18148/zs/2026-2003

Dawei Jin, Shanghai Jiao Tong University, China, daweijin@sjtu.edu.cn

Jun Chen, University of Stuttgart, Germany, junchen113@gmail.com

Wide Scope and Interpretation of *why*: Experimental Evidence for Causal Interrogatives

Abstract: This paper reports three experiment studies investigating the unique syntactic and interpretive properties of the why-adjunct in Mandarin Chinese in comparison to other wh-types. The empirical findings are: (i) Mandarin why-questions demonstrate a different intervention effect from other wh-questions; (ii) multiple wh-sentences with the causal interrogative phrase paired with additional interrogative phrases yield interpretations that are notably less natural than those with other wh-phrases; and (iii) why-questions lack non-interrogative uses, such as free-choice or indefinite meanings. These results converge in showing that the causal adjunct exhibits an exceptional wide scope, consistent with one prominent line of theoretical research.

Keywords: *why*-question, intervention effect, multiple *wh*-question, non-interrogative meaning

Cite as: Jin, Dawei & Jun, Chen. 2026. Wide Scope and Interpretation of *why*: Experimental Evidence for Causal Interrogatives. *Zeitschrift für Sprachwissenschaft* 45(1). https://doi.org/10.18148/zs/2026-2003
Received: 06 November 2024; **accepted:** 28 January 2026; **published:** 25 June 2026

1 Introduction

A longstanding assumption posits that the causal why-adjunct behaves differently from other wh-phrases (Lawler 1971; Cattell 1978; Bromberger 1992; Hegarty 1992; Iatridou & Kroch 1992; Oshima 2007; Tomioka 2009; Jin 2019; Soare 2017; 2021).[1] The core claim is that the why-constituent always has a wide scope reading, tied to its proposition-modifying semantics (Rizzi 1990; 2001; Bromberger 1992; Murphy 2017). Specifically, 'why' is base-generated at the licensing position of the question-typing Q operator, which is above the TP-internal position where other wh-phrases are assumed to merge, based on the idea that the licensing of C-level categories takes place after the licensing of T-level operators.

The exceptional wide scope of 'why' has been diagnosed by its resistance to semantic embedding. Notably, 'why' is never interpreted within the scope of another operator (Lawler 1971; Cattell 1978; Bromberger 1992). The English *why*-question in (1b) must receive the interpretation where *why* stays outside the quantified proposition:

(1) Why did at most a dozen students register for this class?
   a. 'At most a dozen students registered for this class. Why is that?' (*why* > *at most*)
   b. *'What reason is such that at most a dozen students register for this class because of that particular reason?' (**at most* > *why*)

(1b) is compatible with the situation in which no more than a dozen (among the context-relevant students) registered for class. It is nonetheless not compatible with the reading where the situation contains several groups of students that each registered for the class for a different reason, and the speaker seeks a reason that would account for a group of no more than a dozen students.

The restriction to only a wide scope reading is not limited to a particular syntactic position (e.g. the subject position). As (2a)–(2b) further show, *why* cannot be interpreted within the scope of adverbs, either (modified from Tomioka 2009; Jin 2019).

(2) a. Why is Adam **often** on vacation?
   'Adam is often on vacation. Why?' (*why* > *often*)
   *'What reason is such that Adam goes on vacation mostly because of that particular reason?' (**often* > *why*)
   b. Why is Adam **likely** to accept this offer?
   'Adam is likely to accept this offer. Why?' (*why* > *likely*)
   *'What reason is such that it is likely that Adam accepts this offer because of that particular reason?' (**likely* > *why*)

[1] Throughout this paper we use 'why' as a generic representative of the causal adjunct independent of language. We use the italic form *why* to refer to the causal adjunct specifically in English.

A related high attachment proposal is put forward in Rizzi (2001) for *perché* 'why' in Italian, motivated by its interaction with focus. Rizzi argues that a regular wh-phrase is associated with the position of focus projection below CP, compatible with the fact that the co-occurrence of wh-constituents and contrastive focus is degraded.[2] In contrast, as (3) shows, *perché* is able to appear together with contrastive focus (indicated via capitalization) and crucially, co-occurrence *only* takes place when *perché* c-commands the focused constituent (not the other way round), which is indicative of a wide scope interpretation.

(3) a. Perché QUESTO avremmo dovuto dir-gli, non qualcos'altro?
why this have.FUT.1PL must.PST.PTCP say-3SG.DAT not something.else
'Why THIS we should have said to him, not something else?'

b. *QUESTO perché avremmo dovuto dir-gli, non qualcos'altro?
this why have.FUT.1PL must.PST.PTCP say-3SG.DAT not something.else
#'THIS why we should have said to him, not something else?'

Hence the empirical picture points to a relation of causal modification taking place at the clausal level, distinct from other modification at the verbal level.

On the basis of a series of acceptability rating tasks, this paper presents experimental evidence from Mandarin Chinese for the claim of the uniqueness of 'why'. Mandarin *weishenme*, along with similar why-adjuncts in other East Asian languages like Japanese and Korean, has traditionally been at the core of proposals within the external merge, high attachment approach to the why-adjunct (Ko 2005; 2006; Tsai 2008; Stepanov & Tsai 2008; Tomioka 2009; Takita & Yang 2014; Endo 2015; Jin 2016; Miyagawa 2017; Murphy 2017; Cheng 2021). This body of work provides novel, language-specific insights that, until now, have not undergone systematic testing. Taken together, they could build a strong case for a wide scope interpretation. Three testable diagnostics developed in these accounts will be of particular interest to us here. First, 'why' is argued to trigger an intervention violation that is more robust than that induced by the rest of wh-phrases. Second, 'why' cannot be interpreted coherently in multiple wh-questions, in contrast to other wh-phrases. Third, 'why' is uniquely excluded from expressing a range of non-question readings such as universal or existential readings.

We believe controlled experiments play an irreplaceable role in confirming the above generalizations. This is because for the arguments about 'why' being unique by way of its idiosyncratic syntactic, semantic and pragmatic features to work, it is crucial to establish that the judgments for 'why' in the relevant contexts (e.g. an intervening-violating configuration) are qualitatively different from those for the rest of wh-types. The issue is made subtle, however, by the fact that wh-phrases are not homogeneous. Adjuncts *in general* introduce degradation

[2] Rizzi shows that the focus projection in Italian is restricted to the focus expression that overtly expresses a contrast, realized by a stress pattern.

in acceptability judgments, following from properties that distinguish them from arguments, e.g. their inability to anchor an individualizable domain, be referential or be discourse-linked (d-linked). It is thus crucial to be certain that whatever pattern obtained is indeed attributable to the exceptional nature of 'why' rather than a broader argument-adjunct difference. To this end, quantitative methodology is desirable with its ability to detect multiple levels of contrast (Gibson & Fedorenko 2010; Gibson et al. 2013; Schütze 2016; Linzen & Oseki 2018), compared to the more limited scope of introspection-based, judgment-of-minimal-pairs analyses.

The novel contribution in this paper thus lies in obtaining a clear understanding of whether and to what extent the judgment of why-sentences differs from that of other wh-types by conducting an experimental evaluation. Specifically, we conducted three rating-based formal acceptability tasks, using stimuli representing intervention-violating configurations, multiple wh-configurations and non-interrogative contexts.

The rest of this paper is structured as follows. Section 2 presents the relevant background. Section 3 reports the experiments. Section 4 discusses the findings. Section 5 concludes the paper.

2 Theoretical background

In this section, we discuss three diagnostics from previous literature that make the case for the exceptional wide scope of 'why' in Mandarin. Specifically, these diagnostics compare 'why' with other wh-phrases in contexts of intervention effects, multiple wh-questions, and non-interrogative wh-sentences. We then subject these diagnostics to empirical testing by conducting three experiments, which are detailed in Section 3.

2.1 Intervention effects

The distinction between 'why' and other wh-phrases factors into the pattern of intervention effects. We understand wh-intervention in its standard sense as a constraint wherein a wh-restriction cannot have a scopal operator other than its own question operator as its closest c-commanding operator (Hoji 1985), i.e. (4).

(4) *[Q [Op [… *wh* …]] …]

(5) instantiates one such garden-variety intervening configuration, because a scopal element stays closer to the wh-phrase than its Q operator.

(5) ??Zhiyou tamen jia maidao-le shenme?
only they family buy-PRF what
'What did only their family successfully purchase?' *[Q_j [F-Op_i [… F_i …] […*what*$_j$ …]]]

Importantly, while all wh-types are observed to incur a degraded judgment to a certain degree under the intervening configuration, they do not pattern the same. Rather, the literature has often characterized the causal adjunct in Mandarin as inducing the most robust/reliable/severe intervention effect (Tsai 1994; Cheng & Rooryck 2000; Kim 2002; Ko 2005; Soh 2005; Yang 2011; Li 2011; Jin 2016). According to these reported judgments, the following structure with 'why' being c-commanded by a focus phrase is more degraded compared with a corresponding intervening configuration featuring other wh-phrases.

(6) *Zhiyou tamen jia weishenme maidao-le youpiao?
only they family why buy-PRF stamp
Intended: 'For what reason$_i$ did only their family successfully purchase the stamps t_i?'

A predominant approach where this is accounted for relies on the notion that wh-adjuncts are less acceptable in intervention contexts than wh-arguments, and 'why' is particularly so by way of being the paradigmatic adjunct (Cheng & Rooryck 2000; Law 2001; Kim 2002; Soh 2005; Yang 2011). This approach encounters a problem in Japanese and Korean, where the causal adjunct presents another exceptional case to the generalization for wh-intervention yet the pattern is quite opposite to the case in Mandarin. Whereas Mandarin 'why' induces particularly robust intervention, 'why' in Japanese and in Korean are uniquely *exempt* from intervention (Ko 2005; Tomioka 2009). Take (7) for illustration. Here the why-adjunct *naze* is situated in an intervention-violating configuration (it is c-commanded by a focus expression). The structure is nevertheless judged as acceptable, even though other wh-phrases are not acceptable in a comparable configuration.

(7) Daremo naze ko-nak-atta-no? (Japanese)
anyone why come-NEG-PST-Q?
'Why did no one come?'

Both the pattern of 'why' in Mandarin and the one in Japanese and Korean receive a unified explanation, if (instead of classing 'why' as a prototypical wh-adjunct) we adopt the approach where 'why' takes scope differently (Ko 2005; Tomioka 2009; Li 2011; Jin 2016; 2019). In this view, 'why' is attached high, its base position being where it is interpreted, i.e. the scope position of the Q operator. Now a focus phrase needs to be interpreted in the scope of a sentence type/speech act operator such as Q. It thus must also stay below the scope of 'why', as in (8). In other words, even with 'why' at its base position, we arrive at a configuration that induces no intervention, given that said base position is high enough.

(8) [Q_j *why*$_j$ [F-Op_i [… F_i …]]]

This approach allows us to account for the circumvention of 'why' in Japanese and Korean as in (7). We can do so, if we further consider the scrambling characteristic of both languages:

Scopal elements in Japanese and Korean can undergo a free word order operation that affects surface order (i.e. scramble). Although focus may precede Q at the surface level of representation, this happens via scrambling, which is standardly taken to apply on the PF side of the grammar and thus does not modify the LF structure. At LF, the focus operator still scopes below Q. This way, the surface form of (7) has an LF as in (8), and it is the latter that feeds into the semantic representation, giving rise to an unproblematic interpretation.

The unacceptability of 'why' in Mandarin also follows, assuming one difference with Japanese and Korean: Scopal elements in Mandarin are strictly isomorphic (Huang 1982; Ernst 1994; Huang et al. 2009). They cannot scramble. A surface structure with focus c-commanding 'why' such as (6) would require an LF as in (9), assuming scope isomorphism. However, this is uninterpretable.

(9) *[Zhiyou$_{\text{F-Op}}$ [tamen jia]$_{\text{F}}$ [$_{\text{Q}}$ weishenme maidao-le youpiao?]]
only they family why buy-PRF stamp
Intended: 'For what reason$_i$ did only their family successfully purchase the stamps t_i?'

⇝ *[F-Op$_i$ [… F$_i$ …] [Q$_j$ *why*$_j$…]]

When taking account of one parametric variation, the wide scope proposal therefore enables us to account for both the ungrammaticality in Mandarin and the circumvention in Japanese/Korean.

Crucially, the wide scope explanation comes down to a property specific to 'why'. If this explanation holds, the ungrammaticality induced by Mandarin 'why' will be an altogether different phenomenon from the garden-variety wh-intervention represented above in (4), despite their apparent similarity (cf. Jin 2022). Recall that canonical cases of wh-intervention feature a focus operator *scoping between* Q and its wh-restriction. Differing from that, the ungrammaticality represented by the structure in (9) is induced, because there can be no higher scopal operator c-commanding the already high-attached 'why'. There is simply no such position available for the interpretation of focus in Mandarin. Said ungrammaticality is triggered without real 'intervening', as no separation between Q and its wh-restriction is involved. Operating from the view that 'why' does not trigger *bona fide* intervention, the wide scope approach to 'why' thus predicts that acceptability judgment of 'why' differs significantly from that of canonical wh-adjuncts in an intervening context in Mandarin. Alternatively, if 'why' belongs with other wh-adjuncts in terms of position and interpretation (the why-as-adjunct view), it is predicted that acceptability judgment for why-sentences patterns with that of canonical wh-adjuncts.

2.2 Multiple *wh*-questions

Another aspect where the uniqueness of 'why' manifests itself lies in its inability to give rise to a coherent interpretation in a multiple wh-question. A Mandarin constituent question is formed with the wh-phrase staying *in situ*, i.e. occupying the same position as its non-interrogative counterpart. The wh-in-situ property applies, both when the question contains a single wh (e.g. 10), and when multiple wh-s are involved (e.g. 11).

(10) Ni zai na'er xia dan de?
you LOC where place order PRT
'Where did you place your order?'

(11) a. Ni shenme shihou zai na'er xia dan de?
you what time LOC where place order PRT
'When and where did you place your order?'
b. Ni shenme shihou zai na'er zenme xia dan de?
you what time LOC where how place order PRT
'When, where, and how did you place your order?'

Notably, *weishenme* 'why' is argued to stand out, excluded from a multiple-wh reading that is elsewhere available to other wh-phrases (Stepanov & Tsai 2008; Takita & Yang 2014).[3] This is exemplified in (12).[4]

(12) a. *Ni weishenme shenme shihou xia dan de?
you why what time place order PRT
Intended: 'Why and when did you place your order?'
b. *Ni weishenme shenme shihou zai na'er xia dan de?
you why what time LOC where place order PRT
Intended: 'Why, when and where did you place your order?'

The wide scope approach to *weishenme* renders an explanation for its lack of multiple-wh readings (Takita & Yang 2014; Endo 2015; Jin 2019; 2020). Consider first how a single-wh interpretation proceeds. As (13) illustrates, a regular wh-restriction is assumed to be base-generated at the T domain, where it introduces a variable (wh_i) at one of the argument slots of the proposition it is contained in ([$_{TP}$… wh_i…]). At a higher level of compositional interpretation,

[3] One anonymous reviewer points out that the degraded judgment appears to be ameliorated with the position of *shenme shihou* swapped against that of *weishenme*. In our following multiple-wh experiment, we adopted this supposedly rescuing configuration to see the proposed generalization about the exclusion of *weishenme* from a multiple-wh reading can still be maintained. We return to this issue briefly in Section 4.

[4] Here the literature traditionally distinguishes between a single-pair reading (e.g. (11a) expects a unique time-location pair as answer) and a pair-list reading (e.g. (11a) asks that for each individual time, what is the corresponding location) (Engdahl 1986; Dayal 2017). Both interpretations are unavailable to (12).

this propositional radical combines with a C head (i.e. C^0). For a question reading to come about, the C head is to be endowed with a [+wh] feature. Following Dayal (2017), what this C head does semantically is to take in said proposition and outputs a set of propositions, crucial to the question interpretation. Finally, at the level above C head, a question operator Q_i at the specifier of the C domain evaluates this set of alternatives.[5]

(13) $[_{\text{Spec, CP}}\ Q_i\ [_{C'}\ i\ [_{C'}\ C^0\ [+wh]\ [_{TP}\ wh_i\ ]]]]$

In the presence of multiple wh-restrictors, each will need to be evaluated by a separate Q operator. A multiple-wh interpretation may arise, if we assume that all question operators merge simultaneously at the specifier of the C-head (syntactically speaking they undergo absorption). Thus, the multiple-wh structure in (11a) (repeated below as 14a) receives an LF as in (14b).

(14) a. Ni <u>shenme shihou</u>$_i$ <u>zai na'er</u>$_j$ xia dan de?
you what.time LOC where place order PRT
'When and where did you place your order?'

b. $[_{\text{Spec, CP}}\ Q_i\ Q_j\ [_{C'}\ j\ [_{C'}\ i\ [_{C'}\ C^0\ [+wh]\ [_{TP}\ wh_i\ wh_j\ ]]]]]$

A proposition containing two variables is fed to the C head (wh_i and wh_j), which yields a set of open propositions. Composition within the C-domain proceeds in parallel to a single wh-question, barring that there is a recursive application of the process where a Q operator composes with a lambda-abstracted open proposition introduced by an index (Q_i with i and Q_j with j).[6]

Consider now the multiple-wh configuration with 'why' paired up with another wh. With 'why' evaluated at an exceptional wide scope above that of a regular wh, the above mechanism cannot apply. The difference is that 'why' does not leave behind a variable at one of the argument slots of the proposition radical. Rather, it directly resides where the Q operator is and is thus above the C head (Takita & Yang 2014). This yields the LF as in (15).

(15) LF for (12a): $[_{\text{Spec, CP}}\ [Q_j\ j\ why_j]\ Q_i\ [_{C'}\ i\ [_{C'}\ C^0\ [+wh]\ [_{TP}\ \ldots wh_i \ldots]]]]$

(15) cannot be interpreted like (14b), and moreover, cannot give rise to any interpretations due to an interpretive conflict: Because the other wh-restriction is below C, the C head needs to be [+wh]-endowed. It takes in a proposition, and outputs a set of propositions which is then fed to 'why' at the level above C. However, here a conflict arises with regards an independent

[5] In keeping with Dayal's system, we assume that this process is done via lambda abstraction: An index introduces lambda abstraction of the wh-restriction that bears that index (e.g. the index i in (13)). Afterwards, a Q operator composes with the lambda-abstracted open proposition introduced by an index, which introduces quantification over the variable. The formal details are not crucial.

[6] In the pair-list reading the two indices may enter into a functional relation via Skolemization as in Dayal (2017), to be ignored here.

truthfulness presupposition. This presupposition, projected by *and only by* why-questions, states that the content expressed by the modified result clause is true (e.g. Hintikka & Halonen 1995: 647).[7] In our case, the modified result clause of 'why' is a set of propositions (a question, following the Hamblin/Karttunen sense), which by definition is not evaluable with a truth value. Consequently, an interpretation failure ensues. Hence, 'why' is only compatible with a single-wh question (where the C head is not [+wh]-endowed), but must be excluded from multiple-wh questions.

For this analysis to be viable, we need to be certain that 'why' uniquely resists multiple-wh contexts. This is necessary in order to rule out other possibilities. A most obvious one is to consider 'why' as an instance of adjunct wh-phrases being incompatible with a multiple-wh environment. There is independent evidence that this is the case in at least some wh-in-situ structures. In English, for example, the judgment of a multiple-wh configuration may be off with both *how* and *why* staying *in situ* (e.g. Huang 1982; Huang et al. 2009).[8]

(16) a. *Who fainted how?
b. *Who arrived why?

The pattern has been addressed by assuming that wh-adjuncts do not have a set of individuals (or do not easily anchor a domain of individuals from context). We will not go into detail to make the discussion overcomplex here (cf. e.g. Reinhart 1997). Suffice it to say that while the exact pattern in multiple wh-in-situ is not settled, the issue becomes relevant in deciding between accounts. Specifically, we expect to observe in Mandarin a uniquely severe degradation for 'why'

[7] Following the prevalent view since Belnap (1969), a why-question expresses causal modification by relating a proposition (i.e. a cause) to another proposition (i.e. the modified result) (see also Szabolcsi & Zwarts 1993; Hintikka & Halonen 1995; Schwarz & Simolenko 2018). For the causation relation to obtain in the first place, the result itself has to be a given, hence the content expressed by the result clause must be true (Hintikka & Halonen 1995; Tomioka 2009; Larson & Sawada 2012; Schwarz & Simolenko 2018; Fortmann 2019; Zifonun et al. 1997: 2292ff.). Thus, the why-question in (1b), repeated below, asks for the reason why the proposition that at most a dozen students registered for class is true.

(i) Why did at most a dozen students register for this class?
Presuppose: At most a dozen students registered.

Based on the same intuition, 'why' has been treated as selecting for a fact (Mellor 1995; 2004), which is ontologically separate from a proposition. The distinction is not crucial for the current account. Note that this special relation holds independent of the temporal specifications of the clauses involved. This truthfulness presupposition may be conceived as a reflex of the relation that is established between the causal explanation and its result: Causal dependence is characterized by a necessity relation, such that assuming there is a cause/explanation, then its result *must* follow as a consequence. Further assuming that wh-questions in general project an existential presupposition, then it follows that the question *why-p* presupposes that there exists a cause for *p*, from the domain of 'why'. From this we derive the truthfulness presupposition.

[8] The generalization with English how-in-situ is not straightforward. While (16a) is degraded, *who arrived how?* is seen as more toward the acceptable end. Thus rather than rigid ungrammaticality, the pattern of judgments shows significant variation, depending on lexical choices.

in a multiple wh-question, under an account that explains the ungrammaticality of 'why' via its high attachment. On the other hand, with 'why' treated as a garden-variety adjunct similar to 'how', it is predicted that 'why' should pattern with 'how' under multiple wh-configurations.

2.3 Non-interrogative *wh*-sentences

Another behavior in which 'why' has been argued to behave differently from the rest of wh-phrases involves its lack of non-interrogative uses: As in many languages (Kuroda 1965), wh-items in Mandarin can express the readings of universals, existential indefinites, free choice items (FCIs) or negative polarity items (NPIs), yet these functions noticeably exclude 'why'. We focus on two of these uses here, the use of wh as an indefinite and as a free choice item.

Wh-indefinites are widely attested across languages (Haspelmath 1997). Mandarin belongs with a class of languages in exhibiting the so-called bare wh-indefinite use, where indefinites are identical in form with the corresponding interrogatives without introducing additional morphology (Yun 2019). This is in contrast to another class of languages, including Japanese and Korean, where wh-indefinites are derived from interrogatives by attaching an affix. In following with a generalization over bare indefinite languages, wh-indefinite expressions in Mandarin are limited to a narrow scope reading, rather than taking free scope as regular indefinites (Cheng 1991; Li 1992; Lin 1998; 2004; Bruening 2007; Yun 2019). They also differ from regular indefinites by having a non-specific and non-referential interpretation. The following instances of wh-indefinite use are only valid when the speaker does not have a witness to the wh-expression in mind. This obligatorily non-specific reading is often reinforced by the co-occurrence of evidentials like *haoxiang* 'apparently' or illocution-weakening particles like *yixia*, which helps with disambiguation by excluding an interrogative reading (Lü 1942; Chao 1968; Xue 2014).

(17) a. Ta shou-li haoxiang ti-le ge shenme dongxi.
she hand-inside seem carry-PRF CLF what thing
'She appears to be carrying something in her hands.'

b. Jiu kan ta zenme caozuo-le yixia, wo de diannao jiu you neng yong-le.
just see her how operate-PRF one.bit I POSS computer then again can use-PRF
'We just watched her operating for a bit on that computer somehow, and my computer was back in business again.'
(taken from Chinese Web 2017 (zhTenTen17) Simplified, Jakubíček et al. 2013)

It has been shown that *weishenme* 'why' differs from other wh-phrases and cannot be interpreted non-interrogatively in an environment that otherwise licenses an indefinite use, as illustrated in the unacceptable case in (18) (Shi 2021; Niu 2023):

(18) *Benlai qingjia-le, ta haoxiang weishenme you keyi lai-le.
originally ask.for.leave-PRF she seem why after.all can come-PRF
Intended: 'She had originally asked to be off, yet it seems like she can come in after all for some reason.'

Previous research has also observed a free choice use that is derived from wh-interrogative words in Mandarin (Lü 1942; Cheng 1991; Giannakidou & Cheng 2006; Sugimura 2007). The free choice reading comes about, when the wh-element is followed by an universalizing morpheme, such as the distributive operator *dou* and *ye*, with *ye* restricted to negative contexts. Thus, the following corpus-attested examples cannot be interpreted as a question, but rather receive an *any*-like interpretation:

(19) a. Shei dou xihuan ta.
who DOU like her
'Anyone would like her.'
b. Zhiyao xin li you sha, na'er dou shi maerdaifu.
as.long.as heart inside have sand where DOU COP Maldives
'As long as there is a place in the heart for sand, anywhere could be the Maldives.'
(taken from Chinese Web 2017 (zhTenTen17) Simplified, Jakubíček et al. 2013)

The example in (20) is taken to show that *weishenme* 'why' also fails to license a free choice reading, contra other wh-phrases (Shi 2021; Niu 2023).

(20) *Yuangong weishenme bei jiegu dou hui dedao san-bei peichangjin.
employees why PASS sack DOU will receive three-times severance.pay
Intended: 'Employees will be given three times the normal severance pay, regardless of the reason for their dismissal.'

To derive non-interrogative interpretations, it is proposed that the Hamblin alternatives expressed by the wh-phrase are evaluated by different embedding operators, instead of being evaluated by the Q operator (Kratzer & Shimoyama 2002; Giannakidou & Cheng 2006; Beck 2019). This means that in wh-indefinites, an operator introduces existential quantification over the Hamblin alternatives. Universal quantification applies to the set of alternatives in the case of a free choice reading. Once again, this process could be subject to an asymmetry by adjuncthood. The gist is that an adjunct that lacks an internal structure is inherently less amenable to anchoring a closed/enumerable domain of individuals, rendering it unclear what set of alternatives the embedding operator is applied to. It could then follow that 'why' lacks non-interrogative interpretations due to its adjunct status (Sugimura 2007; Xue 2014; Shi 2021; Niu 2023).

Alternatively, deviating from subsuming 'why' under adjuncts, Jin (2016; 2020) and Cecchetto & Donati (2017) argue that a wide scope analysis explains the idiosyncratic distribution of 'why'. Note first that the embedding (existential/universal) operators are assumed to bind into the restrictor set of a wh-phrase below the C domain. With the wh-variable now existentially/universally bound, the C head no longer composes with an open proposition and derives a set of propositions out of it for the question interpretation at higher levels (cf. section 2.2 for the role of C head in wh-interrogatives). Instead, it simply selects a propositional radical with no free variables, like it does with any declarative-type sentence radicals. Crucially, however, the way such non-interrogative reading comes about relies on the wh-variable staying within the domain of these embedding operators. Meanwhile, if we assume that the wh-restrictor of 'why' merges exceptionally high and attaches to the scope of the Q operator, it then follows that the Hamblin alternatives of 'why' never interact semantically with the embedding operators. This way, the only option is for the why-variable to receive binding by the Q operator, making an interrogative reading the only possible interpretation.

In a nutshell, we once again arrive at two contrasting predictions, depending on the view adopted. As with previous contexts, a wide scope view predicts that participants will assign a distinctly low judgment rating to why-sentences in non-interrogative contexts. This prediction is in contrast to the one from the alternative, why-as-adjunct view, which groups 'why' with other wh-adjuncts such as 'how', and therefore predicts comparable rating patterns between these adjunct types.

In sum, suffice it to say that there has been a body of observations pointing to the exceptional nature of Mandarin 'why'. These pertain to a distinctive pattern in intervention violations, as well as the unavailability of multiple-wh and non-interrogative interpretations. These patterns, if established, offer evidence for treating the why-adjunct as being attached high where its scope is resolved. At the same time, experimentally testing these claims using rigid formal methods is needed to make sure that the empirical basis of the characterization of 'why' is as solid as possible, which is particularly so given the non-homogeneous nature among wh-types. We now turn to our new study, which aims to fill in a noticeable empirical lacuna by experimentally investigating the behavioral idiosyncrasies of 'why' across three environments.

3 Experiment

The theoretical implications of the current paper crucially follow from the empirical claim that 'why' is unique. To test this claim, we conducted three formal judgment surveys on the following environments: (1) focus intervention-violating configurations, (2) multiple wh-constructions, (3) non-interrogative wh-constructions (including wh-indefinites and wh-free choice items). The

goal is to see if acceptability ratings on a Likert scale differ significantly by wh-type. In the following, we report our experiments in turn. The stimuli lists used in the three experiments of our survey research, as well as a description of the complete statistical analysis and the R code scripts, are available in the OSF repository.

A note on our choice of wh-types and use of the term 'adjunct' is in order here. It is long noted that East Asian wh-in-situ phrases receive different licensing mechanisms. This observation dates back to Huang (1982), who initially proposed a division based on the familiar distinction between syntactic arguments and adjuncts. Subsequent work on wh-in-situ has nevertheless shown that this is not the most revealing contrast. A more theoretically-relevant distinction is nominal status, i.e. whether a wh-element contains a transparent, divisible nominal component. On this dimension, expressions like *shenme shihou* 'what time' or *shenme difang* 'what place' are adverbial but nominal, as they include a decomposable morphological argument within an otherwise adverbial projection. In contrast, adverbials like *zenme* 'how' and *weishenme* 'why' (on its causal reading) are non-nominal. They do not break down into isolable parts. It is argued that only nominal phrases introduce an individual variable, the right type of variable that can be bound at a distance, whereas non-nominals, which cannot supply an individual variable, necessarily trigger feature movement (Tsai 1994; 1999; 2008; Stepanov & Tsai 2008). Tsai and collaborators thus argue that this distinction modulates acceptability judgments in configurations that involve wh-in-situ and covert movement.

Against this backdrop, the current design considers three predictions: (1) 'why' patterns with any other wh-adjuncts, with no distinction whatsoever between them, or (2) 'why' patterns with non-nominal adjuncts like 'how', or (3) 'why' behaves in a *sui generis* way, distinct from both nominal and non-nominal adjuncts. Hypothesis (1) is included only as a minimal baseline, and the substantive theoretical contrast lies with the other two predictions. To decide between these predictions, in the experiments we chose a nominal adjunct, *shenme shihou* 'what time', to contrast with *zenme* and *weishenme*, the latter two being non-nominal. Such choice also has the benefit of ensuring positional consistency. If we adopt a DP like *shenme* 'what' to serve as the nominal baseline, it would not match the positions of a wh-adverbial. In using 'what time' (instead of 'what'), we preserve the predictive adequacy of the nominality theory and simultaneously remove a potential confound arising from positional mismatches.

3.1 Experiment 1

Competing approaches seek to account for the intervention behavior of 'why'. As mentioned above, we mainly consider two substantive approaches. The first approach treats 'why' as a typical, non-nominal wh-adjunct, attributing its intervention behavior to this status. We define

nominality in terms of whether a wh-expression has an individual(ized) domain, realized as a divisible, transparent internal structure. The temporal adverb *shenme shihou* '(literally) what time' has an internal structure on a par with wh-DPs, whereas the manner adverb *zenme* does not. Given 'what time', 'how' and 'why', this view predicts that 'why' does not differ significantly from 'how' in intervention-violating environments. If we observe any difference between 'why' and 'what time', that difference should be explained by a categorial distinction along the line of nominality, i.e. we expect to observe a comparable difference between 'how' and 'what time'. In contrast, the second approach treats 'why' as unique among wh-phrases due to its wide scope, a property not shared by other wh-elements. This approach posits that the intervention effects associated with 'why' differ from canonical wh-intervention. Given 'what time', 'how' and 'why', it predicts that any difference between 'why' and 'what time' cannot be explained in terms of the difference between 'how' and 'what time'.

Participants A total of 50 individuals from Shanghai Jiao Tong University, a public university in China, participated in the experiment. Our language background survey revealed that all participants were self-reported native speakers of Mandarin Chinese. None reported significant prior exposure to a language beyond Mandarin before age 14. Subjects were each compensated with 20 yuan RMB for their participation. We adopted a filler accuracy threshold of 85%.[9] Eight participants were excluded from analysis for falling below this threshold. Forty-two participants remained in the analysis, comprising 16 males and 26 females (age 25.33±4.26).

Materials The experimental stimuli followed a 3×2 factorial design. Three types of wh-expressions were tested: *weishenme* 'why', *zenme* 'how' and *shenme shihou* 'what time'. Among them, 'how' was chosen to contrast with 'why' as a canonical non-nominal wh-phrase, and 'what time' stood in as a baseline for nominal wh-phrases. These were crossed with a focus phrase consisting of constituent 'only' (*zhiyou* in Mandarin) and its focus associate DP. Crosslinguistically and in Mandarin, focus phrases are reported to give rise to the most robust intervention effect (Kim 2002; Soh 2005; Beck 2006; Yang 2011), which we discuss further in section 4. One sample set of target stimuli is given as follows.

(21) a. *Weishenme* 'why'-question

Zhiyou xuesheng weishenme mai tiyu caipiao?
only student why buy sports lottery.ticket
'Why do only students buy sports lottery tickets?'

[9] Our design included both grammatical and ungrammatical fillers. We treated an ungrammatical filler response as erroneous if it received a score of 4 or above, based on a 7-point Likert scale. Similarly, we treated a grammatical filler response as erroneous if it received a score of 4 or below. We then calculated the overall filler accuracy rate for a given participant.

b. *Zenme* 'how'-question

Zhiyou xuesheng zenme mai tiyu caipiao?
only student how buy sports lottery.ticket
'How do only students buy sports lottery tickets?'

c. *Shenme shihou* 'what time'-question

Zhiyou xuesheng shenme shihou mai tiyu caipiao?
only student what time buy sports lottery.ticket
'When do only students buy sports lottery tickets?'

In our design we opted for a nominal adverbial over a DP (e.g. *shenme* 'what'). This is due to the fact that the latter would occupy the object position as in (22), a consequence of Mandarin being wh-in-situ with SVO order. The problem is that choosing a DP introduces a potential confound for lack of a match in position with 'how' and 'why' in the other conditions.

(22) *Shenme* 'what'-question

Zhiyou xuesheng mai shenme?
only student buy what
'What do only students buy?'

Testing the adverbial 'what time' alongside 'how' thus allows us to probe whether or not 'why' can be subsumed under existing formulations of adjuncthood, while simultaneously eliminating a potential confound introduced by positional mismatches.

In choosing to address the concern with positional confound, our design differed from the one in the recent experiment study of Mandarin intervention in Jin & Yan (2024), which tested wh-DPs. Note in addition that Jin & Yan (2024) probed a much broader empirical scope, contrasting focus phrases against quantifiers of varying monotone classes to determine the robustness of each individual intervener type. Our study adopted a narrowly focused approach, settling upon one intervener type with the aim of a larger sample size and increased statistical power for the exclusive analysis of the behavior of 'why'.

To gauge whether an intervention effect is there, we additionally tested the corresponding controls. Contrary to an intervention-violating target sentence, which has a focus phrase occupy a structurally higher position than the wh-phrase, the c-commanding phrase in the control sentence is a definite plural, which is widely seen as not introducing a structural or interpretational violation in that position (Kotek 2014; 2017; Li & Law 2016). (23) illustrates a sample set corresponding to (21). The c-commanding DP consists of an NP selected by the definiteness marker *zheixie* 'these', the latter a combination of a demonstrative form *zhei* and a pluralized classifier *xie*.[10]

[10] We thank the anonymous reviewer for suggesting the use of the current control design.

(23) a. *Weishenme* 'why'-question control

Zheixie xuesheng weishenme mai tiyu caipiao?
DEM.PL student why buy sports lottery.ticket
'Why do these students buy sports lottery tickets?'

b. *Zenme* 'how'-question control

Zheixie xuesheng zenme mai tiyu caipiao?
DEM.PL student how buy sports lottery.ticket
'How do these students buy sports lottery tickets?'

c. *Shenme shihou* 'what time'-question control

Zheixie xuesheng shenme shihou mai tiyu caipiao?
DEM.PL student what time buy sports lottery.ticket
'When do these students buy sports lottery tickets?'

Including controls thus enabled us to obtain the participants' baseline acceptability for the sentences with an intervening configuration: The structure in (21) gives rise to an intervention effect if a difference approaching statistical significance is observed between the judgment acceptability of said structure and its control structure as in (23).

Target sentences were categorized into three lists and assigned to three groups of participants in a Latin square design. Each participant rated 27 sentences in the main trial. These included nine test stimulus sentences with different lexical items. Stimulus items were pseudo-randomly interspersed with twice as many distractor fillers. These fillers were evenly divided into those that were chosen to be apparently acceptable and those that were apparently unacceptable. Control items followed the same Latin square arrangement, yielding another three lists. With both target and control stimulus items included, 378 data points were sent to statistical analysis.

Procedures The experiment was executed in an online environment using the Tencent Wenjuan survey platform (https://wj.qq.com/).[11] After providing informed consent and completing a preliminary language background survey and familiarization with the instructions, participants commenced the tasks. Each task featured a set of four practice stimuli prior to the introduction of the main trials.

Participants assessed the sentences on a 7-point Likert scale (ascending order, 1 being the least acceptable). Sentences were presented individually on a screen. Participants read a sentence and

[11] As a popular survey platform in China, Tencent Wenjuan served solely as the online hosting interface for our questionnaire. Participants were recruited externally via advertisement postings as well as through social networks, which were then provided with a link directing them to the survey. The platform offered a stable, user-friendly interface familiar to China-based respondents. The platform's features suffice to fulfill our purpose of eliciting Likert-based rating judgments. While it lacks certain advanced functionalities, such as recording response times, these are not essential to our current design.

then indicated their choice by clicking on the corresponding point along a scale underneath that sentence, before proceeding to the next sentence on the following page. The task averaged 290 seconds. Note that all sentences were in Chinese characters. The samples presented here have been transcribed in pinyin Romanization for exposition.

Results Given that an interval-scaled data type was being treated here, and because different participants may use these scales inconsistently, each participant's raw responses were transformed using a z-score procedure (Sprouse et al. 2011; 2012; 2013; Schütze & Sprouse 2014). At the same time, we retained the untransformed (raw) Likert ratings in the analysis to ensure that the direction and magnitude of the key contrasts were not artifacts of the z-score transformation. The main text of this paper reports results based on the untransformed scores. Complete analyses of both data types (z-scored and untransformed) are provided in the online repository. Importantly, the z-score transformation preserved all relationships present in the original data and introduced no distortion.

Figure 1 shows the rating results after z-score transformation.

Table 1 lists model-based point estimates for z-scored ratings by type (target vs. control) and condition, with means (SD) and 95% confidence intervals.

Visual inspection of the rating pattern suggests that control conditions, which instantiate an intervention-free c-commanding scopal expression, were rated better than target conditions across three wh-expressions. For the target conditions, 'what time' and 'how' showed similar distributions, whereas 'why' was rated noticeably lower. We additionally ran an ordinal mixed model (contrast coded against the 'how' condition, Christensen 2019) on the rating data. The model (specified as a cumulative link mixed model, `clmm (Rating ~ wh.category * type + (1|participant) + (1|item))`) predicted ratings from a fixed effect of `wh.category` interacting with `type` (`target` or `control`) and two random effects: By-`participant` and by-`item`. Post hoc pairwise comparison was conducted using the `emmeans` package with Tukey's Honest Significant Difference test (with α-adjustments). The model revealed a significant main effect of `type`, such that target sentences received significantly lower ratings than their corresponding controls. In addition, there was a significant interaction between `wh.category` and `type`, indicating that the effect of wh-category differed between target and control groups. To unpack this interaction, we separately conducted pairwise comparisons within the target and the control conditions. Within the target condition, sentences with 'why' were rated significantly lower than those with 'how' ($\hat{\beta}$: −1.45±0.35, $p < 0.001$) and 'what time' ($\hat{\beta}$: −1.87±0.35, $p < 0.0001$). No significant difference was observed between sentences with 'how' and with 'what time'. Within the control condition, ratings did not differ significantly across wh-types, indicating that wh-category had no detectable effect on the acceptability of control items.

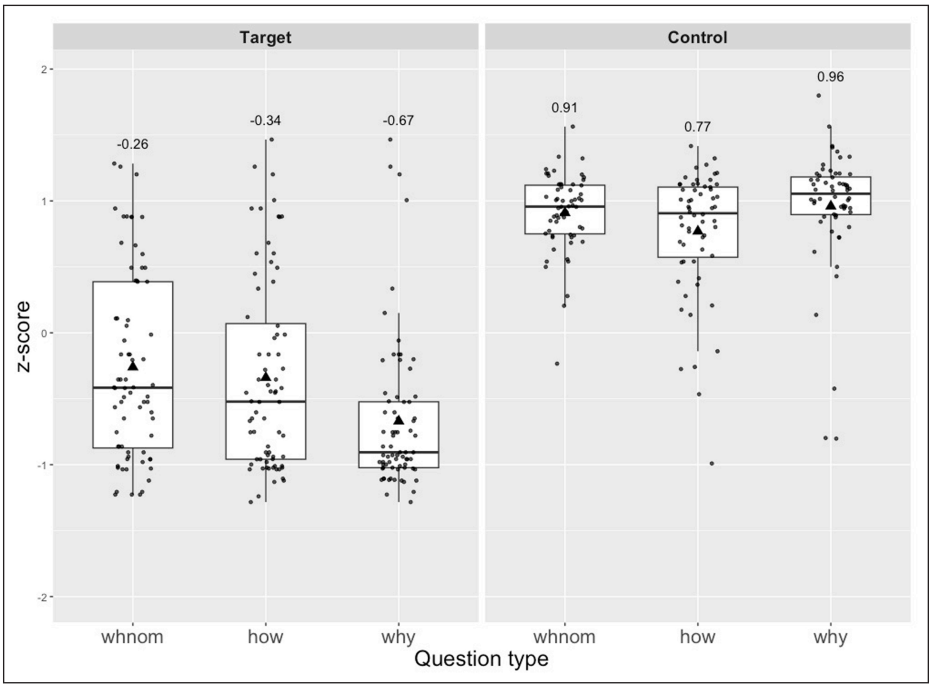

Figure 1: Acceptability judgment ratings across wh-type in the intervention configuration with standardized z-score transformation, represented in a combined boxplot and scatterplot. The left panel represents the target, while the right panel represents the control. In each panel, the three boxes correspond to the three question types: *what time* (coded as `whnom`), *how*, and *why*. Transformed z-scores range from −1.28 to 1.80. Numbers on the plot (as well as the triangles within the respective boxes) represent the calculated means. Horizontal solid lines on the boxes represent median values. The closer the absolute value to zero, the less it deviates from conditioned standard deviation. Positive and negative signs of the numbers represent the direction of a deviation.

Table 1: Z-scored ratings in Experiment 1.

Condition type	Wh-type	z-scored means (SD)	95% CI
Target	whnom	−0.26 (0.71)	[−0.50, −0.02]
Target	how	−0.34 (0.73)	[−0.58, −0.10]
Target	why	−0.67 (0.58)	[−0.91, −0.43]
Control	whnom	0.91 (0.29)	[0.70, 1.13]
Control	how	0.77 (0.48)	[0.55, 0.99]
Control	why	0.96 (0.46)	[0.74, 1,18]

Overall, this experiment has uncovered a significant difference between why-questions and other types of wh-questions with respect to focus-induced intervention. Why-questions elicited more pronounced intervention effects, while a convergence in judgments was observed between 'how' and 'what time'.

An anonymous reviewer wondered about the factor of the possible readings of 'why' and its effect on the acceptability judgment. 'Why' is known to have ambiguous interpretations. Thus, *weishenme* may ask for the cause of an action as well as the purpose of one (the latter based on a *wei* 'for' + *shenme* 'what' parse). Importantly, on the purposive parse *weishenme* is structurally on a par with other nominal wh-phrases, which is expected to lead to an amelioration of the intervention violation. To the extent where purpose 'why' is available to the speakers, its rescuing reading would therefore bias acceptability *upwards*. This means our observed degradation reported above makes for a conservative estimate of how unacceptable causal 'why' is: If participants obtained a mixed interpretation during the task, those obtaining a strictly causal interpretation should assign an even lower score. Our claim about the distinctiveness of 'why' can therefore be maintained when interpretive ambiguity is considered. In fact, the current pattern could even underestimate the effect size of the distinctive behaviors of causal 'why'.

To better understand how interpretive ambiguity factors in boosting acceptability judgments, we conducted a follow-up analysis.[12] We were able to do so, by taking advantage of the fact that the stimuli divided into two (equal-numbered) groups of lexical sets. One group lexically disfavors a purposive reading, whereas the other group allows it. The first, 'exclusive' group featured nonhuman-denoting lexicalizations (e.g. *why do only woodpeckers eat insects?*), where the actions are naturally interpreted causally rather than purposively, because such agents lack full intentional agency and goal-oriented mental representation. In contrast, the second, 'ambiguous' group used human-denoting lexicalizations (e.g. *why do only students buy lottery tickets?*), where the actions can plausibly be understood both in causal terms and in terms of teleologically guided behaviors. Thus, if our experiment participants indeed obtained a purposive reading for *weishenme* in the 'ambiguous' group, and if purposive readings were inflating ratings, we would expect an interaction between *wh*-category and group, that is, compared with the other wh-items, the difference between 'ambiguous' and 'exclusive' *weishenme* is bigger, with 'ambiguous' *weishenme* receiving relatively higher ratings than 'exclusive' *weishenme*.

[12] Another approach brought to bear on this issue is to situate target wh-questions within an appropriate preamble that interpretively biases to one of the two available readings. We did not opt to pursue this approach for the workload it involves; supplying contextualization in a rigid experimental paradigm would lead to a full implementation across all condition types as well as across experiments, going beyond the scope of the current study. See a related analysis in the recent work in Jin & Yan (2025), which restricts the possible readings by supplying a context that settles for one reading per wh-type.

We therefore fitted a mixed-effects model including *wh*-category, group (exclusive vs. ambiguous) and their interaction, and compared it to a simpler model without the interaction. Our model comparison showed no improvement when the interaction was added. In other words, group (reading) did not modulate the pattern across *wh*-categories, that is, the difference in rating between the two groups for 'why' is not bigger than, say, for 'what time'. We take this preliminary result to indicate that, at least for the current experiment, the availability of purposive reading does not significantly impact acceptability judgment. We believe the most likely explanation is that while *weishenme* allows itself to be interpreted ambiguously, not all participants actually obtained a purposive parse of *weishenme*, hence the lack of a significant finding. Alternatively, it is possible that our participants indeed obtained the two available parses, yet the acceptability judgments corresponding to them were not distinct enough to reach statistical significance. Note independently that Kim et al. (2023)'s experiment yielded a finding towards this direction. It would be interesting to explore both directions further and afterwards to find the underlying reasons for them.

In sum, we believe the precise impact of ambiguous interpretation (how big an effect it has on acceptability judgment) remains an open empirical question that needs further investigation. At the same time, to the extent that a purposive reading of *weishenme* is available, it would, if anything, be expected to raise acceptability ratings rather than lower them. Taken together, the evidence is compatible with causal 'why' being less acceptable than the other wh-types in focus intervention contexts.

3.2 Experiment 2

We proceed to our second acceptability judgment task, designed to test whether multiple interrogative constructions involving the why-adjunct exhibit an idiosyncratic behavior. Again, the substantive approaches we consider are as follows. One approach treats 'why' as a typical, non-nominal wh-adjunct. Given 'what time', 'how' and 'why', it predicts that multiple wh-questions with 'why' should pattern similarly to those with 'how'. Both should elicit a comparable decrease in acceptability judgment relative to 'what time'. A second approach treats 'why' as uniquely exhibiting wide scope. It follows from the wide scope property that 'why' cannot yield a coherent interpretation when combined with another wh-element to form a multiple wh-question. Consequently, this approach predicts that the decrease in acceptability observed for 'why' (relative to 'what time') cannot be explained in terms of a similar difference observed between 'how' and 'what time'.

Participants A total of 128 individuals from Shanghai Jiao Tong University, a public university in China, participated in this experiment. The data from eight participants was excluded from analysis for falling below our filler accuracy threshold of 85%. The data of 120

participants was included in the subsequent analysis (42 males and 78 females, 25.14±3.77). Our language background survey revealed that all participants were self-reported native speakers of Mandarin. None reported significant prior exposure to a language beyond Mandarin before age 14. Subjects were each compensated with 10 yuan RMB for their participation.

Materials We created three conditions of target sentences, with `wh.category` manipulating three types of phrases: *weishenme* 'why', *zenme* 'how', and *shenme shihou* 'what time'. In addition to the wh-category type in question, each multiple interrogative question contained another controlled/unvaried wh-phrase *shui* 'who'. The resulting three target conditions are each illustrated with one sample stimulus in (24).

(24) a. Multiple wh-question with *weishenme* 'why'
Shui weishenme bangjia-le fushang Huo xiansheng?
who why kidnap-ASP rich.businessman Huo(Surname) Mister
'Who kidnapped Mr. Huo the business tycoon why?'

b. Multiple wh-question with *zenme* 'how'
Shui zenme bangjia-le fushang Huo xiansheng?
who how kidnap-ASP rich.businessman Huo(Surname) Mister
'Who kidnapped Mr. Huo the business tycoon how?'

c. Multiple wh-question with *shenme shihou* 'what time'
Shui shenme shihou bangjia-le fushang Huo xiansheng?
who what time kidnap-ASP rich.businessman Huo(Surname) Mister
'Who kidnapped Mr. Huo the business tycoon when?'

We also created another three conditions of control sentences across the three *wh*-categories to serve as a baseline of acceptability.[13] These are single wh-sentences that correspond with the target multiple-wh questions, with the unvaried wh-phrase replaced by a non-interrogative DP. A sample set of the control stimuli is given in (25).

(25) a. Single wh-question with *weishenme* 'why'
Nimen weishenme bangjia-le fushang Huo xiansheng?
2PL why kidnap-ASP rich.businessman Huo(Surname) Mister
'Why did you guys kidnap Mr. Huo the business tycoon?'

b. Single wh-question with *zenme* 'how'
Nimen zenme bangjia-le fushang Huo xiansheng?
2PL how kidnap-ASP rich.businessman Huo(Surname) Mister
'How did you guys kidnap Mr. Huo the business tycoon?'

[13] We owe it to an anonymous reviewer for suggesting that we contextualize our target data against appropriate control items in Experiment 2, as well as in Experiment 3 which is to follow.

c. Single wh-question with *shenme shihou* 'what time'
Nimen shenme shihou bangjia-le fushang Huo xiansheng?
2PL what time kidnap-ASP rich.businessman Huo(Surname) Mister
'When did you guys kidnap Mr. Huo the business tycoon?'

Altogether there were nine target sentences, which were then distributed across three groups using a Latin Square design, so that participants saw a unique lexical item in each target condition. Each participant rated three target sentences and six filler sentences. Control items followed the same Latin square arrangement, yielding another three lists. In total, 360 data points entered the statistical analysis.

Procedures The procedures of Experiment 2 were the same as Experiment 1. The stimulus items were pseudo-randomly interspersed with twice as many fillers serving as distractors. The entire experiment, including the background survey and the task, lasted around 4 minutes.

Results **Figure 2** shows the rating results by condition after z-score transformation. Key point estimates from Experiment 2 are provided in **Table 2**.

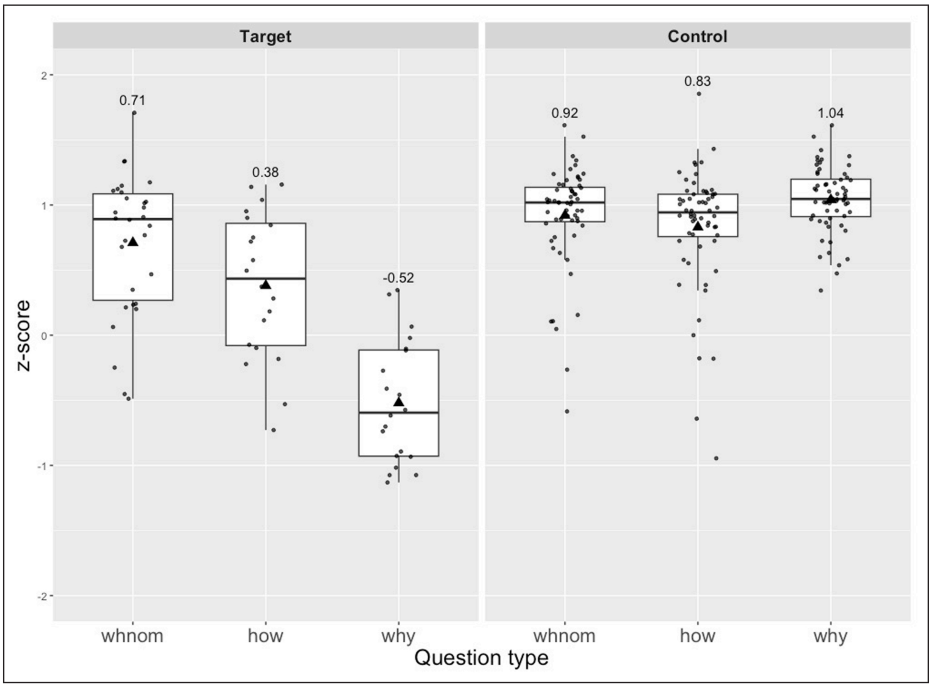

Figure 2: A box plot, augmented with a scatterplot, showing the z-score transformed ratings across wh-type pertaining to the multiple-wh paradigm (left panel: target; right panel: control).

Table 2: Z-scored ratings in Experiment 2.

Condition type	Wh-type	z-scored means (SD)	95% CI
Target	whnom	0.71 (0.54)	[0.39, 1.04]
Target	how	0.38 (0.56)	[−0.01, 0.78]
Target	why	−0.52 (0.47)	[−0.91, −0.11]
Control	whnom	0.92 (0.40)	[0.78, 1.07]
Control	how	0.83 (0.48)	[0.68, 0.97]
Control	why	1.04 (0.25)	[0.90, 1,19]

An inspection of the rating distributions showed a clear difference between target and control items. In the multiple-wh (target) context, 'why' sentences received the lowest rating, followed by 'how' and 'what time' in that order. In contrast, the control sentences exhibited a uniform pattern: all three wh-categories were rated consistently high with no visible pairwise differences, and participant ratings were clustered closely together.

Our ordinal mixed model (`clmm(Rating~ wh.category * type + (1|participant)+ (1|item))`) was fitted to predict Likert scores from a fixed effect of `wh.category` (contrast coded with the 'how' condition) interacting with `type` (`target`, `control`), as well as by-`participant` and by-`item` random effects. Our models revealed a significant main effect of `type`, such that target sentences received significantly lower ratings than their corresponding controls. In addition, there was a significant interaction between `wh.category` and `type`, indicating that the effect of wh-category differed between target and control items. To unpack this interaction, we separately conducted pairwise comparisons within the target and the control conditions. Specifically, sensitivity to wh-category distinctions was present only in the target group. Within the target conditions, post hoc comparisons showed that sentences containing 'why' were rated significantly different from 'what time' conditions ($\hat{\beta}$: −3.51±1.02, $p < 0.001$), as well as 'how' conditions ($\hat{\beta}$: −3.06±0.95, $p \approx 0.003$). The 'what time' condition was rated higher than the 'how' condition, with the difference being marginal. In contrast, the post hoc analysis of the control data did not reveal significant differences among any pairs of these three conditions. Overall, the current experiment revealed that multiple wh-questions involving the why-adjunct are generally unacceptable, in contrast to multiple wh-questions involving 'how' and 'what time'.

3.3 Experiment 3

We now proceed to our third acceptability judgment task, with the aim of testing whether non-interrogative constructions involving the why-adjunct behave differently from those involving

'how' and 'what time'. We conducted two sub-experiments, with the first one testing wh-indefinites and the second one free choice wh-items. In both environments, we test the wide scope approach to 'why' against an alternative, why-as-adjunct approach. Given 'what time', 'how' and 'why', we look into whether 'why' elicits a decrease in acceptability judgment relative to 'what time', and if so, whether said difference is comparable to a similar decrease elicited by 'how'. The approach treating 'why' as a canonical non-nominal predicts that the behaviors of 'why' and 'how' pattern together. In contrast, the approach treating 'why' as uniquely exhibiting wide scope predicts a distinctive, severely more degraded acceptability pattern for 'why'.

Participants A total of 127 participants from Shanghai Jiao Tong University, a public university in China, were recruited for the experiment. The data from 7 participants was excluded from analysis for falling below our filler accuracy threshold of 85%. The data of 120 participants entered statistical analysis (45 males, 75 females, 25.84±3.59). Our language background survey revealed that all participants were self-reported native speakers of Mandarin. None reported significant prior exposure to a language beyond Mandarin before age 14. Subjects were each compensated with 10 yuan RMB for their participation.

Materials We conduced two subexperiments, one for the wh-indefinite paradigm and one for the wh-free choice item paradigm. For each subexperiment, we created three conditions of target sentences, which varied by wh-category. Three types of wh-phrases were manipulated: *weishenme* 'why', *zenme* 'how', and *shenme shihou* 'what time'. 'How' was introduced as a standard case of adjunct against which the behaviors of the why-adjunct were compared, and 'what time' served as a baseline for nominal wh-phrases. A sample set of target stimuli in the indefinite environment is given in example (26).

(26) a. Wh-indefinite construction with *shenme shihou* 'what time'

Dajia dou bu gen xiaoming shuohua. Guji ta shenme shihou
everyone DOU NEG with Xiaoming talk by.estimate he what time
dezui-guo ren.
offend-EXP person
'No one's talking to Xiao Ming. Guess he ruffled some feathers at some point.'

b. Wh-indefinite construction with *zenme* 'how'

Dajia dou bu gen xiaoming shuohua. Guji ta zenme
everyone DOU NEG with Xiaoming talk by.estimate he how
dezui-guo ren.
offend-EXP person
'No one's talking to Xiao Ming. Guess he somehow ruffled some feathers.'

c. Wh-indefinite construction with *weishenme* 'why'

Dajia dou bu gen xiaoming shuohua. Guji ta weishenme dezui-guo ren.
everyone DOU NEG with Xiaoming talk by.estimate he why offend-EXP person
'No one's talking to Xiao Ming. Guess he for some reason ruffled some feathers.'

We additionally created matching control sentences across wh-types in order to establish a baseline of acceptability. The gist was to include a similar question form, this time with a truly interrogative interpretation. A sample set of the control stimuli is given in (27). To enforce an interrogative reading and to exclude a non-interrogative one, a sentence adverb *daodi* 'in the world, on earth' replaced *guji* 'by estimate', which modifies a question speech act (Ernst 1994; 2001). A question-final interrogative particle *ne* is further appended to go with the *daodi*-adverb and ensure that the sentence is overtly marked as the question type (Cheng 1991).

(27) a. Wh-interrogative construction with *shenme shihou* 'what time'

Dajia dou bu gen xiaoming shuohua. Ta daodi shenme shihou dezui-guo ren ne?
everyone DOU NEG with Xiaoming talk he on.earth what time offend-EXP person Q
'No one's talking to Xiao Ming. When in the world did he ruffle some feathers?'

b. Wh-interrogative construction with *zenme* 'how'

Dajia dou bu gen xiaoming shuohua. Ta daodi zenme dezui-guo ren ne?
everyone DOU NEG with Xiaoming talk he on.earth how offend-EXP person Q
'No one's talking to Xiao Ming. How in the world did he ruffle some feathers?'

c. Wh-interrogative construction with *weishenme* 'why'

Dajia dou bu gen xiaoming shuohua. Ta daodi weishenme dezui-guo ren ne?
everyone DOU NEG with Xiaoming talk he on.earth why offend-EXP person Q
'No one's talking to Xiao Ming. Why in the world did he ruffle some feathers?'

A sample set of stimuli in the free choice environment is given in example (28).

(28) a. Wh-free choice construction with *shenme shihou* 'what time'

Wo juede ba, ta name lei, shenme shihou chu cuo-le dou keyi
I feel PRT she that tired what time commit mistake-ASP DOU can
yuanliang.
forgive
'I think given that she's so tired, she can be forgiven no matter when she has made a mistake.'

b. Wh-free choice construction with *zenme* 'how'

Wo juede ba, ta name lei, zenme chu cuo-le dou keyi
I feel PRT she that tired how commit mistake-ASP DOU can
yuanliang.
forgive
'I think given that she's so tired, she can be forgiven no matter how she has made a mistake.'

c. Wh-free choice construction with *weishenme* 'why'

Wo juede ba, ta name lei, weishenme chu cuo-le dou keyi
I feel PRT she that tired why commit mistake-ASP DOU can
yuanliang.
forgive
'I think given that she's so tired, she can be forgiven no matter what reason caused her to make a mistake.'

Wh-question forms with a truly interrogative reading were included as the control, illustrated as in (29).

(29) a. Wh-interrogative construction with *shenme shihou* 'what time'

Ni juede, ta shenme shihou chu cuo-le ne?
you feel she what time commit mistake-ASP Q
'When do you think has she made a mistake?'

b. Wh-interrogative construction with *zenme* 'how'

Ni juede, ta zenme chu cuo-le ne?
you feel she how commit mistake-ASP Q
'How do you think has she made a mistake?'

c. Wh-interrogative construction with *weishenme* 'why'

Ni juede, ta weishenme chu cuo-le ne?
you think she why commit mistake-ASP Q
'Why do you think has she made a mistake?'

While both wh-indefinites and wh-free choice constructions are known to carry special prosody distinct from the interrogative use (Hu 2002; Sugimura 2007), we believe that given that we are dealing with a reading-based task, extra caution is sensible to make sure that participants are truly obtaining a non-interrogative reading for the target sentences and that there is no ambiguity in the available interpretation. To do so, in this design we created a comprehension question for each of the test items: Upon finishing a test sentence, the participant read on the subsequent page a dialogue of the following form in example (30). The A-utterance was repeated from the sentence from the previous page. The B-response offered a confirmation of A's proposition content. The participant was next given a binary forced choice question:

(30) A: Dajia dou bu gen xiaoming shuohua. Guji ta shenme shihou
everyone DOU NEG with Xiaoming talk by.estimate he what time
dezui-guo ren.
offend-EXP person
'No one's talking to Xiao Ming. Guess he rubbed the folks up the wrong way at some point.'

B: Haoxiang shi.
apparently COP
'It appears so.'

Question: Is the response by B to the statement by A natural to you?

A. Yes B. No

The reply of B makes for a felicitous continuation in the ongoing discourse, under the assumption that the target sentence of A is uttered as a non-interrogative/declarative statement. B's reply will not be seen as felicitous, however, with an interrogative construal of the A-utterance: By being cooperative, B is expected to provide new information relative to the wh-portion, not confirmation. In other words, if a non-interrogative reading is what participants consistently obtained, we should expect that they overwhelmingly picked the *Yes*-choice. We could thus reaffirm whether participants successfully obtained the intended interpretation by inspecting the percentage of *Yes*-choices for the target items. This was evaluated against the baselines of filler sentences, which were interspersed with the target stimuli items at a 2:1 ratio in the design. Half of the fillers received a canonically interrogative reading (e.g. 31a), for which we expected that participants overwhelmingly picked the *No*-choice. The other half were declarative sentences (e.g. 31b), for which we expect predominantly *Yes*-choices.

(31) a. Haizi-men zai na'er canjia xialingying huodong?
child-PL LOC where take.part summer.camp activities
'Where do the kids take part in summer camp activities?'

b. Daxiongmao zai caodi shang chi-zhe zhuzi.
giant.panda LOC where like live-PRF
'The giant panda was eating bamboo on the lawn.'

Inspecting the forced choice percentages for both filler types thus offers us a baseline for comparison with the percentage of choices in the target items.

Target sentences followed a Latin square design. 18 target sentences were divided into three lists, with participants rating the target stimuli assigned to three groups accordingly. Each individual rated 6 target sentences and 12 filler sentences. Control items followed the same Latin square arrangement, yielding another three lists. In total, 720 data points entered the statistical analysis (360 for each subexperiment).

The procedures of Experiment 3 were similar to those in Experiment 1 and 2, with the exception that each page containing the stimulus items was followed by another page containing a corresponding forced-choice comprehension question. The stimulus items were pseudo-randomly interspersed with twice as many fillers serving as distractors. The entire experiment, including the background survey, lasted around 8 minutes.

Results As in the previous experiments, **Table 3** reports the key point estimates, separately presented by subexperiment. **Figure 3** shows the ratings by condition for the first (wh-indefinites) subexperiment after z-score transformation.

Figure 4 shows the ratings by condition for the second (wh-free choice) subexperiment after z-score transformation.

Our visual inspection showed that across the three wh-expressions, the 'why' condition received the lowest rating, regardless of whether it occurs in the indefinite construction or the free-choice construction. In comparison, the 'what time' condition and the 'how' condition were rated as generally acceptable in both contexts, with 'what time'-sentences receiving the most natural judgment by the participants. For the control conditions, ratings were generally high across all three wh-categories, with minimal differences between them. Mean z-scores of control items were similar to those of target wh-nominal and target 'how' sentences, but not to target why, indicating that controls patterned with the higher-rated target items. The scatterplots also showed that participant ratings for controls were more tightly clustered than those for targets, reflecting greater uniformity in acceptability judgments.

This pattern was further confirmed by two ordinal mixed models applied to the indefinite and free choice environment separately. Cumulative link mixed models (`clmm(Rating~ wh.category * type + (1|participant)+ (1|item)`) were fitted separately to the two sub-experiments, each applied to its respective data set.

Table 3: Z-scored ratings in Experiment 3.

Condition type		Wh-type	z-scored means (SD)	95% CI
Indefinites	Target	whnom	0.78 (0.58)	[0.54, 1.02]
	Target	how	0.70 (0.47)	[0.46, 0.94]
	Target	why	−0.41 (0.67)	[−0.65, −0.16]
	Control	whnom	0.77 (0.50)	[0.62, 0.92]
	Control	how	0.91 (0.27)	[0.76, 1.06]
	Control	why	0.69 (0.49)	[0.54, 0.84]
Free choice	Target	whnom	0.66 (0.52)	[0.40, 0.92]
	Target	how	0.61 (0.91)	[0.34, 0.87]
	Target	why	−0.42 (0.53)	[−0.69, −0.16]
	Control	whnom	0.64 (0.53)	[0.44, 0.85]
	Control	how	0.56 (0.48)	[0.36, 0.77]
	Control	why	0.62 (0.57)	[0.42, 0.83]

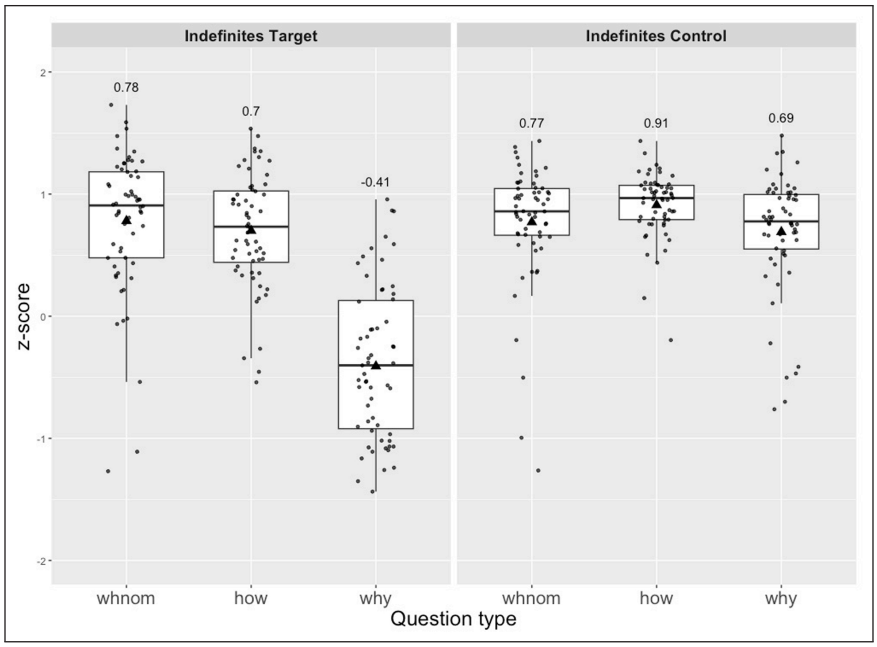

Figure 3: A box plot, augmented with a scatterplot, showing the z-score transformed ratings across wh-type pertaining to the indefinites paradigm (left panel: target; right panel: control).

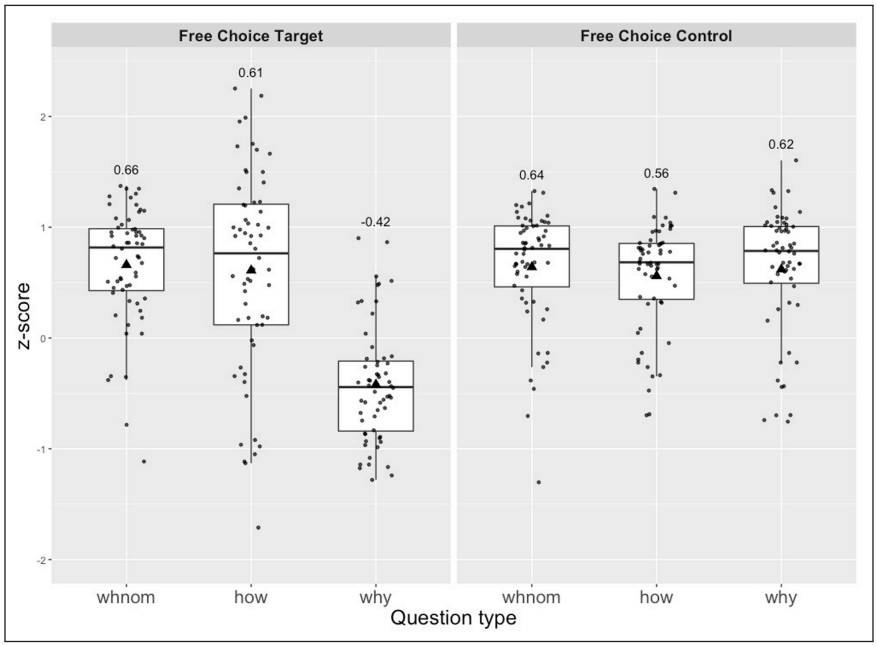

Figure 4: A box plot, augmented with a scatterplot, showing the z-score transformed ratings across wh-type pertaining to the free choice paradigm (left panel: target; right panel: control).

The models predicted the ratings from a fixed effect of `wh.category` (contrast coded against the 'how'-condition) interacting with `type` (`target`, `control`), as well as by-`participant` and by-`item` random effects. For wh-indefinites, our models revealed a significant interaction between `wh.category` and `type`, indicating that the effect of wh-category differed between target and control items. Separately testing the individual pairwise comparisons within the target and the control conditions using post hoc HSD tests (with α-adjustments), our results revealed that compared with 'how'-indefinites, a higher rating for the wh-nominal indefinites condition ($\hat{\beta}$: 0.83±0.39, $p \approx 0.08$) and a significantly lower rating for the why-indefinites condition ($\hat{\beta}$: −3.93±0.53, $p < 0.0001$) were observed. There was no significant difference between each pair of the three control conditions. For the free choice environment, the model also revealed a significant interaction between `wh.category` and `type`. Within the target conditions, post hoc comparisons showed that the 'why' condition is rated significantly lower than the 'how' condition ($\hat{\beta}$: −2.11±0.423, $p < 0.001$), as well as the 'what time' condition ($\hat{\beta}$: −3.03±0.48, $p < 0.001$). There was no significant difference between the 'what time' condition and the 'how' condition. In addition, there was no significant difference between each pair of the three control conditions.

Overall, our results show that 'why' was rated the least acceptable when interpreted as an indefinite and as a free choice item. Between 'what time' and 'how' conditions, a

higher rating was observed for 'what time' across both environments, which did not reach statistical significance.

Our additional inspection of forced choice results revealed a 95% rate of *Yes*-choices for 'what time' and 'how'-sentences in the wh-indefinite and wh-free choice contexts, i.e. all the acceptable target sentences. For a comparison, the 'good' filler sentences (i.e. judged to be acceptable by participants along the Likert-scale) registered a mean accuracy rate of 89% (averaged over declarative and interrogative sentence types, we considered a forced choice response to declaratives to be accurate if it yielded a *Yes*-choice, and one to interrogatives to be accurate if it yielded a *No*-choice). We take the fact that participants overwhelmingly chose 'Yes'-responses to indicate unequivocally that participants in general are capable of incorporating the lexical, grammatical and orthographical information in the immediate context to obtain a non-interrogative reading for the target sentences, at least for all the acceptable items in our experiment.

4 Discussion

The current study presented three experiments to examine how Mandarin speakers rated sentences with causal 'why'. Our rating paradigms compared the Likert-scale-based acceptability judgments of why-sentences with those of two other wh-categories, i.e. *zenme* 'how' and *shenme shihou* 'what time'. Experiment 1 tested the effect of wh-category on sentence acceptability under a focus-intervention configuration. Experiment 2 looked specifically into the between-category difference within a multiple wh-interrogative environment. Experiment 3 investigated whether wh-category had an influence on acceptability under the context that licenses an indefinite as well as free choice reading.

We offer several findings. First, results from Experiment 1 revealed that all three wh-expressions incurred a degraded judgment when c-commanded by focus, in contrast to the control conditions, which were consistently rated above the midrange on the z-transformed scale. Between each target condition and its corresponding control, we observed a significant difference. Still, when we only take account of target sentences, why-sentences received significantly more degraded judgments than those with both 'how' and 'what time'. The acceptability of 'how' and 'what time' were similarly degraded under focus-induced intervention, for which our model found no significant difference. Hence the evidence pointed to a two-way difference ('why' vs. others) within intervention-violating configurations.

Second, results from Experiment 2 showed that why-sentences were unacceptable in multiple wh-in-situ contexts. We did not observe a categorial distinction between 'how' and 'what time': Regardless of whether the phrase was nominal or not, judgments tended towards being

acceptable. We conclude that a two-way distinction between 'why' and other wh-types best accounted for the pattern of judgments within the environment of multiple wh-questions.

An anonymous reviewer observed that the acceptability of 'why' in multiple-wh configurations could be order-sensitive. Specifically, configurations where the controlled wh-element precedes 'why' are more acceptable than those where it follows 'why'. We wish to emphasize that in our design, we consistently used the configuration that the reviewer identifies as having a potential ameliorating effect: the target *weishenme* was preceded by the second wh-element, *shui* 'who' (cf. example 24a). Even with this preferred order, we found a significant effect of wh-type. In other words, under the potentially rescuing 'who-before-why' order, the 'why' condition received distinctively low ratings compared to other wh-categories. This supports our claim regarding the special behavior of 'why'. If there is a *minimal-pair* difference regarding order, we would expect the 'why-before-who' order to yield even worse ratings than those reported here.

Tentatively, the observation regarding the potential amelioration with the second wh-element preceding 'why' suggests that multiple-wh configurations may interact with exceptional wide scope mechanisms in the formal semantics literature (e.g. a scope-out/quantify-in analysis, cf. Krifka 2001; Kitagawa et al. 2003; Fox 2012). While a detailed discussion is beyond the scope of the current, empirically-oriented study, this remains an interesting field to explore to better delineate the special behavior of *weishenme*.

Third, in Experiment 3, as with the previous two experiments, our results revealed an effect of wh-category on the acceptability judgment of target sentences, which were consistent across the two non-question environments (indefinite and free-choice) investigated. Among them, why-sentences were consistently rated low, with a significant difference observed against the ratings of 'how'-sentences as well as 'what time'-sentences. Our results did not yield a finding of significance with regards the comparison of 'how' and 'what time'. Both types elicited generally acceptable judgments. We conclude that 'why' differs from other wh-types in being excluded from non-question readings.

In sum, the results have converged on a significant finding regarding the why-adjunct across the three experimental settings. Importantly, these findings cannot be explained away by invoking a broader class defined on nominality. This is because we controlled our experiments to include 'how' as a baseline of comparison. If a distinction along the line of nominality underlies the distinctive behaviors observed for the why-adjunct, we would expect that 'why' and 'how' behave on par with one another, which did not turn out to be the case: What was revealed was a robust difference of clear effect size between 'why' and 'how' that were consistent across experiments. In each one of them, 'why' induced the more robust violations: 'Why' incurred stronger degradation in judgment than 'how' in intervention-violating environments. In both multiple-wh and non-interrogative environments, a significant decrease in acceptability

compared against 'what time' was obtained for 'why', but not for 'how'. Taken together, we are able to establish that 'why' (instead of *generic* adjuncts) is uniquely incapable of yielding a coherent interpretation.

As section 2 detailed, establishing these empirical patterns constitutes evidence for the wide scope view of the why-adjunct (Ko 2005; 2006; Tsai 2008; Stepanov & Tsai 2008; Tomioka 2009; Takita & Yang 2014; Endo 2015; Jin 2016; Miyagawa 2017; Murphy 2017; Cheng 2021). According to this view, 'why' does not leave a trace. Rather, it is directly base-generated where it takes scope, which is assumed to be the scope position of the Q operator (Lawler 1971; Bromberger 1992; Tomioka 2009; Stepanov & Tsai 2008; Soare 2021). It thus resides in the C domain instead of the T domain, unlike the other wh-adjuncts. Assuming this, the intervention configuration where focus c-commands 'why' is then not a possible one, as the focus operator must stay within the domain of the Q operator. Next, in a multiple-wh configuration, with 'why' entering compositional interpretation from a structure above the C head, it requires a propositional argument specified with a true value, which conflicts with the necessarily unspecified nature of the question radical it ends up composing with. Finally, in both existential and universal wh-readings, the high attachment position of 'why' has the consequence that its Hamblin alternatives cannot be evaluated by an existential or universal operator. As such, the body of experimental evidence lends support to a growing body of theoretical research regarding the scopal behaviors of 'why' (Lawler 1971; Cattell 1978; Bromberger 1992; Hegarty 1992; Iatridou & Kroch 1992; Oshima 2007; Tomioka 2009; Soare 2017; 2021).[14] A caveat is in order, however, as the current experimental evidence directly motivates the case for a wide scope analysis of Mandarin *weishenme*, and likely carries over to its counterparts in other East Asian wh-in-situ languages. It is, in principle, possible that other languages behave differently. In particular, a potential point of variation lies with languages in which wh-elements regularly undergo movement to take scope. For instance, it has been argued that in at least a subset of these languages, the why-adjunct undergoes auxiliary inversion, which is traditionally taken as a diagnostic for movement (Rizzi 2001; Ko 2005; Thornton 2008; Shlonsky & Soare 2011). It is then possible that the why-element in such languages moves to take scope after all, just like the other wh-elements. We therefore leave open the possibility of parametrization, in the sense that the why-element across languages does not uniformly instantiate the same "no-trace, direct-merge" option. Understanding how this parametrization pans out remains an intriguing next step.[15]

In addition to the empirical support for the high attachment of 'why', our experiments yield additional findings. First, we observed a pattern of degraded judgment for the non-nominal

[14] See independent work from island domains (Jin & Yan 2025; Ji & Jin 2025) that converge on the wide scope behaviors of 'why'.

[15] Note that wh-movement languages are not homogeneous with regards inversion: Italian *perché* 'why' is observed to not trigger auxiliary inversion.

wh-adjunct 'how' as well as the nominal 'what time' in an intervening environment. Second, a mild degradation of 'how' was found in the multiple-wh context. We attempt preliminary explanations of these findings and discuss the implications. We will leave it to future experiments to validate these explanations.

Experiment 1 additionally revealed that 'how' and 'what time' elicited degraded judgments when c-commanded by a focus phrase. While the size of degradation was smaller compared to that of 'why', it was still found that participants rated the intervention-violating structures of the two wh-types lower than the control structures and also the middling range of the scale of judgment, which we consider as evidence for a canonical case of focus intervention effect. Our findings more specifically show a uniform, consistent decline in acceptability across nominality lines. In this aspect the findings replicate the experimental results in Jin & Yan (2024) and confirm a widely held view regarding the particular robustness of focus as an intervener (Kim 2002; 2006; Beck 2006; Soh 2005; Hagstrom 2006; Yang 2009; 2011). A key assumption has been that focus triggers intervention regardless of the wh-category, even if a wh-type distinction might be relevant for other scopal elements (like quantifiers) (Tsai 1994; Cheng & Rooryck 2000; Law 2001; Soh 2005). By confirming this assumption, our findings favor an explanation of focus intervention based on the notion that focus and wh are interpreted in much the same way. Wh-phrases range over Hamblin alternatives (Karttunen 1977), just like focus triggers context-determined alternatives (Rooth 1995). Both the Q operator and the focus operator are sensitive to the alternatives within their domain. Such parallelism has the consequence that when an offending focus operator is closer to wh than the Q operator is, the evaluation of wh-alternatives by Q becomes affected of sorts (Beck 2006; Cable 2010; Li & Law 2016). Regardless of the exact theoretical implementation, the idea pertains to the fundamental alternative semantics of wh, hence is independent of wh-type. This way, focus evaluation accounts are insensitive to the categorial difference between wh-types.[16]

An interesting further question has to do with capturing the relatively milder pattern of unacceptability for the non-'why', garden-variety wh-intervention effect. Some accounts argue that the underlying explanation lies in pragmatics or working memory. A pragmatic explanation to wh-intervention (e.g. Tomioka 2007; Eilam 2008; Branan 2017) builds on the assumption that linearly adjacent positions under a canonical information structural configuration are limited to a partition of new information and old information. Focus encodes new information, and in a question-answer sequence, the wh-element corresponds to the focused constituent in the answer. The simultaneous occurrence of a focus phrase and a wh-expression thus creates an anomalous case to the normal way information is packaged. Alternatively, it is proposed that

[16] The empirical picture is, however, more complicated than presented here. Xie (2014) argues that non-exhaustive focus does not trigger intervention effects. It is also proposed that focus adverbs in Mandarin do not act as interveners, unlike focus DPs (Soh 2005; Yang 2009; 2011). We will leave it to future work to tackle different focus types.

parsing a focus expression is a process of dependency resolution (of the association between focus and its operator). With wh-expression interpreted akin to focus, the intervention-violating configuration leads to a nested dependency of multiple foci, which is independently known to be difficult to process (Lewis & Vasishth 2005; Vasishth & Lewis 2006; Beck & Vasishth 2009; Jin & Yan 2024). In each of these two cases, we expect to see more individual differences in the ability to accommodate pragmatic violations or in the cognitive capacity relevant to processing, which results in more variation in judgment. We expect to explore this issue further in the future.

In Experiment 2, 'how' elicited a discernible decline in judgments compared to 'what time' in multiple-wh contexts. Note that this decline did not reach statistical significance, with 'how'-sentences in general rated above the middling range on the z-transformed scale of judgment. Still, we attempt some explanation here. We suspect d(iscourse)-linking plays a role, in line with a vast literature independently attributing differences in acceptability judgment across wh-category to the variation in d-linking status. Here we understand d-linking as a notion lying at the syntax-discourse interface, as outlined in Comorovski (1987) and Pesetsky (1987). A d-linked 'wh'-phrase is, semantically speaking, one that requires the range of felicitous answers to a question be limited to a contextually salient set. As it stands, phrases denoting individuals (including places and times) more easily settle upon a restricted domain than phrases denoting manners. It is argued that acceptability judgments correlate with the extent to which a phrase lends itself to domain restriction (Anagnostopoulou 1994). The fact that 'what time' invites better judgments than 'how' could be because the former introduces a lexical restriction and settles upon a salient domain from which the alternatives are chosen, and the latter is less readily so.

In addition, the d-linking factor may exert a stronger drag on acceptability in multiple-wh contexts than in single-wh ones. On a related note, we observed that in forming a multiple wh-question that solicits information about the manner of action, wh-expressions with an explicit lexical restriction were preferred over *zenme* 'how'. In an informal follow-up consultation with three native speakers not involved in the main experiments, we asked how they would form a multiple 'wh'-question eliciting both an individual person and a manner, based on materials taken from our stimuli from Experiment 2. In these cases, the favored strategy made use of alternative forms as *yong shenme fangshi* 'with what manner' or *yong shenme shouduan* 'with what means', e.g. *Shui yong shenme fangshi bangjia-le fushang Huo xiansheng?* 'Who kidnapped Mr. Huo the business tycoon in what manner?'. While the use of *zenme* 'how' was still more often than not comprehensible, it was not considered the most common way to pose the relevant question in actual speech. This raises the possibility that the infelicity of *zenme* 'how' is Gricean by nature, as the (more ready) availability of alternative forms affects how speakers accept *zenme* in questions. Future comparison with existing studies on the amelioration effect of d-linking or the effect of constructing alternatives in pragmatic reasoning on wh-questions (e.g.

Kluender 1998; Hofmeister & Sag 2010; Goodall 2015) are needed to understand how broadly these pragmatic factors may apply (cf. Chen 2023 for work on Mandarin).

5 Conclusions

This paper presents results from controlled experiments showing that the why-adjunct in Mandarin behaves in a unique way. A total of three formal judgment surveys we conducted in intervention-violating contexts, multiple wh-interrogative contexts, as well as non-interrogative wh-contexts, all point in the same direction: 'why' presents an exception to the generalization that characterizes wh-questions. This result holds when a fine-grained distinction within wh-phrases is considered. We conclude that the evidence is compatible with a high attachment approach to 'why', argued for in Rizzi (1990), Bromberger (1992), inter alia, according to which 'why' does not leave a trace during interpretation as regular wh-expressions do and is restricted to the widest scope.

Appendix

The list of abbreviations in this paper is given as follows.

ACC	accusative	IPFV	imperfective	Q	question particle
CLF	classifier	LOC	locative	REL	relativizer
COP	copula	NEG	negation	RES	resultative
DAT	dative	NOM	nominative	RP	resumptive pronoun
DEM	demonstrative	PL	plural	SG	singular
DOU	universalizing adverb/particle	POSS	possessive	TOP	topic marker
ERG	ergative	PRF	perfect		
EXP	experiential aspect	PRS	present		
FUT	future tense	PRT	particle		
GEN	genitive	PST	past		
HAB	habitual	PTCP	participle		

Table A1: Abbreviations of glossing terms.

Data Availability Statement: The stimuli lists used in the three experiments of our survey research, as well as a description of the complete statistical analysis and the R code scripts, are stored privately in the OSF repository. We plan to make them accessible in the event that the manuscript is accepted for publication.

Competing Interests: The authors have no competing interests to declare.

Funding: This study was funded by The Philosophy and Social Sciences Office of the Government of Shanghai (grant No. 2021BYY007).

Acknowledgements: We wish to express our deep gratitude to the two anonymous reviewers of *Zeitschrift für Sprachwissenschaft* and the handling editor, Professor Oliver Schallert, for their careful reading and detailed comments, which have improved the quality of the manuscript. We further thank Hanbo Yan for her very valuable help on methodology and data analysis. Early stages of this research benefitted directly from discussions with Xingbao Niu and Ali Hussain Birahimani. The manuscript also received valuable feedback and inspiration through discussions with Sigrid Beck, Aron Hirsch, Tatiana Scheffler, Swantje Tönnis, and Matthias Gerner. Xuziyun Pan and Yan Yang kindly assisted with collecting data and recruiting participants. Lihua Xu, Rick Yan, Min Li, and Yu Li offered support at various stages of this project. We are indebted to the above individuals. Finally, we appreciate the audiences at the linguistic meetings at GDUFS, Aarhus University and Wuhan University. As usual, all the errors in this paper are our own.

References

Anagnostopoulou, Elena. 1994. *Clitic dependencies in Modern Greek*. University of Salzburg dissertation.

Beck, Sigrid. 2006. Intervention effects follow from focus interpretation. *Natural Language Semantics* 14(1). 1–56. https://doi.org/10.1007/s11050-005-4532-y.

Beck, Sigrid. 2019. Which alternatives? Talk given at the DgfS XPrag summer school.

Beck, Sigrid & Shravan Vasishth. 2009. Multiple focus. *Journal of Semantics* 26. 159–184. https://doi.org/10.1093/jos/ffp001.

Belnap, Nuel. 1969. Questions: Their presuppositions, and how they can fail to arise. In Kirk Lambert (ed.), *The logical way of doing things*, 23–37. New Haven: Yale University Press.

Branan, Kenyon. 2017. In-situ wh-phrases in superiority violating contexts don't have to be in-situ. In Claire Halpert, Hadas Kotek & Coppe van Urk (eds.), *A pesky set: Papers for David Pesetsky*, 353–359. Cambridge, MA: MITWPL.

Bromberger, Sylvain. 1992. *On what we know we don't know: Explanation, theory, linguistics, and how questions shape them*. Chicago: University of Chicago Press.

Bruening, Benjamin. 2007. Wh-in-situ does not correlate with wh-indefinites or question particles. *Linguistic Inquiry* 38(1). 139–166. https://doi.org/10.1162/ling.2007.38.1.139.

Cable, Seth. 2010. *The grammar of Q*. Oxford: Oxford University Press.

Cattell, Ray. 1978. On the source of interrogative adverbs. *Language* 54(1). 61–77. https://doi.org/10.2307/412999.

Cecchetto, Carlo & Caterina Donati. 2017. Why there is no non-interrogative *why*. Paper presented at 2017 SLE Annual Meeting, Zurich.

Chao, Yuen-Ren. 1968. *A grammar of spoken Chinese*. Berkeley: University of California Press.

Chen, Jun. 2023. *On shi and de in Mandarin: clefts and beyond*. Universität Stuttgart dissertation.

Cheng, Lisa. 1991. *On the typology of wh-questions*. MIT dissertation.
Cheng, Lisa. 2021. What-as-why sentences in Cantonese. In Gabriela Soare (ed.), *Why is why unique?*, 219–248. Berlin: de Gruyter Mouton.
Cheng, Lisa & Johan Rooryck. 2000. Licensing wh-in-situ. *Syntax* 3(1). 1–19.
Christensen, Rune Haubo Bojesen. 2019. Ordinal—-regression models for ordinal data. Manuscript, Technical University of Denmark.
Comorovski, Illeana. 1987. Discourse-linking and the wh-island constraint. *Papers of the North East Linguistics Society* 19(1). 254–274.
Dayal, Veneeta. 2017. *Questions*. Oxford: Oxford University Press.
Eilam, Aviad. 2008. Intervention effects: Why Amharic patterns differently. In Natasha Abner & Jason Bishop (eds.), *WCCFL 27*, 141–149. Somerville, MA: Cascadilla Proceedings Project.
Endo, Yoshio. 2015. Two ReasonPs: What are*(n't) you coming to the US for? In Ur Shlonsky (ed.), *Beyond functional sequence*, 220–231. New York: Oxford University Press.
Engdahl, Elisabet. 1986. *Constituent questions*. Dordrecht: Reidel.
Ernst, Thomas. 1994. Conditions on Chinese A-not-A questions. *Journal of East Asian Linguistics* 3(3). 241–264. https://doi.org/10.1007/BF01733065.
Ernst, Thomas. 2001. *The syntax of adjuncts*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/cbo9780511486258.
Fortmann, Christian. 2019. *Warum*-infinitives in German. *Linguistische Berichte* 258. 179–227. https://doi.org/10.46771/2366077500258_2.
Fox, Danny. 2012. Multiple wh-questions: uniqueness, pair-list and second order questions. MIT Seminar.
Giannakidou, Anastasia & Lisa Cheng. 2006. (In)definiteness, polarity and the role of wh-morphology in free choice. *Journal of Semantics* 23(2). 135–183. https://doi.org/10.1093/jos/ffl001.
Gibson, Ted & Evelina Fedorenko. 2010. Weak quantitative standards in linguistics research. *Trends in Cognitive Sciences* 14(6). 233–234. https://doi.org/10.1016/j.tics.2010.03.005.
Gibson, Ted, Steven Piantadosi & Evelina Fedorenko. 2013. Quantitative methods in syntax/semantics research: A response to Sprouse and Almeida (2013). *Language and Cognitive Processes* 28(3). 229–240. https://doi.org/10.1080/01690965.2012.70.
Goodall, Grant. 2015. The D-linking effect on extraction from islands and non-islands. *Frontiers in Psychology* 5. 1493. https://doi.org/10.3389/fpsyg.2014.01493.
Hagstrom, Paul. 2006. A-not-A Questions. In Henk van Riemsdijk & Martin Everaert (eds.), *The Blackwell companion to syntax*, 173–213. London: Blackwell.
Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford: Clarendon Press.
Hegarty, Michael. 1992. Adjunct extraction without traces. In *Proceedings of the 10th West Coast Conference on Formal Linguistics*, 209–223.
Hintikka, Jaakko & Ilpo Halonen. 1995. Semantics and pragmatics for *why*-questions. *Journal of Philosophy* 92(12). 636–657. https://doi.org/10.2307/3808201.
Hofmeister, Philip & Ivan Sag. 2010. Cognitive constraints and island effects. *Language* 86(2). 366–415.
Hoji, Hajime. 1985. *Logical form constraints and configurational structures in Japanese*. University of Washington doctoral dissertation.
Hu, Fang. 2002. A prosodic analysis of wh-words in Standard Chinese. Paper presented at Speech Prosody 2002, Aix-en-Provence, France, 11–13 April.

Huang, James. 1982. *Logical relations in Chinese and the theory of grammar*. MIT dissertation.
Huang, James, Audrey Li & Yafei Li. 2009. *The syntax of Chinese*. Cambridge, UK: Cambridge University Press.
Iatridou, Sabine & Anthony Kroch. 1992. The licensing of CP-recursion and its relevance to the Germanic verb-second phenomenon. *Working Papers in Scandinavian Linguistics* 50. 498–507.
Jakubíček, Miloš, Adam Kilgarriff, Vojtěch Kovář, Pavel Rychlỳ & Vít Suchomel. 2013. The TenTen corpus family. In *7th International Corpus Linguistics Conference* , 125–127. Valladolid.
Ji, Xu & Dawei Jin. 2025. Comparing *wh-in-situ* and *wh*-fronting island effects. Manuscript, Shanghai Jiao Tong University.
Jin, Dawei. 2016. *The semantics-pragmatics interface and island constraints in Chinese*. University at Buffalo dissertation.
Jin, Dawei. 2019. A semantic account of intervention effects in Chinese *why*-questions. *Linguistics and Philosophy* 52(1). 1–43. https://doi.org/10.1007/s10988-019-09270-x.
Jin, Dawei. 2020. In situ islands are root phenomena. Manuscript, Shanghai Jiao Tong University. https://doi.org/10.13140/RG.2.2.26005.14567/1.
Jin, Dawei. 2022. The intervention effect in Suzhounese questions. *Poznań Studies in Contemporary Linguistics* 58(4). 775–822. https://doi.org/10.1515/psicl-2022-0031.
Jin, Dawei & Hanbo Yan. 2024. Intervention effects in Mandarin Chinese—An experimental study. *Journal of Semantics* 41(2). 121–148. https://doi.org/10.1093/jos/ffae006.
Jin, Dawei & Hanbo Yan. 2025. Reassessing the argument-adjunct asymmetry in wh-in-situ islands in Mandarin: An experimental investigation. *Natural Language and Linguistic Theory* 43. 1327–1363. https://doi.org/10.1007/s11049-024-09638-0.
Karttunen, Lauri. 1977. Syntax and semantics of questions. *Linguistics and Philosophy* 1(1). 3–44.
Kim, Nayeon, Ziying Li & Jiayi Lu. 2023. Island sensitivity of two different interpretations of why in Chinese. *Frontiers in Psychology* 13. 1059823. https://doi.org/10.3389/fpsyg.2022.1059823.
Kim, Shin-Shook. 2002. Focus matters: Two types of intervention effect. Talk presented at WCCFL 21, Santa Cruz.
Kim, Shin-Shook. 2006. *Intervention effects in questions*. University of Frankfurt doctoral dissertation.
Kitagawa, Yoshihisa, Dorian Roehrs & Satoshi Tomioka. 2003. Multiple *wh*-interpretations. In *Generative Grammar in a broader perspective: Proceedings of the 4th GLOW in Asia*, 209–233.
Kluender, Robert. 1998. The limits of syntax. In *On the distinction between strong and weak islands: A processing perspective*, 241–279. Amsterdam: Brill.
Ko, Heejeong. 2005. Syntax of why-in-situ: Merge into [Spec, CP]. *Natural Language & Linguistic Theory* 23(4). 867–916. https://doi.org/10.1007/s11049-004-5923-3.
Ko, Heejeong. 2006. On the structural height of reason wh-adverbials: Acquisition and consequences. In Lisa Cheng & Norbert Corver (eds.), *Wh-movement moving on*, 319–349. Cambridge, MA: MIT Press.
Kotek, Hadas. 2014. *Composing questions*. MIT dissertation.
Kotek, Hadas. 2017. Intervention effects arise from scope-taking across alternatives. *NELS* 47. 1–14.
Kratzer, Angelika & Junko Shimoyama. 2002. Indeterminate pronouns: The view from Japanese. In Chung-min Lee, Ferenc Kiefer & Manfred Krifka (eds.), *Contrastiveness in information structure, alternatives and scalar implicatures*, 123–143. Berlin: Springer.

Krifka, Manfred. 2001. Quantifying into question acts. *Natural Language Semantics* 9(1). 1–40. https://doi.org/10.1023/A:1017903702063.

Kuroda, Shigeyuki. 1965. *Generative grammatical studies in the Japanese language*. MIT dissertation.

Larson, Richard & Miyuki Sawada. 2012. Root transformations and quantificational structure. In Liliane Haegeman, Nye Richard & Aelbrecht Lobke (eds.), *Main clause phenomena: New horizons*, 47–77. Amsterdam: John Benjamins.

Law, Ann. 2001. A-not-A questions in Cantonese. *UCL Working Paper in Linguistics* 13. 295–318.

Lawler, John. 1971. Any questions? In *Chicago Linguistic Society 7*, 163–173.

Lewis, Richard & Shravan Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29. 1–45. https://doi.org/10.1207/s15516709cog0000_25.

Li, Audrey. 1992. Indefinite wh in mandarin chinese. *Journal of East Asian Linguistics* 1(2). 125–155. https://doi.org/10.1007/BF00130234.

Li, Haoze. 2011. *Focus intervention effects in Mandarin*: Chinese University of Hong Kong MA thesis.

Li, Haoze & Jess Law. 2016. Alternatives in different dimensions: a case study of focus intervention. *Linguistics and Philosophy* 39(3). 201–245. https://doi.org/10.1007/s10988-016-9189-0.

Lin, Jo-Wang. 1998. On existential polarity wh-phrases in chinese. *Journal of East Asian Linguistics* 7(3). 219–255. https://doi.org/10.1023/A:1008284513325.

Lin, Jo-Wang. 2004. Choice functions and scope of existential polarity wh-phrase in Mandarin Chinese. *Linguistics and Philosophy* 27(4). 451–491. https://doi.org/10.1023/B:LING.0000024407.76999.f7.

Linzen, Tal & Yohei Oseki. 2018. The reliability of acceptability judgments across languages. *Glossa* 3(1). 1–25. https://doi.org/10.5334/gjgl.528.

Lü, Shuxiang. 1942. *Zhongguo Wenfa Yaolue [A synopsis of Chinese grammar]*. Beijing: The Commercial Press.

Mellor, D.H. 1995. *The facts of causation*. London: Routledge.

Mellor, D.H. 2004. For facts as causes and effects. In J. Collins, N. Hall & L. Paul (eds.), *Causation and counterfactuals*, 309–324. Cambridge, MA: MIT Press.

Miyagawa, Shigeru. 2017. *Agreement beyond phi*. Cambridge: MIT Press.

Murphy, Andrew. 2017. Toward a unified theory of wh-in-situ and islands. *Journal of East Asian Linguistics* 26(1). 189–231. https://doi.org/10.1007/s10831-017-9155-z.

Niu, Xingbao. 2023. *A comparative study of the syntax and semantics of causal interrogative questions in Chinese and Korean*: Shanghai Jiao Tong University MA thesis.

Oshima, David. 2007. On factive islands: Pragmatic anomaly vs. pragmatic infelicity. In Takashi Washio, Ken Satoh, Hideaki Takeda & Akihiro Inokuchi (eds.), *New frontiers in artificial intelligence*, 147–161. Berlin: Springer. https://doi.org/10.1007/978-3-540-69902-6_14.

Pesetsky, David. 1987. Wh-in-situ: Movement and unselective binding. In Eric Reuland & Alice ter Meulen (eds.), *The representation of (in) definiteness*, 98–129. Cambridge, MA: The MIT Press.

Reinhart, Tanya. 1997. Quantifier scope: How labor is divided between QR and choice functions. *Linguistics and Philosophy* 20(4). 335–397.

Rizzi, Luigi. 1990. *Relativized minimality*. Cambridge, MA: The MIT Press.

Rizzi, Luigi. 2001. On the position "Int(errogative)" in the left periphery of the clause. *Current studies in Italian syntax* 14. 267–296.

Rooth, Mats. 1995. *Association with focus*. Amherst, MA: University of Massachusetts dissertation.

Schwarz, Bernhard & Alexandra Simolenko. 2018. On the logical makeup of how- and why-questions. In Sireemas Maspong et al. (ed.), *Proceedings of semantics and linguistic theory 28*, 533–545. Washington DC: Linguistic Society of America.

Schütze, Carson T. 2016. *The empirical base of linguistics: Grammaticality judgements and linguistic methodology*. Berlin: Language Science Press. https://doi.org/10.17169/langsci.b89.100.

Schütze, Carson T. & Jon Sprouse. 2014. Judgment data. In Robert Podesva & Devyani Sharma (eds.), *Research methods in linguistics*, 27–50. Cambridge, UK: Cambridge University Press.

Shi, Hongli. 2021. *A study on the syntax and semantics of non-interrogative question words in Chinese*. Beijing Foreign Studies University dissertation.

Shlonsky, Ur & Gabriela Soare. 2011. Where's 'why'? *Linguistic Inquiry* 42(4). 651–669. https://doi.org/10.1162/LING_a_00064.

Soare, Gabriela. 2017. Why is 'why' Unique? Talk given at the workshop at the 2017 annual SLE meeting.

Soare, Gabriela. 2021. *Why is 'why' Unique?: Its syntactic and semantic properties*. Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110675160.

Soh, Hooi Ling. 2005. Wh-in-situ in Mandarin Chinese. *Linguistic Inquiry* 36(1). 143–155. https://doi.org/10.1162/ling.2005.36.1.143.

Sprouse, Jon, Shin Fukuda, Hajime Ono & Robert Kluender. 2011. Reverse island effects and the backward search for a licensor in multiple wh-questions. *Syntax* 14(2). 179–203. https://doi.org/10.1111/j.1467-9612.2011.00153.x.

Sprouse, Jon, Carson T. Schütze & Diogo Almeida. 2013. A comparison of informal and formal acceptability judgments using a random sample from linguistic inquiry. *Lingua* 134. 219–248. https://doi.org/10.1016/j.lingua.2013.07.002.

Sprouse, Jon, Matt Wagers & Colin Phillips. 2012. A test of the relation between working-memory capacity and syntactic island effects. *Language* 88(1). 82–123. https://doi.org/10.1353/lan.2012.0004.

Stepanov, Arthur & Wei-Tien Dylan Tsai. 2008. Cartography and licensing of *wh*-adjuncts: a cross-linguistic perspective. *Natural Language & Linguistic Theory* 26(3). 589–638. https://doi.org/10.1007/s11049-008-9047-z.

Sugimura, Hirofumi. 2007. Xiandai Hanyu Yiwen Daici Ye/dou Jiegou de Yuyi Fenxi [A semantic explanation of the universal use of modern Chinese interrogative pronouns]. *Shijie Hanyu Jiaoxue* 17(3). 284–301.

Szabolcsi, Anna & Frans Zwarts. 1993. Weak islands and an algebraic semantics for scope taking. *Natural Language Semantics* 1. 235–284. https://doi.org/10.1007/BF00263545.

Takita, Kensuke & Barry Chung-Yu Yang. 2014. On multiple wh-questions with 'why' in Japanese and Chinese. In Mamoru Saito (ed.), *Japanese syntax in comparative perspective*, 206–227. New York: Oxford University Press.

Thornton, Rosalind. 2008. Why continuity. *Natural Language & Linguistic Theory* 26(1). 107–146. https://doi.org/10.1007/s11049-007-9031-z.

Tomioka, Satoshi. 2007. Pragmatics of LF intervention effects. *Journal of Pragmatics* 39(9). 1570–1590. https://doi.org/10.1016/j.pragma.2007.03.002.

Tomioka, Satoshi. 2009. Why questions, presuppositions, and intervention effects. *Journal of East Asian Linguistics* 18(4). 253–271. https://doi.org/10.1007/s10831-009-9053-0.

Tsai, Wei-Tien Dylan. 1994. On nominal islands and LF extraction in Chinese. *Natural Language & Linguistic Theory* 12(1). 121–175. https://doi.org/10.1007/BF00992747.

Tsai, Wei-Tien Dylan. 1999. The *how*s of *why* and the *why*s of *how*. *UCI Working Papers in Linguistics* 5. 155–184.

Tsai, Wei-Tien Dylan. 2008. Left periphery and how-why alternations. *Journal of East Asian Linguistics* 17(2). 83–115. https://doi.org/10.1007/s10831-008-9021-0.

Vasishth, Shravan & Richard Lewis. 2006. Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language* 82. 767–794.

Xie, Zhiguo. 2014. Focus, (non-)exhaustivity, and intervention effects in wh-in-situ argument questions. *The Linguistic Review* 30(4). 585–617. https://doi.org/10.1515/tlr-2013-0019.

Xue, Xiaoying. 2014. An investigation into the non-interrogative uses of question words in Chinese. *Journal of Hunan University of Science and Technology* 17(3). 128–132.

Yang, Barry Chung-Yu. 2009. *Intervention effects and the covert component of grammar*. National Tsinghua University dissertation.

Yang, Barry Chung-Yu. 2011. Intervention effects and wh-construals. *Journal of East Asian Linguistics* 21(1). 43–87. https://doi.org/10.1007/s10831-011-9080-5.

Yun, Jiwon. 2019. Meaning and prosody of wh-indeterminates in Korean. *Linguistic Inquiry* 50(3). 630–647. https://doi.org/10.1162/ling_a_00318.

Zifonun, Gisela, Ludger Hoffmann & Bruno Strecker. 1997. *Grammatik der deutschen Sprache*. Berlin/New York: de Gruyter.