

This is a pre print version of the following article:

Language and counterfactual reasoning in Chinese, English and Chinese L1 -English L2 reasoners / Bassetti, Bene. - In: INTERNATIONAL JOURNAL OF BILINGUALISM. - ISSN 1367-0069. - 26:1(2022), pp. 82-103. [10.1177/13670069211022850]

Terms of use:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

08/05/2026 04:03

(Article begins on next page)

Language and counterfactual reasoning in Chinese_{L1}, English_{L1} and Chinese_{L1}-English_{L2} reasoners

Bene Bassetti
University of Birmingham, UK

Abstract

Aims: No recent studies have investigated language effects on counterfactual reasoning in bilinguals. This paper investigates the impact of bilinguals' native language and language of testing on counterfactual reasoning, addressing two questions:

1. Do older Chinese reasoners, educated before English became a school subject, draw different inferences, or use different cues to draw inferences, compared with English peers and younger Chinese_{L1} reasoners? Does knowing English affect their reasoning?
2. Do Chinese reasoners draw different inferences, or use different cues, if tested in Chinese or English?

Design: Experiment 1: The explanatory variables are First Language (between-group: Chinese, English), Age Cohort (between-group: young, older), Inferential Chain Length (within-group: Short, Long). Experiment 2: The explanatory variables are Language of Testing (between-group: Chinese, English) and Inferential Chain Length (within-group: Short, Long). The outcome is the consequent probability rating. Open questions investigate cues used to draw inferences.

Analysis: The sample comprised 188 participants. Generalized linear mixed-effects models were used for quantitative data, thematic analysis for qualitative data.

Findings: Older Chinese speakers rate long-chain consequents as more probable than English peers. Chinese and English reasoners use different cues to make inferences, as do Chinese reasoners tested in Chinese_{L1} or English_{L2}.

Originality: This is the first paper to compare Chinese reasoners educated before and after English entered the school curriculum, and to investigate inferential chain length effects on Chinese counterfactual reasoning. It introduces a novel task (consequent evaluation), and adopts a mixed-method approach to investigate both the product and process of reasoning, using quantitative and qualitative data respectively.

Significance: The study provides new evidence and interpretation for the old debate about language effects on counterfactual reasoning in cognitive psychology; shows that conditional reasoning is a fruitful topic for linguistic relativity and bilingual cognition research; testifies that qualitative data allows to detect differences in thinking processes.

Keywords: counterfactual reasoning, linguistic relativity, bilingualism, Chinese, English as a Second Language

Author Note

Correspondence concerning this article should be addressed to Bene Bassetti, School of Education, University of Birmingham, Birmingham B15 2TT, UK.

Tel: (+44) (0)121 415 58221

E-mail: b.bassetti@bham.ac.uk

Acknowledgments

Pilot studies for this project were funded by a British Academy Small Grant (SG-50522). The author is grateful to: Meihui Lu for contributing to materials preparation and data collection; Alex Panicacci and Liang Qing for collecting data; Gary Feng, Lisa Gabern Liu and Dredre Gentner for sharing research materials; Caiwen Wang for sourcing translations of materials; Paul Noordhof and Nick Chater for discussing counterfactuals; Chi-He Elder and Luna Filipovic for insightful comments on a previous version of this article. Pilot results were presented at the Conference on Language, Communication and Cognition, Brighton, 2008, the International Conference on Thinking, London, 2012, and the International Symposium on Bilingualism, Alberta, 2019.

Declaration of Conflicting Interest.

The Author declares that there is no conflict of interest.

Introduction

Counterfactual reasoning, that is to say reasoning about what could have happened, plays a pervasive role in human thought, from reflecting on the consequences of past actions to evaluating scientific evidence. Some languages – such as English – have linguistic devices that mark unequivocally the counterfactual mode, other languages – such as Chinese – do not. Based on various studies of counterfactual reasoning in Chinese_{L1} and English_{L1} speakers, Bloom (1981) argued that Chinese reasoners have difficulty with counterfactual reasoning, and perform better if tested in English_{L2}, because the Chinese language does not mark counterfactuality. Following a decade of criticisms and failures to replicate, Bloom's proposal was rejected. However, work by both Bloom and his critics was marred by theoretical and methodological shortcomings (see below). The present paper then investigated whether Chinese participants with a linguistic, cultural and educational background similar to Bloom's participants would reason counterfactually differently from English native-speaking peers, and whether Chinese comprehenders reason differently when tested in English or Chinese.

Counterfactuals and counterfactual reasoning in English

In the English language, counterfactual conditionals differ from other conditionals in both form and meaning, and this difference has a psychological reality for native English (English_{L1}) speakers, as discussed below.

Looking at form, while all English conditionals use the conjunction *if* in the antecedent clause and an optional *then* in the consequent clause, counterfactuals also use tense shift, and the modal *would* in the consequent (or consequents, in the case of inferential chains). For instance, (a) is not counterfactual:

(a) *If it rained, plants grew*

but the additional layer of past in (b) marks a counterfactual (see Ippolito, 2013):

(b) *If it had rained, plants would have grown.*

Looking at meaning, English_{L1} speakers only use subjunctive conditionals if they believe that the antecedent is false (Lewis, 1973). The speaker of (a) above does not know whether it rained, the speaker of (b) believes that it did not rain. Crucially, research consistently shows that English_{L1} listeners infer the nonfactuality not only of the antecedent, but also of the consequent (see Byrne, 2016, for a review). In an early study (Fillenbaum,

1974), after hearing a counterfactual statement such as *If he had caught the plane he would have arrived on time*, almost half of American EnglishL1 participants falsely recalled the negated consequent *He did not arrive on time*. Various researchers have since confirmed that EnglishL1 comprehenders infer the falsity of both consequent and antecedent using offline (Byrne & Tasso, 1999; Thompson & Byrne, 2002) and more recently online tasks, and revealed factors that can affect inferences and wide individual variation in counterfactual implication processing (for a review, see Kulakova & Nieuwland, 2016).

Counterfactuals and counterfactual reasoning in Chinese

The Chinese language has no dedicated linguistic device (syntactic construction or lexical item) to distinguish counterfactuals from other conditionals. Conditionals are generally marked with the conjunction *ruguo* in the antecedent clause (there are other conjunctions, and *ruguo* can be omitted), and the optional conjunction *jiu* in the consequent clause. However, there is no lexical or grammatical structure that is dedicated to indicating counterfactuality, which is instead ‘marked by a combination of linguistic structures and relies on pragmatic inference’ (Jing-Schmidt, 2017, p. 32). Indeed, various linguistic features – such as tense markers *le* and *zao* – can contribute to the counterfactual interpretation of a conditional (Feng & Yi, 2006; Jiang, 2019).

Counterfactuality can be communicated in languages that have no equivalent of the subjunctive (Byrne, 2016). Indeed, although Chinese has no distinct device to mark counterfactuality, some aspects of counterfactual thinking do not differ between Chinese and American English speakers, such as counterfactual regrets (Chen et al., 2006) or age of onset of counterfactual thinking (Erbaugh, 1985). Also, counterfactual reasoning is well documented among Chinese native speakers when reasoning with *yaobushi* (‘had it not been the case that’; Hsu, 2014), a specialised marker that is exclusively used to negate ‘down-to-earth, contingent events or states’ that are known to be true, but ‘no abstract thoughts’ (Jiang, 2019, p. 284). However, the absence of a dedicated counterfactual markers may influence other aspects of counterfactual reasoning, as discussed in the wide debate about Chinese counterfactual reasoning among cognitive psychologists in the 1980s.

Bloom’s studies and its critics

Bloom (1981) tested whether the lack of overt counterfactual marking in Chinese could result in differences in counterfactual reasoning between Chinese and American English native speakers. The most convincing part of Bloom’s research investigated the inferences drawn from the so-called Bier story, a counterfactual story about a fictional 18th century philosopher called Bier. The story can be summarized in a false, and explicitly denied, antecedent, followed by an inferential chain of four consequents:

Bier did not know Chinese. If Bier had been able to read Chinese, he:

- A. would have discovered that Chinese philosophers looked at relationships between natural phenomena
- B. would have been influenced by Chinese philosophers
- C. would have created a new philosophical theory, including both individual phenomena and their relationships
- D. would have influenced Western philosophy with this new theory.

American and Chinese native speakers read the story and performed a multiple-choice comprehension task, whereby they decided which, if any, of a series of restatements of the consequents was true, and then explained their answer. Almost all (97%) American university students answered correctly, compared with 63% of Chinese ones. Among Chinese non-student adults only 46% answered correctly when tested in ChineseL1, but this raised to 86% when a subgroup was later tested in EnglishL2. Bloom concluded that Chinese speakers

reason counterfactually ‘less directly, with a greater investment of cognitive effort and hence less naturally’ than English speakers when dealing with abstract or complex contexts as in the Bier story (p. 22). This was due to language and not to Chinese speakers’ inability or unwillingness to reason counterfactually, because Chinese_{L1} comprehenders performed better when tested in English_{L2} than in Chinese_{L1}. The crucial issue of what Bloom considered a correct answer is discussed below, after a review of the studies that followed Bloom’s lead.

Bloom’s work sparked a debate that resulted in a rejection of his findings and claims (Lucy, 1992). Most researchers criticised Bloom’s methods, particularly the language and the content of the Bier story (Au, 1983; Liu, 1985), while Lardiere (1992) criticised his interpretation of his results and suggested a cultural rather than linguistic explanation, showing that reasoners from various Arabic-speaking countries refused to engage with counterfactual reasoning tasks for cultural reasons, in spite of having a counterfactual marker. Various studies then failed to replicate Bloom’s findings, however no study used the same story with similar participants, as they often tested participants with good knowledge of English_{L2}, or used a simplified version of the Bier story (see criticism in Bloom, 1984, and response in Au, 1984) or a different story (Wu, 1994). There has been almost no research on Chinese counterfactual reasoning since. An unpublished study (Yeh & Gentner, 2005) found that Chinese_{L1}, but not English_{L1}, reasoners perform better with stories about known than unknown events (e.g., *if antibiotics had never been discovered*, vs. *if Michael had gone out with his girlfriend*), meaning that Chinese reasoners rely on real-world knowledge to clarify whether a story is counterfactual, or to make inferences. Liu (2018) found that English_{L2} proficiency may correlate with speed of processing of counterfactual sentences in Chinese_{L1}. There is also indirect evidence that counterfactuality may be difficult to Chinese native speakers, as they have well-documented difficulty in learning and using English_{L2} counterfactuals (Chou, 2000; Conroy & Linda, 2013). In spite of the lack of interest among researchers, Bloom’s work is still cited (and refuted) in discussions of linguistic relativity research, whether in dedicated monographs (Deutscher, 2010; Everett, 2013) or in cognitive psychology textbooks (Friedenberg & Silverman, 2011; Galotti, 2017).

Bloom’s study is worth investigating again, in order to address some issues with his own research, as well as his followers’, as follows.

1) Correct answer. To Bloom, the only correct answer was the rejection of the consequent (‘Bier couldn’t speak Chinese and therefore hadn’t accomplished any of the things referred to’, Bloom, 1981, p. 30). However, this is not a valid inference, with a counterfactual no inference is allowed about the truth of the consequent. Indeed, consequents may even true, because the premise is not a necessary condition, and non-monotonic reasoning is allowed, meaning that it is possible to introduce an alternative antecedent, i.e. an additional premise that enables the consequent to be true regardless of the falsity of the antecedent (Byrne, 1989). For instance, a missionary may have explained Chinese philosophy to Bier. Bloom scored such answers as incorrect. A new study should avoid scoring consequent rejection as the correct answer, as the correct answer in terms of formal logic is that the truth value of the consequent cannot be inferred.

2) Bloom and his critics only focussed on participants’ rejection of the truth of consequents, with no attention to their reasoning processes. However, cross-linguistic differences, including differences between monolinguals and second language users, may appear in the process of reasoning, even when the product (the answer) is the same. For instance, Bassetti et al. (2018) found that Chinese_{L1}-English_{L2} bilinguals and English native speakers used different calendar calculation strategies, even though they gave the same answers. For this reason, it is crucial to collect qualitative data, whereby reasoners explain the reasoning that led them to choose a response.

3) Inferential chain length. Bloom argued that Chinese readers struggled with the complexity and abstractness of the Bier story. However, what Bloom called ‘complexity’ was in fact inferential chain length. Real-life (as opposed to formal logic) reasoning is often probabilistic and pragmatic (Oaksford & Chater, 2010). In probabilistic reasoning, as the inferential chain becomes longer, the consequent’s probability may become less and less related to the truth of the antecedent, so that consequents may become more probable the further down the chain they are. For instance, in the Bier story, the last consequent (Bier influencing Western philosophy with a theory that links natural phenomena) could have happened without Bier knowing Chinese (false antecedent), but the first consequent (Bier discovering that Chinese philosophers linked natural phenomena) was more reliant on the antecedent being true. In English, all consequents are marked as counterfactual by the use of tense shift and modals, and are all equally interpreted as being contrary-to-fact. In the absence of marking, it is possible that Chinese reasoners could consider each consequent’s probability, and be influenced by the consequent’s distance from the false antecedent. Since all answers other than rejections of the truth of all consequents were classified by Bloom as incorrect and not further investigated, it is impossible to know whether participants had reasoned probabilistically, accepted some consequents but not others, added an alternative antecedent, evaluated consequents as improbable rather than false, or used other strategies. A new study should then not treat all consequents in the same way, but compare performance in a short-chain and a long-chain consequent.

In conclusion, the Chinese language does not overtly distinguish counterfactuality from conditionality, and for this reason Chinese comprehenders rely on linguistic and non-linguistic cues to decide the level of factuality of a statement. Previous research that investigated differences in counterfactual reasoning between Chinese and English native speakers yielded mixed and contested but mostly null results, but it was marred by methodological issues. This study then aimed at replicating Bloom’s study with participants that were linguistically and culturally similar to his own, but addressing the issues reported above.

The present study

The present study adopted a mixed-methods approach to investigate the effects of linguistic background and language of testing on counterfactual reasoning. Experiment 1 investigated native Chinese reasoners who were comparable to Bloom’s (1981) participants in terms of linguistic and educational background, comparing them with native English peers, and with younger Chinese and English reasoners, and Experiment 2 compared Chinese reasoners tested in Chinese or English. The study was a conceptual replication of Bloom’s (1981) study of Chinese and English speakers’ counterfactual reasoning, using the Bier story previously used in this line of research, but with a different task and dependent variable to address some shortcomings of previous research as described below. First, a consequent evaluation task was created in order to measure participants’ probability rating of consequents. This is because everyday conditional reasoning is probabilistic (Evans, 2012), and probability ratings allow for more fine-grained distinctions than the binary true/false judgments of previous studies. Second, in order to test for effects of inferential chain length, a short- and a long-chain consequent were compared. Third, in order to clarify the reasoning behind participants’ responses to the counterfactual reasoning task, the study elicited qualitative data by asking participants to explain the reasons for their consequent probability ratings.

The first aim of the present study was to run a conceptual replication of Bloom’s (1981) study of Chinese and English speakers’ counterfactual reasoning, testing participants with similar background to Bloom’s and the same materials, the Bier story, but with different

task and dependent variable. To this end, Experiment 1 compared Chinese and English native speakers' reasoning about the same counterfactual story (Bloom's Bier story). After reading the story in their respective L1, participants performed a consequent evaluation task. If language affects counterfactual reasoning, as Bloom claimed, English_{L1} reasoners should rate consequents as false, inferring the falsity of the consequents from the falsity of the antecedent due to the pragmatic implicatures of the English language, and Chinese_{L1} reasoners should consider the consequents as more probable than English reasoners.

Given that previous studies could not replicate Bloom's findings with different participants, and Bloom (1984) attributed this failure to the testing of participants who knew the English language, this study investigated Chinese reasoners who were born in the People's Republic of China by 1965, and therefore had been schooled before English became a school subject. They were then compared with English native speakers of similar ages. If the older Chinese and English groups differ, this would confirm Bloom's claim of differences in reasoning between Chinese and English native speakers with participants that are comparable to his original one. To further test this, the two older groups were compared with two groups of young Chinese and English reasoners, tested in Chinese and English respectively.

A new task was introduced, so that participants would evaluate the probability of consequents, instead of evaluating its truth or falsity as in previous studies. This is because in light of the issues highlighted above, this study adopted a probabilistic approach to reasoning (Evans, 2012), assuming that natural language reasoning is not binary as in formal logic, but is based on evaluations of the probability of consequents given the antecedent, and therefore using a probabilistic approach allows for a more real-life form of reasoning than requesting a true/false response. In the Consequent Evaluation Task, participants assessed a rephrasing of the consequent by selecting one of five statements, which correspond to different levels of probability of the consequent, namely 'true', 'probable', 'undecidable', 'improbable', 'false'. This yielded an ordinal measure, with increasing levels of improbability, ranging from 'true' (the consequent is interpreted as factual, therefore as having the highest level of probability) to 'false' (the consequent is interpreted as counter-to-facts, therefore having the lowest probability level). Unlike previous studies, where the rejection of the consequent was considered the only correct answer, no answer was scored as correct or incorrect.

Inferential chain length was introduced as an explicatory variable. This is because, as discussed above, Bloom argued that Chinese speakers had difficulty with complex stories, which actually meant long inferential chains. Participants then evaluated two consequents with different positions in the inferential chain (second and fourth consequent). The prediction was that English reasoners should consider both consequents false, as both are marked with modals and tense shift. Chinese reasoners, for whom consequents are not marked for counterfactuality, may consider the short chain consequent less probable, as chances of it happening without the antecedent being true are lower, whereas the long-chain consequent would be considered more probable, as various alternative causes could lead to the truth of the more remote consequent without the truth of the antecedent.

Finally, unlike previous studies, this study investigated not only the product of the reasoning (the probability rating), but also the process of reasoning, that is to say how the inference was made. This was achieved by systematically collecting and analysing reasoners' explanations of their responses to the consequent evaluation task. This is for two reasons. First, qualitative data can explain the experimental results. Open answers may reveal whether Chinese speakers refuse to engage with the task for cultural reasons (Lardiere, 1992). For instance, they may refuse to reason within the logic of the task, perhaps rejecting the truth of the premise that in Bier's time Chinese works had not been translated, or putting forward an alternative antecedent. Second, if language indeed affects thought, this does not necessarily

mean that responses will be different, but perhaps the same response may be obtained differently. For instance, participants may rely only on linguistic cues or on other sources of information such as real-world knowledge. Qualitative data can shed light on such differences. This is particularly important when researching the effects of knowledge of more than one language, as bilinguals and L2 learners have more than one language and culture at their disposal and so a wider toolbox than monolinguals. They can reach the same conclusion as monolinguals, but do it differently, as shown for instance in the different calendar calculation strategies used by Chinese-English bilinguals and English native speakers (see e.g., Bassetti et al., 2018). Such cross-linguistic differences would be hidden if only quantitative data is collected.

Experiment 1

Method

Participants. Of the total 188 participants entered in the first analysis, 48 were eliminated prior to the main analysis for refusing to reason within the boundaries of the story (see Results). The final sample then included 140 participants, divided in four groups: 38 older Chinese native speakers, 27 older English speakers, 41 young Chinese speakers and 34 young English speakers. All Chinese participants were living in China, but the older ones had been schooled before and the younger ones after English became a school subject (year of birth: $Med_{OlderChinese} = 1956$, range: 1939-1965; $Med_{YoungChinese} = 1993$ [1990-98]). The English groups had similar ages to Chinese groups ($Med_{OlderEnglish} = 1956$ [range: 1939-68]; $Med_{YoungEnglish} = 1995$ [1983-99]). All participants had completed high-school: young participants were university students; older participants were mostly university graduates (Chinese = 79%; English = 76%). Among older Chinese participants, roughly half had studied scientific and half non-scientific subjects, whereas among the young group 80% had studied non-scientific subjects.

The two Chinese groups differed in knowledge of English. All the young Chinese had passed the TEM-4 (the English test required for university admission), and their median self-rating was ‘very proficient’ (85% were ‘rather’ or ‘very proficient’, the rest were equally distributed above or below). Half of the older participants had never studied English or self-rated as ‘very unproficient’, and 39% self-rated as ‘rather unproficient’ (the rest were ‘rather proficient’, excluding one ‘native-like’). The twenty older Chinese speakers who reported a year of onset of acquisition for English had started learning it in the 1960s (55%) or 1970s (40%; one had started earlier). In terms of other languages, many young English participants reported low levels of proficiency in French, while some older Chinese participants had studied Japanese or Russian.

Participants were recruited in suitable locations (universities, pubs) or via email using direct approach, snowballing, and personal contacts. Due to the difficulty of recruiting and testing older participants, some participants completed the questionnaire in hardcopy and others received it by email. Participation was voluntary and unpaid (some participants received up to £1 worth of gifts or charity donations).

Materials

As participants were tested in their native language, materials consisted of the English and Chinese versions of the Bier story, adapted as described below. Au’s (1983) version of the story was preferred to Bloom’s (1981) original, because the latter was written in a language suitable for Hong Kong Chinese readers of the late Seventies, which differs from

contemporary Standard Chinese in lexicon, grammar and script. In order to clarify the counterfactual nature of the Chinese *if*-clause *ruguo ta kandedong Zhongwen de hua* (literally: ‘if he can read Chinese’), the story explicitly negated the antecedent stating *Unfortunately Bier could not read Chinese*. The text was slightly adapted (see Supplementary Materials), to reflect advice from proofreaders (two Chinese applied linguists, and three English native speakers) and the work of four professional translators who translated the story from English into Chinese for this project. Two amendments are worth reporting here. First, *if* was translated as *ruguo*. Although Au had used *jiaru*, *ruguo* was used by all but one translator, and is generally used to translate counterfactuals in English language textbooks in China (Zhang, 2009). Second, *yiding* (‘certainly’) was added to the second consequent to increase the similarity of *hui* (‘can’) with the English *would* (rather than *could*) *have*, and *certainly* was added to the English version for consistency.

The English story was 172 words long, the Chinese story was 267 hanzi long, equivalent to 178 words (Sun et al., 1985). The English story is provided in the Appendix; all materials in the Supplementary Materials, OSF (<https://osf.io/jsvk5>) and iris (www.iris-database.org).

Tasks and procedure

Consequent evaluation task. The task required the evaluation of the probability of two statements: the rephrasing of a short-chain consequent (the story’s second consequent *he certainly would have been influenced by Chinese philosophers*, negatively reworded as *Bier was not influenced by Chinese philosophers*), and the rephrasing of a long-chain consequent (the fourth consequent *would have influenced Western philosophy*, positively reworded with the addition of the specific nature of the influence as *Bier led European philosophers to notice the interrelationships among natural phenomena*).

There were four more statements. Three control statements were used to ensure that participants had understood the story and the task (e.g., *Bier was a German philosopher*). Participants’ ability and willingness to reason within the boundaries of the story was tested with the statement *In the 18th century Chinese works had already been translated into European languages*, a positively-worded rephrasing of the negated premise (with the correct answer being *They had not*). Questions were arranged in four different orders.

Each statement (including control ones) was evaluated by selecting one of five options, which corresponded to true/probable/undecidable/improbable/false, but were phrased explicitly in order to avoid misunderstandings, for example *He was influenced* (= the consequent is True) and *He was not influenced* (= False). Participants were instructed to select one option *on the basis of the text* (see Evans, 2002).

Open questions. After each of the two statements in the consequent evaluation task, the reasoning behind participants’ probability evaluations was elicited with *Please explain your answer (compulsory)*, followed by a box for answering.

Procedure. Participants first completed the consequent evaluation task, which was presented as a reading comprehension task, then a short questionnaire about biographical and linguistic backgrounds including questions about education level and language learning history.

Analysis. Consequent ratings were coded in ascending order of improbability (i.e., descending order of probability), from 1 (= ‘true’, e.g., *He led them*) to 5 (= ‘false’, e.g., *He did not lead them*; 2 = probable, 3 = ‘undecidable’, 4 = ‘improbable’). The three control items and the negated premise were coded as correct or incorrect, with all probabilistic answers coded as incorrect, e.g., the only correct answer was agreeing that Bier was German.

The sample of 188 did not include participants ($n = 11$) who had answered incorrectly more than one of the three control questions, and had therefore been eliminated for failing to understand or engage with the story or the task, or for a tendency to rate as probable events that were presented as true in the story.

The influences of participants' first language, participants' age cohort, and inferential chain length on probability ratings of counterfactual statement were tested using a cumulative link mixed model (CLMM) from the *ordinal* package (Christensen, 2019) using R-3.5.1 (R Core Team, 2018) and RStudio 1.1.456 (RStudio Team, 2016). A cumulative link mixed model was used in order to include a random intercept to account for participant variation, and because the outcome variable was ordinal (probability rating with five levels). The initial model was specified using a design-driven approach. In line with the research questions, the model included the main effects and interactions between first language (Chinese, English), age cohort (young, older) and inferential chain length (short, long), and random intercepts for participants. The assumption of proportional odds was tested using a likelihood ratio test. The random structure was checked by comparing the model with and without it. Fixed factors significance was tested using the *Anova* function in the *RVAideMemoire* package (Hervé, 2015), and p values are reported in the text.

Qualitative data was coded and analysed using MAXQDA 2018 (VERBI, 2017). Due to missing answers, open question respondent numbers were: Chinese_{older} = 36; English_{older}: 24; Chinese_{young} = 41; English_{young} = 27. In a hybrid inductive-deductive approach to thematic analysis, some themes were borrowed from the counterfactual reasoning literature (e.g., 'alternative antecedents') while others emerged from the data. The thematic analysis was complemented by frequency analyses of lexical choices. Quotations of participants' explanations are presented under Results (translations by the author).

In order to compare the length in words of Chinese- and English-language answers, the number of hanzi in Chinese answers was divided by 1.5, using the established '1.5 factor' (Sun et al., 1985), which states that on average 1.5 hanzi correspond to one word in the English translation of the same text.

Results

Willingness to reason within the boundaries of the story

First, to test whether older Chinese_{L1} reasoners may be less willing than English peers and younger participants to reason within the boundaries of the story, we investigated their willingness to accept the falsity of the negated premise (answering *In the 18th century Chinese works had already been translated into European languages with They had not*). Among older Chinese_{L1} reasoners, 37% failed to accept the falsity of the negated premise, compared with 13% of older English_{L1} and 23% of young reasoners. Accuracy in the response to the negated premise (*They had not* = 1, all other answers = 0) was entered in a logit mixed-effect model that included as fixed effects L1 and age cohort and their interaction, and random intercepts for participants. The interaction ($\chi^2 = 20.59, p < 0.001$) revealed that older Chinese_{L1} reasoners had lower predicted odds of accepting the falsity of the negated premise ($b = -2.38, SE = 0.56, z = -4.27, p < 0.001$).

Next, we tested whether failure to accept the falsity of the premise led participants to consider the consequents more probable. About half of the 22 older Chinese participants who had rejected the falsity of the negated premise had also rejected the falsity of the consequents, rating them as true, probable or undecidable. Consequent probability ratings were entered in a model that included as fixed effects L1, age cohort, inferential chain length, accuracy in the negated premise question and their interactions, and random intercepts for participants. There

was a main effect of negated premise accuracy ($\chi^2 = 29.59, p < 0.001$), and crucially the four-way interaction ($\chi^2 = 4.00, p = 0.046$) revealed that older ChineseL1 reasoners who had rejected the negated premise falsity – those who believed that translations may have existed – had higher predicted odds of rating the long consequent as probable ($b = -4.87, SE = 2.36, z = -2.06, p = 0.039$).

To investigate whether knowing English may increase older Chinese speakers' willingness to reason within the boundaries of the story, the older Chinese group's responses to the negated premise were entered in a linear regression model with self-rated English proficiency as an ordinal predictor. Higher English proficiency was associated with a higher likelihood of accepting the falsity of the negated premise ($\chi^2 = 11.74, p = 0.038$). Finally, the long-chain consequent probability evaluations of older Chinese who had not accepted the falsity of the negated premise were entered in a linear regression model with self-rated English proficiency as an ordinal predictor. Higher English proficiency was associated with improbability ratings of the long-chain consequent among this group ($\chi^2 = 94.63, p < 0.001$).

Participants who had rejected the falsity of the negated condition were then eliminated from further analysis, leaving the final sample of 140 analysed below.

Consequent ratings

Figure 1 shows Chinese and English reasoners' consequent ratings by age cohort and inferential chain length. The 'false' rating (i.e., inferencing the falsity of the consequent) was the median across groups and conditions, but descriptively it was more frequent with the short- than the long-chain consequent (74% of all answers vs 62%), and among English than Chinese speakers (73% of English group's answer, 64% of the Chinese group's answers). Only a small minority of participants chose the response that was correct in terms of formal logic, i.e., 'undecidable', but these responses were roughly four times more frequent among Chinese than English participants (14% vs 3% of responses).

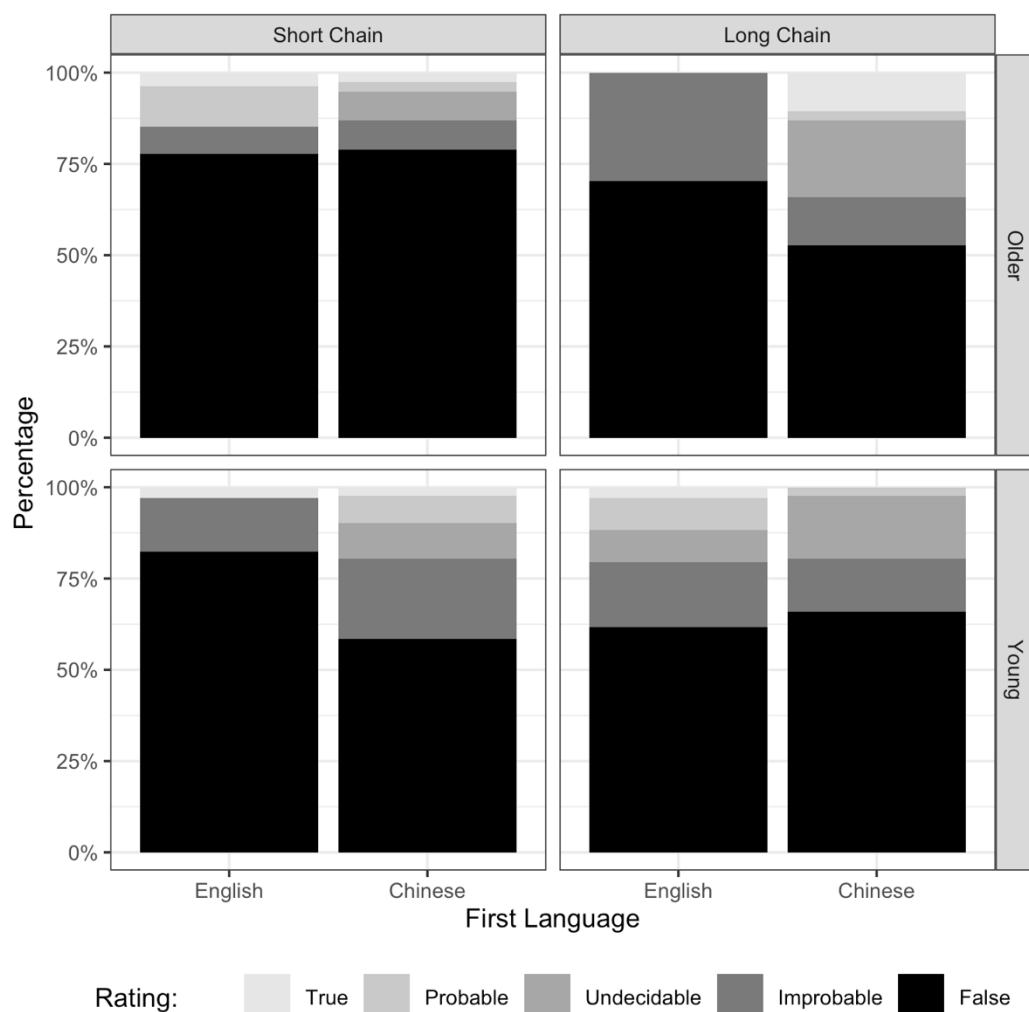


Figure 1. Percentage of probability ratings in the Consequent Evaluation Task by first language (Chinese, English), age cohort (young, older) and inferential chain length (short, long).

The final model (Table 1) included as fixed effects first language, age cohort, inferential chain length and their interactions, and random intercepts for participants. The Anova test showed a three-way interaction of L1, age cohort and inferential chain length ($\chi^2 = 7.73, p = 0.005$). The main effect of inferential chain length ($\chi^2 = 4.69, p = 0.030$) was justified by the three-way interaction, and there was no main effect of first language ($\chi^2 = 3.30, p = 0.069$). The model then shows that older Chinese native speakers had higher predicted odds of rating the long-chain consequent as probable.

Fixed effects	Log odds	SE	z	p
L1 (Chinese)	-1.41	0.65	-2.17	0.030 *
Age cohort (older)	-0.57	0.76	-0.75	0.452
Chain length (long)	-1.36	0.61	-2.20	0.028*
L1 (Chinese): Age cohort (older):	1.57	0.97	1.62	0.106
L1 (Chinese): Chain length (long)	1.69	0.78	2.17	0.030*
Age cohort (older):Chain length (long)	1.38	0.92	-1.50	0.133
L1 (Chinese): Age cohort (older): Chain length (long)	-3.24	1.19	-2.72	0.007**

Table 1. Fixed effects from the model fitted to participants' consequent probability ratings using the formula $\text{Response} \sim \text{L1} * \text{Age cohort} * \text{Inferential chain length} + (1 | \text{Participant})$

Note 1. As the baseline response was 'True', a positive beta value reflects the log odds of higher improbability (or lower probability)

Note 2. Significance level: * $p < .05$, ** $p < .01$

Qualitative data. In explaining the reasons of their inferences, Chinese_{L1} reasoners produced more complex answers, drawing from a wider variety of cues, than English_{L1} reasoners.

1) *Answer length.* Chinese_{L1} participants produced longer answers than English_{L1} participants (Figure 2a).

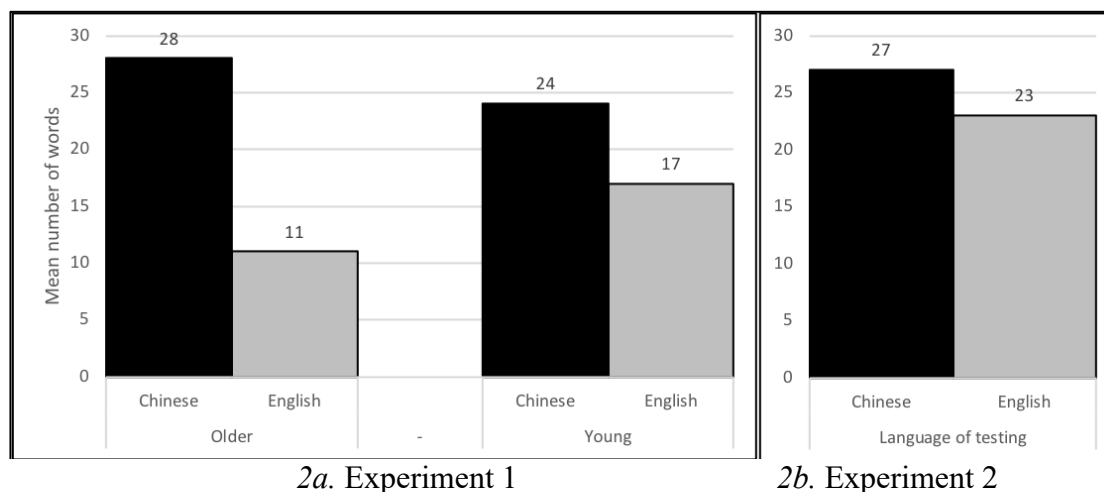


Figure 2. Mean answer length by group.

2a: Experiment 1: older Chinese, older English, young Chinese, young English.

2b: Experiment 2: Chinese-tested Chinese, English-tested Chinese.

2) *Falsity of antecedent and consequents.* Across groups, the most frequently mentioned reason for consequent probability ratings was the falsity of the antecedent, often accompanied by the falsity of the negated condition ('Bier did not know Chinese, Chinese works had not been translated', ChMa03). However, the falsity of consequents was mentioned by English_{L1} respondents much more often than by Chinese_{L1} respondents, particularly with long-chain consequents (Figure 3a). When discussing short-chain consequents, participants negated the truth of the first consequent ('[Bier] was not aware of their [Chinese philosophers'] focus on interrelationships', EnYo40); when discussing the long-chain consequent, they mostly negated the second consequent ('He was not influenced himself', EnMa16).

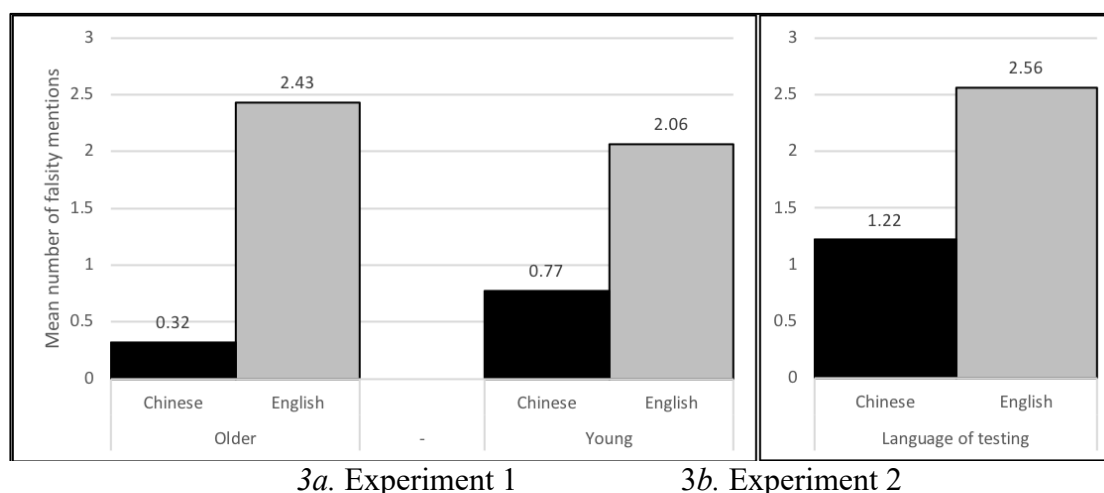


Figure 3. Mean number of mentions of long-chain consequent falsity by group.
 3a: Experiment 1: older Chinese, older English, young Chinese, young English.
 3b: Experiment 2: Chinese-tested Chinese, English-tested Chinese.

3) *Different approaches to the task.* As detailed below, older Chinese respondents were the most likely to reason outside the logical scope of the reasoning task (alternative antecedents, other linguistic and non-linguistic cues), young Chinese respondents approached the task as a test of logical reasoning, and English_{L1} respondents produced short and simple explanations.

3a) *Alternative antecedents* – alternative conditions that could have enabled the consequent although the antecedent was false – were absent in the English groups, but 9% of both Chinese groups produced at least one, usually positing that Bier could have heard about Chinese philosophy through oral transmission (‘through communication with other scholars who had read Chinese documents’, ChYo16). Occasionally, answers were elaborate: ‘One day, Bier saw a book written in Chinese ... which had text and pictures. Looking at the pictures, he had a feeling that the book dealt with the relationship between natural phenomena. He asked someone who knew Chinese to tell him what the book was about, and he had this sudden revelation, that phenomena were related’ (ChMa08).

3b) *Other linguistic and non-linguistic cues.* Some older Chinese reasoners relied on a variety of linguistic and non-linguistic cues: expectations about how philosophy works (e.g., if he did not notice this he was not a real philosopher, $n = 2$); expectations of an injunction for a story to be relevant (e.g., why would this story talk about Bier if he had not achieved anything, $n = 2$); linguistic cues (e.g., the word *kexi* ‘unfortunately’ shows that Bier’s achievements were false, $n = 2$).

3c) *Young Chinese reasoners: A logical reasoning test.* Many young Chinese respondents approached the task as a logical reasoning test. Twenty-four percent stated that *ruguo* indicates counter-to-facts events (‘*ruguo* in the text means it did not happen’), and 20% used at least one term related to logic (‘inferring’, ‘logic’), unlike all other groups.

3c) *English_{L1} reasoners: Stating the obvious.* English_{L1} participants produced simple answers, as if the answer was obvious (‘he could not speak Chinese!’, EnMa01). A quarter answered with a short statement in the subjunctive mood (‘Would have happened if he had been influenced’, EnYo33). Only one considered that the antecedent was not a necessary condition for the consequents, a point made by eight young Chinese_{L1} respondents. A few refused to infer beyond what the text said (i.e., to infer the falsity of the consequent), but they did not explain why.

4) *Lexical choices* (Figure 4a) revealed more linguistic markers of causality among EnglishL1 reasoners and more markers of degree of probability among ChineseL1 reasoners.

4a) *Causality markers* (causally linking consequents to the antecedent) were more common among EnglishL1 reasoners, who produced eight different causal conjunctions (*as, because, so*), whereas Chinese speakers generally used only the conjunction *yinwei* ('because'). For example, 'He had no access to Chinese works as he could not speak Chinese... so it is unlikely that he was influenced...' (EnMa16).

4b) *Degree of probability*. ChineseL1 respondents were much more likely to evaluate the level of probability of consequents, and produced 13 different linguistic markers to qualify low levels of probability (e.g., *kenengxing bijiao xiao*, 'rather small probability'). Among English reasoners, probability markers were absent, apart from remarking the impossibility of consequents ('*It will be impossible for Bier...*', EnYo30).

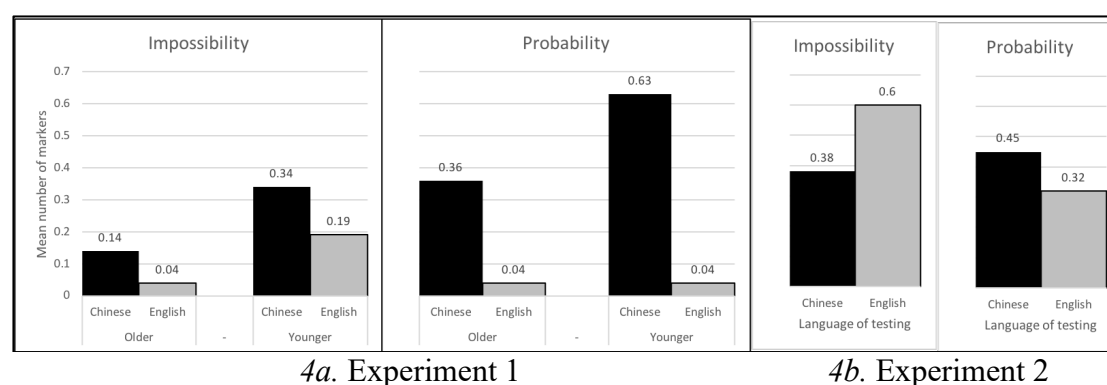


Figure 4. Mean number of impossibility markers and probability markers by group. 4a. Experiment 1: older Chinese, older English, young Chinese, young English. 4b. Experiment 2: Chinese-tested Chinese, English-tested Chinese.

Discussion

Among older Chinese participants, more than a third refused to engage with the task, as they failed to accept the negated premise that there were no western language translations of Chinese texts at the time. This supports a cultural rather than linguistic explanation, in line with Lardiere's (192) suggestions. Indeed, many of these participants also considered the long-chain consequent possible, showing that perhaps some of participants in previous studies who apparently did not reason counterfactually may have refused to engage with the task rather than having difficulties with counterfactual reasoning. However, the majority of older Chinese participants accepted the falsity of the premise, showing that a culture-induced refusal to engage with similar tasks could explain only a small part of the differences in counterfactual reasoning reported in Bloom.

Across all groups, most participants rated consequents as false in the consequent evaluation task, and explained their choice in open questions with reference to the falsity of the antecedent. This extends to native Chinese speakers the finding that English native speakers generally infer the falsity of consequents (Byrne & Tasso, 1999; Thompson & Byrne, 2002). However, consequent falsity appears to be more obvious to English than Chinese reasoners, because in open answers each English participant spontaneously mentioned it twice on average, while Chinese participants only rarely did so, particularly the older ones.

The few who concluded that the truth value of the consequent of a counterfactual cannot be inferred, rating consequents as 'undecidable', were almost exclusively Chinese.

Perhaps this inference may be more available to ChineseL1 reasoners than it is to EnglishL1 reasoners because the Chinese language lacks a dedicated counterfactual marker and the pragmatic implicatures of the English language. Interestingly, 9% of Chinese respondents (both older and young) produced an alternative antecedent, confirming that Chinese reasoners are more likely to reason beyond the straightforward causal relationship between the falsity of antecedent and consequent.

Finally, quite a few reasoners across groups rated consequents as improbable rather than false, meaning that not all English speakers denied the truth of consequents. This shows that, if the task does not force reasoners to select 'false' by requiring a binary response, they may prefer to rate events in terms of probability level.

The statistical analysis showed that the older Chinese participants were likely to rate the long-chain consequent as more probable than both older English and younger Chinese reasoners. The most likely explanation is that there are cultural and educational differences between older Chinese participants on the one hand, and English native and young Chinese participants on the other. Young Chinese participants had been studying English as a school subject for years, including subjunctive conditionals, and had been exposed to a more westernised education and testing system. Other explanations are less likely. This cannot be attributed to older Chinese participants' inability to reason counterfactually, because they did not differ from other groups with short-chain consequents; it cannot be attributed to linguistic differences between the Chinese and English languages, because young and older Chinese participants behaved differently; it cannot be due to differences in intelligence, because all participants had at least met the entry requirements for university education; finally, it cannot be due to effects of aging on reasoning, because older English and Chinese participants behaved differently.

The largest differences between Chinese and English reasoners were however not the actual inferences, but the process of making inferences, as revealed by open answers. English native speakers mostly thought that denial of the consequent naturally follows from the subjunctive mood, as they gave short and simple responses, and often produced linguistic markers of causality. This confirms that the subjunctive mood throughout the story indicates to them that all events are counter-to-fact, and that false consequents follow from false antecedents and from each other. This is not an obvious inference to Chinese speakers, and indeed a quarter of young Chinese participants felt the need to clarify that *ruguo* indicated counter-to-fact events (in this story, as in general it indicates a conditional). Chinese speakers produced longer answers because, in the absence of a dedicated counterfactual marker, they considered more, and more varied, cues. It is unclear whether this may at least partly be due to a cultural preference for more complex answers, not limited to counterfactual reasoning. They also produced a variety of probability level markers, reflecting that the Chinese language has a rich vocabulary for this (Feng & Yi, 2006), and possibly showing that Chinese reasoners consider more fine-tuned differences in probability levels.

Looking at the effects of English language knowledge on older Chinese reasoners, first of all, knowledge of English correlated with willingness to engage with the task. Second, among those who did not engage with the task knowledge of English correlated with lower probability rating of long-chain consequents. It is possible that studying English in this age cohort may be related to openness to western culture in general, or a western-style approach to counterfactual reasoning in particular.

Experiment 2

Experiment 1 found some differences in counterfactual reasoning between Chinese and English native speakers tested in their respective native language. Experiment 2 then investigated whether such differences may be due to the language of the story, by testing

whether Chinese reasoners with knowledge of EnglishL2 would behave differently if tested in Chinese or in English in the counterfactual reasoning task used in Experiment 1.

Method

Sixty Chinese undergraduate students were tested in either Chinese or English ($n = 30$ each). Story and task were the same as in Experiment 1. Participants also read a filler story and performed a consequent rating task in their other language (English for those who read the Bier story in Chinese, and vice versa). The filler story (from Yeh & Gentner, 2005) was about a fictional Eastern tribe, and it was prefactual (Byrne & Egan, 2004), with conditionals referring to events that might happen in the future. Participants were tested by their English language teacher in their classroom, using four versions with different order of questions. For English-language materials, they received a bilingual word list and a Chinese translation of questions. Participants were tested in Chinese first, then the paper was removed and they answered the English task and background questionnaire.

Results

Consequent ratings

As shown in Figure 5, the median consequent probability rating was ‘false’ across groups and conditions. The final model included as fixed effects language of testing, inferential chain length and their interaction, and random intercepts for participants. There were no main effects or interactions. There were also no correlations between consequent ratings and measures of English proficiency (TEM4 score, high-school final English mark) or academic achievement (high-school final mark).

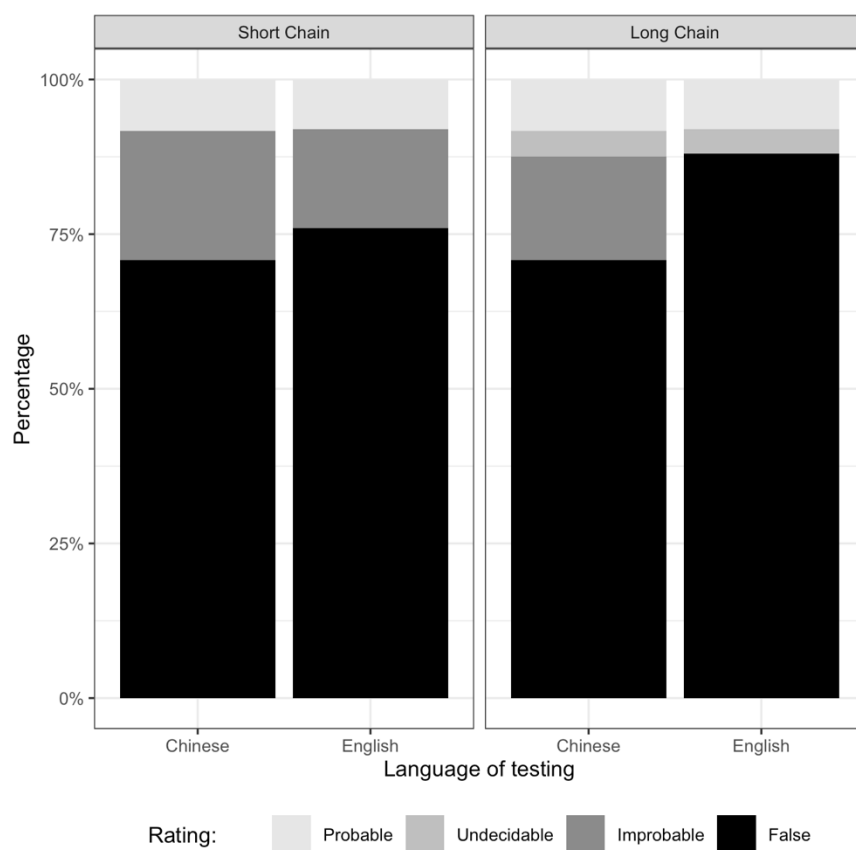


Figure 5. Percentage of probability ratings in the Consequent Evaluation Task by language of testing (Chinese, English) and inferential chain length (short, long).

Qualitative data. The qualitative analysis revealed some differences between the two groups.

1) *Answer length.* The two groups' answers were of similar length (Figure 2b). This was similar to Chinese participants in Experiment 1.

2) *Falsity of antecedent and consequents.* Participants tested in English asserted the falsity of long-chain consequents about twice as much as those tested in Chinese (Figure 3b; this was similar to English participants in Experiment 1). Unlike long-chain consequents, antecedents were considered false equally across groups.

2) *Mentioning (but not using) the subjunctive mode.* Among those tested in English, 16% explicitly mentioned the term *subjunctive* (or *xuni*), which for all but one meant that events did not happen, e.g. 'The passage use subjunctive mood when describing Chinese philosophy's influence on Bier. So he did not directly influenced by Chinese philosophers' (BiEn19); 'The text used the *if* subjunctive mood, showing that it is inconsistent with facts, so [Bier] did not make them notice' (BiEn28). Chinese-tested respondents did not mention the term *xuni* (with one exception). Unlike the English participants in Experiment 1, only one produced an answer in the subjunctive mood.

3) *The meaning of if/ruguo.* A third of Chinese-tested respondents, and 12% of English-tested ones, explained that *if/ruguo* means counter-to-fact, e.g., 'The text uses many *ruguo*, if Bier had understood Chinese, then he would have developed a new theory ..., showing that he did not notice the relationship between natural phenomena, and he could not make others notice it' (BiCh22).

4) *Reasoning outside the logical scope of the task.* Alternative antecedents were mentioned by 14% of participants, regardless of language of testing (Chinese: 13%; English: 16%). Reliance on other linguistic or non-linguistic cues was minimal, as only three respondents used real-world knowledge, and three relied on the linguistic cue *unfortunately* ('the 'unfortunately' in the third line tells us that Bier, like the other philosophers, did not notice [the interrelationships], BiCh17). Reasoning outside the logical scope of the task was far more common among those who rated counterfactuals as probable or improbable (57%), compared with the majority who rated them as false.

5) *Lexical choices.* English-tested respondents produced twice as many statements of direct causality (e.g., *because/yinwei*) and linguistic markers of impossibility (e.g., *impossible/wufa*) as Chinese-tested peers, and slightly lower numbers of linguistic markers of probability (e.g., *maybe/keneng*; see Figure 4b).

Discussion

Language of testing did not affect inferencing, as the vast majority of ChineseL1-EnglishL2 reasoners rated the consequents as false. It appears that the overt marking of counterfactuality in the English text has no effects, contrary to Bloom's (1981) finding forty years ago that Chinese reasoners were more likely to assert the falsity of the consequent if tested in English than in Chinese. Indeed, with long-chain consequents in particular, Experiment 2 participants rated consequents as false more often than any group in Experiment 1, including English native speakers. Only a tiny minority answered 'undecidable', and few used the range of probability levels offered in the task, meaning that for this age cohort, tested within a university environment, the expected answer is the consequent falsity.

This may be a washback effect of English language teaching and testing, as participants had studied English as a compulsory school subject, were taught that English subjunctive conditionals mean that the consequent is false, and had to master English conditionals to pass

compulsory language exams. Indeed, both textbooks and language tests train Chinese students to consider English subjunctives as counter-to-fact statements. The past counterfactual structure features highly in textbooks that prepare Chinese students for important English language tests such as TEM-4 (Shi, 2006). These textbooks explain that with *if+subjunctive* both the *if*-clause and the main clause are about contrary-to-facts events, sometimes even adding causality prepositions. For instance, Zhang (2009) glosses *If you had come here earlier, we would have finished the work now* with ‘in reality you did not come early and the work is not finished’ [translation by the author]. Chinese students may then assume that English counterfactual stories imply the falsity of the consequent, and use the same approach with both English and Chinese materials when tested within an English-language context, as in this study. This effect may have been stronger in Experiment 2 than in the previous one because participants were tested during an English language session with their English language teacher.

Although the two groups gave very similar consequent probability ratings, qualitative data revealed some differences in their inferencing processes. Table 2 shows similarities and differences between English-tested Chinese participants in Experiment 2 on the one hand, Chinese-tested Chinese participants (across experiments) and EnglishL1 participants in Experiment 1. English-tested Chinese participants were generally in-between, as they sometimes behaved like Chinese-tested Chinese people – producing long answers, explaining that *if/ruguo* marks counter-to-fact events, and producing alternative antecedents, which no English people did – and sometimes behaving like English natives – producing various denials of the truth of consequents, many causality markers, and more markers of impossibility than of probability. However, they also displayed a distinctive behaviour, not found either in English participants or in Chinese-tested Chinese participants, as they explicitly reported using the subjunctive as a clue, by mentioning the English term *subjunctive* or the Chinese equivalent *xuni*. These bilinguals were then using the tools provided by their second language, and in doing so displayed a peculiar behaviour. Evidence of behaviours peculiar to bilinguals is far less common than evidence of in-between behaviours, but it has both been theorised (e.g., Bassetti & Cook, 2011) and demonstrated empirically (e.g., Park & Ziegler, 2014).

Item	EN-tested Chinese NSs		Group			
	vs.	vs.	CH-tested		EN-tested Chinese NSs	English NSs
	CH-tested Chinese NSs	English NSs	Chinese NSs (Exp 1)	Chinese NSs (Exp 2)		
Word length	+	-	24	27	23	17
Denials of consequent truth	-	+	0.77	1.22	2.56	2.06
Causality markers	-	+	0.32	0.67	1.20	1.00
Impossibility vs. probability markers	-	+	0.34 / 0.63	0.38 / 0.45	0.60 / 0.32	0.19 / 0.04
Alternative antecedents	+	-	9%	13%	16%	0%
Explaining <i>if</i> /subjunctive as counter-to-facts	+	-	24%	33%	28%	0%
Mentions of subjunctive	-	-	0%	0%	16%	0%
Answering with a subjunctive	-	-	NA	NA	4%	25%

Figure 6. Left-hand column: Similarities (plus sign) and differences (minus sign) between English-tested Chinese native speakers (Experiment 2), Chinese-tested Chinese native speakers (Experiment 1 young group, Experiment 2), and English native speakers (Experiment 1 young group). Right-hand column: Descriptive statistics by group: Chinese-tested Chinese native speakers, English-tested Chinese L1 (bold and greyed), and EnglishL1 reasoners.

Note. ‘Answering with a subjunctive’ is not possible in Chinese (NA).

General discussion

The study revealed both similarities and differences between native English and Chinese speakers – particularly older ones – and between Chinese reasoners tested in Chinese or English. To summarise, while most participants across groups inferred the falsity of the consequent from the falsity of the antecedent, the older Chinese reasoners overall rated the long-chain consequent as more probable than English reasoners or younger Chinese reasoners did. The older Chinese were also less willing to reason within the logical boundaries of the counterfactual reasoning task, as a third of them doubted the falsity of the negated premise, and this behaviour was statistically more frequent among those with no or only minimal knowledge of English. In general, these results support both Bloom's view that Chinese speakers with knowledge of English are more likely to reject the truth of consequent – a linguistic explanation – and Lardiere's (1992) cultural explanation of a refusal to engage with the task. Whatever differences Bloom may have tapped into – linguistic and/or cultural – do not appear to exist anymore, as all Chinese students attend a more westernised educational system, where it is compulsory to study the English subjunctive, which textbooks and exams present as

marking counter-to-fact statements in both antecedents and consequents. Indeed, there were no differences in answers between Chinese reasoners tested in Chinese or in English, showing that Chinese university students tested in an academic setting simply reject the truth of all consequents, regardless of the presence or absence of overt counterfactual marking.

At the same time, the Chinese participants did not reason in the same way as English peers. Qualitative data shows that overall English native speakers naturally inferred the falsity of the consequent without much reflection and with frequent mentions of causal links. Instead, Chinese speakers were more likely to refuse to make inferences (by choosing 'undecidable'), to propose an alternative antecedent (almost 10% of reasoners in Experiment 1), to rely on a variety of linguistic and non-linguistic cues, to indicate subtle differences in levels of probability. They also often felt the need to state explicitly that *ruguo* indicates counter-to-fact statements (in the context of that story). This shows that the counterfactuality of *ruguo* has to be established and stated, unlike the counterfactuality of *if*+subjunctive for English speakers. It looks like, in the absence of overt marking, counterfactuality in Chinese is identified using a variety of cues, as argued by Jing-Schmidt (2017), but interestingly this absence of overt marking also results in more complex reasoning and more nuanced answers.

Finally, looking at implications for bilingual cognition research, qualitative data (Fig. 6) shows that the reasoning processes of Chinese speakers who were tested in English were in-between those of English speakers and of Chinese speakers tested in Chinese. This confirmed the convergence typically found in bilinguals, but there was also evidence of a peculiar approach not found in English reasoners or Chinese-tested Chinese reasoners. This confirms the bilinguals' ability to use creatively the tools provided by both their languages.

Conclusions

This study makes at least two contributions to research on linguistic relativity and on bilingual cognition. First, the study shows that conditional reasoning is a promising research topic. From its early days, linguistic relativity research has investigated the effects of language on thought about continua that are carved up differently by different languages, such as the colour spectrum which is divided in different colour categories across languages. Hypotheticality – the probability of realisation of the events presented in a conditional – is also a continuum, which different languages cut up into different categories (Comrie, 1986).

Although both hypotheticality and colour are continua, language effects may be more evident in conditional reasoning than in colour perception, because hypotheticality is abstract, and therefore more likely to be affected by language, compared with more basic processes such as colour perception. Future research could use more updated materials and on-line tasks, use non-linguistic materials and tasks (Lucy, 1992), include measures of relevant individual differences (IQ, working memory), and attempt to disentangle effects of language and of culture, if this is indeed possible. Yet, studies such as the present one present a promising avenue for research.

Second, the study contributes to discussions of research methodology, by arguing that research on language and cognition in general – and on bilingual cognition in particular – should complement quantitative data with qualitative data. In this study, qualitative analysis revealed subtle cross-linguistic differences in the process of reasoning, even when there were no quantitative differences in inferences. Reasoning research should then investigate not only the product (the response), but also the process of reasoning. Asking participants for their introspection about their reasoning processes is a promising approach. This is particularly important in the case of bilinguals, who have access to a repertoire of more than one language and culture, and may reach the same conclusion as monolinguals but in different ways.

Appendix.

The English version of the Bier story.

Bier was an Eighteenth century German philosopher who wanted very much to investigate the natural laws of the universe. At that time, European philosophers who investigated natural phenomena often only investigated individual natural phenomena and did not pay attention to the interrelationships between them. At the time there was already some contact between China and Europe, and Chinese works could be found in Europe, but none had been translated. Unfortunately Bier could not read Chinese. If he had been able to read Chinese, he would have discovered that when Chinese philosophers investigated natural phenomena, they focused on the interrelationships between phenomena. If Bier had read Chinese philosophy, he certainly would have been influenced by Chinese philosophers, have synthesised Chinese and Western philosophies, and have created a new philosophical theory which not only explained natural phenomena as individual entities, but which also made clear their interrelationships. This theory would not only have overcome a weakness in the Western philosophy of that time, but also would have influenced Western philosophy, bringing it closer to science.

References

- Au, T. K.-F. (1983). Chinese and English counterfactuals: The Sapir-Whorf hypothesis revisited. *Cognition*, 15, 155-187. [https://doi.org/10.1016/0010-0277\(83\)90038-0](https://doi.org/10.1016/0010-0277(83)90038-0)
- Au, T. K.-F. (1984). Counterfactuals: In reply to Alfred Bloom. *Cognition*, 17(3), 289-302. [https://doi.org/10.1016/0010-0277\(84\)90012-X](https://doi.org/10.1016/0010-0277(84)90012-X)
- Bassetti, B., et al. (2018). The linguistic transparency of first language calendar terms affects calendar calculations in a second language. *Acta Psychologica*, 186, 81-89. <https://doi.org/10.1016/j.actpsy.2018.04.006>
- Bloom, A. (1981). *The linguistic shaping of thought: A study in the impact of language on thinking in China and the West*. Lawrence Erlbaum.
- Bloom, A. (1984). Caution--the words you use may affect what you say: a response to Au. *Cognition*, 17(3), 275-287. [https://doi.org/10.1016/0010-0277\(84\)90011-8](https://doi.org/10.1016/0010-0277(84)90011-8)
- Byrne, R. M. (1989). Suppressing valid inferences with conditionals. *Cognition*, 31(1), 61-83. [https://doi.org/10.1016/0010-0277\(89\)90018-8](https://doi.org/10.1016/0010-0277(89)90018-8)

- Byrne, R. M. (2016). Counterfactual thought. *Annual review of psychology*, 67, 135-157. <https://doi.org/10.1146/annurev-psych-122414-033249>
- Byrne, R. M., & Egan, S. M. (2004). Counterfactual and prefactual conditionals. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 58(2), 113. <https://doi.org/10.1037/h0085791>
- Byrne, R. M., & Tasso, A. (1999). Deductive reasoning with factual, possible, and counterfactual conditionals. *Memory & cognition*, 27(4), 726-740. <https://doi.org/10.3758/bf03211565>
- Chen, J., et al. (2006). Culture and counterfactuals: On the importance of life domains. *Journal of Cross-Cultural Psychology*, 37(1), 75-84. <https://doi.org/10.1177/0022022105282296>
- Chou, C.-L. (2000). Chinese speakers' acquisition of English conditionals: Acquisition order and L1 transfer effects. *Second Language Studies*, 19(1), 57-98.
- Christensen, R. H. B. (2019). *ordinal: Regression models for ordinal data*. In <http://www.cran.r-project.org/package=ordinal>
- Comrie, B. (1986). Conditionals: A typology. In E. C. Traugott (Ed.), *On conditionals*. CUP.
- Conroy, M. A., & Linda, C. (2013). Language choices in L2 English sentence production: Why speakers could have used modal perfect but didn't. *International Review of Applied Linguistics in Language Teaching*, 51(4), 379-406. <https://doi.org/10.1515/iral-2013-0016>
- Deutscher, G. (2010). *Through the language glass: Why the world looks different in other languages*. Metropolitan Books.
- Erbaugh, M. S. (1985). The acquisition of Mandarin. In D. I. Slobin (Ed.), *The Crosslinguistic Study of Language Acquisition* (Vol. 3, pp. 373-455). Erlbaum.
- Evans, J. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin*, 128(6), 978. <https://doi.org/10.1037/0033-2909.128.6.978>
- Evans, J. (2012). Questions and challenges for the new psychology of reasoning. *Thinking & Reasoning*, 18(1), 5-31. <https://doi.org/10.1080/13546783.2011.637674>
- Everett, C. (2013). *Linguistic relativity: Evidence from across languages and cognitive domains*. de Gruyter Mouton.
- Feng, G., & Yi, L. (2006). What if Chinese had linguistic markers for counterfactual conditionals? Language and thought revisited. *The 28th Annual Conference of the Cognitive Science Society*, 1281-1286. <https://doi.org/10.1.1.497.2311>
- Fillenbaum, S. (1974). Information amplified: Memory for counterfactual conditionals. *Journal of Experimental Psychology*, 102(1), 44. <https://doi.org/10.1037/h0035693>
- Friedenberg, J., & Silverman, G. (2011). *Cognitive science: An introduction to the study of mind*. Sage.
- Galotti, K. M. (2017). *Cognitive psychology in and out of the laboratory*. Sage.
- Hervé, M. (2015). *RVAideMemoire: diverse basic statistical and graphical functions*. In <https://CRAN.R-project.org/package=RVAideMemoire>
- Hsu, C.-F. (2014). Semantic-based mental representation of Chinese counterfactuals: Evidence from a psycholinguistic study of Yaobushi. *Language and Linguistics*, 15(3), 391-410. <https://doi.org/10.1177/1606822X14520664>
- Ippolito, M. (2013). *Subjunctive conditionals: A linguistic analysis*. MIT Press.
- Jiang, Y. (2019). Chinese and counterfactual reasoning. In C.-R. Huang, Z. Jing-Schmidt, & B. Meisterernst (Eds.), *The Routledge Handbook of Chinese Applied Linguistics* (pp. 276-293). Routledge.
- Jing-Schmidt, Z. (2017). What are they good for? A constructionist account of counterfactuals in ordinary Chinese. *Journal of Pragmatics*, 113, 30-52. <https://doi.org/10.1016/j.pragma.2017.03.004>

- Kulakova, E., & Nieuwland, M. S. (2016). Understanding counterfactuality: A review of experimental evidence for the dual meaning of counterfactuals. *Language and Linguistics Compass*, 10(2), 49-65. <https://doi.org/10.1111/lnc3.12175>
- Lardiere, D. (1992). On the linguistic shaping of thought: Another response to Alfred Bloom. *Language in Society*, 21(2), 231-251. <https://doi.org/10.1017/S004740450001527X>
- Lewis, D. (1973). *Counterfactuals*. Blackwell.
- Liu, L. G. (1985). Reasoning counterfactually in Chinese: Are there any obstacles? *Cognition*, 21(3), 239-270. [https://doi.org/10.1016/0010-0277\(85\)90026-5](https://doi.org/10.1016/0010-0277(85)90026-5)
- Liu, Y.-T. (2018). Linguistic relativity in L2 acquisition: Chinese-English bilinguals' reading of Chinese counterfactual statements. *Language and Linguistics*, 19(1), 117-155. <https://doi.org/10.1075/lali.00005.liu>
- Lucy, J. A. (1992). *Language diversity and thought: A reformulation of the linguistic relativity hypothesis*. CUP.
- Oaksford, M., & Chater, N. (Eds.). (2010). *Cognition and conditionals: Probability and logic in human thinking*. OUP.
- Park, H., & Ziegler, N. (2014). Cognitive shift in the bilingual mind: Spatial concepts in Korean-English bilinguals. *Bilingualism: Language and Cognition*, 17, 410-430. <https://doi.org/10.1017/S1366728913000400>
- R Core Team. (2018). *R: A language and environment for statistical computing*. In R Foundation for Statistical Computing. <https://www.R-project.org>
- RStudio Team. (2016). *RStudio: Integrated development for R*. Boston, MA. <http://www.rstudio.com>
- Shi, J. (2006). *Siji yufa yu cihui [TEM-4 grammar and vocabulary]*. Nanjing Daxue Chubanshe.
- Sun, F., et al. (1985). Comparative patterns of reading eye movement in Chinese and English. *Perception and Psychophysics*, 37, 502-506. <https://doi.org/10.3758/BF03204913>
- Thompson, V. A., & Byrne, R. M. J. (2002). Reasoning counterfactually: Making inferences about things that didn't happen. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 28(6), 1154-1170. <https://doi.org/10.1037//0278-7393.28.6.1154>
- VERBI. (2017). *MAXQDA, software for qualitative data analysis*. Germany.
- Wu, H.-f. C. (1994). *If triangles were circles, ... a study of counterfactuals in Chinese and in English*. Crane Publishing.
- Yeh, D., & Gentner, D. (2005). Reasoning counterfactually in Chinese: Picking up the pieces. *Proceedings of the Twenty-seventh Annual Meeting of the Cognitive Science Society*, 2410-2415. <https://escholarship.org/uc/item/7wt3q4jt>
- Zhang, Y. (2009). *SAT wenfa [SAT grammar book]*. Shanghai Yiwen Chubanshe.

Supplementary materials. English and Chinese materials.

READING COMPREHENSION

Please read the text below and answer the questions on the basis of the text by ticking one option.

Bier was an Eighteenth century German philosopher who wanted very much to investigate the natural laws of the universe. At that time, European philosophers who investigated natural phenomena often only investigated individual natural phenomena and did not pay attention to the interrelationships between them. At the time there was already some contact between China and Europe, and Chinese works could be found in Europe, but none had been translated. Unfortunately Bier could not read Chinese. If he had been able to read Chinese, he would have discovered that when Chinese philosophers investigated natural phenomena, they focused on the interrelationships between phenomena. If Bier had read Chinese philosophy, he certainly would have been influenced by Chinese philosophers, have synthesised Chinese and Western philosophies, and have created a new philosophical theory which not only explained natural phenomena as individual entities, but which also made clear their interrelationships. This theory would not only have overcome a weakness in the Western philosophy of that time, but also would have influenced Western philosophy, bringing it closer to science.

- 1 Bier was very interested in the natural laws of the universe.
 He was It is probable that he was It is improbable that he was
 He was not Undecidable
- 2 Bier was a German philosopher.
 He was It is probable that he was It is improbable that he was
 He was not Undecidable
- 3 Bier was not influenced by Chinese philosophers.
 He was not influenced It is improbable that he was influenced
 It is probable that he was influenced He was influenced Undecidable
Please explain your answer (compulsory):

...

- 4 In the 18th century China and the West had no contact.
 They had no contact It is improbable that they had contact
 It is probable that they had contact They had contact Undecidable
- 5 In the 18th century Chinese works had already been translated into European languages.
 They had It is probable that they had It is improbable that they had
 They had not Undecidable
- 6 Bier led European philosophers to pay attention to the interrelationships among natural phenomena.
 He led them to pay attention It is probable that he led them to pay attention
 It is improbable that he led them to pay attention He did not lead them to pay attention
 Undecidable
Please explain your answer (compulsory):

...

阅读理解

阅读下列短文，并根据文章内容回答问题。在所选答案前的序号上画圈（如①、有兴趣）。

十八世纪时，有一位德国哲学家名字叫比尔。他很喜欢研究宇宙间的自然法则。当时欧洲的哲学家在研究自然现象时，往往只研究个别的自然现象，而不着重各现象之间的关系。那时候中国跟欧洲已经有交通往来，中国的著作亦有流传到欧洲的，可是都还没有被翻成外文。可惜比尔不懂中文。如果他看得懂中文的话，他就会发现中国哲学家研究自然现象时，特别注重各自然现象之间的关系。如果比尔读过中国哲学的话，他就一定会受到中国哲学家的影响，融合中国和西方的哲学，创出一个新的哲学理论。这个理论不但注重个别的现象，也会强调各现象之间的关系。这理论不仅能补充当时西方哲学之不足，并且也一定会影响西方哲学，使其更接近科学。

1 比尔对宇宙间的自然法则很有兴趣。

- 1、有兴趣 2 可能有兴趣 3 不太可能有兴趣 4 没有兴趣 5 不可判断

2 比尔是一位德国哲学家。

- 1 是 2 可能是 3 不太可能是 4 不是 5 不可判断

3 比尔没有受到中国哲学家的影响。

- 1、没有受到中国哲学家的影响 2、不太可能受到了中国哲学家的影响
3、可能受到了中国哲学家的影响 4、受到了中国哲学家的影响 5、不可判断

请说明你选择的理由（必填）：

...

4 十八世纪的时候，中国跟欧洲没有交通往来。

- 1、没有交通往来 2、不太可能有交通往来 3、可能有交通往来
4、有交通往来 5、不可判断

5 十八世纪的时候，中国的著作已被翻译成欧洲语言。

- 1、已被翻译成欧洲语言 2、可能已被翻译成欧洲语言
3、不太可能已被翻译成欧洲语言 4、没有被翻译成欧洲语言 5、不可判断

6 比尔使欧洲哲学家注意到各自然现象间的关系。

- 1、使他们注意到了 2、可能使他们注意到了 3、不太可能使他们注意到了
4、没有使他们注意到 5、不可判断

请说明你选择的理由（必填）：

...