國立清華大學語言學研究所

博士論文

UNDERSTANDING LINGUISTIC ABILITY IN CHINESE CHILDREN WITH WILLIAMS SYNDROME

威廉氏症候群中文語言能力之探索

研究生：許靜芬 Ching-Fen Hsu

指導教授：曾志朗博士 Dr. Ovid J.L. Tzeng

連金發博士 Dr. Chin-Fa Lien

民國 九十四 年 二 月

© Copyright By Ching-fen Hsu 2005

All Right Reserved

DEDICATION

To my strength, help, and wisdom source, the creator of heaven and earth,

God Almighty,

‘Do not be afraid or discouraged because of this vast army.

For the battle is not yours, but God’s’

-----2 CHRONICLES 20:15

TSING HUA UNIVERSITY

ACKNOWLEDGEMENT

坐在書桌前寫致謝詞的我心中充滿無限的感恩。回想多年前，一位來自後山的孩子，隻身來到一個陌生的城市，連公車票怎麼投，火車票怎麼買都不會的鄉下女孩，如今能一個人獨立的來去世界各地，這當中的學習與成長，真的要感謝很多人。我要獻上感謝的第一位是我的指導教授曾志朗老師，謝謝他一直以來的鼓勵與支持，還記得在曾老師還當教育部長時，有一次我走進他的辦公室，告訴他，我要放棄學業回家鄉去，他抬頭看了我一眼，然後說，你連學業都沒有完成，能做什麼？後來，我仔細想了想，的確想不出我能做的，就又乖乖的回去讀書。常常曾老師的一句話，會讓我思想良久，他總是在關鍵時刻給我一句關鍵的話，就因為如此，我今天才會在這裡寫博士論文的致謝詞，真的非常感謝他。另外，曾志朗老師對學生的照顧，真的令我印象深刻，儘管再忙只要是和學生的meeting，一定有時間，他總是為學生著想鋪路，從我碩士班到現在，一路的幫忙與指導，除了感謝，還是感謝。另外，我要謝謝洪蘭老師，從她身上，我學到如何當一位好的研究者，老師看書的速度很快，常常在她的課堂上能吸收最新的科學知識，而且總是能與生活結合，讓我深深感受到神經科學的廣博與樂趣。更重要的是，洪老師帶給我的生命教育，什麼是一位研究者的社會責任與回饋，在她身上，我看到一個美好的典範。謝謝兩位老師開啟了我的研究之路，創造了優質的研究環境，讓我願意繼續在這條路上努力，不僅看到台灣研究環境還有進步的空間，也在這樣的空間裡看見希望。

我要謝謝國科會提撥經費給博士生到國外短期進修，我的論文就是在這一年之內完成的，也謝謝馬里蘭大學公園分校（University of Maryland at College park）的教授 Stephen Crain，願意當我的共同指導教授(co-advisor)，在我待在美國的那一段時間，耐心聽我的研究，給我許多的建議與幫助，更是答應在百忙之中，從美國來到台灣為我口試，心中真的非常感激。還有 Andrea Zukowski 教授，由於

她的專長也是在威廉氏症候群的研究，因此與她有相當多的討論，常常透過與她的討論，使我對於自己的研究有更多的瞭解，也經常在與她討論後直接奔向圖書館，借一堆書回來看，補足欠缺的知識，對於她的幫助也是銘刻於心。記得 2004 年的八月，正逢兩年一次的全美威廉氏症協會(Williams Syndrome Association)在密西根州舉辦，她更慷慨的為我支付來回的機票與住宿等費用，讓我得以參與這個盛會，是我畢生難忘的經驗之一。另外，我也要謝謝 Ivano Caponigro 教授，他是一位相當有教學熱誠的老師，很鼓勵我，也幫助我許多，還記得有一次，原本只要與他 meeting 三十分鐘的，卻是延長到將近兩個小時，最後，他告訴我，要不是因為他實在肚子餓得不得了（當時已是下午三點，他還沒吃中飯），他很願意再繼續討論下去。還有，我也要謝謝 Howard Lasnik 教授， 非常有耐心的聽我的研究，並且給我建議，他真的是一位非常有耐心的老師。美國的四月通常是梅雨季，雨下個不停，也記得我與 Howard 在一個下雨的午後有 meeting，當時的我因為連日的雨而覺得心情糟透了，我一進他的辦公室就問他，為什麼雨會下個不停，只見他笑笑的安慰我說，因為四月的雨帶來五月的花啊，隨即他邊上網查閱天氣預報，邊告訴我說，讓我們來看看雨什麼時候停…。就是這樣的支持與耐心的幫助，讓我在美國一年半的時間裡，受益良多。也結交了許多好朋友，像是 Ellen Lau，總是不厭其煩的幫我修改論文，甚至，在寒流來襲的新年，與 Henny 帶我到唐人街看舞龍舞獅，陪伴我過中國新年，還有 Robert Fiorentino 及他的可愛日本女友 Utako Minai，常常為我加油打氣，也謝謝 Luisa Meroni 及她的先生 Andrea Gualmini 在我搬新家銜接不上的那一個禮拜收留我住在他們那裡，這一切的一切真的都讓我好感激，永生難忘，謝謝你們，溫暖了一個家遠在太平洋的那一邊且在美國無依無靠的台灣女孩的心。

也謝謝許多朋友給我精神上的支持鼓勵與實際上的幫助，像是凡真，文儀，Sister Nora，于牧師及師母，他們的女兒 Catherine，瓊如，小秋，宜青，中玉，傅雲慶醫師，連金發老師，Min Min，怡如，慧如，惠琦，福和國中李曉薔老師，周主任，清江國小柯娜雯老師，林耀誠老師，如蕙姐，林宗宏老師，郭保麟醫師，

罕病基金會的宏盈，簡淑，逸羚，uncle Robert，uncle Stephen，uncle Charles，傳琳姐，曉燕姐，協和助道會的弟兄姊妹們，淑仁，還有在美國的房東譚太太，及住在 Dewberry Lane 的室友們，謝謝你們。

最要感謝我的父母，謝謝你們的支持與幫助，對於我的決定你們從來就沒有反對過，總是默默支持我，關心我，真的謝謝。看著你們漸多的白髮，心中的不捨，絕不是任何言語能表達的，今年是爸爸的六十大壽，很高興我能把畢業證書當成禮物，希望他快快樂樂，開心的享受生活。謝謝你們，我愛你們。

最後，我要謝謝參加實驗的所有威廉寶寶，及他們的家長，如果沒有你們的支持，這個研究也不可能完成，對我而言，你們是天使，希望未來的路，你們能走的更好，祝福你們平安喜樂。還有要謝謝罕見疾病基金會的贊助，使得這本論文有好的品質，用好的面貌呈現，也在此一併致謝。

ABSTRACT

This study investigated the hypothesis of selective impairment on form and meaning in language processing on individuals with Williams Syndrome. It has been known that individuals with WS have spared grammatical knowledge even with mental retardation (average IQ of 55 or below) and poor cognition. Past research also showed that individuals with WS preserved normal verbal working memory and such intact verbal ability was thought to be responsible for their relatively good language performance (Wang & Bellugi, 1994; Jerrold, Baddely, & Hewes, 1999; Vicari, Brizzolara, Carlesimo, Pezzini, & Volterra, 1996; Vicari, Carlesimo, Brizzolara, and Pezzini, 1996; Karmiloff-Smith, Grant, Berthoud, Davies, Howlin, & Udwin, 1997; Robinson, Mervis, & Robinson, 2003; Laing, E., Grant, J., Thomas, M. S .C. & Karmiloff-Smith, A., in press). With a good verbal memory but deficit cognitive ability, individuals with WS are hypothesized to rely heavily on memory abilities in learning their language. This may explain the finding that grammatical knowledge of WS individuals is strong while their semantic understanding might be weak (Zukowski, 2001; Grant, Valian, & Karmiloff-Smith, 2002). In other words, individuals with WS might have selective impairment on form and meaning in their linguistic ability. In order to examine this issue, three projects were involved.

The first was studying on counterfactual conditions with negation in Chinese. Counterfactuals, which are mismatched in form expression and meaning understanding, served as the ideal probes to test this hypothesis. The second project was still on counterfactuals, but with a clearer conditional marker yaobushi in sentence initial position. In doing so, it was claimed that the unambiguous reading of counterfactuals might cause more efficient process in individuals with WS. The third

project was about proposition integration in semantics. It was interested to investigate the relationship between form and meaning in memory of individuals with WS. Are they more sensitive to grammatical forms than sentence meaning?

A general issue of developmental delay or deviant is also raised whenever populations with mental retardation are studied. The supporters of developmental delay argued that individuals with intellectual disabilities would show similar pattern on neuropsychological profiles to their mental age matches. They are just slow in development (Zigler & Balla, 1982; Zigler, 1969). On the other hand, the supporters of developmental deviant claimed that individuals with intellectual disabilities would show a dissimilar pattern to their mental age matches. They are different in development (Ellis & Cavalier, 1982; Ellis, 1969). In this set of studies, this general issue was also under investigation.

TSING HUA UNIVERSITY

In the study of counterfactuals with negation, the constituent comparison model in sentence verification was employed (Clark and Chase, 1972; Carpenter and Just, 1975). According to this model, the number of mental operations could be calculated. We hypothesized that people might form various possible representations mentally of a complicated sentence like counterfactuals. In this case, three representations are possibly formed in mind. Two of them were form-based representations and the other was meaning-based one. Different SOA were manipulated in sentence presentation: 0-SOA (i.e. simultaneous task) and 5-sec-SOA (i.e. delayed task). Three groups in different ages were recruited for each task from the sixth graders to college students. As for individuals with WS, the SOA was unlimited. We hypothesized that under time limitation, a form-based representation was easier to be induced. The results showed that all the three age groups were highly consistent in the performance of counterfactuals. They all formed a meaning-based representation at both SOA. However, the sixth graders showed a slightly different pattern from other two groups. Meanwhile, a clear age effect was observed. The sixth graders had slowest response latency among three groups and also erred most. Individuals with WS showed a meaning-based representation as their chronological age matches of college students and mental age matches of the sixth graders.

In the study of counterfactual conditionals with yaobushi, the design and procedure were parallel to the study of counterfactual conditionals with negation. Different participants in three age groups were recruited in each task. Meanwhile, two new individuals with WS participated in this study with unlimited SOA. The results showed that all three age groups were highly consistent in forming a meaning-based representation of counterfactuals at both SOA. A clear age effect and task effect were also observed. Meanwhile, the response latency in this study was faster than the one in the study of counterfactual conditional with negation in all age groups, suggesting an unambiguous reading of counterfactuals with yaobushi were processed more efficiently and successfully. As for the individuals with WS, they also formed a meaning-based representation as their chronological age controls and mental age controls.

In the study of proposition integration in semantics, the basic rationale was based on proposition entailment. A Bransford and Franks' (1971) paradigm in memory recognition was conducted. There were different superset sentences broken down into sub-sentences with different number of propositions and presented auditorily to college students and individuals with WS. Half of the sentences were presented in the training section and the other half were presented in the recognition section.

Participants were asked to make judgment of each sentence and assigned a confidence rating value to each judgment. The results showed that individuals with WS showed similar patterns as their normal controls in false positive rates. However, their means of confidence ratings were significantly higher than controls. In order to argue against the possible confound of yes-bias tendency, scrambled sentences were lumped together in recognition with the other half sentences from supersets. The results showed that individuals with WS recognized the scrambled sentences as they had never heard before and assigned very negative values as their normal controls.

In conclusion, individuals with WS are developmental delay, but not deviant in nature. In these three projects, they showed similar patterns as their chronological age matches in proposition integration and also in counterfactual conditionals with yaobushi. Further, they showed a similar pattern as their mental age matches in counterfactual conditionals with negation. These results indicated that they have spared logical reasoning ability and spontaneous proposition integration in their mental model. Thus, it is concluded that individuals with WS do not have selective impairment on form and meaning in their linguistic ability.

TABLE OF CONTENTS

LIST OF FIGURES

TSING HUA UNIVERSITY

LIST OF TABLES

TSING HUA UNIVERSITY

LIST OF APPENDEX

CHAPTER I

LITERATURE REVIEW: LINGUISTIC AND CONGITIVE PROFILES OF WILLIAMS SYNDROME

A What is Williams Syndrome?

In 1961, a cardiologist in New Zealand named J. C. R. Williams identified supravalvular aortic stenosis (SAVS) as a characteristic of this population. It was named Williams Syndrome after this identification. It is also known as Williams-Beuren syndrome, which memorialized a doctor in Germany who discovered infantile hypercalcemia and IIH a type of retardation involving abnormal calcium metabolism related to this syndrome (Semel & Rosner, 2003). Four levels can be explored when talks about Williams Syndrome: the neurological level, the cognitive level, the linguistic level, and the anatomical level. The neurological level covers the epidemiology and general diagnosis information in genome. The cognitive level refers to visual-spatial ability, planning, arithmetic, category-sorting ability, Piagetian concepts and face processing. The linguistic level covers their syntactic abilities, semantic advantages, morphosyntactic impairments and the interaction between these domains. The anatomical level includes a survey of Williams Syndrome's brain structures and anatomical findings. The strengths and weaknesses in each of these levels are reviewed accordingly.

B Neurological Level of Williams Syndrome

Individuals with Williams Syndrome (WS) compose a genetically disordered population with a discovery ratio of 1 in 7,500 live births (Morris, 2004), with equal incidence in boys and girls. This syndrome is caused by wide band of gene deletion (i.e. missing) in chromosome 7, including elastin, which is a kind of protein and offers nutrition to soft tissues of the heart (Korenberg, Chen, Hirota, Lai, Bellugi, Burian, Roe & Matsuoka, 2000). This deficit is a kind of heredity, which mothers and fathers are likely to be the origin of parent equally, and resulted from a different order of genes in mothers and/or fathers. Due to this genetic imperfection, individuals with this syndrome suffer several cardiovascular difficulties such as supravalvular aortic stenosis (SVAS). This difficulty causes individuals with WS to almost always have high blood pressure. From early infantile stage, they are characteristically hypercalcemic, which leads to difficulty excreting. This physical condition affects the emotional well-being of WS children. Parents of WS children often go through a hard time feeding and taking care of them. Besides, WS children are developmentally delayed individuals overall in body, for example, infants' height and weight are smaller than the normals in the 5th percentile), language, visual-motor ability and abnormal sensitivities to sounds/ hyperacusis (Lenhoff, Wang, Greenberg, Bellugi, 1997). WS have mild to moderate mental retardation and their average intelligent quotient is 55 (range 40 to 90 in WISC-R). In addition, verbal IQ and performance IQ are quite below the normal range, which is similar to individuals with Downs Syndrome (DS) on both scores. They have many facial features that are unlike normal individuals such as full prominent lips, a stellate iris pattern, prominent ear lobes, a wide mouth, small and widely spaced teeth, a medial eyebrow flare, a flat nasal bridge, a short nose, anteverted nares, and a hoarse voice (Bellugi, Lichtenberger, Jones, Lai, and George, 2000). Unlike Down syndrome, since these facial characteristics are not easy to detect at first glance, only a well-trained doctor or geneticist can recognize them as a WS individual from these facial clues alone. However, they resemble themselves within a population. The most direct test to diagnose this syndrome is to use fluorescent in situ hybridization (a FISH test, for short), which is a kind of elastin deletion probe. This test has an almost one hundred percent degree of accuracy in testing during prenatal or postnatal stages. However, this syndrome is not hereditary. According to the survey of Williams Syndrome Associations (WSA) in America and Britain, all parents who have children with this syndrome are normal. This syndrome is the result of gene mutation.

C Cognitive Ability of Williams Syndrome

There are five sub-domains discussed in this level from the basic concept in Piagetian tests of quantity/quality eternity to distinguished top-down or bottom-up processing strategies, and other spared cognitive abilities observed, for example face processing superiority and outstanding social behavior are under discussion.

TSING HUA UNIVERSITY

Piagetian Concept Impairment

Other missing genes in chromosome 7 such as LIMK1, GTF21, FZD3 have severe impact on WS in many aspects of cognition. For example, poor performances on the Piagetian test of conservation. The experimenter shows two glasses of water with the same volume in them at first, and changes one of the two glasses into a flat container of different shape with unchanged volume. The whole procedure is executed in front of WS children. Then the experimenter asks which one had more water in it? WS children typically reply that the higher glass (the original one) contained more water in it. But the truth is both of the containers are equal in their volume since the water in the flat container comes from the glass with equal volume. This simple test reflects that WS children lack of the concept of quantity eternity. This conservation concept is fully developed in normal children of age 7 (Gleitmen, 1990). However, WS children do not have this concept even in their adolescence. According to Bellugi et al. (2000), this poor performance is demonstrated in both individuals with WS and DS children.

Local Preference: Seeing Trees Before Forest

Some perceptual required tasks such as free drawing or copying also demonstrated this poor ability in performance. However, individuals with WS and DS show completely different patterns. When WS are asked to copy a diamond shape constituted with small x's, they cannot copy the global configuration of that diamond. They can only replicate the small x in two or three different lines either vertically or horizontally. The same pattern is observed in the picture of a letter D composed of small y. WS children did not draw the big letter D, but only small y in straight lines. However, DS children do quite well in replicating the configuration of the diamond without copying the details. The same pattern of this double dissociation is also observed in free drawing of a house and a bicycle. WS children draw the components of a house and a bicycle in detail, including such features as the windows, the door, the roof, the sidewalk, and even the swimming pool in the house picture, and the petals, the wheels, and the chain in the bicycle picture are clearly drawn, but they are not coherently organized. From their drawings, it is hard to tell what the objects are. However, drawings of DS children show the completed contour of the house and the bicycle. Another visual-motor ability test on WS children is to ask them to perform the block design in Wechsler Intelligence Scale for Children-Revised (WISC-R). The experimenter shows them a model of block design composed of four blocks, each of them has segmented colors in black and white. WS children are required to arrange another four blocks to mimic this shown model. However, the results showed that WS children have deficits in arranging these blocks, requiring many more steps than their DS counterparts and perform overall in a disjointed and fragmented way. DS children can arrange the blocks in accord with the global configuration of the shown model. But they also have problems in duplicating the details. These studies demonstrate that WS children have deficits in visual-spatial perception and visual-motor production as DS children, another type of genetic disorder, in different patterns. These two groups show a clear pattern of double dissociation in top-down and bottom-up processing strategies as they can see either forest or trees.

Face Processing Superiority

Individuals with WS show an overall advantage in face processing. According to Bellugi et al.(2000), in Benton Upright Faces Test, in which participants are presented a face model and then followed six pictures with different faces (including the same one as the model) in different orientation and light, WS children are required to recognize which one of the six is exactly the same one as the original. This test requires a good ability in recognition under various possible faces. The results show that WS children perform significantly better than DS children in recognition. In the Warrington Face Memory Test, both WS and DS children are presented one picture as a probe, then two pictures (i.e. the original face with another distracted one) follow for them to match. The results also demonstrate an advantage in face processing with WS children. They perform at a level almost eighty percent of that of normal controls in this matching task.

Their counterparts, DS children, show poor performance: only sixty percent correct. Even in the most difficult tasks like the Mooney closure task, in which stimuli are obscure and reduced to simple rough sketches with different gray scales, participants are required to identify not only the object or figure on the Mooney picture, but also to identify properties of it. For example, if the figure is young or old. The results show that WS individuals have significantly better performance in telling the properties of the obscure pictures than DS individuals. From these studies on investigating WS individuals' face processing ability, WS children show a remarkable and better processing than DS counterparts.

Further confirmation comes from a direct contrast of simple visual-spatial percept like Benton Judgment of Line Orientation and face perception like Benton Face Recognition Test. In the former, participants are shown part of a model composed of lines with different orientations. This shown part constitutes two lines selected and fitted at a certain angle in the model. Participants are asked to identify which two lines in the model are selected based on the shown angle. WS children perform near forty percent correct, which is very far away from the normal range. However, when a Face Recognition Task is encountered, in which a face model is shown to participants and followed six pictures with various shadow, orientation, and lighting, WS children perform within correct range of the normals at eighty percent. This contrast in performance in visual-spatial line orientation and face recognition with different orientation supports the superiority in face processing of WS. This with-in sub-domain dissociation is demonstrated clearly.

Hypersociability

Unlike autistic children, individuals with WS do not avoid interacting with people. Actually, they like approaching people, even strangers whenever possible. People who have experiences in talking with WS for the first time always are impressed by their intimacy. They are characteristically of friendly and hyper-sociable. Jones et al. (2000) conducted a study at Salk Institute at UC San Diego on WS in measuring their extraordinary social ability. There were 26 WS participants, another

26 matched normal controls chronologically, and still another 12 matched normal controls mentally in this study. All participants are shown both positive and negative photos which are selected from a norm. Twenty-one photos are presented for each pole. Participants are required to rate the shown photo on a five-point scale from +2 to -2 to represent the most positive feeling to the most negative feeling. The results show that WS participants give overall higher rating points for both positive and negative photos than the other two groups of normal controls. They show a high approachability bias. To the WS population, people were nice and approachable. As Jones et al. point out, WS gave a high rating to a photo with smiling face, but normal controls gave it a minus rating because it was smirking and frowning. In other words, it seems hard for WS individuals to recognize subtle social cues; they only relying on superficial signals. Since WS are distinguished from other cognitively impaired populations such as DS and autism in their hypersociability, there is a question that remains to be answer. Is this distinguished ability gene-based, pre-wired naturally before language development, or is it caused by a language boost, a by-product of a language advantage? Jones et al. designed a parental separation task trying to answer this question. Individuals with WS and normal chronologically age matched controls were included as participants. They were leaded to a playing field with parents and experimenters. After 3-5 minutes free play with toys, parents of the children were instructed to say "goodbye" to their child and left the playing field. At the same time, experimenters begin to code the child's emotional expressions. They analyze the frequency and intensity of children's facial expressions and voice changes. That is, how often and in what degree the children expressed their frustration, anger, happiness, and other feelings by face and intonation. After one half to one whole minute, left parents went back to the field to reunite with their child. The results showed that WS children showed less emotion and had much weaker facial

expressions and voice variation. They were calm and quiet waiting for their parents. Even when asking questions to the experimenter, they were not out of control. It implied that WS infants and children do not fear to be alone if necessary. Compared to the normal children, they gave much more frequent emotional facial expressions and vocal changes such as whispers, whinge, cries, screams, and anger. Moreover, WS individuals do not avoid eye contact with people. In a Barrier Task, in which WS toddlers are supposed to reach for a toy behind a plastic barrier, the WS toddlers made eye contact directly with the experimenter during the whole process and ignored the toy completely. They smiled at the examiner and engaged other people in the room. The results demonstrated that WS individuals are pre-wired with regard to the social ability. They prefer looking into other people's eyes and faces. They show a steady and calm attitude during the separation of parents. In other words, the unusual hypersociability observed is not a by-product of language development. They are born with it.

D Linguistic Ability of Williams Syndrome

Linguistic preservation of individuals with WS is a well known ability relative to their cognitive ability. Their syntactic knowledge and semantic representation are explored below. The concept of *sparing* and *impairment* is also clarified. In semantic processing, both lexical semantics and sentential comprehension are included. However, it will be argued that WS individuals are particular in developmental trajectories, which are dissimilar to the normal controls in their language proficiency. WS individuals are delayed, but not deviant in their linguistic knowledge. Further, a within dissociation in language module of WS individuals is proposed.

Syntactic Knowledge

In contrast to impaired cognition, WS individuals seem to preserve quite well linguistic abilities mainly based on research from Bellugi and her colleagues at the Salk Institute of UC San Diego in La Jolla (REFERENCES). According to their research, WS children can comprehend highly complicated syntactic structures such as reversible passive sentences. The experimenter showed two pictures to the participants and simultaneously described a sentence. Participants were asked to select one of the two pictures matching the semantics of the described sentence. For example, "The horse is chased by the girl" was spoken to participants and two pictures followed having two possibilities. One picture was of a girl chased by a horse and the other was a horse chased by a girl (the matched picture). Since both participants are animate, they can be AGENT or THEME in their thematic roles. Therefore, participants had to completely understand the grammatical structure otherwise they could not select the correct picture. The results showed that individuals with WS performed almost perfectly on this test (i.e. their correct percentage was at ceiling) while individuals with DS were only guessing (i.e. their correct percentage was at chance level), suggesting WS individuals preserve good ability in comprehending passive sentences with reversible participant roles. Besides, WS individuals also can complete complex tasks like grammatical judgment in detecting and correcting ill-formed sentences. In an experiment involving the production of negation questions (Zukowski, 2001), WS children performed in a similar pattern with normal controls, suggesting a nearly normal development in negative formation. The author hypothesizes that the difficulty in production of negation questions in children is caused by the negation raising problem, and not the auxiliary raising difficulty based on Thornton's observation. In this task, five sentence types were designed with ten tokens for each type. The first sentence type included the simplest questions (yes/no), for example, *Do you like hamburgers?* The second type included affirmative Wh-object questions, for instance, *What kind of pets do you have?* The third type included negative Wh-object questions such as *What flavor don't you like?* The fourth type included affirmative Wh-adjunct questions, for example, *Where does your tarantula sleep?* The fifth type included negative Wh-adjunct questions like *Where don't you want to live?* All of these expected elicitations from children were probed with a telephone interview game with experimenters and participants. At the beginning of this game, children were instructed to ask questions about certain topics which were designed beforehand. Each of them pretended to talk to a person who is famous, like Elvis or Bill Clinton, on the telephone. During a session, the experimenter sat next to the child feeding questions to make the conversation continue. This elicitation created a much more natural atmosphere in collecting negation question data. Each sentence type was elicited by different presuppositions. For example, for the simple yes/no question, the probe sentence would be "*I wonder if she likes hamburgers. Ask her for me*". For affirmative Wh-object questions, the probe could be "*I don't know what kind of pets she has. Ask her what kind*". For negative Wh-object questions, the experimenter would say "*She doesn't like one flavor? Ask her what flavor?*" For affirmative Wh-adjunct questions, the probe sentence would be "*I wonder where tarantula sleeps. Ask her where*". For negative Wh-adjunct questions, the child would be asked in a question like this "*I heard that there is one place that she doesn't like to live. Ask her where*". Through this game, negation questions were easily elicited from children. A real scenario is quoted in the following (Zukowski, 2001:81).

Experimenter: I wanted to ask Elvis about his vacation this year, but actually I don't know if he got a vacation. Ask him if he did.

WS child: Did you have a nice vacation?

Elvis: Oh yeah it was excellent.

Experimenter: Oh I wonder who he visited, ask him who.

WS child: Who did you visit?

Elvis: I visited my cousins.

Experimenter: You know what, I heard that there's one cousin that he doesn't like. Ask him which one.

WS child: Which one don't you like?

Elvis: I don't like Bob.

Experimenter: He doesn't like Bob? Ask him why.

WS child: Why?

Elvis: Why what?

WS child: Why don't you like Bob?

Elvis: Because I have to share a room with him and he snores.

TSING HUA UNIVERSITY

If the child did not reply with an expected target question but produced a short Wh-question such as the example shows in the fourth lines from the bottom (i.e. *why* in this case), the accomplice who plays the role of somebody in this task has to induce the full question sentence instead of continuing this dialogue (e.g. *why what* in this example). There were 11 WS children and another 11 mental age matched normal children in this study. The results showed that both WS children and mentally matched controls performed excellently in the simple yes/no questions (95% for WS children and 100% for normal children). Besides, both of these groups have high percentage of correctness in affirmative Wh-questions, including Wh-objects and Wh-adjuncts, too. In these affirmative conditions, WS children have 86% correct and normal children have 99%. However, both of them performed poorly in negative conditions (45% for WS children and 47% for normal controls), suggesting a preference to produce negation questions in correct word order. Instead of the correct targets, WS children are prone to neglect the necessity of raising the auxiliary. There are 31% errors in failing to raise the auxiliary. For example, they would say "*Why you didn't kiss your tarantula?*" instead of saying "Why didn't you kiss your tarantula?" The same mistake is also observed in normal controls with a relatively low error rate: 9%. Though there is a discrepancy between WS children and non-disordered normal controls in this type of errors, the overall error patterns in failing to produce negation questions of these two groups are similar (54% for WS children and 51% for normal controls). In addition to the responses of failing to raise the auxiliary, both of them reduplicated the auxiliary in a sentence. For instance, a WS child said "*Where can your dogs can't sleep?*" and a normal child said "*What place do you don't wanna live?*" In this error type, normal controls had a higher percentage (18%) than WS children (11%). Moreover, both of them raised the auxiliary in a sentence but left the negation marker in situ. This error pattern is exactly compatible with the hypothesis proposed by Thornton and Zukowski that children undergo difficulty in negation raising but not participant-auxiliary inversion. For example, a WS child responds "*Where do you not wanna live?*" and a normal child responds "*Why do you not kiss your tarantula?*" Similarly, normal controls scored two-times as high (24%) than WS children (12%). The different strategies showed in these two groups indicated different processing of negation questions, but did not show any differences in their abilities. Both of them performed almost equally in producing ungrammatical sentence structures, suggesting they are inclined to produce correct negation question forms in different patterns. Thus, Zukowski concludes that WS children have near normal ability as mental age matched controls.

Zukowski (2001) conducted another experiment on the production of relative clauses in English. Ten WS children and chronologically age matched normal controls participated in this study. WS participants had higher mental age than the normals. The experimenter read a scenario to each child and asked a question to elicit responses of relative clauses. Another accomplice took part in this study, usually the child's mother, sitting opposite to the child. At the beginning of this task, a lead-in description is stated to the child as the following scenario. "*So, here's how this game works. We're going to look at these pictures. And your mom has a picture just like ours. So there are two girls, and they are both playing with trucks. One girl is jumping over her truck, and the other girl is sitting on her truck. And what's going to happen is.. My friend Max the mouse is going to show up, and he's going to look at one of the trucks. And we have to tell your mom, which truck is Max looking at, so she can put a copy of Max on her sheet. And she'll show it to us and we'll tell her whether she got it right or not. Ok? So this girl is jumping over the truck, and this girl is sitting on a truck. Let's see where Max shows up (picture changes after pushing the space bar)*". Then the experimenter asks a question like this "Can you see Max? Ok, tell your mom, which truck is Max looking at"? In this study, there were two types of sentences designed with participant and object gaps. In each gap type, Zukowski also manipulated the number of animate participants. Each gap-type sentence had one or two animate participants. As the scenario described, participants were shown base pictures on a computer screen, and after pushing the space bar, they look at the picture with the same objects/participants inside either in different colors or joining another participants like a mouse or a bird. All the stimuli in the four conditions were put into four kinds of relative embedded clauses, namely, full sentence and noun phrase in matrix participant or object positions. Therefore, it is a 2 (participant gap vs. object gap) x 2 (one animate vs. two animate) x 4 (full sentence vs. noun phrase in matrix sentence with participant/object position) x 4 (four sentences in each condition) design. For example, "*The cow that the girl is pointing to*" for the Object Gap Condition with two animate participants in the noun phrase has the appropriate response, "*Max is looking at the cow that the girl is pointing to.*" The results showed that individuals with WS are 77 percent correct in Participant Gap responses relative to 82 percent correct with normal controls. On the contrary, WS individuals perform at only 11 percent correct in Object Gap responses relative to 51 percent correct with normal controls. Though the correct percentage is not high for WS children, 9 out of 10 participants can produce at least one object gap target sentence. Zukowski argues that WS children do have the requisite linguistic knowledge for generating relative clauses. Their poor performance in relative clauses results from the difficulty in implementing this knowledge when producing them. As an analogy to normal adults, they also show difficulty in comprehending triple center-embedded clauses such as *The truck [that the girl [that the man spoke to ___] jumped over ___] turned red* (2001:69). However, normal adults are not viewed as a syntactically impaired population. They show implementation difficulty in language competence and WS populations should be evaluated in the same way.

Absolute Sparing and Relative Impairment

It is worth paying attention to the concepts of absoluteness and relativity. According to Mervis, Robinson, Bertrand, Morris, Klein-Tasman, Armstrong (2000), the so called *absolute sparing* can be defined as abilities that are commensurate with chronologically age matched individuals in the normal population. The *relative sparing* is defined as abilities that show advantages within a specific domain of the same population such as the superiority in face processing and poor visual-spatial ability in WS; or *relative sparing* can be defined across different population comparisons. For example, the comparison of WS and DS are conducted using various cognitive abilities. On the contrary, absolute impairment is a particular cognitive ability that is never acquired. Relative impairment is the comparison of a matched population, ability, or norm. Following these definitions, let us rethink the grammatical ability of WS individuals.

Grant, Valian & Karmiloff-Smith (2002) conducted another relative clause experiment to argue against the claim of intact grammatical ability in WS. This was a simple elicited imitation task. There were 14 WS children and adults participating in this study. There were another three groups of normal controls: 5, 6, and 7 years old, respectively. Each normal control group had 11 children. WS participants have higher mental age near 9 years of age from British Picture Vocabulary Test (BPVT), a test of vocabulary similar to the Peabody Picture Vocabulary Test in America. The authors assume that imitation errors were a function of complexity of structure, not of utterance length. Four types of relative clauses were designed. (1) SS sentence type; the modified noun is a participant in both the major sentence and embedded clause. For example, *the boy chasing the horse is fat*. (2) SO sentence type; the modified noun is a participant in the major sentence and an object in the embedded clause. The object could be a direct object or a prepositional object. For example, *the cat the cow chases is black* for the former and *the book the pencil is on is red* for the latter. (3) OS sentence type, the modified noun is an object in the major sentence and a participant in the embedded clause. For instance, *the dog chases the horse that is brown* for the direct object example and *the square is in the star that is blue* for the prepositional object. (4) OO sentence type; the modified noun is an object in the main clause and at the same is the object in the embedded clause. For instance, *the dog is chasing the cow the boy sees* for the former and *the pencil is on the shoe the girl has* for the prepositional object.

According to their imitated errors, several categories are classified. The first category is syntactic change, referring to additions or substitutions of relative pronouns such as *who/which* for *that*. The second category is lexical change, suggesting any word changes such as *girl* for *woman* or *a* for *the*. The third category is morphological change, suggesting any minor changes in words like pluralizing a noun or changing the verb tense. The results show that these four groups have the same pattern in responses. The difficulty order for these four sentence types are SO, OS, OO, SS from hardest to easiest. The most important observation is that WS children and adults perform the pattern like the 5 year-old group regardless of having the highest mental age among these groups. So, vocabulary limitation cannot be a reason to argue against this result. Furthermore, the authors claim that memory limitation and motor inability for the difficulty of WS to produce relative clauses also cannot hold because these individuals can imitate filler sentences well. Therefore, this is a robust result, showing a relative preservation in language of WS individuals.

Another task followed which support their claim. The authors added the relative pronoun *that* to all sentence types to have participants imitate. They assumed that the completeness of sentence structures would help individuals with WS process sentences easier, especially for SO sentences. For example, *the cat [**that** the cow chases] is black*. The results show that WS do make fewer errors across all sentence types when the overt maker is added. It implies that WS are sensitive to syntactic structure. When the structure is more overt, it is easier to extract the meaning of the sentence to make imitation succeed. The experiments from Grant et al. clearly explain one point that WS individuals are not intact in their linguistic knowledge, though a final conclusion is still far too complex to make. Therefore, it would be erroneous to conclude that language advantage of WS means that their language is intact. People with WS do impress others in their language communication given their mental retardation. However, this ability is not equivalent to their chronological age. In other words, syntactic modularity in WS population deserves further consideration.

Semantic Fluency

WS individuals have good verbal fluency and knowledge of lexical semantics. In a lexical production task on semantic fluency (Bellugi, et al., 199x), WS, DS, and normal controls were instructed to give all the animal names as fast as they could in a minute. The results showed that WS children and normal controls gave many different animal names in this limited time, but DS children gave much fewer. Besides, WS children produce significantly more uncommon and low frequency lexical items such as hippopotamus, dragon and brontosaurus than normal controls and DS groups. This exceptional ability in lexical knowledge gets further support in a comparison of the scores on Peabody Picture Vocabulary Test (PPVT) on both WS and DS individuals when the latter have higher full IQ scores in WISC-R than the former. Bellugi et al. also observed that the numbers of examples in this task rise across age span from age 5 to 40 and a sharp increase in fluency happened around 11 years old.

Other experiments demonstrated that WS have unique semantic organization. In a homonym task, participants were presented a triad with the target as a homonym with its two meanings: the primary and the secondary meaning. Participants were required to respond orally which two have similar meanings and should go together. DS individuals and normal controls had the homonym target go with the primary meaning more frequent than the secondary one. But, WS individuals gave judgments to both meanings of the homonym target equally. Besides, in a definition task, WS also showed rich vocabulary knowledge in define homonyms. For example, when the experimenter gave the same word to both WS and DS children, WS easily gave clear definitions for it, but DS could not.

Morphosyntactic Disadvantage

A language advantage is also demonstrated in sentence repetition on WS individuals (Bellugi et al., 2000). There were four groups participating in this study, including children with WS, children with focal lesions, children with language impairment, and normal children. All the participants were required to repeat the sentences that were heard. The results showed that WS (age range from 4 to 12) did not have any differences in repetition from children with focal lesion and language impaired counterparts who were not genetically disordered, but significantly different from the normal group. In another sentence repetition task, a clear developmental trend on morphosyntactic ability was observed. There were three groups of WS, DS, and normal controls from 2 to 16 years old participating in this study. All the participants were required to repeat what they heard from the experimenter. Their production errors were analyzed in detail. The results showed that WS children have morphosyntactic errors at the beginning of age 5 to the end of age 16 with a decreasing number of errors. This developmental pattern is very similar to normal controls who have errors from age 3 to age 10. But, DS children are quite delayed in producing these errors probably from the beginning of age 10 and continue to age 16 without decreasing error rates. This study demonstrated a relative advantage of WS population on morphosyntactic knowledge over DS.

A similar observation is also made in Karmiloff-Smith, Grant, Berthoud, Davies, Howlin, & Udwin (1997). Since past research results are based on sub-items of the Test for Reception of Grammar (TROG in short, Bishop, 1983) with few participants, the authors argued that a more careful conclusion about WS language ability should be made according to TROG full-scale analysis with larger numbers of participants. Participants were required to point an appropriate picture when they heard a lexical item or a sentence. For the tests of lexical items, participants had to identify nouns like shoe/apple, verbs like eating/sitting, adjectives like long/red), negatives like the boy is *not* running, and noun phrases like the big cup. In addition, for the tests of morphosyntax, participants had to differentiate comparatives (the knife is longer than the pencil), relative clauses (the boy chasing the cat is fat), and post-modified participants (the horse but not the boy is standing). The results showed that the mean TROG test age (6;3) for WS participants was significantly lower than their chronological age (18;2).

A further analysis showed that WS do not have problems with the lexical items, but they have difficulties on their morphosyntactic comprehension. A further examination on gender agreement of French-speaking children and normal controls was tested. Fourteen French-speaking WS children participated in this study, having higher verbal IQ both on vocabulary and syntax than normal controls. The authors manipulate two independent variables, i.e. gender and congruence. It was a 2 (feminine vs. masculine) x 2 (concordance vs. discordance) design. In French, there are two kinds of natural gender, i.e. feminine and masculine. All the elements in a sentence have to agree with gender. For example, *un tapis* means "the carpet" (a masculine article and a corresponding noun in the same gender) and *une chaise* means "the chair" (a feminine article and an agreed inflection in *-e* attached to the following noun). In other words, gender agreement can be realized by the articles and word endings clearly.

The authors not only included real words as stimuli, but also invented novel words as the names of objects and animals. Each pair of stimuli in an article and a noun wais presented as pictures with different colors. When these pictures with different colors were shown to participants, the experimenter asked them to repeat the pair of words (i.e. an article and a noun) correctly. After correct repetition, the experimenter hid a ring under one of the shown pictures (e.g. a picture with green ant in this case) and asked participants where the ring was. Sentence completion was elicited to shorten the response of participants, for example, the experimenter said *I hide my ring…?* with a rising intonation in French and participants were supposed to reply an answer such as *Under the green ant* in French with agreement across-the-board. The results showed that WS participants and normal controls performed similarly. They made more errors on the discordant condition and fewer ones on the concordant condition. However, WS individuals showed significantly higher errors than controls. In addition, WS individuals also performed more errors on article-adjective agreement in their responses. For example, they could not reply correctly giving the response *un tapis vert* (THE-CARPET-GREEN, article-noun-adjective) to mean "the green carpet". That is, they failed in making adjectives agree with cues on articles. On the contrary, normal controls performed well on this agreement. In another further manipulation, the authors used neutral article *deux* (two) as a probe to investigate whether a gender cue on the word ending of a noun was valid or not. It meant that in this case participants could only respond based on word endings. For example, given *deux chaise* (two chairs), participants had to reply *la chaise verte* (the green chairs). The results showed that WS performed at chance. In other words, WS fail to make use of cues on word endings. A more detailed analysis of error types in repetition revealed the nature of WS in their morphosyntactic ability. They gave the largest proportion of errors on agreement rather than lexical errors such as missing articles, missing adjectives, or missing nouns. On the other hand, normal controls made more lexical errors than agreement errors. This pattern seems to demonstrate a double dissociation of morphosyntax in WS individuals and normal controls.

Sociolinguistic Ability

Based on the observations of hypersociability and relative preservation of language ability of WS population, it is not surprising to notice the interaction of these two domains. They show rich affectivity in linguistic discourse. They do not have a fear of strangers as in other cognitive disorders like autism, Downs Syndrome, and even normals. In a story-telling task conducted by Jones, Bellugi, Lai, Chiles, Reilly, Lincoln, and Adolphs (2000), WS were instructed to tell a Frog Story, a story about a boy and his dog searching for the missing frog. The researchers made several evaluations to judge the narratives of WS based on units of propositions. The results showed that WS individuals used significantly more linguistic expressions than age-matched DS counterparts and mental age-matched normal controls in affective states, character speech (e.g. role playing), sound effects, audience hookers (e.g. what do you know?), and emphatic markers (e.g. suddenly; wow). It seems that WS are experts in using linguistic terms to enhance their narratives. One of the examples excerpted from Bellugi et al. is listed below.

"Suddenly when they found the frogs…There was a whole family of frogs…And ah he was amazed! He looked…and he said "Wow! Look at these…a female and a male frog and also lots of baby frogs. Then he take one of the little frogs home. So when the frog grow up, it will be his frog….The boy said, "Good bye, Mrs. Frog…good bye, many frogs. I might see you again if I come around here again. Thank you Mr. Frog and Mrs. Frog for letting me have one of your baby frogs to remember him".

In an interview of WS children at Bellugi's lab, they were asked questions about themselves and their family. A similar result was also observed. WS children showed much more affective states, emphatic markers, evaluative comments and character speech turns than DS counterparts and normal controls. The most interesting part was that WS children took their turns as the host to ask questions. But DS individuals and normal controls did not have this pattern. In addition, in a social ability questionnaire designed by the Salk Institute testing social approach behavior and social emotional behavior, the groups of WS, DS and autism showed different social scores (Bellugi et al. 2000).

This questionnaire was also based on a parental report, which evaluated their children in approachability toward other people relative to children of the same age, in what degree other people wanted to engage their children and how correct their children were at predicting emotions of other people. The results showed that WS children have the highest scores both in approachability and emotion-predicting ability. For example, they often go up to strangers, asking questions actively, introducing themselves and showing warm feelings. The least scores were shown among autistic children. Their parents often reported that autistic children treat all people similarly, avoiding eye contact, preferring to be alone whenever possible. DS individuals are between these two extreme groups as a medial group in these social-ability tests. They are reported as a shy and physically-withdrawing population. The same pattern is also reflected in global social ability scores (i.e. the averaged scores of approachability and emotion-predicting ability). In sum, WS children have outstanding fluency in narratives and sociolinguistic behaviors when communicating with other people, but not other populations with developmental delay such as their DS and autistic counterparts.

General Review of Linguistic Ability

So far, we have investigated the strengths and weaknesses of WS individuals generally. On their ability of language, they have good performance in syntactic comprehension and production. In semantic understanding, they also show high fluency in lexical semantics and rich vocabulary in their mental lexicon. However, WS individuals are not as good as normal controls in some aspects of language. Instead, they are relatively preserved compared to other genetic disorders like DS individuals and the Prader-Willi Syndrome population, but not absolutely spared in relation to normal people. They are observed to have aberrant semantics. It shows not only in their use of special knowledge of lexical items, but also in their interpretation of sentences.

In Bellugi et al.'s counterfactual experiment on WS, despite their high correctness in grammatical structures as normal controls, WS individuals show a responding gap especially on semantics in relation to normal controls. It seems that WS do have a relative ability on syntactic level, but they still have distinguished and peculiar semantic knowledge in some degree, which needs further investigation. Moreover, their preference in emphasizing superficial structure in language is parallel to the processing of face recognition. This is supported by an experiment of WS individuals showing significantly higher rating scores towards both positive and negative face models compared to normal controls. In other words, they perform the same strategy in processing visual information and auditory speech. Furthermore, since WS's morphosyntactic knowledge is incomplete, they do not acquire the easiest and earliest linguistic parts in development such as gender agreement and other basic concepts as normal children. Therefore, WS individuals are inferred having different learning paths than the native speakers. In fact, they are much like second language learners as Karmiloff-Smith claims.

E A Neural-Based Hypersociability of Williams Syndrome

Galaburda, A. & Bellugi, U. (2000) analyze over hundred cases of WS individuals' brains and find that a marked indentation of the temporal-parietal regions in the area of the sulcus. Besides, they also observe that the whole posterior-parietal regions and occipital regions are small when compared to normal brains. This deficit seems to account for the visual-spatial deficits of WS children in copying drawings and the poor performance in visual-motor integrations. Further evidence comes from the observation that the central sulci pattern in WS individuals is different from the one in normal people; they have abnormal medial-dorsal cortices. In addition, there is a dramatic reduction and unusual shape of amygdala observed in WS individuals' brains, especially in the dorsal portion of the lateral nucleus (LNd). Moreover, the temporal horn (TH) is more dorsal than the normal. In other words, amygdala is deficient in WS population both in shape and in size, resulting in unusual ability in social interaction with people. This is also compatible to the demonstration of focal bilateral amygdala damage patients giving abnormally positive ratings to unfamiliar faces or negative faces typically received from controls, though WS give abnormally positive ratings across all faces. Maybe this is the possible reason to explain their fearless and overfriendly affectations to strangers.

Bellugi et al. (1996) also conduct a MRI anatomy analyses on both WS and DS populations and normal controls. The results show that the anterior cortex of WS individuals is relatively larger than the one in DS counterparts, but smaller than normal. However, the volume size of neo-cerebellum of WS is enlarged and biggest among these three groups. This preservation seems to account for the hyper fluent language ability of WS children relative to the poor performance on language tasks of autistic children with shrunken neo-cerebellums, which is believed to be responsible for processing high functional cognitive operations in the brain. There is also a relative preservation of mesial-temporal lobe in WS brains, which is related to the emotional feelings when conjoined with some frontal areas. Compared to DS individuals' preservation, lenticular nuclei (i.e. a place where basal ganglia is located) have a larger proportion of cerebellum volume than WS and normal controls. This difference between DS and WS individuals accounts for the better motor skills of the former than the latter.

So far, the cognitive profile of WS population has been reviewed. Generally, WS children have defeated cognitive abilities in spatial orientation, planning, simple concept of conservation, and algorithm. However, they have spared capability in face processing as normal people, which is demonstrated from the absolute performance in Benton upright faces test, Warrington face memory test, and Mooney closure test. Their poverty in visual-motor integration is also observed when asking them to make drawings, and copy sketches of models. They are also characteristically hyper-sociable. From early infant experiments, their hypersociability is not a by-product of language development. WS toddlers prefer to have eye contact and engage others with freely and easily, suggesting an innate hypersociability. Biologically, WS population is observed to have amygdala deficit in reduced size and uncommon shape, which is believed to have high correlation with the overfriendly preference to people. In my point of view, WS population's relative preservation in language proficiency and spared superiority in face processing are the by-products of hypersociability, which results from the deficient amygdala. Language advantages and face processing superiority have a close relationship and are very helpful to interact with people in communication. Due to the preservations of these by-products, the WS population performs relatively better than other genetically disordered groups like DS counterparts. This inference (or assumption) is parallel to other cognitive deficits demonstrated in relation to gene deletion such as IQ-defeated genes, visual-spatial deficit genes, heart disease gene (i.e. elastin). That is, a neural-based assumption toward hypersociability of WS population is proposed.

Based on this assumption, their relative good performance in language finds a reasonable and neurological explanation given genetic disordered mental retardation. However, this assumption needs further investigation. Since the WS population and autistic individuals are distinguished by their social ability, their patterns are totally different. Both of them suffer the amygdala deficits, but they seem to be located at two poles in the same spectrum. Autistic individuals are completely withdrawn from people relative to the hypersociability of WS individuals. Therefore, the final conclusion is too far to be reached now. But, the most important contribution of WS population is that cognitive profiles can be distinguished clearly into separated domains. In a word, they have weaknesses and strengths within cognition such as spatial orientation and face processing.

F Memory Ability of Individuals with Williams Syndrome

Wang and Bellugi (1994) first demonstrated the dissociation between verbal and visual-spatial short-term memory on individuals with Williams Syndrome. They used digit span as the test for verbal short-term memory and corsi block as a measure for visual-spatial short-term memory. They performed a double dissociation on short term storage for phonological and for visual-spatial information compared with individuals with Down syndrome. Jarrold, Baddeley, Hewes, (1999) reexamined this finding with more careful control groups. In Wang and Bellugi, they matched full IQ scores of WS and DS individuals. However, Jarrold et al. argued that this match had confound. The dissociation may result from the deficient verbal ability of DS individuals and the impaired visual-spatial ability of WS individuals. Thus, Jarrold et al. took both verbal and non-verbal IQ on these two genetic groups and covaried out the effect of any differences in these measures. Meanwhile, they recruited moderate learning disability individuals as a control group. Their results replicated Wang and Bellugi. Individuals with WS performed better on digit span tasks than on spatial tasks. On the contrary, individuals with DS performed the reverse pattern. Vicari, Brizzolara, Carlesimo, Pezzini, & Volterra (1996) also confirmed this dissociation. They further pursued a question to see if there is selective impairment within verbal working memory. They used immediate recall tasks testing individuals with WS and normal mental age matched controls. Their rationale is that if the phonological competence of individuals with WS is intact while semantic competence is deficient, then selective impairment between spared short-term and impaired long-term systems in working memory is expected. The results confirmed their predictions. Individuals with WS showed a significant difference recall rate in primacy effect, but no difference in recency effect when compared with normal controls. These results indicate a clear dissociation between verbal short-term memory and verbal long-term memory, suggesting further an intact phonological loop and impaired lexical-semantic system.

Another study examining verbal short-term memory more directly, Vicari, Carlesimo, Brizzolara, and Pezzini (1996) demonstrated spared and impaired functions within verbal working memory. Six word lists were presented to children with WS and mental age matched controls. Two of the word lists were composed of two-syllables with high frequency words or low frequency words; another two were composed of four-syllables with high or low frequency words; still another two were composed of acoustically similar or dissimilar words. Participants were asked to repeat these words after presentation of the experimenter. Vicari and colleagues found that children with WS showed the same word length effect (i.e. two-syllable words were repeated more accurately than four-syllable words) and phonological similarity effect (i.e. acoustically dissimilar words were repeated more accurately than similar words) as normal controls. However, children with WS showed less of a frequency effect (i.e. difference in accuracy in repeating high vs. low frequency words) than normal controls. We concluded that the contribution of the phonological loop towards word span effects in children with WS and normal controls was the same (pp921). The frequency effect was interpreted as a hyper-phonological strategy used by children with WS relative to normal controls. We hypothesized that an impaired contribution of long-term memory to short-term memory caused this effect in children with WS. In other words, normal controls used both phonological recoding and semantic information from long-term memory to recall words; however, children with WS used only a phonological recoding strategy to recall both high and low frequency words. Since long-term memory of individuals with WS is impaired, they rely more on short-term memory, which is comparable with the results observed so far.

Karmiloff-Smith, Grant, Berthoud, Davies, Howlin, & Udwin (1997) conducted an experiment of morphosyntactic study on individuals with WS. The results showed that there was a nonword repetition advantage. Individuals with WS performed with extremely high accuracy in repeating the nonce words invented by the experimenters when compared to normal people. This advantage in repeating novel words by WS individuals demonstrated an unusual verbal working memory ability. According to the authors, it implies that WS individuals seem to just encode the phonological form of the word, but not the meaning of the form. However, normal controls always got confused in repeating these nonce words. They often asked to know the meaning of the nonce words, but WS individuals didn't. Authors concluded that WS individuals have a deficit in morphosyntactic knowledge and they further claimed a different learning path of WS children from normal controls. In other words, WS individuals seem to rely on working memory in learning language as second language learners. This hypothesis seemed to partially explain the observation of a longitudinal comparison between WS and DS populations (Singer-Harris, Bellugi, Bates, Jones, and Rossen, 1997). Since WS and DS individuals are genetically disordered and developmentally delayed, when they begin to differ from each other in their language ability is interesting.

Singer-Harris et al. recruited fifty-four WS and thirty-nine DS individuals attending a test of MacArthur Communicative Development Inventory (CDI), a questionnaire for parents to report various questions about their children. The results showed that there is no significant difference reported in the onset of first words between WS and DS individuals. After the second year, these two populations showed a very different pattern in language development. WS children began to learn lots of vocabulary relative to their DS counterparts. WS parents also reported that their children could produce many words without understanding the meanings of words. On the other hand, DS parents reported that their children often have good comprehension but have difficulty in producing words. Singer-Harris et al. concluded that language ability of WS and DS children diverges with the improvement of grammar. As grammar emerges, WS population has relatively good language ability compared to DS individuals. Based on the findings of Karmiloff-Smith et al. and Singer-Harris et al., a possible explanation of WS population showing surprising linguistic ability in development given their mental retardation is because of their good verbal short-term memory.

Even though a reduced frequency effect was demonstrated on children with WS, verbal working memory is still relatively spared. Robinson, Mervis, and Robinson (2003) showed significant correlations across the board between working memory and grammatical ability on children with WS. They used forward digit span, backward digit span, and nonword repetition as verbal working memory indexes, and two inventories for grammatical ability: PPVT-R (Peabody Picture Vocabulary Test-Revised), which tested receptive vocabulary, and TROG (Test for the Reception of Grammar), which measured grammatical comprehension. All these measures were tested on children with WS and typically developing children with mental age matched. Partial correlations between the memory measures and the raw scores of each block in TROG were calculated. The results showed that, perhaps surprisingly, none of the memory measures were correlated with grammatical ability on typically developing children. On the contrary, all these measures were significantly correlated with the raw scores of TROG on children with WS. From these results, two conclusions were made: (1) due to a stronger relation between working memory and grammatical ability, children with WS seem to rely on working memory in learning language more than normal developing children, suggesting a high possibility in rote memorization of vocabulary; (2) the manipulation of items in working memory, rather than simple rote short-term storage of verbal items, is the key component in acquiring grammar. These conclusions were comparable with the observations in research on children with WS that their verbal IQ is often higher than mental age matched children, but their grammatical ability is much more delayed than in their mental age-matched counterparts.

The results of Robinson and colleagues (2003) are compatible with the findings of nonword repetition advantage in Karmiloff-Smith et al.'s study (1996) as well as the results from the longitudinal study on the development of first words on children with WS and children with DS in Singer-Harris et al.'s study (1997). Both nonword repetition and growth of vocabularies might be the results of rote memory on children with WS because of their spared verbal working memory. They can pronounce lexical items quite well, but do not exactly understand the meanings. In other words, due to the verbal working memory advantage of children with WS, it is possible that they dissociate form and meaning on lexical items in certain degree.

An anecdote described in a paper from Bellugi et al. (2000) yields some insight. A WS child said *I have to evacuate the glass* as she empties a glass of water (p.13).

She made an incorrect word choice to express the meaning, though they were in the right semantic field. This implies that WS could not use the appropriate word for right contexts. Is it possible that individuals with WS dissociate grammatical knowledge and meaning understanding on sentential level? In the production study of relative clauses (Zukowski, 2001) confirms the semantic problems on sentential level. In Object Gap Conditions, one can recall that WS children only produce 11 percent correct responses relative to 51 percent correct in normal controls. In overall responses, there was a 47 percent error rate for WS children relative to only 23 percent incorrect for normal children in responding to participant gap type relative clauses. That is, WS individuals produced fewer correct responses and made many more errors than normal controls. Among these errors, both of them produced a high percentage of argument mapping errors. For example, when a question "***which truck*** *turned red*?" is asked to the children, many of them replied that "***the girl*** *that's jumping over the truck turned red*" instead of the real target "***the truck*** *that the girl is jumping over turned red*". Similarly, when another question "***which car*** *is Max looking at?*" was asked to them, they responded that "Max is looking at ***the pigeon*** *that is flying over the car*" instead of the actual target "Max is looking at ***the car*** *that the pigeon is flying over*". WS children produced grammatical sentences, but they mistook the participant of the relative clause as the participant of the matrix clause or noun phrase. This error type is observed much more frequently in WS than in normal controls. Therefore, WS individuals seem to preserve quite good ability in building up surface structures, but reflect problematic semantics understanding. This dissociation is thus termed syntax-semantics mismatch or form-meaning dissociation in the following chapters.

There is another observation about syntax-semantics mismatch on sentential level of individuals with WS. Bellugi et al. (2000) tested WS children on counterfactual questions, a complex grammatical structure with inferences in logic. The experimenter asked participants counterfactual questions observing their responses and analyzing their producing structures on both grammatical structures and semantics. For instance, “What if you were a bird?” was uttered to participants with WS and DS. The results showed that individuals with WS performed preserved grammatical knowledge to counterfactual questions. For example, they responded *You could fly, you could have babies, fly north or south, east or west; I'd fly through the air being free; I'd fly through the air and soar like an airplane and dive through trees like a bird; I would fly where my parents could never find me, bird wants to be independent; I would fly and if I like a boy, I would land on his head and chirping*. They used correct subjunctive moods that were consistent with the counterfactual questions. On the contrary, individuals with DS produced ungrammatical structures in short and illogical ways like *Bird seeds; you'd be strong; I don't fly; fly in the air; I not a bird, you have wing*. Though individuals with WS performed much better than their genetic counterparts, which was similar to the controls, their semantics were not as good as their performance on grammatical structures compared to normal controls.

However, there have not been any studies that have directly and clearly tested this dissociation on individuals with WS. In the following research projects, the major theme is to investigate the hypothesis of form-meaning dissociation by using different stimuli with different modalities in Chinese.

CHAPTER II

MANUSCRIPT: THE LINGUISTIC ABILITY OF LOGICAL REASONING: EVIDENCE FROM COUNTERFACTUAL CONDITIONALS WITH NEGATION IN CHINESE CHILDREN WITH WILLIAMS SYNDROME

A Abstract

Individuals with Williams Syndrome are reported as having good verbal working memory relative to visual-spatial working memory (Wang and Bellugi, 1994; Jarrold, C., Baddeley, A.D., Hewes, A.K, 1999; Vicari, Carlesimo, Brizzolara, and Pezzini, 1996; Robinson, Mervis, and Robinson, 2003), and this verbal working memory advantage is highly correlated with grammatical abilities. For this reason, it has been hypothesized that individuals with Williams Syndrome rely heavily on verbal memory in learning language, which in turn, makes WS individuals have much better linguistic ability than their Downs Syndrome counterparts. If this hypothesis is correct, WS individuals may show selective impairment on form and meaning in their language ability. This hypothesis is compatible with parental reports that individuals with Williams Syndrome very often use inappropriate words in conversation, while the form remains grammatical.

Counterfactuals, which depict contrary-to-fact conditions, are appropriate stimuli for testing this hypothesis because they contain a mismatch between form and meaning. The constituent comparison model (CCM) is employed in this study (Carpenter and Just, 1975), which allows us to see what kind of representation is formed, namely, form-based representation or meaning-based representation. A sentence verification paradigm was conducted with two different stimuli-of-asynchrony (SOA): 0 seconds and 5 seconds. It is predicted that at longer

SOAs, a meaning-based representation will be formed and at short SOAs, a form-based representation will be formed. The results showed that at both SOA conditions college students formed meaning-based representations rather than form-based representations. Meanwhile, we also wanted to know at what age this ability to process complex counterfactual conditionals develops; thus we also did the same procedure with the eighth graders and the sixth graders.

The results showed that both the eighth graders and the sixth graders performed a meaning-based representation of counterfactual conditionals like college students, but the sixth graders showed a slightly different pattern from other two groups based on the hypothesized calculations of mental operations in CCM. With these results in hand, we finally performed the study on the population of interest, WS, and found that WS individuals showed the same pattern as the sixth graders. We conclude that WS individuals do not show a selective impairment on form and meaning in their linguistic ability. In other words, they are developmental delayed, but not deviant. At the same time, we demonstrate that counterfactual conditionals are indeed difficult sentence structures for which the ability to process develops from late childhood and into adulthood.

B Linguistic Relativity or Linguistic Universal

In 1981, Alfred Bloom conducted a serious of experiments on counterfactual conditionals both on English-speaking and Chinese-speaking adults. He hypothesizes that, since there are very clear linguistic markers in English denoting counterfactual realm (i.e. subjunctive mood), speakers of languages which lack such linguistic markers in counterfactual realm like Chinese would have difficulty in logical reasoning.

To test his hypothesis, stories written in English and in Chinese separately were tested in Taiwan, Hong Kong, and the U.S.. After reading this story, participants were required to answer a multiple choice comprehension question in which all the counterfactual implications were listed as alternatives for participants to make choice. If participants thought none of these answers were appropriate, they had to give reasons why they chose it. When this story was tested on American students and non-students, 54 out of 55 participants, namely 98%, chose the correct answer, suggesting a successful counterfactual inference. However, when the same story was distributed to Taiwanese students and non-students, only 2 out of 28 (7%) and 5 out of 75 (6%) responded correctly. The same finding was observed in Hong Kong students: 1 out of 17 (6%) responds correctly. Bloom ran the study again with a less complex version of the story to see if this would increase counterfactual responses in Chinese participants. The correct responses for Taiwanese students, Taiwanese non-students, and Hong Kong students did increase (63%, 46%, and 50%, respectively); however, their accuracy rates were still much lower than their American counterparts (96%). Thus, Bloom concluded that language structures had strong impact on thought processes like counterfactual reasoning.

In another set of experiments, Bloom also tested a single logic question on English-speaking and Chinese-speaking participants (see Bloom, 1981, p31). The logic question was something like: "If all circles were large and this small triangle '△' were a circle, would it be large? (假如所有的圓圈都很大，如果這個小三角形'△'是圓圈，那麼這個三角形是不是很大？)" 95 out of 115 American students (83%) accepted this premise and responded yes. Meanwhile, only 44 out of 173 Chinese students in Taiwan (25%) responded in the same way. Most of them in fact questioned the premise saying that "No! How can a triangle be a circle? How can this small circle be large? What do you mean?" Even if the concreteness of the counterfactual question increases, for example, "If all chairs were red and this table were a chair, would it be red?" the yes response did not go higher for Chinese-speaking participants. Bloom reports that he was told by his participants these questions were "un-Chinese" (see Bloom 1981, p.13). Bloom recounts an anecdote about a Taiwanese who overstayed his visa in Albany's court and had tremendous difficulty in understanding when the judge asked hypothetical questions like "If you weren't leaving tomorrow, you would be deportable" and also "If you have to be deported, where would you wish to be deported to?" (Bloom, 1981)

Bloom (1981, 1984) claims that linguistic markers on counterfactual conditionals in Chinese results in cognitive consequences are absent. The consequences reflect difficulties in developing cognitive schema specific to counterfactual thinking. In English, the subjunctive linguistic markers allow a direct linkage between language and the external world. That is, the subjunctive linguistic markers are like keys which can open the counterfactual box. However, because they don't have these keys, Chinese speakers would have an indirect linkage, or translation relationship, between language and the external world.

Bloom says that Chinese could reason counterfactually only in at-the-moment situations. Based on his observations, when a situation happened and a statement is described at the right moment, the statement can be interpreted as a counterfactual conditional. For example, if two persons were late in train station and saw the train being driven away, a statement like "(literally) If we not late, train then would not drive away ASP-complete)" (如果我們沒有遲到，火車就不會開走了, "If we hadn't been late, the train would not have been driven away") could be interpreted as a counterfactual conditional. Therefore, Bloom reformulates his hypotheses as "Do linguistic categories exclusively determine thoughts?" to "When and in what ways do linguistic categories shape thoughts?" and also "Do categorical differences across language necessarily entail cognitive differences correspondingly?" to "Which linguistic differences entail corresponding cognitive differences (Bloom, 1981)? Thus far, his theory is basically a weak version of Sapir-Whorf Hypothesis.

This hypothesis is consistent with Li and Thompson's analyses (1992). They claim that the Chinese conditional can only be understood in real situational context. For example, it is not surprising to hear a conversation including a sentence "(literally) if you see I sister, you certainly might know she become pregnant ASP-complete)" (如果你看到我妹妹，你一定知道她懷孕了) between two Chinese persons in which they could understand each other quite well, though this sentence may have three possible interpretations: (1) real situation---"If you see my sister, you will know that she is pregnant"; (2) hypothetical or imaginable situation---"If you saw my sister, you would know she is pregnant"; and (3) counterfactual situation---"If you had seen my sister, you would have known that she was pregnant" or "If you had seen my sister, you would know that she was pregnant". Therefore, for Chinese, how can a conditional statement be interpreted as a counterfactual one? According to Li and Thompson, the only way to disambiguate the three possible interpretations is based on the shared knowledge between speakers and listeners, namely, situational context. Therefore, Li & Thompson reach the same conclusion as Bloom does.

Au (1983) criticizes the story in Bloom's studies as too unidiomatic to understand for Chinese. She conducted another series of studies trying to replicate Bloom's findings on counterfactual conditionals in English and Chinese, using a different, more humanly relevant story. Au distributed this story written in English and in Chinese to native Chinese speakers in Hong Kong with 12 years experience studying English as a second language. The results showed that Au's participants responded counterfactually ranging from 96% to 100% in English version and also in the Chinese version, suggesting that the Chinese have little difficulty in counterfactual reasoning. Later, Au also added the original Bloom story to the study and found a

12% difference in Chinese but not English between the two stories, and argued that the difference could be due to the difference in the quality of the stories. When Au added a revised version of the Bloom story, she found no difference in Chinese (97% vs. 100%). Furthermore, the counterfactual response percentages in Au's studies for the Bloom story are higher than the response percentages observed in Bloom's studies (63% for Taiwanese students and 50% for Taiwanese non-students in Chinese version, and 52% for Taiwanese students and 86% for Taiwanese non-students in English version). Therefore, Au argues that Bloom's findings on counterfactual differences in English and Chinese resulted from the unidiomatic writing in the Chinese Bloom story, not from the difference in concreteness and abstractness. Au concludes that the absence of counterfactual linguistic markers cannot cause any difference in logical reasoning, advocating a viewpoint on linguistic universality.

Aimed at Au's findings, Bloom (1984) responds with two points. First, Au's participants have been contaminated in English for a long time (almost over eleven years). This learning English as a second language experience allows these participants to construct the English subjunctive and associate it to counterfactual schema without a problem. In other words, those participants had developed an English counterfactual schema into their Chinese world. Thus, it was not difficult for them to reason counterfactually. Second, Au's story is too concrete to elicit the difference between English and Chinese. Bloom claims that Chinese would have difficulty in counterfactual reasoning on abstract inferences, but not in concrete situations. Bloom replicated his studies by using the Bloom story in moderately complex version and also the highly complex version on native English speakers and native Chinese speakers. 97% English-speaking participants accepted counterfactual inference right away no matter which complex version of the story was received. However, only 53% out of 83 Taiwanese native speakers responded counterfactually to the moderately complex version, and even more dramatically, only 7% out of 103 Taiwanese native speakers responded counterfactually to the highly complex version. Further support comes from Bloom's report on a parallel but small sample study on Hong Kong students. The results showed that only 10 out of 20 (50%) and 1 out of 17 (59%) counterfactual responses to moderately complex version and the highly complex version, respectively. Therefore, Bloom concludes that the presence of linguistic markers on counterfactuals do help thinkers to reach the inferences.

Au (1984) retorts that Chinese can do counterfactual reasoning without knowing English subjunctive. For example, in one of her studies, she asked participants to translate a short paragraph containing counterfactual inferences from Chinese into English before reading the Bloom story. This short paragraph was like this: *Mrs. Wong doesn't know English. If Mrs. Wong knew English, she would be able to/could/would read English books.* Only 5 out of 43 (12%) native Chinese speakers in Hong Kong gave acceptable translations, suggesting that few participants know English subjunctive. However, 39 out of 43 (93%) native Chinese speakers responded counterfactually. Au concludes that Chinese can do logical reasoning no matter they learn English subjunctives or not. In other words, subjunctive markers are tangential to counterfactual thought.

Liu (1985) points out the possibility of "passive competence" on the part of Au's participants, which means that though the participants cannot write acceptable or correct translation with subjunctive mood, they might still be able to comprehend it through reading. Thus, Liu conducted a series of studies trying to reconcile the findings in Bloom and Au from two points. First, due to the incomparability in English proficiency of Bloom's Chinese participants in Taiwan and Au's Chinese participants in Hong Kong, Liu recruited students who haven't had experience in learning English as second language, namely, the fourth graders to sixth graders. She also recruited students that had different amounts of experience in learning English as a second language, the seventh graders to the eleventh graders, who are supposed to have learnt English from one to five years. Liu wanted to see whether students' performance would vary according to their English proficiency. Second, by recruiting students who did not understand English and those who were familiar with English, Liu also wanted to see whether concreteness or abstractness of the content influenced students' performance.

Second, since there were eight grades of students recruited in her study, she could also investigate whether counterfactual ability varies according to grades. Both Au and Bloom's story (in revised version) were tested on 521 participants in the end[1]. Each participant was required to comprehend these two stories and answer two questions. One was the counterfactual influence question and the other was a factual question[2] designed by Liu. By doing so, she claimed to make a basic distinction between these two stories to see whether these two stories were different even without requiring counterfactual reasoning. The results showed that the counterfactual responses of the fourth and fifth graders were significantly lower than the counterfactual responses of the sixth graders, suggesting a clear developmental trend.

Third, the counterfactual responses of the fourth to the seventh graders were significantly lower than the counterfactual responses of the eighth to the eleventh graders, suggesting another developmental trend. In other words, since there is no difference between the eighth graders to the eleventh graders' performance and also there is a very distinct difference between the fourth/fifth graders to the sixth graders

---

[1] Originally, there were 744 participants in total who participated in Liu's study (1985). However, according to Liu's criteria, there were 223 participants excluded from data analysis due to their inappropriate explanation to the correct answer (cf. Bloom and Au's criteria in data analysis) or the wrong answer.

(who do not have any experience in learning English as a second language), it is clear that participants' English proficiency cannot be a factor influencing counterfactual ability in reasoning. As to this finding, Liu confirms Au's conclusion with different populations.

However, Liu finds out an effect of concreteness and abstractness of the content. The counterfactual responses of the sixth graders to the eleventh graders are significantly higher towards Au's more concrete story than the counterfactual responses towards Bloom's more abstract story. Though the fourth and the fifth graders' counterfactual responses are not significantly different between these two stories, there is a trend suggesting a higher percentage in counterfactual response towards the more concrete story than the abstract story. Moreover, the percentage of incorrect answers for the factual question in the Bloom story (19%) was generally higher than the percentage in the Au story (4%). Therefore, Liu concludes that the concreteness or abstractness of the content is a possible factor influencing counterfactual reasoning. Furthermore, a clear age effect in counterfactual reasoning can be observed in this study. There are two distinct stages: from the fourth/fifth grade to the sixth grade and from the seventh grade to the eighth grade. After the eighth grade, there is not any difference between the counterfactual responses with the eleventh grade. Therefore, Liu suggests that competence in counterfactual reasoning is about age 14 (1985:251). To sum up, Liu concludes that years in learning English as a second language of participants cannot be a factor in counterfactual reasoning in Chinese that the content of the story is influential in response accuracy, and also that logical reasoning ability can be different according to age[3]. Meanwhile, the results that Bloom found only explain one thing, that is, that Chinese are prone to interpret

[3] In Au's study 5 report (1984), an age effect is also observed. There is a significant correlation between correct responses and participant's age (from the fourth grade to the seventh grade).

conditionals in the non-past, rather than in the past time frame.

C Counterfactual without Past Tense: A Brief Introduction

In English, tense and verb inflections make conditional interpretations very clear. Usually, there are three kinds of conditionals: reality, hypothetical unreality, and counterfactual unreality. If a present tense is put in an if-clause, then the conditional is used in situations of reality, as example (1) shows below. However, if a past tense is put in if-clause, then the conditional would be used in hypothetical situations as in example (2). Also, if a past tense is in if-clause and a combination of would + past participle is in consequence clause, the conditional would be used in counterfactual situation as example (3). Finally, if a past participle is put in an if-clause, the conditional is used in counterfactual situations as example (4) shows below. In other words, the correspondence between tense inflections on verbs and conditional interpretations is clear and direct.

I. Reality:

(1) If you heat water into 100 degree, it boils.

II. Unreality: hypothetical/imaginable

(2) If we moved, we could have a garden.

(3) If we moved, we could have had a garden.

III. Unreality: counterfactual

(4) If I had arrived by three, I would have met him.

However, in Chinese, the story is totally different. Here, there are also three possible interpretations of conditionals as in English. For example, if a sentence like example (5) is uttered, it refers to situations of reality. On the other hand, if a sentence like example (6) is uttered, it could be hypothetical or counterfactual. Since there is no tense and verb morphology at all, the interpretation of conditional is ambiguous in Chinese. Taking sentence (7) as an example, the same sentence can be interpreted to have three different readings. Thus, counterfactual thinking is argued to be context-dependent (Li and Thompson, 1992), which is based on the background knowledge shared between listeners and speakers. In other words, the counterfactual interpretation is possible only when the context is sufficient. However, in this study, it is going to be argued that without context, the counterfactual conditional reading is also possible.

(5) Reality:

Ex: 如果 你 踩 煞車， 車子 就 會 停 下來。

Ruguo ni cai shache chezi jiu hui ting xialai

If you step on brake car then will stop FM

‘If you step on the brake, the car will stop.’

Note: FM stands for functional marker, representing a continuous action.

(6) Unreality:

Ex: 如果 你 聽 我的 話， 就 不 會 吃 苦 了。

Ruguo ni ting wode hua jiu bu hui chi ku ASP-complete

If you listen my words then not will eat bitter ASP-complete

(Hypothetical) ‘If you listened to me, you would not suffer.’

(Counterfactual) ‘If you had listened to me, you would not have suffered.’

(7) Ex: 如果 你 看到 我 妹妹，你 一定 知道 她 懷孕 了

Ruguo ni kan-dao wo meimei, ni yiding zhidao ta huaiyun le.

if you see I sister, you definitely know she pregnant ASP

I. Reality

'If you see my sister, you will know that she is pregnant.'

II. Unreality: hypothetical/imaginable

'If you saw my sister, you would know that she is pregnant.'

II. Unreality: counterfactual

'If you had seen my sister, you would have known that she was pregnant.'

'If you had seen my sister, you would know that she was pregnant.'

'If you saw my sister, you would have known that she is pregnant'

According to Eifring (1988), there are many lexical items as conditional markers in Chinese like *jiaru* (假如), *ruguo* (如果), *jishi* (即使), *jiusuan* (就算). All these conditional markers are equivalent to English *if*. For example,

(ruguo-bushi)

(8) Ex: 如果他不是到美國去唸書，我就會天天看到他.

Ruguo ta bushi dao meiguo qu nienshu, wo jiou huei tientien kandao ta.

if he not arrive America go study, I then would everyday see he

'If he had not gone to America to study, I would have seen him everyday.

(ruguo-meiyou)

(9) Ex: 如果他沒有到美國去唸書，我就會天天看到他.

Ruguo ta meiyou dao meiguo qu nienshu, wo jiou huei tientien kandao ta.

if he not arrive America go study , I then would everyday see he

'If he had not gone to America to study, I would have seen him everyday.

(jishi-bushi)

(10) Ex: 即使他不是到美國去唸書，我也不會天天看到他.

Jishi ta bushi dao meiguo qu nienshu, wo ye buhuei tientien kandao ta.

if he not arrive America go study, I also not would everyday see he

'If he had not gone to America to study, I would not have seen him everyday.

(jusuan-bushi)

(11) Ex: 就算他不是到美國去唸書，我也不會天天看到他.

Jiusuan ta bushi dao meiguo qu nienshu, wo ye buhuei tientien kandao ta.

if he not arrive America go study, I also not would everyday see he

'If he had not gone to America to study, I would not have seen him everyday.

Other conditional markers exist in Chinese like *jiaru* (假如), *jiashi* (假使), *jiading* (假定), *jiashe* (假設), *jihuo* (即或), *jibien* (即便), *zungran* (縱然), *zungshi* (縱使), *zhiyao* (只要). In modern Chinese, all these conditional markers are used very frequently. Based on the frequency corpus of Academia Sinica, the frequency of all the conditional markers is listed in Table 5. Some have argued that the choice between them is determined by the likelihood of the events or premises described (Bloom, 1981). For example, if *jiaru* is used in conditional sentence, it is more unlikely that the event would happen. However, if *ruguo* is used, it is more neutral without committing the listener to have any presupposition on the described event. Probably this is one reason why their frequency is so different (341 vs. 2000).

Table 1 Frequency of Conditional Markers in Chinese

Counterfactual	Glossary	frequency	Counterfactual	Glossary	frequency

markers			markers		
如果	ruguo	2000	假使	jiashi	63
只要	zhiyao	1828	縱使	zungshi	59
即使	jishi	1043	即便	jibian	41
假如	jiaru	341	縱然	zungran	40
以為	yiwei	740	要不是	yaobushi	38
就算	jiusuan	231	假定	jiading	37
假設	jiashe	170	即或	jihuo	2

Though there are many conditional markers in Chinese, the steps in producing or comprehending conditional concepts are not different from other languages like English. The first step is to build up a possible world, which refers to the possibility of the event happening in the world. To build up this possible world, a conditional marker is usually put in sentence initial position. Different languages have different conditional markers, for example, *if* in English and many lexical items in Chinese as listed in Table 5. The second step is to differentiate time frame, which means determining when the described event would happen in a possible world. In English, it is very clear whether the sentence is intended for a reality situation as in (1), a hypothetical situation as in (2), or a counterfactual situation as in (3). This differentiation is easy because of tense inflections on verb morphology. However, in Chinese, as (7) listed, there are three possible interpretations toward the same sentence. Thus, the time differentiation in Chinese is very unclear, because there is no tense inflection on verbs at all. There is one solution to uniquely differentiate time: using temporal adverbs. If a temporal adverb is added in conditionals like *tomorrow*, *yesterday*, or *next year*, it would be very clear whether the conditional is in situations of reality, hypothetical, or counterfactual. However, is a counterfactual conditional reading possible without temporal adverbs?

Counterfactual conditionals have two clauses, namely, the if-clause and the consequence clause. The if-clause usually refers to a premise or a condition, the consequence clause refers a proposition to fulfill the premise. In a counterfactual situation, subjunctive mood is used to negate or falsify the premise described in the if-clause and to exclude the true condition of the consequence clause. For example, *If Hoover were now President, America would be at war* (Quine, 1980:21). Since Hoover is not a President now, America is not at war. Through the negation of the premise, the proposition followed is not true for now. Following this logic, though Chinese does not have any morphosyntactic items to represent this contrary-to-fact concept like subjunctive mood in English, Chinese does have negation markers to falsify a premise like *meiyou* (not). Thus, a prefix conditional *ruguo* (if) and a negation *meiyou* (not) in if-clause and an aspectual maker *le* (denoting an event mentioned already happened in the past) embedded in consequence clause thus coerce a counterfactual conditional interpretation in Chinese. For example,

(12) Ex: 如果我沒有遲到，車子就不會開走了.

Ruguo wo meiyou chidao, chezi jiou bu hui kai zou le.

If I not late car then not would drive away Asp-completed

'If I hadn't been late, the car would not have been driven away.'

Subjunctive conditionals are used only where the first component is definitely believed to be false (Quine, 1980). Similarly, a negation marker is used only when the premise is set up in unreality in Chinese. This assumption is related to the hypotheses we are interested in here.

First, can counterfactual reasoning in Chinese be determined without context? Are discourse cues necessary for Chinese to reason counterfactually? Can Chinese-speaking individuals show counterfactual reasoning without being heavily contextually biased? If all the discourse cues are taken away, are Chinese-speaking individuals still able to process counterfactually? In this study, we are interested in whether participants can do counterfactual reasoning without presenting any contextual information.

Second, does counterfactual reasoning ability differ developmentally? According to Liu's studies, if there are two distinction stages in counterfactual reasoning, why is the competence in logical reasoning achieved at age 14? Though, in Liu's findings, the counterfactual responses of the eighth graders are not different from the ninth, the tenth, the eleventh graders, isn't it possible that there is another distinction stage, for example from the twelfth graders to the freshmen/sophomores of the university? Therefore, three different age groups are recruited and investigated in their understanding of counterfactual conditionals: the sixth graders, the eighth graders, and college students.

Third, which representation of counterfactuals will be formed under time limitation (i.e. 0-SOA)? Based on Carpenter's study in English (1973), there are two possible representations for complex structures like counterfactual conditionals in processing, which are in correspondence with the presentation duration. Is the representation of a Chinese counterfactual conditional presented in a certain amount of time the same as the representation presented in a longer period? In other words, do Chinese-speaking individuals form these two types of representations? Is it possible that they will have residual difficulty in on-line logical reasoning due to the lack of subjunctive mood marking?

Fourth, and finally, which pattern will Williams Syndrome individuals show on a logical reasoning task like counterfactual conditionals? Is their pattern similar to their chorological age group or their mental age group? We expect that individuals with WS cannot do logical reasoning in accordance with their relatively deficient semantics.

To address these questions, Chinese counterfactual conditionals were employed in a series of reaction-time studies. These studies were conducted on three unimpaired groups across different ages. Meanwhile, Williams Syndrome individuals were recruited as a fourth group to examine whether or not they have spared logical reasoning ability.

D Form Representation or Meaning Representation

Carpenter (1973) conducted two experiments to investigate how people extracted linguistic information from complex structures like counterfactual conditionals. Since there is a mismatch between form and meaning of counterfactuals, it is hypothesized that two possible representations could be formed in processing: a representation based on grammatical structure and a representation based on meaning. In other words, when people process sentences like counterfactuals, they might form two representations mentally, a more complex representation and a simpler representation. If this is the case, it could be that there is an immediate structure-based processing of the sentence, but sufficient time is given, a simpler representation based on the meaning of a sentence would be constructed from its surface structure. For example, in English, an if-clause like *if John had died* may result in the formation of representations of [false (John, died)] or (John, lived) according to the structure complexity or the semantic content, respectively.

Carpenter's study focused on the question of whether a positive proposition embedded in a higher falsifying polarity or a simpler form with its true meaning was represented. A verification task (Clark and Chase, 1972; Carpenter and Just, 1975) was employed. Twelve participants were required to make a truth value judgment about whether the meaning of a test sentence was congruent with the meaning of a counterfactual target sentence. For example, there were two test sentences like *John lived* and *John died* for a target clause of a counterfactual sentence like *if John had died*. We hypothesized that the response latency would be shorter if the target and test representations matched, and that the response latency would be longer if a mismatch was encountered. Further, we hypothesized that for a short presentation, a complex representation would be constructed first which matched the grammatical structure of the target sentence. Later on, when sufficient time was available, a simpler representation would be constructed, which would necessarily mismatch the grammatical structure of the target sentence.

Based on this hypothesis, Carpenter manipulated different stimulus-of-asynchrony (SOA) to design two tasks: simultaneous task and delayed task. In the simultaneous task, the target sentence with subjunctive mood was presented at the same time as a test sentence on the computer screen. After a very short time (~2 s), the target sentence disappeared on the screen and only the test sentence remained. Participants were predicted to make their truth value judgment of the test sentence based on the representation of grammatical structure of the target sentence. In this situation, the reaction time of *John died* would be shorter than *John lived* due to the congruent mental representation of logical form in predicates between the test sentence and the target sentence (i.e. [*false (John, died*)]. In contrast, in the delayed task, the target sentence was presented on the computer screen first, and after 5 seconds the test sentence was presented. Participants were then predicted to make truth value judgment of the test sentence based on the representation of the simpler form. In this situation, the reaction time of *John lived* would be shorter than *John died* because the mental representation between the test sentence and the target sentence (i.e. [*John, lived*]) was congruent. The basic logic underlying both experiments was that once the representation was congruent in comparison, a faster mental operation would be executed and a shorter reaction time would be observed.

Two factors were manipulated on the target sentence: sentence type (i.e. factual vs. counterfactual sentences), and sequence of the sentence type (i.e. if-clause vs. consequence clause). Thus, there were four types of target sentences included: (1) factual-factual target sentence, e.g. *Mary stayed, since Judy live,* where the real interpretation of this example would be [(Mary, stayed)] and [(Judy, lived)]; (2) factual-counterfactual target sentence, e.g., *Mary stayed, since Judy would have lived,* where the interpretation should be [[(Mary, stayed)] and [(Judy, died)]; (3) counterfactual-factual target sentence, e.g. *Mary would have stayed, but Judy lived,* where the interpretation should be [(Mary, left)] and [(Judy, lived)]; (4) counterfactual-counterfactual target sentence, e.g. *Mary would have stayed, if Judy had lived,* where the interpretation should be [(Mary, left)] and [(Judy, died)]. For each target sentence, there was a following test sentence intended for truth value judgment. All the test sentences were affirmative and differed on truth value, which meant whether a test sentence matched or mismatched the representation of a target sentence. For example, a test sentence like *Mary stayed* was false for a target sentence *Mary would have stayed, if Judy had lived*; and *Judy died* was true for the same target sentence.

There were four test sentences for each target sentence. Among these four, two (one true and one false) were testing for if-clause and another two (one true and one false) were testing for consequence clause. For factual clauses, true test sentences were predicted to have faster response latency than false test sentences. The same predictions applied for factual sentences in both simultaneous task and delayed task because the representations did not change according to different SOA. However, the predictions were different for counterfactual target sentences in tasks due to the possible representations of counterfactual conditionals based on grammatical structure (i.e. form representation) or meaning representation. In the simultaneous task, while the SOA was zero, the representation would be congruent with the complex form of grammatical structure in predicate. As the chart of detailed mental operations in Table 1 shows, the condition with false responses was predicted to be faster than the condition with true responses because the FA condition had fewer mental operations than the TA condition (if mental operations are defined as how many stages participants would go through in verification processing, the minimum requirement in this case was 4, which was represented in letter K). However, the prediction reversed in delayed task. When the SOA was 5 seconds, the representation was predicted to be congruent with the simpler form as in the representations of factual sentences. Another chart with detailed mental operations is displayed in Table 2. The condition with true responses was expected to have faster response latency than the condition with false responses because the TA condition involved fewer mental operations than the FA condition (in this case, the minimum mental operations were 2).

The results confirmed Carpenter's predictions. To sum up, for factual target sentences in both tasks, true test sentences were faster than false test sentences. In contrast, for counterfactual target sentences, in the simultaneous task, true test sentences were slower than false test sentences. Meanwhile, the results reversed in delayed task. These results confirmed the prediction of congruence in comparison between target sentences and test sentences, and indicated that under time pressure, a complex representation would be formed based on grammatical structure; however, after sufficient time was presented, a simpler representation was constructed based on meaning.

Table 2 Representations and Mental Operations for the Affirmative Conditions in Counterfactual Sentences in Simultaneous Task (complex form)

Stimulus and representation	**True Affirmative (TA)**	**False Affirmative (FA)**
Target sentence	Mary would have stayed.	Mary would have stayed.
Test sentence	Mary left.	Mary stayed.
Target sentence rep.	[NEG, (stayed, M)]	[NEG, (stayed, M)]
Test sentence rep.	(left, M)	(stayed, M)
	– ☒	– + ☒
	– + ☑	+ +
index = false ☒	+ +	
index = true ☑		
	response = true	response = false
	K + 1 comparisons	K comparisons (K=4)

TSING HUA UNIVERSITY

Table 3 Representations and Mental Operations for the Affirmative Conditions in Counterfactual Sentences in Delayed Task (simple form)

Stimulus and representation	**True Affirmative (TA)**	**False Affirmative (FA)**
Target sentence	Mary would have stayed.	Mary would have stayed.
Test sentence	Mary left.	Mary stayed.
Target sentence rep.	(left, M)	(left, M)
Test sentence rep.	(left, M)	(stayed, M)
	+	– ☒
		+
index = false ☒		
index = true ☑	response = true	response = false
	K comparisons (K=1)	K + 1 comparisons

The logic of the studies that I conducted was similarly based on the multiple stages model of sentence processing, or constituent comparison model (CCM), which was proposed by Clark and Chase (1972) and modified by Carpenter and Just (1975). According to this model, several stages can be identified in verification between sentences. The first stage is to form a representation of the target sentence, the second stage is to form a representation of the test sentence, the third stage is to compare these two representations in mind, and the fourth stage is to judge whether these two representations are the same or different. In this model, mental operations go from the inner propositions to the outermost proposition when two representations are compared. The more congruent the representations are, the fewer operations are necessary for processing. The fewer the mental operations executed, the shorter the reaction times should be.

In this model, the logical form of the sentence is hypothesized to have the structure of (predicate, argument), which is viewed as a unit in comparison. Later on, when the comparison moves to the left hand side, the embedding negation marker is viewed as another unit in comparison. If two units are compared, a sign is assigned to mark the mental operation. The sign could be positive, which means that the two representations match with each other, or the sign could be negative, which means that the two representations are mismatched with each other. Once a mismatch is encountered, the comparison restarts from the beginning. Of course, the comparison will not be an endless loop. When the comparison gets a negative sign, it becomes positive in the second comparison. Meanwhile, the truth value index will change its default true value to a false value.

In Table 3, true affirmative sentences require fewer mental operations than false affirmative sentences. A hypothetical letter K is assigned for the least condition, and other compared stages are added if they are more than K. In this case, when the test sentences are true affirmatives, the comparison is the least because the number of mental operations involved is the fewest, K (K equals to 1, in this case). Test sentences with false affirmatives, due to one mismatch in comparison of predicates between the target sentence and the test sentence, require one more mental operation, thus, the number of operations is K + 1. In Table 4, we see that true negative sentences require one more comparison stage than false negative sentences because of the predicate mismatch. Thus, the truth value index changes twice. Compared to the affirmative sentences shown in Table 3, negative sentences involve more mental operations because the mismatch is in the embedding polarity, rather than in the embedded predicates. Once a mismatch is found in the embedding position, the comparison restarts from the beginning. So, more mental operations are required in test sentences with negation. The hypothesized representations and mental operations for the factual target sentences and test sentences with affirmatives are given in Table 3 and test sentences with negatives are given in Table 4 below. Thus, the ordering of test conditions in the end for factual target clauses is TA < FA < FN < TN according to the number of mental operations.

Table 4 Representations and Mental Operations for the Affirmative Conditions in Factual Sentences

Stimulus and Representation	**True Affirmative (TA)**	**False Affirmative (FA)**
Target sentence	I was late.	I was late.
Test sentence	I was late.	I was on time.
Target sentence rep.	(late, I)	(late, I)
Test sentence rep.	(late, I)	(on time, I)
	+ ☑	– ☒
		+
index = false ☒	response = true	response = false
index = true ☑	K comparisons	K + 1 comparisons
	(K=1)	
	我遲到了,........	我遲到了,..........
	我遲到了,........	我準時到,..........

Table 5　Representations and Mental Operations for the Negative Conditions in Factual Sentences

Stimulus and representation	**True Negative (TN)**	**False Negative (FN)**
Target sentence	I was late.	I was late.
Test sentence	I was not on time.	I was not late.
Target sentence rep.	(late, I)	(late, I)
Test sentence rep.	[NEG, (on time, I)]	[NEG, (late, I)]
	– ☒	– + ☒
	– + ☑	+ +
index = false ☒	+ +	
index = true ☑		
	response = true	response = false
	K + 4 comparisons	K + 3 comparisons
	我遲到了,…………….	我遲到了,………….
	我沒有準時到,……….	我沒有遲到,………

For target sentences with counterfactuals, three possible mental representations could be formed. Each of the representations may result in a different ordering of test conditions because the number of mental operations involved is different. Two of the representations are formed based on the grammatical structures of counterfactual sentences, which are called ‘complex form one’ and ‘complex form two’, respectively. The other representation is formed based on the exact meaning, which is called ‘simple form’. For the complex form one representation, it is hypothesized that a conditional marker *if* is represented as an embedding marker, accompanying with a negative marker as another embedding marker. Under this representation, test sentences with false negatives have the least comparison stages, K (K equals to 6, in this case). Test sentences with true negatives have one more mental operation involved due to the mismatch of the predicates, thus, K + 1 comparisons are needed. Test sentences with true affirmatives are harder because they need K + 2 mental operations. Last, test sentences with false affirmatives are predicted to be the hardest condition because three mismatches exist: predicate, polarity, and conditional marker. Thus, these sentences require the most mental operations, K + 3. The truth value index changes three times before the response. A detailed representation and the mental operations required in comparison between counterfactual target sentences and test sentences with affirmatives are given in Table 5 and test sentences with negatives are given in Table 6 below. The ordering of test conditions is FN < TN < TA < FA in the representation of complex form one.

TSING HUA UNIVERSITY

Table 6 Representations and Mental Operations for the Affirmative Conditions in Counterfactuals Sentences in Simultaneous Task (complex form one)

Stimulus and representation	**True Affirmative (TA)**	**False Affirmative (FA)**
Target sentence	If I hadn't been late.	If I hadn't been late.
Test sentence	I was late.	I was on time.
Target sentence rep.	{IF, [NEG, (late, I)]}	{IF, [Neg, (late, I)]}
Test sentence rep.	(late, I)	(on time, I)
index = false ☒	– + ☒	– ☒
index = true ☑	– + + ☑	– + ☑
	+ + +	– + + ☒
		+ + +
	response = true	response = false
	K + 2 comparisons	K + 3 comparisons
	如果我沒有遲到,………	如果我沒有遲到,……….
	我遲到了,………………	我準時到,……………….

Table 7　Representations and Mental Operations for the Negative Conditions in Counterfactuals Sentences in Simultaneous Task (complex form one)

Stimulus and representation	**True Negative (TN)**	**False Negative (FN)**
Target sentence Test sentence Target sentence rep. Test sentence rep. index = false ☒ index = true ☑	If I hadn’t been late. I was not on time. {IF, [Neg, (late, I)]} [Neg, (on time, I)] – ☒ – + + ☑ + + + response = true K+1 comparisons	If I hadn’t been late. I was not late. {IF, [Neg, (late,I)]} [Neg, (late,I)] – + + ☒ + + + response = false K comparisons (K = 6)
	如果我沒有遲到,……… 我沒有準時到,…………	如果我沒有遲到,……. 我沒有遲到,…………..

On the other hand, if the representation is complex form two, it is hypothesized that only a negative marker is represented as an embedding marker, accompanied by a flipped predicate that has the opposite meaning to the original predicate in counterfactuals. In this case, the predicate *late* is mentally flipped to *on time*. Under this representation, test sentences with true affirmatives are the easiest condition in verification between target sentence and test sentence because of the match in both predicates and negation markers. Test sentences with false negatives are the second easiest condition because of the match in negation. Further, test sentences with affirmatives are the harder conditions due to the mismatch of negation (both TA and FA) and also of predicate (TA). Thus, the ordering of test conditions in complex form two is TN < FN < FA < TA, which is different from the ordering of complex form one. However, they have one point in common, that is, negations triumph affirmatives in both orderings.

Table 8 Representations and Mental Operations for the Affirmative Conditions in Counterfactuals Sentences in Simultaneous Task (complex form two)

Stimulus and representation	**True Affirmative (TA)**	**False Affirmative (FA)**
Target sentence	If I hadn't been late.	If I hadn't been late.
Test sentence	I was late.	I was on time.
Target sentence rep.	[NEG, (on time, I)]	[Neg, (on time, I)]
Test sentence rep.	(late, I)	(on time, I)
index = false ☒	– ☒	– + ☒
index = true ☑	– + ☑	+ +
	+ +	
	response = true	response = false
	K + 3 comparisons	K + 2 comparisons

	如果我沒有遲到,……… 我遲到了,………………	如果我沒有遲到,………. 我準時到,……………….

Table 9 Representations and Mental Operations for the Negative Conditions in Counterfactuals Sentences in Simultaneous Task (complex form two)

Stimulus and representation	**True Negative (TN)**	**False Negative (FN)**
Target sentence	If I hadn’t been late.	If I hadn’t been late.
Test sentence	I was not on time.	I was not late.
Target sentence rep.	[Neg, (on time, I)]}	[Neg, (on time, I)]
Test sentence rep.	[Neg, (on time, I)]	[Neg, (late, I)]
		– ☒
	+ + ☑	+ +
index = false ☒		
index = true ☑	response = true	response = false
	K comparisons (K = 2)	K + 1 comparisons
	如果我沒有遲到,………	如果我沒有遲到,…….
	我沒有準時到,…………	我沒有遲到,…………..

The last and also the only possible mental representation of counterfactual conditionals is based on exact meaning rather than grammatical structures. Under this representation, there is neither conditional marker nor negative marker as embedding markers. Further, there is no mental flip in representation. In other words, the representation of counterfactuals is just like the representation of factuals. Under this representation, the ordering of test conditions is TA < FA < FN < TN as Table 3 and Table 4 displayed.

An interesting question, then, is which representation of counterfactuals will be formed in working memory in Chinese. Will Chinese speakers form a mental representation based on the grammatical structures when participant to time limitation, as English speakers seem to? Or will Chinese have an unchanged representation in form or meaning no matter whether they get a longer SOA or shorter SOA? Due to the involvement of logical reasoning in counterfactual understanding, we might think that mental representation of counterfactuals would differ as a function of age. Moreover, since individuals with Williams Syndrome might have a selective impairment on form and meaning, as section B introduced, we hypothesize that this genetically disordered population might represent counterfactuals in complex form one or two instead of in a simple form representation. In order to answer these questions, two tasks, incorporated within seven experiments, were employed on three different developmental ages and also on genetically disordered individuals with WS.

As section D mentioned, in Chinese, there is no tense inflection on verbs in denoting subjunctive mood. Thus, a negation marker *not* (沒有) was added in sentences to form a counterfactual conditional, which is parallel to the meaning of negating or falsifying the premise raised in if-clause. Similarly, two conditions (i.e. TA and FA) were tested in counterfactual understanding. If Carpenter's congruency hypothesis is correct, the condition with TA should be responded faster than condition with FA. For example (repeated (12) below),

(13) Ex: 如果我沒有遲到，車子就不會開走了.

Ruguo wo meiyou chidao, chezi jiou bu hui kai zou le.

If I not late car then not would drive away Asp-completed

'If I hadn't been late, the car would not have been driven away.'

Test sentence (TA condition): 我 遲到 了

Wo chidao le

I   *late*   *Asp-completed*

'I was late'

Test sentence (FA condition): 我 準時 到

Wo zhunshi dao

I   *on time*   *arrive*

'I was on time'

However, there is a confound. While we hypothesized that the test condition with TA should demonstrate a faster response latency than the test condition with FA because of predicate match, the comparison basis between these two conditions is in fact different. For condition TA, one mental flip occurs, from affirmative to negative (i.e. 遲到 to 沒有遲到); but for condition FA, two mental flips on both predicate and polarity occur, namely, from 遲到 to 準時到 and from affirmative to negative. In other words, this comparison is beyond the congruency hypothesis, but involves a polarity match. In order to keep the comparison fair and clear, another factor, polarity, was added in the test sentences. Therefore, four test sentences for each target clause were designed. Each of the four test sentences matched with the target sentences in different degree: condition TA (i.e. [(late, I)], 我遲到了) matched with the target clause in predicate, condition TN (i.e. [NEG, (on time, I)], 我沒有準時到) matched with negation, condition FA (i.e. [(on time, I)], 我準時到) matched with none of the representation of target clause, and condition FN (i.e. [(NEG, (late, I)], 我沒有遲到) matched with both predicate and negation. In addition, since the the four conditions match with respect to different parts of the counterfactual representation, we can examine the interesting question of which part is the most important cue in match, namely, predicate, negation, predicate and negation, or none. In this study, when the test sentences completely match with the complex representation in both predicate and polarity, or when the test sentences do not match any part of the complex representation, they are counted as testing the congruence of form. In contrast, when the test sentences partially matched with the complex representation in either predicate or polarity, they are counted as testing the congruence of meaning. That is, when test sentences are testing the match of predicate forms, the truth value of them is false while when test sentences are testing the match of predicate meanings, the truth value is true.

E Language and Thought Experiment I: Simultaneous Task of Counterfactual Conditionals with Negation

Participants: College Students

Twenty-two college students from National Tsing Hua University were included (mean age = 19.8, range from 18 to 23, 12 females and 10 males). All participants participated for course credit in Introductory Linguistics. They were right-handed users (except one participant) and none of them were reported to have medical problems. They were native Chinese speakers and had been studying English for more than six years, since the first grade of the eighth school (i.e. the seventh grade).

Participants: The Eighth Graders

Twenty-nine[4] students in the second grade (i.e. the eighth grade in the States) of Fu He The eighth School participated in this study (mean age = 14, range from 13 to 15, 16 females and 13 males). They were rewarded with a present after finishing the study. All participants were right-handed users and none of them were reported to

[4] Twenty Nine participants' data were analyzed in the end (rather than thirty) because one participant's accuracy was too low (13 out of 128 trials). This low accuracy may result from reversed button clicking. As the procedure mentions, participants were instructed to click left button for true responses and right button for false responses. It is highly possible that this participant misunderstood left and right buttons. Thus, his data was thrown away.

have having medical problems. They were native Chinese speakers and had been studying English for more than two years, since the first grade of the eighth school.

Participant: The Sixth Graders

Twenty-one[5] students in the sixth grade of Qing Jiang The sixth School participated in this study (mean age = 12, range from 11 to 13, 16 females and 5males). They were rewarded with a present after finishing the study. All participants were right-handed users and none of them were reported having medical problems. They were native Chinese speakers and had not had any experience in studying English[6].

Design

A verification task was employed which was basically parallel to Carpenter's (1973) paradigm. Two independent variables were designed for target sentences: sentence type, which meant whether it was a factual or a counterfactual sentence; and sequence of sentence type, which meant whether the tested sentence was the first clause or the second clause. Thus, there were four types of target sentences in this study: (1) factual-factual target sentence (FF); (2) factual-counterfactual target sentence (FC); (3) counterfactual-factual target sentence (CF); and (4) counterfactual-counterfactual target sentence (CC). There were four different sentences with different scenarios, which were related to real situations in daily life. For example, to take an example from CC, a sentence like *If I hadn't been late, the car would not have been driven away* was designed as a target sentence. Another two

[5] Because three participants had very low accuracy, their data were not included in analysis (17 out of 128 trials, 28 out of 128 trials, and 62 out of 128 trials). It is possible that they misunderstood the left and right buttons as the the eighth school student reported just now. The alternative would be that these participants did not understand the task. Thus, there were 21 participants' data in the final analysis.

[6] It is quite hard to say for sure whether these sixth grade students have been studying English or not. In Taiwan, the sixth schools have started teaching English as a second language since the third grade. If this was so for these students, then they have studied English for three years, though only for one hour a week.

independent variables were designed for test sentences: truth value, which referred to the match or mismatch between a target sentence and a test sentence, and polarity, which referred to the sentence polarity—whether it was affirmative or negative. Thus, there were four types of test sentences: (1) true affirmative (TA) like *I was late*; (2) false affirmative (FA) like *I was not late*; (3) true negative (TN) like I was not on time; (4) false negative (FN) like *I was on time*. In the end, we had a 2 (factual vs. counterfactual) x 2 (clause one vs. clause two) x 4 (sentence scenario) x 2 (true vs. false) x 2 (affirmative vs. negative) x 2 (probe clause one vs. probe clause two) design. There were 128 experimental trials in this study. There were 64 fillers included. For each filler target sentence, there were eight test sentences for it (four for clause one and four for clause two). Participants were required to judge whether the test sentence was true or false based on the truth condition presupposed of the target sentence.

TSING HUA UNIVERSITY

Materials

As was mentioned in section two regarding counterfactual logic in Chinese, it has been argued that Chinese does have counterfactual conditionals. The expression of the counterfactual is not on morpho-syntactic morphemes like those for subjunctive moods, but on lexical items like the coercion of conditional marker *if* (ruguo), the negation *not* (meiyou), and sentence-final completed aspect (le). Through the combination of these three lexical items, a counterfactual conditional interpretation is possible in Chinese. Under this coercion, it is very clear that the agent (I) in this sentence was late and the car was not there (i.e. *If I hadn't been late, the car would not have been driven away*). There were four types of these kinds of coerced sentences in this study and all of them were based on a natural relationship in daily life like patient and doctor, typhoon and airport, cat and mice. These four target sentences are listed in Appendix 1, Appendix 3, Appendix 5, and Appendix 7, respectively.

For each target sentence, there were four types of test sentences which probed participants' understanding of the target sentences. To take the sentence in (1) as an example, two test sentences with affirmative polarity were presented and each of them had different truth values, namely, one for true and one for false. For example, *I was late* was the test sentence with a true value in the affirmative and *I was on time* was the test sentence with a false value in the affirmative. Another two test sentences with negative polarity were presented to participants and each of them also had different truth values. For example, *I was not on time* was the test sentence with a true value in the negative, and *I was not late* was the test sentence with a false value in the negative. Meanwhile, the probed test sentences were not only for clause one (if-clause), but also for clause two (consequence clause). For example, *the car was driven away* (true value in affirmative), *the car was there* (false value in affirmative), *the car was not there* (true value with negative), *the car was not driven away* (false value with negative). All the test sentences for clause one and clause two of each target sentence are listed in Appendix 2, Appendix 4, Appendix 6, and Appendix 8.

Fillers were related to experimental trials, but the meanings were opposite. For example, while the experimental trial was describing the relation between being late and the car's leaving, the filler would be *I was on time, and the car was still there* (我準時到了，車子還在原地) and a different proposition was also included like *I was not late, the car was still there* (我沒有遲到，車子也還在原地). There were two fillers for each experimental target sentence. For each filler target sentence, there were eight test sentences (i.e. four probing the if-clause and four probing the consequence clause), which were the same as the test sentences for corresponding experimental target sentence (cf. Appendixes 2, 4, 6, 8). All the filler sentences were listed in Appendix 9.

Procedure

A fixation point was presented on the computer screen for 500 ms. After the fixation, a target sentence and a test sentence were displayed on the screen simultaneously. After 2000 ms, the target sentence disappeared and the test sentence remained on the screen. Only after the judgment was made did the test sentence disappear. Participants were required to judge the truth value of the test sentence based on comprehension of the target sentence. If the test sentence was consistent with the target sentence, participants were instructed to press the left button of the mouse as soon as possible. If the test sentence was not consistent with the target, they were to press the right button of the mouse immediately. There were three blocks in this study, and each of them contained 64 trials. Between each section, there was a break. All participants did 10 practice trials first to confirm their understanding of the task. Three random lists were assigned to participants, which were counterbalanced. All college students were tested in a quiet laboratory of National Tsing Hua University, and at most three were tested at the same time. All the eighth graders and the sixth graders were tested in the computer rooms of their own schools, which were also very quiet.

Predictions

For factual target clauses, according to the constituent comparison model and Carpenter’s (1973) findings, the test sentences with different truth values and polarities should show the following ordering in processing from the easiest to the hardest: true affirmatives (TA), false affirmatives (FA), false negatives (FN), and true negatives (TN). In other words, the condition ordering would be like this: TA < FA < FN < TN. If this ordering was expressed in terms of number of mental operations involved (K), it thus would be: K, K+1, K+4, K+5. Based on the calculation of CCM, in this case, K equaled to two.

For counterfactual target clauses, the predictions are task dependent. Due to the possible different representations between target clauses and test sentences, we hypothesized that in the simultaneous task condition (i.e. 0-SOA), a more complex representation would be formed containing a conditional marker and also a negation marker (i.e. {IF, [NEG, (late, I)]}). In this situation, since test sentences with negatives (i.e. FN and TN) either completely matched the representations of target clauses or partially matched (with the negative polarity), the response latency was predicted to be shorter. Meanwhile, test sentences with affirmatives (i.e. FA and TA) completely mismatched or partially matched with the predicate of the representation of target clauses; therefore the response latency was predicted to be longer, because more mental operations would be involved. In this situation, test sentences with different truth values and polarities were predicted to show the following ordering in processing from the easiest to the hardest: false negatives (FN), true negatives (TN), true affirmatives (TA), and false affirmatives (FA). In other words, the condition ordering is: FN < TN < TA < FA. If this ordering was expressed in terms of numbers of mental operations involved (K), it thus would be: K, K+1, K+3, K+4. Therefore, it was expected that Carpenter's congruency hypothesis would be observed: Test sentences with true affirmatives (TA) should be responded to faster than test sentences with false affirmatives (FA) because test sentences with TA are the matched condition. Of course, another possibility in the representation formed also exists: alternative two, in which only the negation marker would be formed and the predicate is flipped (e.g. [NEG, (on time, I)]). However, we think that alternative one is more likely than alternative two. Alternative three, which does not contain any embedding markers like conditional or negation (e.g. (late, I)), is also a possible representation.

Compared with the English counterfactual design, the Chinese design is more complicated to interpret because three representations could be formed in processing counterfactual conditionals rather than only one. If we take a CC target sentence as an example, *If I had not been late, the car would not have been driven away* (如果我沒有遲到，車子就不會開走了), one of the representations might be {IF, [NEG, (late, I)]}, which retains the conditional marker and the negation marker. In this situation, the ordering of the four test conditions is FN < TN < TA < FA, as in Table 5 and Table 6. The second possible representation is [NEG, (on time, I)], which flips the predicate into the opposite one. In this situation, the ordering of the four test conditions would be TN < FN < FA < TA. The third possible representation is [(late, I)], which is without any marker. In this situation, the ordering of these four test conditions would be TA < FA < FN < TN.

Which representation is the one people generally form in processing Chinese counterfactual conditionals? Parallel to Carpenter's study on English counterfactuals (1973), we expect to see in this study on Chinese counterfactuals that test sentences with true affirmatives would be responded to faster than test sentences with false affirmatives when the SOA is 5 seconds. When the SOA is 0, if the TA condition is responded faster than the FA condition, it would suggest that either the first or the third representations are formed. However, if the ordering of these two test conditions are reversed (i.e. FA < TA), then it will be easy to tell that the second representation has been formed.

Results: College Students Data

The latencies and error rates to respond to factual and counterfactual clauses are shown in Table 10 below.

Counterfactual vs. factual target clauses

Mean response times to counterfactual target clauses and factual target clauses were 4848ms and 4360ms, respectively. These two response latencies were significantly different ($F (1, 2505) = 72.67$, $p < .0001$). Meanwhile, participants made more errors on test sentences probing counterfactual target clauses than test sentences probing factual target clauses ($F (1, 2526) = 968.90$, $p < .0001$).

For counterfactual target clauses, there was also a significant interaction between truth values and polarities ($F (1, 1198) = 24.29$, $p < .0001$).

For factual target clauses, there was an interaction between truth values and polarities which reached significance ($F (1, 1280) = 14.59$, $p < .0001$). The same pattern was found for error rates ($F (1, 1301) = 10.64$, $p = .0011$).

The same pattern was found on error rates ($F (1, 1219) = 17.30$, $p < .0001$). The interaction of the four conditions on counterfactual and factual target clauses was not significant ($F (3, 2499) = 1.18$, $p = .31$), suggesting that the response latency of factual target clauses was not in general higher than counterfactual target clauses. The main effect of target clause was significant ($F (1, 2499) = 78.06$, $p < .0001$) and the main effect of four conditions was also significant ($F (3, 2499) = 27.01$, $p < .0001$). The difference between counterfactual target clauses and factual target clauses was highly significant within each condition, implying faster response latency in general for counterfactual clauses compared to factual clauses.

The interaction of error rates of four conditions on both target clauses was also not significant ($F (3, 2520) = 0.07$, $p = .97$). The main effect of error rates on target clauses was significant ($F (1, 2520) = 965.21$, $p < .0001$), but the main effect of error rates on four conditions was not significant ($F (1, 2520) = 0.20$, $p = .89$). The difference on error rates between counterfactual and factual target clauses was reliably significant ($p < .0001$), implying that college students erred more on counterfactual target clauses than on factual target clauses.

Counterfactual vs. factual clause position

Response latency to counterfactual target clauses in the if-clause was 4835ms and in the consequence clause the latency was 4880ms. This latency difference did not reach significance (F (1, 1220) = 0.12, p = .72), suggesting that clause positions for counterfactual targets did not influence in processing. The same pattern was found in their error rates (F (1, 1221) = 0.59, p = .44), suggesting that participants did not make more errors based on clause position. One possibility is that college students felt counterfactual target clauses were already very difficult, and it was because of this 'ceiling' of difficulty that no clause position effect was found. In contrast, since factual target clauses were easier, there was more space for clause positions to cause difference in processing.

Response latency to factual target clauses in if-clause was just 4225ms and in the consequence clause the latency was 4514ms. This latency difference was significant (F (1, 1302) = 13.71, p = .0002), implying that clause positions for factual targets did have an influence in processing. Their error rates did not show any difference (F (1, 1303) = 0.00, p = .97), suggesting that participants did not make more errors because of the clause positions.

Counterfactual vs. factual sentence type

Response latencies for counterfactual target clauses in four experimental sentence types were 4713ms, 5009ms, 4835ms, and 4880ms, respectively. A one-way ANOVA did not show significant difference among the four conditions (F (3, 1218) = 1.39, p = .24), suggesting that none of the counterfactual target clauses in these four experimental sentences (i.e. FF, FC, CF, CC) caused more difficulty than others. The same pattern was found on their error rates in one-way ANOVA (F (3, 1219) = 0.53, p = .66).

Response latencies for factual target clauses in four experimental sentence types were 4215ms, 4425ms, 4343ms, and 4492ms, respectively. A one-way ANOVA results did not show significant difference among them (F (3, 1300) = 2.32, p = .07), suggesting that none of the factual target clauses in these four experimental sentences (i.e. FF, FC, CF, CC) caused more difficulty than others. They were all similar in degree of difficulty. The same pattern was found in error rates in a one-way ANOVA (F (3, 1301) = 0.10, p = .95).

Test sentences in counterfactual targets

For counterfactual target sentences at 0 SOA, the condition ordering was the following: TA < FN < FA < TN, which is slightly matched with alternative three: TA condition was the easiest, TN condition was the hardest, and FA/FN conditions were in between. The results were parallel to the results observed to factual target clauses, suggesting that participants had formed a simpler representation for counterfactual clauses. Thus, we infer that under time limitation in presentation, college participants could still form a representation based on the sentence's meaning rather than its superficial structure.

Participants responded fastest to test sentences with true value in affirmatives (4407ms), next was to test sentences with false value in negatives (4901ms), next was to test sentences with false value in affirmatives (5005ms), and the last was to test sentences with true value in negatives (5188ms). A one-way ANOVA showed that the difference between these four conditions was significant (F (3, 1198) = 14.91, p < .0001). A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. Among all the comparisons, almost all the comparisons were significant, but two comparisons weren't (FA vs. FN, p = .46; FA vs. TN, p = .15). The difference between TA and other groups was significant at the $p < .0001$ level and the difference between FN and TN also was significant, $p = .03$. Though the ordering showed that the matched condition (FN) for representations of counterfactual target clauses had faster response latency than mismatched condition (FA) in 104ms, the difference was still not significant, suggesting an indifference influence of polarity in test sentences with false responses.

Parallel to Carpenter's findings on counterfactual clauses and modified by Chinese stimuli, as predicted, test sentences with matched predicates were responded to significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA ($p < .0001$). Error rates for factual target clauses in four conditions did not reach significant difference to one another ($F(3, 1301) = 0.09$, $p = .96$), and neither did error rates for counterfactual target clauses in four conditions ($F(3, 1219) = 0.14$, $p = .93$).

Test sentences in factual targets

As predicted, for factual target clauses at 0 SOA, the condition ordering from the easiest to the hardest was like the following: TA < FA < FN < TN. That is, participants responded fastest to test sentences with true value in affirmatives (4046ms), next was to test sentences with false value in affirmatives (4352ms), next was to test sentences with false value in negatives (4453ms), and the last was to test sentences with true value in negatives (4652ms). A one-way ANOVA showed that the difference between these four conditions was significant ($F(3, 1280) = 13.68$, $p < .0001$). A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed.

The results showed almost all the comparisons were significant, although one comparison was not (FA vs. FN, $p = 0.33$). Though false affirmatives (FA) were responded faster than false negatives (FN) by 100ms, the difference was not significant, suggesting that the influence of polarity in test sentences with false responses was unreliable.

Parallel to Carpenter's findings on factual clauses, test sentences with matched representations in predicates were responded to significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA ($p = .0020$).

Affirmative vs. negative test sentences

For counterfactual target clauses, test sentences with affirmatives were responded faster than test sentences with negatives (4685ms and 5040ms, respectively) (F (1, 1220) = 11.17, $p = .0009$). The difference in their error rates was also not significant (F (1, 1221) = 0.10, $p = .7541$).

Test sentences with affirmatives were responded to faster than test sentences with negatives for factual target clauses (4198ms and 4548ms, respectively) (F (1, 1302) = 20.53, $p < .0001$). However, the difference in error rates for the two types was not significant (F (1, 1303) = 0.06, $p = .8038$).

Table 10 Response Latency (in ms) and Error Rates in Simultaneous Task with Negation for College Students

Type of Clause		TA	FA	FN	TN	Total
Factual	RT	4046.18	4352.28	4453.11	4652.84	4360.42
	Errors	7.27%	7.20%	7.24%	7.16%	7.22%
Counterfactual	RT	4407.87	5005.75	4901.16	5188.98	4848.69
	Errors	12.94%	12.68%	12.70%	12.73%	12.77%

Results: The Eighth Graders Data

The latencies and error rates for responses to factual and counterfactual clauses are shown in Table 11 below.

Counterfactual vs. factual target clauses

Participants' mean response latencies to counterfactual target clauses and factual target clauses were 4733ms and 4402ms, respectively. These two response latencies were significantly different (F (1, 2868) = 54.55, p < .0001). Meanwhile, participants made more errors on test sentences probing counterfactual target clauses than test sentences probing factual target clauses (F (1, 2896) = 308.88, p < .0001).

For counterfactual target clauses, there was a significant interaction between truth value and polarity (F (1, 1341) = 5.87, p < .015. The same significant interaction was also seen in error rates (F (1, 1369) = 4.03, p < .04).

For factual target clauses, there was also a significant interaction between truth value and polarity (F (1, 1493) = 21.28, p < .0001). The same pattern was found in error rates (F (1, 1521) = 20.12, p = .0001).

The interaction of four conditions on counterfactual and factual target clauses was not significant (F (3, 2862) = 1.53, p = .2042). The main effect of target clause was significant (F (1, 2862) = 57.41, p < .0001) and the main effect of the four conditions was also significant (F (3, 2862) = 35.21, p < .0001). The difference between factual target clauses and counterfactual target clauses was highly significant for each condition, implying faster response latencies in general for factual clauses than counterfactual clauses (except the comparison on TN condition (p = .0521). This indifference finding on TN comparison of factual and counterfactual target clauses seemed to imply that for the eighth graders, test sentences with true value in negatives were all difficult to them, though the difference of response latency was big (i.e. 161ms).

The interaction of error rates of four conditions on both target clauses was not significant ($F (3, 2862) = 1.19$, $p = .31$). The main effect of error rates on target clauses was significant ($F (1, 2862) = 3688.33$, $p < .0001$), but the main effect of error rates on four conditions was not significant difference ($F (1, 2862) = 0.44$, $p = .72$). The difference of each condition on error rates between counterfactual and factual target clauses reached highly significance ($p < .0001$), implying that the eighth graders erred more on counterfactual target clauses than on factual target clauses as college students.

Counterfactual vs. factual clause position

Response latency to counterfactual target clauses in if-clause was 4740ms and in consequence clause was 4743ms. The difference of their latency did not reach significance ($F (1, 1370) = 0.00$, $p = .96$), implying clause positions for counterfactual targets did not make influence in processing because counterfactual target clauses were all difficult to participants. The same pattern was found on their error rates ($F (1, 1371) = 0.01$, $p = .93$), suggesting that participants did not make more errors because of the clause positions.

Response latency to factual target clauses in if-clause was 4241ms and in consequence clause was 4584ms. The difference of their latency reached significance ($F (1, 1522) = 30.52$, $p = .0001$), implying clause positions for factual targets did have influence in processing. Since factual target clauses were all easy to participants, clause positions caused difference. Their error rates did not show any difference ($F (1, 1523) = 0.15$, $p = .70$), suggesting that participants did not make more errors because of the clause positions.

Test sentences in counterfactual targets

For counterfactual target sentences at zero SOA, the condition ordering was the following: TA < FA < FN < TN, which was what would be predicted if alternative three were the representation being formed. The results were exactly parallel to the results observed for factual target clauses (i.e. the TA condition was the easiest, FA was the second easiest condition and FN was the third, and TN condition was the hardest). These results seemed to indicate that participants had already formed a simpler, semantic representation for counterfactual clauses. Thus, we can infer that under time limitation, younger participants (fourteen-year-olds) could form a representation based on the sentence's meaning, not based on its grammatical structure.

Participants responded fastest to true, affirmative test sentences (4447ms), next was to false, affirmative test sentences (4733ms), next was to false, negative test sentences (4927ms), and the last was to true, negative test sentences (4930ms). A one-way ANOVA showed that the difference between these four conditions was significant ($F(3, 1341) = 11.71$, $p < .0001$). A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. The difference was mainly attributed from comparisons of TA and other groups (TA vs. FA, $p = .0019$; TA vs. FN, $p < .0001$; TA vs. TN, $p < .0001$). The difference between FN and TN was not significant, $p = .72$, suggesting that when test sentences were already difficult, as in negatives, the truth value did not make any difference to participants. Though the ordering showed that responses to the matched condition (FN) for counterfactuals were slower than responses to the mismatched condition (FA) by 193.78 ms, the difference was not significant ($p = .1265$), suggesting no evidence for the influence of polarity in responses to false test sentences.

Parallel to Carpenter's findings on counterfactual clauses although modified by the use of Chinese stimuli as predicted, test sentences with matched representations in predicates were responded to significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA ($p < .0019$). Error rates for factual target clauses across the four conditions were not significantly different from one another ($F(3, 1521) = 0.77$, $p = .51$). But, the difference in error rates for counterfactual target clauses across the four conditions did reach significance ($F(3, 1369) = 4.69$, $p = .0029$).

Test sentences in factual targets

As predicted, for factual target clauses at 0 SOA, the condition ordering from the easiest to the hardest was the following: TA < FA < FN < TN. That is, participants responded fastest to test sentences with true value in affirmatives (4057ms), next was to test sentences with false value in affirmatives (4351ms), the third was to test sentences with false value in negatives (4513ms), and the last was to test sentences with true value in negatives (4769ms). A one-way ANOVA showed that significant differences existed between these four conditions ($F(3, 1493) = 26.50$, $p < .0001$).

A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. The results showed that all the comparisons were significant.

Parallel to Carpenter's findings on factual clauses, test sentences with matched representations in predicates were responded to significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA ($p = .0002$).

Counterfactual vs. factual sentence type

Response latencies for counterfactual target clauses in four experimental sentences were 4738ms, 4869ms, 4585ms, and 4779ms, respectively. A one-way ANOVA results showed significant difference among them ($F(3, 1368) = 2.65$, p

= .0474), and the major difference came from the comparison of sentence 2 and sentence 3. This difference indicated that participants had more difficult time in responding counterfactual target clauses in sentence 2 than in sentence 3. The same pattern was found on their error rates in one-way ANOVA (F (3, 1369) = 0.10, p = .9616).

Response latencies for factual target clauses in four experimental sentences were 4380ms, 4488ms, 4463ms, and 4313ms, respectively. A one-way ANOVA results did not show significant difference among them (F (3, 1520) = 1.65, p = .17), suggesting that none of the factual target clauses in these four experimental sentences (i.e. FF, FC, CF, CC) caused more difficulty than others. The same pattern was found on their error rates in one-way ANOVA (F (3, 1521) = 0.44, p = .72).

Affirmative vs. negative test sentences

For counterfactual target clauses, test sentences with affirmatives were responded faster than test sentences with negatives (4570ms and 4928ms, respectively) (F (1, 1370) = 24.83, p = .0001). The difference in their error rates was also not significant (F (1, 1371) = 0.53, p = .46).

Test sentences with affirmatives were responded faster than test sentences with negatives for factual target clauses (4197ms and 4642ms, respectively) (F (1, 1522) = 52.79, p < .0001). The difference in their error rates was not significant (F (1, 1523) = 0.33, p = .56).

Table 11 Response Latency (in ms) and Error Rates in Simultaneous Task with Negation for the Eighth Graders

Type of Clause	TA	FA	FN	TN	Total

Factual	RT	4057.65	4351.61	4513.64	4769.32	4402.47
	Errors	16.65%	16.13%	15.62%	16.57%	16.26%
Counterfactual	RT	4447.19	4733.39	4927.17	4930.53	4733.50
	Errors	24.77%	22.72%	21.96%	24.71%	23.66%

Results: The Sixth Graders Data

The latencies and error rates to respond to factual and counterfactual clauses were shown in Table 12 below.

Counterfactual vs. factual target clauses

Participants responded to counterfactual target clauses and factual target clauses were 5601ms and 4978ms, respectively. These two response latencies were significantly different (F (1, 2169) = 40.42, p < .0001). Meanwhile, participants made more errors on test sentences probing counterfactual target clauses than test sentences probing factual target clauses (F (1, 2189) = 521.19, p < .0001).

For counterfactual target clauses, there showed a non-significant interaction between truth values and polarities (F (1, 1010) = 0.37, p = .54). The same pattern was found on error rates (F (1, 1030) = 0.02, p = .88). The interaction of four conditions on counterfactual and factual target clauses was not significant (F (3, 2163) = 1.18, p = .31). The main effect of target clause was significant (F (1, 2163) = 42.45, p < .0001) and the main effect of four conditions was also significant (F (3, 2163) = 10.60, p < .0001). The difference of each condition between counterfactual target clauses and factual target clauses was highly significant, implying in general faster response latency on factual clauses than counterfactual clauses.

For factual target clauses, there was an interaction between truth values and polarities and this interaction reached significant difference (F (1, 1133) = 6.71, p < .009). The same pattern was found on error rates (F (1, 1153) = 6.88, p = .008).

The interaction of four conditions on error rates on both target clauses was also not significant (F (3, 2163) = 0.03, p = .99). The main effect of error rates on target clauses was significant (F (1, 2163) = 1693.20, p < .0001), but the main effect of error rates on four conditions was not significant difference, F (1, 2163) = 0.37, p = .77. The difference of each condition on error rates between counterfactual and factual target clauses reached highly significance (p < .0001), implying that the sixth graders erred more on counterfactual target clauses than on factual target clauses as college students.

Counterfactual vs. factual clause position

Response latency to counterfactual target clauses in if-clause was 5676ms and in consequence clause was 5624ms. The difference of their latency did not reach significance (F (1, 1031) = 0.11, p = .73), implying clause positions for counterfactual targets did not make any influence in processing. The same pattern was found on their error rates (F (1, 1032) = 0.11, p = .74), suggesting that participants did not make more errors because of the clause positions. Both indifference clause effects on factual and counterfactual target clause indicated the insensitivity of clause positions for youngest children.

Response latency to factual target clauses in if-clause was 4915ms and in consequence clause was 5113ms. The difference of their latency did not reach significance (F (1, 1154) = 1.70, p = .19), implying clause positions for factual targets did not have influence in processing for the sixth graders. Their error rates did not show any difference (F (1, 1155) = 0.00, p = .97), suggesting that participants did not make more errors because of the clause positions.

Test sentences in counterfactual targets

For counterfactual target sentences at zero SOA, the condition ordering is like the following: TA < FA < TN < FN, which was exactly the same as alternative three. Again, the results were parallel to the results observed of factual target clauses (i.e. TA condition was the easiest, FA condition was the second, and FN/TN were in the lower ranks), suggesting that participants had formed a simpler representation for counterfactual clauses. Thus, we infer that under time limitation in presentation, youngest participants like twelve-year-old children could still form a representation based on the sentence's meaning, but not based on its grammatical structure like college students and the eighth graders.

Participants responded fastest to test sentences with true value in affirmatives (5191ms), next was to test sentences with false value in affirmatives (5602ms), next was to test sentences with true value in negatives (5780ms), and the last was to test sentences with false value in negatives (6146ms). A one-way ANOVA showed that the difference between these four conditions was significant ($F(3, 1010) = 6.96$, $p = .0001$). A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. Among all the comparisons, almost all the comparisons were significant, but two comparisons weren't (TN vs. FN, $p = .27$; FA vs. TN, $p = .24$). The indifference between FN and TN indicated when test sentences were in negative polarity, truth values did not make difference to participants.

Though the ordering showed the matched condition (FN) for representations of counterfactual target clauses in both predicates and polarity had much bigger response latency (366ms) than partially matched condition (TN) only in polarity, the difference was still not significant. This result seemed to indicate that youngest participants responded counterfactual target clauses by combining polarity and mismatched predicate to get the real meaning, thus TN condition were responded faster than FN condition.

Parallel to Carpenter’s findings on counterfactual clauses and modified by Chinese stimuli, as predicted, test sentences with matched representations in predicates were responded significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA ($p = .04$). Error rates for factual target clauses in four conditions did not reach significant difference to one another (F (3, 1153) = 0.41, $p = .74$), so did error rates for counterfactual target clauses in four conditions (F (3, 1030) = 1.23, $p = .29$).

Test sentences in factual targets

As predicted, for factual target clauses at zero SOA, the condition ordering from the easiest to the hardest was like the following: TA < FN < FA < TN. That is, participants responded fastest to test sentences with true value in affirmatives (4579ms), next was to test sentences with false value in negatives (5078ms), next was to test sentences with false value in affirmatives (5139ms), and the last was to test sentences with true value in negatives (5312ms). A one-way ANOVA showed that the difference between these four conditions was significant, F (3, 1133) = 4.80, $p = .0025$. A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. The major difference came from the comparisons of TA and other groups (TA vs. FA, $p = .0058$; TA vs. FN, $p = .0098$; TA vs. TN, $p = .0005$). The results did not show significant difference between FA and FN, $p = .84$. Though false negatives (FN) were responded faster than false affirmatives (FA) by 60ms, the difference in polarity of test sentences was not significant. Meanwhile, the comparison between FN and TN was also not significant, $p = .35$, suggesting an indifference influence of truth values in test sentences with negative polarity.

For the sixth graders, test sentences with false responses and test sentences with negative polarity were all difficult to them, thus, any manipulations based on these test sentences were all in the same degree of difficulty.

Parallel to Carpenter's findings on factual clauses, test sentences with matched representations in predicates were responded significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA ($p = .005$).

Counterfactual vs. factual sentence type

Response latencies for counterfactual target clauses in four experimental sentence types were 5607ms, 5879ms, 5732ms, and 5391ms, respectively. A one-way ANOVA results did not show significant difference among them ($F(3, 1029) = 1.75$, $p = .15$), suggesting that none of the factual target clauses in these four experimental sentences (i.e. FF, FC, CF, CC) caused more difficulty than others. They were all similar in degree of difficulty. The same pattern was found on their error rates in one-way ANOVA ($F(3, 1030) = 0.05$, $p = .98$).

Response latencies for factual target clauses in four experimental sentence types were 4887ms, 5117ms, 5183ms, and 4862ms, respectively. A one-way ANOVA results did not show significant difference among them ($F(3, 1152) = 1.14$, $p = .33$), suggesting that none of the factual target clauses in these four experimental sentences (i.e. FF, FC, CF, CC) caused more difficulty than others. They were all similar in degree of difficulty. The same pattern was found on their error rates in one-way ANOVA ($F(3, 1153) = 0.04$, $p = .99$).

Affirmative vs. negative test sentences

For counterfactual target clauses, test sentences with affirmatives were responded faster than test sentences with negatives (5377ms and 5949ms, respectively) (F (1, 1031) = 13.54, p = .0002). The difference in their error rates was not significant (F (1, 1032) = 1.18, p = .27).

Test sentences with affirmatives were responded faster than test sentences with negatives for factual target clauses (4840ms and 5193ms, respectively) (F (1, 1154) = 5.41, p = .02). The difference in their error rates was not significant (F (1, 1155) = 0.34, p = .55).

Table 12 Response Latency (in ms) and Error Rates in Simultaneous Task with Negation for The Sixth Graders

Type of Clause		TA	FA	FN	TN	Total
Factual	RT	4579.70	5139.08	5078.45	5312.17	4978.55
	Errors	13.69%	13.15%	13.15%	13.25%	13.32%
Counterfactual	RT	5191.10	5602.14	6146.53	5780.40	5601.30
	Errors	22.19%	21.42%	20.57%	21.71%	21.53%

General Discussion of Age Effect

Counterfactual vs. factual target clauses in different ages

A significant effect of age group was observed on response latencies of counterfactual and factual target clauses across the three age groups (F (2, 7542) = 4.90, p = .007). The main difference came from the comparisons of the sixth graders and other two groups (i.e. the sixth vs. the eighth on counterfactuals, p = .0003; the sixth vs. college on counterfactuals, p = .0034; the sixth vs. the eighth on factuals, p = .01; the sixth vs. college on factuals, p = .01). The interaction of response latency of factual and counterfactual target clauses on four conditions for three age groups was also significant ($F (17, 7524) = 1.64$, $p = .04$). Once more, the major contribution came from the comparisons of the sixth graders and other two groups on four conditions of both factual and counterfactual target clauses.

A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed, showing almost all the comparisons reached significance between them. Though the trend showed that college students performed the least reaction times for all conditions, the sixth graders performed the longest reaction times for all conditions and the eighth graders were in between, there were two comparisons which did not reach significant difference. That is, (1) factual TA condition of the sixth and college students ($p = .05$), suggesting that test sentences with true affirmatives were easy to youngest and oldest participants, and also (2) factual TN condition of the sixth and the eighth graders, suggesting that test sentences with true negatives were all difficult to younger participants like the eighth and the sixth graders, but not for college students.

Test sentences in counterfactual targets in different ages

The interaction between truth values and polarity on counterfactual target clauses for three groups reached significant difference ($F (7, 7536) = 7.44$, $p < .0001$), so did the interaction on factual target clauses ($p < .0001$). The interaction between clause positions and different age groups on counterfactual target clauses was not significant ($F (2, 3555) = 0.18$, $p = .83$), suggesting that no specific clause position was more difficult or easier for specific age group. The main effect of age group was significant ($F (2, 3555) = 4.92$, $p = .007$), showing that the sixth graders performed in general longer reaction times than college students and the eighth graders both on if-clauses and consequence clauses of counterfactual target clauses. However, none of the comparisons between college students and the eighth graders in clause positions reached significance, suggesting that these two groups did not show any difference in processing counterfactual target clauses in terms of clause positions.

Test sentences in factual targets in different ages

The interaction between clause positions and different age groups on factual target clauses was not significant ($F(2, 3912) = 0.61$, $p = .54$). The main effect of age group was significant ($F(2, 3912) = 5.50$, $p = .004$) and the main effect of clause positions was also significant ($F(2, 3912) = 27.02$, $p < .0001$). The pattern was exactly parallel to the findings on counterfactual target clauses. Also, the interaction between age group and different target sentences was not significant on factual target clauses ($F(6, 3906) = 1.63$, $p = .13$), suggesting that there was no specific target sentence caused more difficult processing. The main effect of age group was significant ($F(2, 3906) = 5.47$, $p = .004$) and the main effect of target sentence was also significant ($p = .04$).

Counterfactual vs. factual sentence type in different ages

The interaction between age group and different target sentences was not significant on counterfactual target clauses ($F(6, 3549) = 1.86$, $p = .08$), suggesting that there was no specific target sentence caused more difficult processing. The main effect of age group was significant ($F(2, 3549) = 4.96$, $p = .0071$) and the main effect of experimental target sentence was also significant ($p = .005$). None of the comparisons between college students and the eighth graders reached significant difference. However, the comparisons between the sixth graders and other two groups on experimental target sentence 1, 2, and 3 reached significance. Again, none of the comparisons between college students and the eighth graders reached significance. The major contribution of difference came from the comparisons between the sixth graders and other two groups on target sentence 1, 2, and 3 (also the difference of target sentence 4 between the younger two groups).

Affirmative vs. negative test sentences in factual sentences in different ages

The interaction between age group and polarity on factual target clauses was not significant ($F(2, 3912) = 0.42$, $p = .65$). The main effect of age group was significant, $p = .004$ and the main effect of polarity was also significant ($p < .0001$). All the comparisons were significant between the sixth graders and other two groups. Meanwhile none of the comparisons between college students and the eighth graders was significant.

Affirmative vs. negative test sentences in counterfactual sentences in different ages

The interaction between age group and polarity on counterfactual target clauses was also not significant ($F(2, 3555) = 1.15$, $p = .31$). The main effect of age group was significant ($p = .006$) and the main effect of polarity was also significant ($p < .0001$). All the comparisons were significant between the sixth graders and none of the comparisons between college students and the eighth graders was significant.

These results indicated that though test sentences with affirmatives were all easier to participants than test sentences with negatives, the sixth graders still showed the slowest response latency. The same pattern was found on test sentences with negatives. Though they were all difficult to participants, the sixth graders still showed the longest reaction time. Though there was a trend for faster response latencies in general on both test sentences (i.e. affirmatives and negatives) on college students than the eighth graders, none of the comparisons reached significance.

Table 13 Response Latency (in ms) and Error Rates in Simultaneous Task with

Negation on factual and counterfactual target clauses for three age groups

Type of Clause	Group	TA	FA	FN	TN	Total
Factual	College	4046.18	4352.28	4453.11	4652.84	4360.42
		7.27%	7.20%	7.24%	7.16%	7.22%
	The eighth	4057.65	4351.61	4513.64	4769.32	4402.47
		16.65%	16.13%	15.62%	16.57%	16.26%
	The sixth	4579.70	5139.08	5078.45	5312.17	4978.55
		13.69%	13.15%	13.15%	13.25%	13.32%
Counterfactual	College	4407.87	5005.75	4901.16	5188.98	4848.69
		12.94%	12.68%	12.70%	12.73%	12.77%
	The eighth	4447.19	4733.39	4927.17	4930.53	4733.50
		24.77%	22.72%	21.96%	24.71%	23.66%
	The sixth	5191.10	5602.14	6146.53	5780.40	5601.30
		22.19%	21.42%	20.57%	21.71%	21.53%

Summary

As we can see in Table 14, which summarizes the findings of experiment I over the three age groups, most of our findings was consistent with our predictions. The ordering of difficulty for the factual conditions was the same pattern as predicted: the TA condition had the shortest response latency, the TN condition had the longest, and the FA/FN conditions are in between. The same pattern was also observed for the counterfactual conditions (although the sixth age group was slightly different). All these patterns for both factual and counterfactual across the three age groups demonstrated significant differences in reaction times, but not in error rates.

Factual target clauses showed shorter response latencies and lower error rates than counterfactual target clauses across the board, implying that processing is more difficult in contrary-to-fact sentences. Interactions of truth value and polarity in response latency and error rates are clearly observed in factual target clauses and in counterfactual target clauses for the college and the eighth groups. The interaction of four conditions on factual and counterfactual target clauses in response latency was not significant in all three age groups. However, The main effect of four conditions and The main effect of target clauses were significant in three age groups ($p < .0001$), due to the fact that in each condition on factual target clauses was responded to faster than its counterpart on counterfactual target clauses. The same situation was also observed in error rates: Participants erred more on counterfactual target clauses than on factual target clauses.

There was no evidence of a clause effect for counterfactual target clauses, suggesting that an overall difficulty in processing these sentences overrode any potential differences due to the position of the counterfactual clause. However, for factual target clauses, college students and the eighth graders demonstrated a significant clause effect. Interestingly, this effect does not show on the youngest group, the sixth graders, perhaps because processing factual target clauses is harder for them.

Generally speaking, the experimental sentence type (FF, FC, CF, or CC) did not cause differential processing difficulty for both factual and counterfactual target clauses across the three age groups (except the eighth graders on counterfactual target clauses), suggesting overall difficulty or easiness degree in processing target clauses.

There was a very clear polarity effect on both factual and counterfactual target clauses in response latency, which means that affirmative clauses are processed faster than negative clauses. This effect is very robust across all three age groups. However, this polarity effect does not show up on error rates.

Table 14 Summary Findings of Three Groups in Simultaneous Experiment of Counterfactual Conditionals

		COLLEGE	THE EIGHTH	THE SIXTH
Factual Ordering		TA < FA < FN < TN	TA < FA < FN < TN	TA < FN < FA < TN
Factual p-value	(RT)	$p < .0001$	$p < .0001$	$p = .0025$
	(Error Rates)	$p = .9664$	$p = .5116$	$p = .7433$
Counterfactual Ordering		TA < FN < FA < TN	TA < FA < FN < TN	TA < FA < TN < FN
Counterfactual p-value	(RT)	$p < .0001$	$p < .0001$	$p = .0001$
	(Error Rates)	$p = .9335$	$p = .0029$	$p = .2978$
Factual vs. Counterfactual p-value	(RT)	$p < .0001$	$p < .0001$	$p < .0001$
	(Error Rates)	$p < .0001$	$p < .0001$	$p < .0001$
Factual interaction of truth values and polarities	(RT)	$p < .0001$	$p < .0001$	$p < .0097$
	(Error Rates)	$p = .0011$	$p = .0001$	$p = .0088$
Counterfactual interaction of truth values and polarities	(RT)	$p < .0001$	$p < .0156$	$p = .5458$
	(Error Rates)	$p < .0001$	$p < .0449$	$p = .8853$
Interaction of the	(RT)	$p = .3153$	$p = .2042$	$p = .3148$

four conditions on factual and counterfactual	(Error Rates)	p = .9746	p = .3127	p = .9918
Factual clause effect	(RT)	p = .0002	p = .0001	p = .1930
	(Error Rates)	p = .9779	p = .7009	p = .9776
Counterfactual clause effect	(RT)	p = .7274	p = .9617	p = .7389
	(Error Rates)	p = .4424	p = .9397	p = .7409
Factual sentence type effect	(RT)	p = .0741	p = .1751	p = .3319
	(Error Rates)	p = .9572	p = .7263	p = .9901
Counterfactual sentence type effect	(RT)	p = .2448	p = .0474	p = .1560
	(Error Rates)	p = .6647	p = .9616	p = .9860
Factual polarity effect	(RT)	p < .0001	p < .0001	p = .0202
	(Error Rates)	p = .8038	p = .5659	p = .5590
Counterfactual polarity effect	(RT)	p = .0009	p = .0001	p = .0002
	(Error Rates)	p = .7541	p = .4660	p = .2776

F Language and Thought Experiment II: Delayed Task of Counterfactual Conditionals with Negation

Participants: College Students

Twenty-two college students from National Yang Ming University were included (mean age = 20.8, range from 18 to 25, 10 females and 12 males). None of them were reported as having medical problems. All participants were rewarded with a payment of one hundred New Taiwanese dollars.

Participants: Junior High School Students

Thirty-two junior high school students in Fu He Junior High School participated in this study (mean age = 14.04, range from 13 to 14, 5 females and 10 males, 17 participants were missing to report their gender and age). They were rewarded with a present after finishing the study. All participants were right-handed users and none of them were reported as having medical problems.

Participants: Elementary School Students

Twenty-nine elementary school students in Qing Jiang Elementary School participated in this study (mean age = 12.3, range from 12 to 13, 21 females and 8 males). They were rewarded with a present after finishing the study. All participants were right-handed users and none of them were reported having medical problems.

Design and Materials

Parallel to the simultaneous task described above, a verification paradigm was employed. Participants were required to judge whether the test sentence was true or false based on the truth condition presupposed of the target clause. The same numbers of experimental trials were presented to participants and all the target sentences are listed in Appendix 1, 3, 5, and 7. The same test sentences were presented to participants and all the test sentences probing if-clauses or consequence clauses are listed in Appendix 2, 4, 6, and 8. Meanwhile, all the fillers are listed in Appendix 9. The only difference from simultaneous task to delayed task was stimuli of asynchrony (SOA). A detailed description was given in Procedure below.

Procedure

A fixation point was presented on the computer screen for 500 ms. After this attraction of participants' attention, a target sentence was presented on the screen for 5 seconds. After this period of time, the target sentence disappeared and a test sentence was displayed to probe the truth values of one of the target clauses (i.e. if-clause or consequence clause). If the test sentence matched the description of the target clause, participants were instructed to press the left button of a mouse as soon as possible. If any mismatch was found, they pressed the right button of the mouse immediately.

The study contained four blocks, and each of them contained 48 trials. Between each section there was a break. All participants did 8 practice trials first to confirm their understanding of this task. Three random lists were assigned to participants, which were counterbalanced. All college participants were tested in a sound-proof room of Laboratory of Cognitive Neuropsychology in National Yang Ming University. All junior high students and elementary students were tested in the computer rooms of their own schools, which were quiet.

Predictions

For factual target clauses, because the same representations should be formed no matter what the SOA, it was predicted that the results should not be different from the findings in simultaneous task. According to the CCM, the test sentences should pattern from easiest to hardest like the following: true affirmatives (TA), false affirmatives (FA), false negatives (FN), and true negatives (TN). It was also predicted that among these results, Carpenter's findings would be confirmed. That is, true test sentences in affirmatives (TA) should be responded to faster than false test sentences in affirmatives (FA). In this situation, parallel to Carpenter's study, test sentences with true value in affirmatives (TA) should be responded faster than the test sentences with false value in affirmatives (FA).

Results: College Students Data

The latencies and error rates to respond to factual and counterfactual clauses were shown in Table 15 below.

Counterfactual vs. factual target clauses

Participants responded to counterfactual target clauses and factual target clauses are 1957ms and 1770ms, respectively. These two response latencies were significantly different ($F(1, 2517) = 22.69$, $p < .0001$). Meanwhile, participants made more errors on test sentences probing counterfactual target clauses than test sentences probing factual target clauses ($F(1, 2517) = 1767.67$, $p < .0001$).

For counterfactual target clauses, there was also a significant interaction between truth values and polarities ($F(1, 1211) = 27.82$, $p < .0001$). The same pattern was found on error rates ($F(1, 1232) = 23.00$, $p < .0001$). The interaction of four conditions on factual and counterfactual target clauses was significant ($F(3, 2511) = 2.75$, $p = .04$), suggesting in general faster response latency on factual target clauses than on counterfactual target clauses. The difference of each condition between factual target clauses and counterfactual target clauses was highly significant (except

TA condition).

For factual target clauses, there was an interaction between truth values and polarities and this interaction reached significant difference ($F(1, 1279) = 6.78$, $p = .009$). The same pattern was found on error rate, ($F(1, 1300) = 5.53$, $p = .01$).

The interaction of four conditions on error rates on both target clauses was not significant ($F(3, 2511) = 0.58$, $p = .62$). The main effect of error rates on target clauses was significant ($F(1, 2511) = 1767.28$, $p < .0001$), but the main effect of error rates on four conditions was not significant difference ($F(3, 2511) = 0.05$, $p = .98$). The difference of each condition on error rates between factual and counterfactual target clauses reached highly significance ($p < .0001$), implying that college students erred more on counterfactual target clauses than on factual target clauses.

Counterfactual vs. factual clause position

Response latency to counterfactual target clauses in if-clause was 1897ms and in consequence clause was 1991ms. The difference of their latency did not reach significance ($F(1, 1233) = 1.84$, $p = .17$), implying clause positions for counterfactual targets did not make any influence in processing. The same pattern was found on their error rates ($F(1, 1234) = 0.09$, $p = .76$), suggesting that participants did not make more errors because of the target clause positions. Probably counterfactual target clauses were difficult to participants, thus the clause positions did not make any difference. However, factual target clauses were easy to participants, thus the clause positions made difference in processing.

Response latency to factual target clauses in if-clause was 1698ms and in consequence clause was 1842ms. The difference of their latency reached significance ($F(1, 1301) = 7.07$, $p = .007$), implying clause positions for factual targets did have influence in processing. Their error rates did not show any difference ($F(1, 1302) =$

0.01, $p = .91$), suggesting that participants did not make more errors because of the target clause positions.

Test sentences in counterfactual targets

For counterfactual target sentences at 5-second SOA, the condition ordering was like the following: TA < FA < FN < TN. The results were comparable to the prediction of alternative three, the simpler form. Participants responded fastest to test sentences with true affirmatives (1535ms), next was to test sentences with false affirmatives (1957ms), next was to test sentences with false negatives (2065ms), and the last was to test sentences with true value in negatives (2288ms). A one-way ANOVA showed that the difference between these four conditions was significant ($F(3, 1211) = 27.58$, $p < .0001$).

A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. Among all the comparisons, almost all the comparisons were significant, but one comparison wasn't (FA vs. FN, $p = .18$). Though false affirmatives (FA) were responded faster than false negatives (FN) by 107ms, suggesting the difference of polarity in test sentences with false responses was unreliable. The difference between TA and other groups was all significant at $p < .0001$ level and the difference between FN and TN also was significant ($p = .01$).

Parallel to Carpenter's findings on counterfactual clauses and modified by Chinese stimuli, as predicted, test sentences with matched representations in predicates were responded to significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA ($p < .0001$). Error rates for factual target clauses in four conditions did not reach significant difference to one another ($F(3, 1300) = 0.60$, $p = .61$), and neither did error rates for counterfactual target clauses in four conditions ($F(3, 1232) = 0.22$, $p = .88$).

Test sentences in factual targets

As predicted, for factual target clauses at 5-second SOA, the condition ordering from the easiest to the hardest was like the following: TA < FA < FN < TN. That is, participants responded fastest to test sentences with true affirmatives (1525ms), next was to test sentences with false affirmatives (1676ms), next was to test sentences with false negatives (1904ms), and the last was to test sentences with true negatives (2002ms). A one-way ANOVA showed that the difference between these four conditions was significant ($F (3, 1279) = 20.36$, $p < .0001$).

A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. The results showed almost all the comparisons were significant, but one comparison wasn't (FN vs. TN, $p = .17$). Though false negatives (FN) were responded faster than true negatives (TN) by 98ms, the difference was not significant, suggesting that the influence of truth values in test sentences with negative polarity was unreliable. It seemed that test sentences with negatives were difficult to college students, thus the truth values did not make any difference.

Parallel to Carpenter's findings on factual clauses, test sentences with matched representations in predicates were responded to significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA ($p = .01$).

Counterfactual vs. factual sentence type

Response latencies for counterfactual target clauses in four experimental sentence types were 1955ms, 2011ms, 1972ms, and 1842ms, respectively. A one-way ANOVA results did not show significant difference among them ($F (3, 1231) = 1.11$, $p = .3441$), suggesting that none of the counterfactual target clauses in these four experimental sentences (i.e. FF, FC, CF, CC) caused more difficulty than others. They were all similar in degree of difficulty. The same pattern was found on their error rates in one-way ANOVA (F (3, 1232) = 0.10, p = .9590).

Response latencies for factual target clauses in four experimental sentence types were 1717ms, 1866ms, 1777ms, and 1721ms, respectively. A one-way ANOVA results did not show significant difference among them (F (3, 1299) = 1.68, p = .16), suggesting that none of the factual target clauses in these four experimental sentences (i.e. FF, FC, CF, CC) caused more difficulty than others. They were all similar in degree of difficulty. The same pattern was found on their error rates in one-way ANOVA (F (3, 1300) = 0.29, p = .82).

Affirmative vs. negative test sentences

For counterfactual target clauses, test sentences with affirmatives were responded faster than test sentences with negatives (1740ms and 2170ms, respectively) (F (1, 1233) = 40.70, p < .0001). The difference in their error rates was also not significant (F (1, 1234) = 0.02, p = .89).

For factual target clauses, test sentences with affirmatives were responded to faster than test sentences with negatives (1600ms and 1952ms, respectively) (F (1, 1301) = 44.70, p < .0001). The difference in their error rates was not significant (F (1, 1302) = 0.38, p = .53).

Table 15 Response Latency (in ms) and Error Rates in Delayed Task with Negation for College Students

Type of Clause		TA	FA	FN	TN	Total
Factual	RT	1525.66	1676.13	1904.98	2002.99	1770.54

	Errors	4.24%	16.22%	7.31%	6.95%	7.36%
Counterfactual	RT	1535.64	1957.19	2065.12	2288.00	1957.15
	Errors	11.96%	11.62%	11.83%	11.66%	12.05%

Results: The Eighth Graders Data

The latencies and error rates to respond to factual and counterfactual clauses were shown in Table 16 below.

Counterfactual vs. factual target clauses

Participants responded to counterfactual target clauses and factual target clauses were 2234ms and 1993ms, respectively. These two response latencies were significantly different (F (1, 3439) = 32.20, p < .0001). Meanwhile, participants made more errors on test sentences probing counterfactual target clauses than test sentences probing factual target clauses (F (1, 3439) = 3243.13, p < .0001).

For factual target clauses, there was an interaction between truth values and polarities and this interaction reached significant difference (F (1, 1794) = 16.10, p < .0001). The same pattern was found on error rates (F (1, 1825) = 11.44, p = .0007). For counterfactual target clauses, there was also a significant interaction between truth values and polarities (F (1, 1608) = 4.27, p < .03). Their error rates were found marginally significant (F (1, 1639) = 3.34, p < .06). The interaction of four conditions on factual and counterfactual target clauses was significant (F (3, 3433) = 0.55, p = .64). The difference of each condition between factual target clauses and counterfactual target clauses was highly significant (except TN condition, p = .06).

The interaction of four conditions on error rates on both target clauses was not significant (F (3, 3433) = 0.81, p = .49). Main effect of error rates on target clauses was significant ($F(1, 3433) = 3226.31$, $p < .0001$), but main effect of error rates on four conditions was not significant difference ($F(3, 3433) = 0.43$, $p = .72$). The difference of each condition on error rates between factual and counterfactual target clauses reached highly significance ($p < .0001$), implying that junior high school students erred more on counterfactual target clauses than on factual target clauses as college students.

Counterfactual vs. factual clause position

Response latency to counterfactual target clauses in if-clause was 2093ms and in consequence clause was 2298ms. The difference of their latency reached significance ($F(1, 1640) = 7.56$, $p = .0060$), implying clause positions for counterfactual targets had influence in processing. The pattern found on their error rates was not significant ($F(1, 1641) = 0.08$, $p = .77$), indicating that participants did not make more errors because of the clause positions. Contrary to college students, junior high school students did show clause position effect for counterfactual target clauses.

Response latency to factual target clauses in if-clause was 1828ms and in consequence clause was 2128ms. The difference of their latency reached significance ($F(1, 1826) = 24.53$, $p < .0001$), implying clause positions for factual targets did have influence in processing. Their error rates did not show any difference ($F(1, 1827) = 0.01$, $p = .93$), indicating that participants did not make more errors because of the target clause positions.

Test sentences in counterfactual targets

For counterfactual target sentences in 5-second SOA, the condition ordering was like the following: TA < FA < FN < TN, which was exactly the same as the prediction of CCM. Participants responded fastest to test sentences with true affirmatives

(1932ms), next was to test sentences with false affirmatives (2107ms), the third was to test sentences with false negatives (2354ms), and the last was to test sentences with true negatives (2451ms). A one-way ANOVA showed that the difference between these four conditions was significant ($F(3, 1608) = 14.71$, $p < .0001$).

A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. Among all the comparisons, almost all the comparisons were significant. The difference between TA and other groups was all significant and the difference between FA and FN also was significant ($p = .004$). The results were parallel to the prediction and only one comparison did not reach significance, FN vs. TN ($p = .41$). Though false negatives (FN) were responded faster than true negatives (TN) in 96ms, it was still not significant, indicating an indifference influence of truth values in test sentences with negatives.

Parallel to Carpenter's findings on counterfactual clauses and modified by Chinese stimuli, as predicted, test sentences with matched representations in predicates were responded significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA ($p = .03$). Error rates for factual target clauses in four conditions did not reach significant difference to one another ($F(3, 1825) = 0.20$, $p = .89$), so did error rates for counterfactual target clauses in four conditions ($F(3, 1639) = 0.76$, $p = .51$).

Test sentences in factual targets

As predicted, for factual target clauses in 5-second SOA, the condition ordering from the easiest to the hardest was like the following: TA < FA < FN < TN. That is, participants responded fastest to test sentences with true affirmatives (1638ms), next was to test sentences with false affirmatives (1890ms), the third was to test sentences with false negatives (2139ms), and the last was to test sentences with true negatives

(2299ms). A one-way ANOVA showed that the difference between these four conditions was significant ($F (3, 1794) = 30.80$, $p < .0001$). A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. The results showed all the comparisons were highly significant.

Parallel to Carpenter's findings on factual clauses, test sentences with matched representations in predicates were not responded significantly faster than test sentences with mismatched representations in predicates, namely, $TA < FA$ ($p = .0005$).

Counterfactual vs. factual sentence type

Response latencies for counterfactual target clauses in four experimental sentence types were 2167ms, 2282ms, 2222ms, and 2133ms, respectively. A one-way ANOVA results did not show significant difference among them ($F (3, 1638) = 0.75$, $p = .5252$), indicating that none of the counterfactual target clauses in these four experimental sentences (i.e. FF, FC, CF, CC) caused more difficulty than others. They were all similar in degree of difficulty. The same pattern was found on their error rates in one-way ANOVA ($F (3, 1639) = 0.54$, $p = .6533$).

Response latencies for factual target clauses in four experimental sentence types were 1913ms, 1919ms, 2073ms, and 2006ms, respectively. A one-way ANOVA results did not show significant difference among them ($F (3, 1824) = 1.61$, $p = .1859$), indicating that none of the factual target clauses in these four experimental sentences (i.e. FF, FC, CF, CC) caused more difficulty than others. They were all similar in degree of difficulty. The same pattern was found on their error rates in one-way ANOVA ($F (3, 1825) = 0.10$, $p = .96$).

Affirmative vs. negative test sentences

For counterfactual target clauses, test sentences with affirmatives were responded faster than test sentences with negatives (2016ms and 2403ms, respectively) ($F(1, 1640) = 27.20$, $p < .0001$). The difference in their error rates was also not significant ($F(1, 1641) = 1.22$, $p = .27$).

For factual target clauses, test sentences with affirmatives were responded faster than test sentences with negatives (1760ms and 2217ms, respectively) ($F(1, 1826) = 56.92$, $p < .0001$). The difference in their error rates was not significant ($F(1, 1827) = 0.35$, $p = .55$).

Table 16 Response Latency (in ms) and Error Rates in Delayed Task with Negation for The Eighth Graders

Type of Clause		TA	FA	FN	TN	Total
Factual	RT	1638.39	1890.04	2139.39	2299.23	1993.18
	Errors	10.54%	10.48%	10.45%	10.31%	10.67%
Counterfactual	RT	1932.03	2107.49	2354.08	2451.03	2234.33
	Errors	18.98%	18.39%	18.07%	18.33%	19.12%

Results: The Sixth Graders Data

The latencies and error rates to respond to factual and counterfactual clauses were shown in Table 17 below.

Counterfactual vs. factual target clauses

Participants responded to counterfactual target clauses and factual target clauses were 2609ms and 2308ms, respectively. These two response latencies were significantly different ($F(1, 2964) = 37.31$, $p < .0001$). Meanwhile, participants made more errors on test sentences probing counterfactual target clauses than test sentences probing factual target clauses ($F (1, 2964) = 3177.27$, $p < .0001$).

For counterfactual target clauses, there was also a significant interaction between truth values and polarities ($F (1, 1364) = 6.74$, $p = .009$). The same pattern was found on error rates ($F (1, 1392) = 4.72$, $p = .02$). The interaction of four conditions on factual and counterfactual target clauses was significant ($F (3, 2958) = 2.61$, $p = .04$), indicating in general faster response latency on factual target clauses than on counterfactual target clauses. The difference of each condition between factual target clauses and counterfactual target clauses was highly significant (except TN condition, $p = .25$).

For factual target clauses, there was an interaction between truth values and polarities and this interaction reached significant difference ($F (1, 1566) = 24.59$, $p < .0001$). The same pattern was found on error rates ($F (1, 1594) = 21.55$, $p < .0001$).

The interaction of four conditions on error rates on both target clauses was not significant ($F (3, 2958) = 1.71$, $p = .16$). Main effect of error rates on target clauses was significant ($F (1, 2958) = 3151.76$, $p < .0001$), but main effect of error rates on four conditions was not significant difference ($F (3, 2958) = 1.92$, $p = .12$). The difference of each condition on error rates between factual and counterfactual target clauses reached highly significance ($p < .0001$), implying that elementary students erred more on counterfactual target clauses than on factual target clauses as other groups.

Counterfactual vs. factual clause position

Response latency to counterfactual target clauses in if-clause was 2511ms and in consequence clause was 2693ms. The difference of their latency reached significance ($F (1, 1393) = 4.66$, $p = .03$), implying clause positions for counterfactual targets made difference in processing. The same pattern was found on their error rates (F (1, 1394) = 0.50, p = .48), indicating that participants did not make more errors because of the clause positions.

Response latency to factual target clauses in if-clause was 2142ms and in consequence clause was 2461ms. The difference of their latency reached significance (F (1, 1595) = 20.47, p < .0001), implying clause positions for factual targets did have influence in processing. Their error rates did not show any difference (F (1, 1596) = 0.73, p = .39), indicating that participants did not make more errors because of the clause positions.

For the sixth graders, clause positions caused processing difference no matter on factual or counterfactual target clauses.

Test sentences in counterfactual targets

For counterfactual target sentences in 5-second SOA, the condition ordering was like the following: TA < TN < FA < FN, which was comparable with the prediction of alternative three. Participants responded fastest to test sentences with true affirmatives (2223ms), next was to test sentences with true negatives (2680ms), the third was to test sentences with false affirmatives (2749ms), and the last was to test sentences with false negatives (2844ms). A one-way ANOVA showed that the difference between these four conditions was significant (F (3, 1364) = 17.11, p < .0001). A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. The major difference was attributed to the comparison of TA and other groups in p < .0001 level.

Parallel to Carpenter's findings on counterfactual clauses and modified by Chinese stimuli, as predicted, test sentences with matched representations in predicates were responded significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA ($p < .0001$). Error rates for factual target clauses in four conditions did not reach significant difference to one another ($F(3, 1954) = 0.68$, $p = .5658$), but the difference of error rates for counterfactual target clauses in four conditions reached significance ($F(3, 1392) = 3.21$, $p = .02$).

Test sentences in factual targets

As predicted, for factual target clauses in 5-second SOA, the condition ordering from the easiest to the hardest was like the following: TA < FA < FN < TN. That is, participants responded fastest to test sentences with true affirmatives (1933ms), next was to test sentences with false affirmatives (2358ms), the third was to test sentences with false negatives (2360ms), and the last was to test sentences with true negatives (2585ms). A one-way ANOVA showed that the difference between these four conditions was significant ($F(3, 1566) = 19.32$, $p < .0001$). A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. The results showed almost all the comparisons were significant, but one comparison wasn't (FA vs. FN, $p = .79$).

Parallel to Carpenter's findings on factual clauses, test sentences with matched representations in predicates were responded significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA ($p < .0001$).

Counterfactual vs. factual sentence type

Response latencies for counterfactual target clauses in four experimental sentence types were 2691ms, 2573ms, 2615ms, and 2543ms, respectively. A one-way ANOVA results did not show significant difference among them ($F(3, 1391) = 0.57$, $p = .63$), indicating that none of the counterfactual target clauses in these four experimental sentences (i.e. FF, FC, CF, CC) caused more difficulty than others. They were all similar in degree of difficulty. The same pattern was found on their error rates in one-way ANOVA (F (3, 1392) = 0.62, p = .60).

Response latencies for factual target clauses in four experimental sentence types were 2267ms, 2301ms, 2348ms, and 2280ms, respectively. A one-way ANOVA results did not show significant difference among them (F (3, 1593) = 0.25, p = .85), indicating that none of the factual target clauses in these four experimental sentences (i.e. FF, FC, CF, CC) caused more difficulty than others. They were all similar in degree of difficulty. The same pattern was found on their error rates in one-way ANOVA (F (3, 1594) = 0.27, p = .85).

Affirmative vs. negative test sentences

Test sentences with affirmatives were responded faster than test sentences with negatives for counterfactual target clauses (2468ms and 2757ms, respectively) (F (1, 1393) = 11.86, p = .0006). The difference in their error rates was also not significant (F (1, 1394) = 0.91, p = .34).

For factual target clauses, test sentences with affirmatives were responded faster than test sentences with negatives (2137ms and 2473ms, respectively) (F (1, 1595) = 22.70, p < .0001). The difference in their error rates was not significant (F (1, 1596) = 0.15, p = .69).

Table 17 Response Latency (in ms) and Error Rates in Delayed Task with Negation for The Sixth Graders

Type of Clause		TA	FA	FN	TN	Total
Factual	RT	1933.06	2358.57	2360.79	2585.67	2308.60
	Errors	13.18%	12.54%	12.39%	13.00%	13.85%

Counterfactual	RT	2223.06	2749.93	2844.17	2680.71	2609.10
	Errors	23.35%	21.62%	20.75%	22.94%	23.98%

General Discussion of Age Effect

Counterfactual vs. factual target clauses in different ages

A response latency of counterfactual and factual target clauses did not show interaction among three age groups ($F(2, 8920) = 1.51$, $p = .2204$). The main effect of age group was significant ($F(2, 8920) = 5.63$, $p = .0036$) and the main effect of target clause was also significant ($F(2, 8920) = 88.21$, $p < .0001$). The main difference came from the comparisons of the sixth graders and other two groups (i.e. the sixth vs. college on counterfactuals, $p = .0004$; the sixth vs. the eighth graders on counterfactuals, $p = .02$; the sixth vs. college on factuals, $p = .003$).

The interaction of response latency of factual and counterfactual target clauses on the four conditions for three age groups was significant ($F(17, 8902) = 2.05$, $p = .006$). The major contribution came from the comparisons of the sixth graders and college students on the four conditions of both factual and counterfactual target clauses.

A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed, showing almost all the comparisons reached significance between them. Almost none of the comparisons between factual and counterfactual target clauses reached significance, but only one comparison did (i.e. counterfactual TA condition, $p = .03$). Though the trend showed that the eighth graders performed the shorter reaction times than the sixth graders, not all comparisons showed significant difference. Only test sentences with false responses for counterfactual target clauses showed the difference and test sentences with false affirmatives for factual target clauses showed difference.

These results indicated that for younger participants like the eighth and the sixth graders, test sentences with true responses for counterfactual target clauses were equally easy for them to process and also most of the test sentences for factual target clauses.

Test sentences in counterfactual targets in different ages

The interaction between truth values and polarity on counterfactual target clauses for three groups reached significant difference ($F(7, 8914) = 12.43$, $p < .0001$), so did the interaction on factual target clauses ($p < .0001$). The interaction between clause positions and different age groups on counterfactual target clauses was not significant ($F(2, 4189) = 0.51$, $p = .59$), suggesting that no specific clause position was more difficult or easier for specific age group. The main effect of age group was significant ($F(2, 4189) = 5.01$, $p = .006$), showing that the sixth graders performed in general longer reaction times than college students both on if-clauses and consequence clauses of counterfactual target clauses. However, none of the comparisons between college students and the eighth graders on clause positions reached significance, suggesting that these two groups did not show any difference in processing counterfactual target clauses in terms of clause positions. Further, comparisons between college students and the eighth graders reached marginal significance (i.e. difference for clause 1 was $p = .0528$ and difference for clause 2 was $p = .05$).

Test sentences in factual targets in different ages

The interaction between clause positions and different age groups on factual target clauses was not significant ($F(2, 4645) = 2.57$, $p = .07$). The main effect of age group was significant ($F(2, 4645) = 5.66$, $p = .003$) and the main effect of clause positions was also significant (F (2, 4645) = 60.94, p < .0001). The pattern was exactly parallel to the findings on counterfactual target clauses. Also, the interaction between age group and different experimental sentences was not significant on counterfactuals (F (6, 4183) = 0.91, p = .48), suggesting that there was no specific experimental sentence caused more difficult processing. The main effect of age group was significant (F (2, 4183) = 5.00, p = .006) and main affect of experimental sentence was not significant (p = .15).

Counterfactual vs. factual sentence type in different ages

The interaction between age group and different target sentences was not significant on factual target clauses (F (6, 4639) = 1.04, p = .40), suggesting that there was no specific experimental sentence caused more difficult processing. The main effect of age group was significant (F (2, 4639) = 5.60, p = .003) and main affect of experimental sentence was not significant (p = .18). None of the comparisons between college students and the eighth graders reached significant difference. However, the comparisons between the sixth graders and college students on all experimental sentences reached significance. Three comparisons between college students and the eighth graders reached significance (i.e. sentence 1 on counterfactuals and sentence1, 2 on factuals).

Affirmative vs. negative test sentences in factual sentences in different ages

The interaction between age group and polarity on factual target clauses was not significant (F (2, 4645) = 1.49, p = .22). The main effect of age group was significant (p = .003) and the main effect of polarity was also significant (p < .0001). All the comparisons were significant between the sixth graders and college students. Meanwhile none of the comparisons between college students and the eighth graders was significant. However, there was only one comparison between the eighth and the sixth, which reached significance (i.e. test sentences with affirmatives).

Affirmative vs. negative test sentences in counterfactual sentences in different ages

An interaction between age group and polarity on counterfactual target clauses was also not significant ($F(2, 4189) = 0.58$, $p = .56$). The main effect of age group was significant ($p = .007$) and the main effect of polarity was also significant ($p < .0001$). All the comparisons were significant between the sixth graders and college students while none of the comparisons between college students and the eighth graders was significant. There was also another comparison between the eighth graders and the sixth graders on test sentences with affirmatives which reached significance, but not test sentences with negatives.

These results indicated that the sixth graders showed the slowest response latency on both test sentences with affirmatives and negatives. However, for younger participants like the eighth and the sixth graders, they showed indifference processing on test sentences with negatives. Though there was a trend for faster response latencies in general on both test sentences (i.e. affirmatives and negatives) on college students than the eighth graders, none of the comparisons reached significance.

Table 18 Response Latency (in ms) and Error Rates in Delayed Task with Negation on Factual and Counterfactual Target Clauses for Three Age Groups

Type of Clause	Group	TA	FA	FN	TN	Total
Factual	College	1525.66	1676.13	1904.98	2002.99	1770.54
		7.21%	7.33%	7.31%	6.95%	7.36%
	The eighth	1638.39	1890.04	2139.39	2299.23	1993.18

		10.54%	10.48%	10.45%	10.31%	10.67%
	The sixth	1933.06	2358.57	2360.79	2585.67	2308.60
		13.18%	12.54%	12.39%	13.00%	13.85%
Counterfactual	College	1535.64	1957.19	2065.12	2288.00	1957.15
		11.96%	11.62%	11.83%	11.66%	12.05%
	The eighth	1932.03	2107.49	2354.08	2451.03	2234.33
		18.98%	18.39%	18.07%	18.33%	19.12%
	The sixth	2223.06	2749.93	2844.17	2680.71	2609.10
		23.35%	21.62%	20.75%	22.94%	23.98%

Across Task Comparison in Counterfactuals with Negation

Counterfactual vs. factual target clauses

A response latency of counterfactual and factual target clauses in experiments with different SOA showed a significant interaction among three age groups (F (7, 17E3) = 10.49, p < .0001). The main effect of age group, experiments with different SOA, and clause types were all significant at .0001. All the comparisons of target clauses in three different age groups in different experiments reached significance, suggesting a clear task effect on each age group.

The interaction of response latency of factual and counterfactual target clauses on the four conditions for three age groups in different experiments was also significant (F (47, 16E3) = 270.09, p < .0001). A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed, showing all the comparisons reached significance. Task difference caused processing difference in each group.

Test sentences in counterfactual targets

The interaction between truth values and polarity on counterfactual target clauses for three groups in different experiments reached significant difference ($F(23, 17E3) = 529.76$, $p < .0001$), so did the interaction on factual target clauses ($p < .0001$).

The interaction between clause positions and different age groups on counterfactual target clauses in different experiments was significant ($F(11, 7813) = 514.65$, $p < .0001$), suggesting that clause positions received different processing in different experiments.

Test sentences in factual targets

The interaction between clause positions and different age groups on factual target clauses was not significant ($F(11, 8626) = 572.04$, $p < .0001$). The pattern was exactly parallel to the findings on counterfactual target clauses.

Counterfactual vs. factual sentence type

The interaction between age group and different experimental sentences was significant on counterfactuals in different experiments ($F(23, 7801) = 247.52$, $p < .0001$).

Affirmative vs. negative test sentences in factual sentences

The interaction between age group and polarity on factual target clauses was significant ($F(11, 8626) = 584.02$, $p < .0001$). All the comparisons were significant within each group.

Affirmative vs. negative test sentences in counterfactual sentences

An interaction between age group and polarity on counterfactual target clauses was also significant ($F(11, 7813) = 531.70$, $p < .0001$). These results indicated that different experiments with different SOA cause processing difference, which was reflected clearly on each age group.

Summary

Table 19 summarizes the findings of experiment II across the three age groups. The patterns exhibited for factual and counterfactual target clauses matched our predictions. Since SOA was 5 seconds, both factual and counterfactual orderings were predicted to be the same, which was the case. The affirmative with true response (TA) condition had the shortest response latency, the negative with true response (TN) condition had the longest, and false responses in affirmative or negative were in between. However, elementary students showed a slightly different ordering in which FA/FN followed TA and TN.

All the orderings for both factual and counterfactual across the three age groups were significantly different in reaction times, but not in error rates (except for the elementary group). The interaction of factual target clauses and counterfactual target clauses in four conditions was significant, suggesting that factual target clauses generally showed shorter response latencies and lower error rates than counterfactual target clauses across the board. However, this pattern did not show up on junior high students. Thus, it seemed that for this age group factual and counterfactual target clauses were at the same level of difficulty. Interactions of truth value and polarity in response latency and error rates were clearly observed on factual target clauses and on counterfactual target clauses.

A clause effect was observed on factual target clauses and also on counterfactual target clauses. Thus, it seems to make a difference whether the first clause (i.e. if-clause) or the second clause (i.e. consequence clause) is factual or counterfactual.

This is not a surprising result for the long SOA task. That is, after 5 seconds SOA, counterfactual target clauses are like factual target clauses in processing. Surprisingly, the college group did not show this effect as younger students did, perhaps because for them processing counterfactual target clauses is hard enough that subtle differences of clause order are masked.

Finally, as predicted, experimental sentence types did not cause processing difficulty on either factual and counterfactual target clauses across three age groups, suggesting the degree of overall difficulty or easiness in processing target clauses. Meanwhile, a polarity effect was very obvious on both factual and counterfactual target clauses in response latency, meaning that affirmative clauses were processed faster than negative clauses. This effect is very robust across all three age groups. At the same time, this polarity effect did not show up on error rates.

Table 19 Summary Findings of Three Groups in Delayed Experiment of Counterfactual Conditionals

		COLLEGE	THE EIGHTH	THE SIXTH
Factual Ordering		TA < FA < FN < TN	TA < FA < FN < TN	TA < FA < FN < TN
Factual p-value	(RT)	p < .0001	p < .0001	p < .0001
	(Error Rates)	p = .61	p = .89	p = .56
Counterfactual Ordering		TA < FA < FN < TN	TA < FA < FN < TN	TA < TN < FA < FN
Counterfactual	(RT)	p < .0001	p < .0001	p < .0001

p-value	(Error Rates)	p = .8813	p = .5193	p = .0222
Factual vs. Counterfactual p-value	(RT)	p < .0001	p < .0001	p < .0001
	(Error Rates)	p < .0001	p < .0001	p < .0001
Factual interaction of truth values and polarities	(RT)	p = .009	p < .0001	p < .0001
	(Error Rates)	p = .01	p = .0007	p < .0001
Counterfactual interaction of truth values and polarities	(RT)	p < .0001	p < .03	p = .009
	(Error Rates)	p < .0001	p < .06	p = .02
Interaction of the four conditions on factual and counterfactual	(RT)	p = .04	p = .64	p = .04
	(Error Rates)	p = .62	p = .49	p = .16
Factual clause effect	(RT)	p = .007	p < .0001	p < .0001
	(Error Rates)	p = .91	p = .93	p = .39
Counterfactual clause effect	(RT)	p = .17	p = .006	p = .03
	(Error Rates)	p = .76	p = .77	p = .48
Factual sentence type effect	(RT)	p = .16	p = .18	p = .85
	(Error Rates)	p = .82	p = .96	p = .85

Counterfactual sentence type effect	(RT)	p = .34	p = .52	p = .63
	(Error Rates)	p = .95	p = .65	p = .60
Factual polarity effect	(RT)	p < .0001	p < .0001	p < .0001
	(Error Rates)	p = .53	p = .55	p = .69
Counterfactual polarity effect	(RT)	p < .0001	p < .0001	p = .0006
	(Error Rates)	p = .89	p = .27	p = .34

G Williams Syndrome Study

Participants

Two Williams Syndrome individuals were recruited. Each participant was rewarded a present after finishing this study. The detailed information of their age and intelligent IQs was listed below.

Table 20 General Information and Scores of WISC-R of WS individuals

Williams Syndrome	Chronological Age	WISC-R VIQ/PIQ/FIQ			Gender
LMH	17;6	84	66	72	Male
CYJ	18;4	50	53	48	Male

Design and Materials

The design for target clauses and test sentences for factuals and counterfactuals was the same as the design for unimpaired participants. All the materials (including experimental trials and fillers) were also the same, which could be referenced from Appendix 1 to Appendix 9, and all the trials were presented randomly. The only different point from the tasks given to unimpaired participants was that there was no time limitation of SOA for participants with Williams Syndrome. The detail was given in Procedure below.

Procedure

We tried to make this task like a game, which would not make participants with Williams Syndrome feel bored. Thus, they were told that they were going to read a story on computer. In order to save time, the story usually was very short. But, since computer very often made mistakes about the content of the story, after presentation of the story, the computer would make a statement about the story to participants. So, participants were instructed to be very patient and kind to teach the computer the right statement of the story. By doing so, the computer would not make mistakes next time. If the computer said something wrong, which was incompatible with the content of the story, participants were encouraged to respond 'correct' or 'incorrect' to the computer. However, the only way to communicate with a computer was clicking a mouse. So, if the computer made a correct statement, participants should click the left button of the mouse; on the other hand, if the computer made an incorrect statement about the story, participants should click the right button of the mouse. After clicking the button, a new story would show on the computer screen. This instruction made participants feel interested about the task very much because they had never thought of teaching a computer before. They were also very proud that they were smarter than computer.

Actually, the procedure was like the following. In the beginning of the task, a fixation point with a beep sound was displayed for 500ms. After this reminder of a story, a target sentence was presented on the computer screen. Participants with Williams Syndrome were instructed to read and comprehend the sentence. They could click any buttons of the mouse after they fully understood the sentence. Then, a test sentence which may probe the if-clause or the consequence clause would appear on the computer screen. They should make judgment whether the test sentence was compatible with the target clause. That is, they had to decide whether the meaning of a test sentence matched with the meaning of a target clause. If they matched, participants should click the left button of the mouse; if they mismatched, participants should click the right button of the mouse. After clicking the button, a new experimental trial would appear on the computer screen.

TSING HUA UNIVERSITY

Prediction

For factual target clauses, since representations of form and meaning are the same, it is hypothesized that children with Williams Syndrome would follow predictions of CCM (Carpenter, 1973). In other words, it is expected to see the ordering of the four conditions of test sentences like this: TA < FA < FN < TN. That is, affirmative target clauses should receive faster response latency and lower error rates than negative target clauses. On the contrary, for counterfactual target clauses, since there might be a discrepancy between form constructing and meaning understanding for children with Williams Syndrome (Karmiloff-Smith, et al, 1997; Zukowski, 2001; Mervis, et al, 2002), it is hypothesized that a representation based on form would be constructed. That is, the ordering of the four conditions of test sentences would be like: FN < TN < TA < FA as alternative one. It is also possible that the ordering would be TN < FN < FA < TA as alternative two. In this situation, it is inferred that a complex representation is formed because the matched test sentences in predicates and/or in polarity receive faster response latency or higher accuracy than mismatched test sentences. On the other hand, if children with Williams Syndrome form a simpler representation based on meaning like unimpaired participants, the ordering of counterfactual would be exactly like the predictions made for factual target clauses, namely, alternative three.

Results and Discussions

Since participants with Williams Syndrome may have their own characteristics in performance, their data would be presented by individual first and by group later. The latencies and error rates to respond to counterfactual and factual clauses were shown in Table 21 below.

Participant LMH

Counterfactual vs. factual target clauses

LMH responded to factual target clauses and counterfactual target clauses were 1239ms and 1381ms, respectively. These two response latencies were not significantly different ($F(1, 100) = 1.66$, $p = .20$). LMH made more errors on test sentences probing counterfactual target clauses than test sentences probing factual target causes.

For counterfactual target clauses, there was also a significant interaction between truth values and polarities ($F(1, 39) = 0.13$, $p = .71$).

For factual target clauses, there was an interaction between truth values and polarities and this interaction was not significantly different ($F(1, 55) = 3.37$, $p = .07$).

The interaction of the four conditions on counterfactual and factual target clauses was not significant (F (3, 94) = 0.56, p = .64), suggesting that response latency on factual target clauses was not significantly faster than response latency on counterfactual target clauses. Meanwhile, the difference of each condition between counterfactual and factual target clauses was not significant. The main effect of target clauses was not significant (F (1, 94) = 3.33, p = .07), but the main effect of the four conditions reached significance (F (3, 94) = 9.64, p < .0001). All the comparisons were not significant.

Counterfactual and factual clause position

Response latency to counterfactual target clauses in if-clause was 1330ms and in consequence clause was 1429ms. The difference of their latency also did not reach significance (F (1, 41) = 0.32, p = .57), suggesting clause positions for counterfactual targets also did not make any difference in processing. The same pattern was found on their error rates (F (1, 41) = 1.10, p = .30), suggesting that participants did not make more errors because of the clause positions.

Response latency to factual target clauses in if-clause was 1236ms and in consequence clause was 1243ms. The difference of their latency did not reach significance (F (1, 57) = 0.00, p = .96), suggesting clause positions for factual targets did not have influence in processing for LMH. Their error rates also did not show any difference (F (1, 57) = 0.87, p = .35), suggesting that participants did not make more errors because of the clause positions.

For LMH with Williams Syndrome, clause positions did not cause processing difference no matter on factual or counterfactual target clauses.

Test sentences in counterfactual targets

For counterfactual target sentences, the condition ordering was like the following:

TA < TN < FA < FN, which did not match any predictions. It seemed that truth value was a more important factor to this child with WS. LMH responded fastest to test sentences with true affirmatives (1021ms), next was to test sentences with true negatives (1354ms), the third was to test sentences with false affirmatives (1566ms), and the last was to test sentences with false negatives (1788ms). A one-way ANOVA showed that the difference between these the four conditions was significant ($F (3, 39) = 5.26$, $p = .003$). A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. The major difference was attributed to the comparisons of TA and two other groups (i.e. TA vs. FA, $p = .02$; TA vs. FN, $p = .0005$) and also the comparison of TA and TN ($p = .04$).

Parallel to Carpenter's findings on counterfactual clauses and modified by Chinese stimuli, as predicted, test sentences with matched representations in predicates were responded significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA ($p = .02$). The main effect of the four conditions of error rates for counterfactual target clauses reached significance ($F (3, 39) = 3.46$, $p = .02$). The major difference came from the comparisons of FA vs. FN ($p = .0241$), FA vs. TA ($p = .003$) and FA vs. TN ($p = .007$).

Test sentences in factual targets

For factual target clauses, the condition ordering from the easiest to the hardest was like the following: TA < TN < FA < FN. That is, participants responded fastest to test sentences with true affirmatives (856ms), next was to test sentences with true negatives (1336ms), the third was to test sentences with false affirmatives (1397ms), and the last was to test sentences with false negatives (1419ms). A one-way ANOVA showed that the difference between these four conditions was significant, $F (3, 55) = 4.77$, $p = .005$. A proc mixed model with a post hoc test of least significance means

(LSMEANS) by using Tukey method was employed.

The results showed the major difference came from the comparisons between TA and other conditions, i.e. TA vs. FA ($p = .0037$), TA vs. FN ($p = .0022$), TA vs. TN ($p = .006$).

Parallel to Carpenter's findings on factual clauses, test sentences with matched representations in predicates were responded significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA ($p = .003$). The difference of the four conditions on error rates for factual target clauses did not reach significant difference, $F (3, 55) = 1.23$, $p = .30$ and all the comparisons of error rates were not significant different to one another.

Counterfactual vs. factual sentence type

Response latencies for counterfactual target clauses in four experimental sentence types were 1256ms, 1347ms, 1432ms, and 1467ms, respectively. A one-way ANOVA results did not show significant difference among them ($F (3, 39) = 0.32$, $p = .81$), suggesting that none of the counterfactual target clauses in these four experimental sentences caused more difficulty than others. They were all similar in degree of difficulty. The same pattern was found on their error rates in one-way ANOVA ($F (3, 39) = 1.07$, $p = .37$).

Response latencies for factual target clauses in four experimental sentence types were 1425ms, 1147ms, 1134ms, and 1262ms, respectively. A one-way ANOVA results did not show significant difference among them ($F (3, 55) = 0.97$, $p = .41$), suggesting that none of the factual target clauses in these four experimental sentences caused more difficulty than others. They were all similar in degree of difficulty. The same pattern was found on their error rates in one-way ANOVA ($F (3, 55) = 1.15$, $p = .33$).

Affirmative vs. negative test sentences

Test sentences with affirmatives were responded faster than test sentences with negatives for counterfactual target clauses (1202ms and 1551ms, respectively) (F (1, 41) = 4.49, p = .04). The difference in their error rates was also not significant (F (1, 41) = 1.21, p = .27). Test sentences with affirmatives were responded faster than test sentences with negatives for factual target clauses (1099ms and 1375ms, respectively) (F (1, 57) = 4.24, p = .04), suggesting test sentences with affirmatives were easier than test sentences with negatives for factual target clauses to LMH. The difference in their error rates was not significant (F (1, 57) = 1.15, p = .28).

For LMH, the effect of polarity did make difference no matter on counterfactual or factual target clauses as unimpaired participants.

TSING HUA UNIVERSITY

Table 21 Response Latency (in ms) and Error Rates in Delayed Task with Negation for Participant with Williams Syndrome (LMH)

Type of Clause		TA	FA	FN	TN	Total
Factual	RT	856.79	1397.67	1419.29	1336.99	1239.66
	Errors	0.0583	0.2898	0.2708	0.1071	7.81%
Counterfactual	RT	1021.01	1566.03	1788.24	1354.78	1381.31
	Errors	0.2292	0.4922	0.4306	0.3651	32.81%

Participant CYJ

The latencies and error rates to respond to factual and counterfactual clauses were shown in Table 22 below.

Counterfactual vs. factual target clauses

For counterfactual target clauses, the interaction between truth values and polarities was also not significant (F (1, 30) = 0.03, p = .86). Participants responded to factual target clauses and counterfactual target clauses were 2295ms and 2404ms, respectively. These two response latencies were significantly different (F (1, 74) = 0.16, p = .68).

For factual target clauses, there was an interaction between truth values and polarities and this interaction reached significant difference (F (1, 38) = 4.04, p = .05).

The interaction of the four conditions on factual and counterfactual target clauses was not significant (F (3, 68) = 1.01, p = .39), suggesting the response latency of test sentences for factual target clauses was not faster than test sentences for counterfactual target clauses.

Counterfactual vs. factual clause position

Response latency to counterfactual target clauses in if-clause was 2126ms and in consequence clause was 2681ms. The difference of their latency did not reach significance (F (1, 32) = 2.70, p = .11), suggesting clause positions for counterfactual targets also did not make difference in processing. The same pattern was found on their error rates (F (1, 32) = 1.06, p = .31), suggesting that participants did not make more errors because of the clause positions.

Response latency to factual target clauses in if-clause was 2257ms and in consequence clause was 2338ms. The difference of their latency did not reach significance (F (1, 40) = 0.04, p = .83), suggesting clause positions for factual targets did not have influence in processing. Their error rates did not show any difference (F (1, 40) = 0.95, p = .33), suggesting that participants did not make more errors because of the clause positions.

For CYJ, clause positions did not cause processing difference no matter on counterfactual or factual target clauses.

Test sentences in counterfactual targets

For counterfactual target sentences, the condition ordering was like the following: FA < TN < FN < TA. This pattern does not match any of the alternatives. CYJ responded fastest to test sentences with false affirmatives (2366ms), next was to test sentences with true negatives (2374ms), the third was to test sentences with false negatives (2410ms), and the last was to test sentences with true affirmatives (2453ms). A one-way ANOVA showed that the difference between these the four conditions was significant ($F(3, 30) = 0.01$, $p = .99$). Contrary to Carpenter's finding, mismatched predicates (i.e. FA) were responded faster than matched predicates (i.e. TA), but the difference was not significant ($p = .85$). The difference of error rates for counterfactual target clauses in the four conditions was not significant ($F(3, 30) = 0.49$, $p = .69$).

Test sentences in factual targets

For factual target clauses, the condition ordering from the easiest to the hardest to participant CYJ was like the following: FN < TA < TN < FA. This ordering was very different from the prediction: TA < FA < FN < TN. It seemed that the ordering of truth value and polarity were all reversed. Williams Syndrome child CYJ responded fastest to test sentences with false negatives (1644ms), then to test sentences with true affirmatives (2134ms), the third was to test sentences with true negatives (2415ms), and the last was to test sentences with false affirmatives (2894ms). A one-way ANOVA showed that the difference between these the four conditions was not significant ($F(3, 38) = 1.58$, $p = .20$). A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed, showing almost all the comparisons were not significant except one comparison was (FA vs. FN, p = .04).

Parallel to Carpenter's findings on factual clauses, test sentences with matched representations in predicates did not respond significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA (p = .14). Error rates for factual target clauses in the four conditions did not reach significant difference to one another (F (3, 38) = 1.71, p = .18).

Counterfactual vs. factual sentence type

Response latencies for counterfactual target clauses in four experimental sentence types were 2831ms, 2573ms, 2594ms, and 1873ms, respectively. A one-way ANOVA results did not show significant difference among them (F (3, 30) = 1.52, p = .22), suggesting that none of the counterfactual target clauses in these four experimental sentences caused more difficulty than others. The same pattern was found on their error rates in one-way ANOVA (F (3, 30) = 3.33, p = .03).

Response latencies for factual target clauses in four experimental sentence types were 2339ms, 2535ms, 2289ms, and 1961ms, respectively. A one-way ANOVA results did not show significant difference among them (F (3, 38) = 0.41, p = .74), suggesting that none of the factual target clauses in these four experimental sentences caused more difficulty than others. They were all similar in degree of difficulty. The same pattern was found on their error rates in one-way ANOVA (F (3, 38) = 1.94, p = .13).

Affirmative vs. negative test sentences

Test sentences with affirmatives were responded faster than test sentences with negatives for counterfactual target clauses (2412ms and 2393ms, respectively) ($F(1, 32) = 0.00$, $p = .95$). The difference in their error rates was also not significant ($F(1, 32) = 0.94$, $p = .33$).

Test sentences with affirmatives were responded faster than test sentences with negatives for factual target clauses (2431ms and 2131ms, respectively) ($F(1, 40) = 0.60$, $p = .44$). The difference in their error rates was not significant ($F(1, 40) = 0.95$, $p = .33$).

To sum up, the effect of polarity did not make difference to CYJ no matter on factual target clauses or counterfactual target clauses.

Table 22 Response Latency (in ms) and Error Rates in Delayed Task with Negation for Participant with Williams Syndrome (CYJ)

Type of Clause		TA	FA	FN	TN	Total
Factual	RT	2134.46	2894.40	1644.00	2415.82	2295.95
	Errors	12.50%	43.75%	56.25%	25.00%	34.37%
Counterfactual	RT	2453.41	2366.54	2410.85	2374.51	2404.16
	Errors	37.50%	43.75%	50.00%	56.25%	46.88%

Group Data Analysis

The average latencies and error rates to respond to factual and counterfactual clauses of two Williams Syndrome children were listed in Table 23.

Counterfactual vs. factual target clauses

Participants responded to counterfactual target clauses and factual target clauses were 1891ms and 1763ms, respectively. These two response latencies were not significantly different ($F(1, 174) = 0.02$, $p = .89$). Meanwhile, participants made more errors on test sentences probing counterfactual target clauses than test sentences probing factual target clauses ($F(1, 175) = 1768.97$, $p < .0001$).

For counterfactual target clauses, the interaction between truth values and polarities was not significant ($F(1, 72) = 0.00$, $p = .99$). The same pattern was found on error rates ($F(1, 73) = 0.02$, $p = .88$).

For factual target clauses, there was an interaction between truth values and polarities and this interaction reached significant difference ($F(1, 96) = 5.33$, $p = .02$). The same pattern was not found on error rates ($F(1, 97) = 0.04$, $p = .84$). The interaction of the four conditions on factual and counterfactual target clauses was not significant ($F(3, 169) = 0.97$, $p = .40$), suggesting the response latency on factual target clauses was not faster than the latency on counterfactual target clauses. Meanwhile, the main effects of clause type and the four conditions were also not significant (for clause type, $p < .3183$; for the four conditions, $p < .09$). Thus, each condition was not significant difference between factual and counterfactual target clauses.

The interaction of the four conditions on error rates on both target clauses was significant, $F(3, 169) = 0.37$, $p = .77$, suggesting participants did not make more errors on counterfactual target clauses than on factual target clauses. The main effect of clause type was significant ($p < .0001$), but the main effect of the four conditions was not ($p = .6343$). Thus, the difference of each condition on error rates between factual and counterfactual target clauses reached highly significance ($p < .0001$), suggesting that children with Williams Syndrome erred more on counterfactual target clauses than on factual target clauses on every condition as other groups.

Counterfactual vs. factual clause position

Response latency to counterfactual target clauses in if-clause was 1687ms and in consequence clause was 1975ms. The difference of their latency did not reach significance (F (1, 74) = 2.75, p = .10), suggesting clause positions for counterfactual targets did not make difference in processing. The same pattern was found on their error rates (F (1, 75) = 0.01, p = .91), suggesting that participants did not make more errors because of the clause positions.

Response latency to factual target clauses in if-clause was 1652ms and in consequence clause was 1709ms. The difference of their latency did not reach significance (F (1, 98) = 0.04, p = .83), suggesting clause positions for factual targets did not have influence in processing. Their error rates did not show any difference (F (1, 99) = 0.03, p = .85), suggesting that participants did not make more errors because of the clause positions.

For children with Williams Syndrome, clause positions did not cause processing difference no matter on counterfactual or factual target clauses.

Test sentences in counterfactual targets

For counterfactual target sentences, the condition ordering was like the following: TA < TN < FA < FN. This result seemed to indicate that truth value factor was a more important factor for children with Williams Syndrome to make judgment. Participants responded fastest to test sentences with true affirmatives (1617ms), next was to test sentences with true negatives (1730ms), the third was to test sentences with false affirmatives (2016ms), and the last was to test sentences with false negatives (2064ms). A one-way ANOVA showed that the difference between these the four conditions was not significant (F (3, 72) = 1.01, p = .39). The same finding was observed on error rates of counterfactuals (F (3, 73) = 0.48, p = .69).

Contrary to Carpenter’s findings on counterfactual clauses, test sentences with matched representations in predicates were not responded significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA (p = .31).

Test sentences in factual targets

For factual target clauses, the condition ordering for children with Williams Syndrome from the easiest to the hardest was like the following: TA < FN < TN < FA. Participants responded fastest to test sentences with true affirmatives (1453ms), next was to test sentences with false negatives (1494ms), the third was to test sentences with true negatives (1799ms), and the last was to test sentences with false affirmatives (2009ms). A one-way ANOVA showed that the difference between these the four conditions was not significant (F (3, 96) = 2.27, p = .08). The same finding was also observed on error rates of factuals (F (3, 97) = 0.31, p = .81).

Parallel to Carpenter's findings on factual clauses, test sentences with matched representations in predicates were responded significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA (p = .01).

Counterfactual vs. factual sentence type

Response latencies for counterfactual target clauses in four experimental sentence types were 1676ms, 2028ms, 1960ms, and 1651ms, respectively. A one-way ANOVA results did not show significant difference among them (F (3, 72) = 0.66, p = .58), suggesting that none of the counterfactual target clauses in these four experimental sentences caused more difficulty than others. The same pattern was found on their error rates in one-way ANOVA (F (3, 73) = 0.98, p = .40).

Response latencies for factual target clauses in four experimental sentence types were 1730ms, 1791ms, 1648ms, and 1542ms, respectively. A one-way ANOVA results did not show significant difference among them (F (3, 96) = 0.39, p = .76), suggesting that none of the factual target clauses in these four experimental sentences caused more difficulty than others. The same pattern was found on their error rates in one-way ANOVA (F (3, 97) = 0.33, p = .80).

Affirmative vs. negative test sentences

Test sentences with affirmatives were responded faster than test sentences with negatives for counterfactual target clauses (1777ms and 1893ms, respectively) (F (1, 74) = 1.03, p = .31). The difference in their error rates was also not significant (F (1, 75) = 0.38, p = .53). Test sentences with affirmatives were responded faster than test sentences with negatives for factual target clauses (1688ms and 1668ms, respectively) (F (1, 98) = 0.04, p = .84). The difference in their error rates was not significant (F (1, 99) = 0.31, p = .57).

Table 23 Response Latency (in ms) and Error Rates in Delayed Task with Negation for Children with Williams Syndrome

Type of Clause		TA	FA	FN	TN	Total
Factual	RT	1453.04	2009.97	1494.19	1799.35	1763.78
	Errors	11.33%	11.72%	10.60%	10.83%	20.63%
Counterfactual	RT	1617.85	2016.32	2064.96	1730.47	1891.89
	Errors	37.02%	45.63%	39.38%	36.90%	40.25%

General Discussion of Individual Difference

Table 24 listed individual data of the latencies and error rates to respond to factual and counterfactual clauses of two Williams Syndrome children.

Counterfactual vs. factual target clauses

An interaction of response latency of factual and counterfactual target clauses showed significant difference between two participants ($F(3, 174) = 21.61$, $p < .0001$) and each comparison was highly significant ($p < .0001$). This finding showed that Williams Syndrome child LMH in general performed faster reaction times on both target clauses.

The interaction between truth values and polarity on counterfactual target clauses for three groups reached significant difference ($F(7, 69) = 5.70$, $p < .0001$). Almost all the comparisons were significant except FN condition, suggesting that test sentences with false responses in negatives for probing counterfactual target clauses were all difficult to these two children with WS. This difficulty may result from the completely congruent of representations between target clauses and test sentences.

An interaction of truth values and polarity on factual target clauses was also observed ($F(7, 93) = 7.38$, $p < .0001$) and three comparisons reached significance except one comparison (i.e. FN).

Counterfactual vs. factual clause position

The interaction between clause positions and different participants on counterfactual target clauses was significant ($F(3, 73) = 12.42$, $p < .0001$), suggesting that LMH in general responded faster to test sentences for both clauses than CYJ. A post hoc comparison by using Tukey method showed that both if-clauses and consequence clause were highly significant (i.e. for if-clauses, $p = .0025$; for consequence clauses, $p < .0001$).

The interaction between clause positions and participants on factual target clauses was also significant ($F(3, 97) = 11.12$, $p < .0001$). These findings indicated that LMH performed overall faster response latency on both clauses than CYJ and the difference of each comparison was highly significance at .0001. The pattern was exactly parallel to the findings on counterfactual target clauses.

Also, the interaction between participants and different experimental sentences was significant on counterfactuals ($F (7, 69) = 5.97$, $p < .0001$). The interaction between participants and different experimental sentences was also significant on factual target clauses ($F (7, 93) = 5.38$, $p < .0001$). Generally speaking, WS child LMH showed faster reaction times on experimental sentences of factual and counterfactual target clauses than the other WS child CYJ except one experimental sentence (i.e. sentence type 4).

Affirmative vs. negative test sentences

An interaction between participants and polarity on counterfactual target clauses was also significant ($F (3, 73) = 11.31$, $p < .0001$). Once more, all the comparisons between affirmatives and negatives on counterfactual target clauses of LMH and CYJ were significantly different.

The interaction between participants and polarity on factual target clauses was not significant ($F (3, 97) = 12.22$, $p < .0001$). All the comparisons between affirmatives and negatives on factual target clauses of LMH and CYJ were significantly different.

These results indicated that LMH showed much faster response latency on both test sentences with affirmatives and negatives. However, for CYJ, this WS child showed not only an indifference processing on test sentences with affirmatives and negatives, but also slower response latency than the other child CYJ.

Table 24 Response Latency (in ms) and Error Rates in Delayed Task with Negation for Individuals with Williams Syndrome

Type of Clause	Participant	TA	FA	FN	TN	Total
Factual	LMH	856.79	1397.67	1419.29	1336.99	1239.66
		0%	18.75%	12.50%	0%	7.81%
	CYJ	2134.46	2894.40	1644.00	2415.82	2295.95
		12.50%	43.75%	56.25%	25.00%	34.37%
Counterfactual	LMH	1021.01	1566.03	1788.24	1354.78	1381.31
		12.50%	56.25%	37.50%	25.00%	32.81%
	CYJ	2453.41	2366.54	2410.85	2374.51	2404.16
		37.50%	43.75%	50.00%	56.25%	46.88%

Summary

The orderings of the four conditions on factual and counterfactual target clauses are very different for two children with WS. The first child, LMH, showed the same orderings on both factual and counterfactual target clauses; however, the second child, CYJ, did not show the same pattern on two target clauses. However, LMH's performance is more like the sixth graders, in which TA condition is the easiest, TN condition is the hardest, and FA/FN conditions are in between. But, this similarity is only on counterfactuals.

Their performance on factuals is different. The sixth graders performed exactly as the prediction on factuals based on CCM, but LMH did not perform the pattern. Meanwhile, the p-values of factuals and counterfactuals in response latency are significant and the p-value in error rates is significant only on counterfactuals. Neither the other WS child nor the group data reached significance in both dependent variables. Interestingly, none of the p values on the comparisons of factual and counterfactual target clauses reached significance, suggesting that for WS participants factual clauses are not easier than counterfactuals in processing. But, the group data in error rates is significant difference.

The interaction of truth values and polarities on factuals is significantly different in group analysis, but none of the individuals showed the significant pattern. The indifference results are also observed on counterfactuals. Interaction of the four conditions on both factual and counterfactual target clauses is not significant in p value no matter in group or in individual analysis. These results are different from the ones observed in delayed task on unimpaired participants.

For unimpaired participants, they performed differently on each condition of factual and counterfactual target clauses. But, none of the individuals with WS showed this difference. Besides, as predicted, individuals with WS did not show any clause effect or sentence type effect on both target clauses as unimpaired participants (except the error rates of counterfactuals of CYJ). On the most important index of linguistic processing in this task is polarity effect, only LMH showed this effect, but not CYJ.

WS child LMH is more like unimpaired participants and his pattern is similar to the sixth graders. As to another WS child, CYJ, his pattern is very different from unimpaired participants, even from child with WS, LMH.

Table 25 Summary Findings of Two Children with Williams Syndrome in Delayed Experiment of Counterfactual Conditionals

		LMH	CYJ	GROUP
Factual Ordering		TA < TN < FA < FN	FN < TA < TN < FA	TA < FN < TN < FA
Factual p-value	(RT)	p = .005	p = .20	p = .08

HUA UNIVERSITY

	(Error Rates)	p = .30	p = .18	p = .81
Counterfactual Ordering		TA < TN < FA < FN	FA < TN < FN < TA	TA < TN < FA < FN
Counterfactual p-value	(RT)	p = .003	p = .99	p = .39
	(Error Rates)	p = .02	p = .69	p = .69
Factual vs. Counterfactual p-value	(RT)	p = .20	p = .68	p = .89
	(Error Rates)	---	---	p < .0001
Factual interaction of truth values and polarities	(RT)	p = .07	p = .05	p = .02
	(Error Rates)	---	---	p = .84
Counterfactual interaction of truth values and polarities	(RT)	p = .71	p = .86	p = .99
	(Error Rates)	---	---	p = .88
Interaction of the four conditions on factual and counterfactual	(RT)	p = .6420	p = .3934	p = .40
	(Error Rates)	---	---	p = .77
Factual clause effect	(RT)	p = .96	p = .83	p = .83
	(Error Rates)	p = .35	p = .33	p = .85
Counterfactual	(RT)	p = .57	p = .11	p = .10

clause effect	(Error Rates)	p = .30	p = .31	p = .91
Factual sentence type effect	(RT)	p = .41	p = .74	p = .76
	(Error Rates)	p = .33	p = .13	p = .80
Counterfactual sentence type effect	(RT)	p = .81	p = .22	p = .58
	(Error Rates)	p = .37	p = .03	p = .40
Factual polarity effect	(RT)	p = .04	p = .44	p = .84
	(Error Rates)	p = .28	p = .33	p = .57
Counterfactual polarity effect	(RT)	p = .04	p = .95	p = .31
	(Error Rates)	p = .27	p = .33	p = .53

H Comparison of Individual of the Sixth Graders and Individuals with Williams Syndrome

Due to the similarity in performance of WS child LMH and the sixth graders in counterfactuals, it is curious whether there are the sixth individuals performed this pattern. Thus, data of twenty nine the sixth graders was analyzed individually. The results showed there are roughly twelve patterns found. These patterns along with participant numbers were listed in Table 26 below. There are only four out of twenty nine the sixth graders performed the exact ordering as LMH: TA < TN < FA < FN (i.e.

Pattern (2), Table 26). However, if labeled the sequence of TA < TN as the same pattern, then twenty four out of twenty nine (83%) students showed this pattern. On the contrary, if the sequence of FA < FN is counted as the same pattern, only seventeen out of twenty nine (59%) students showed this pattern.

Since CCM predicted condition with true affirmatives should be the easiest in processing and condition with true negatives should be the hardest based on involved mental operations (and conditions with false truth values are in between), the sequence of TA < TN is counted as the same pattern no matter FA/FN in what order. Thus, the pattern of LMH performed is like the majority pattern performed by the sixth graders. In sum, LMH did show a normal pattern as his mental age controls, suggesting a meaning representation of counterfactuals in processing. However, this match in counterfactuals did not show in factual results. WS child LMH showed the same pattern as counterfactuals on factual target clauses, which is different from the results observed on the sixth graders.

For the sixth graders, they performed as the CCM predicted that condition TA is the easiest, TN is the hardest, and FA/FN is in between. Though the ordering is different, the basic pattern (i.e. TA < TN) is the same. Meanwhile, WS child LMH showed very consistent pattern on both factual and counterfactual target clauses in which condition TA is in the first order and condition TN is the second order. It seems to mean that false responses caused bigger effect than true responses on both factuals and counterfactuals for LMH. To sum up results of LMH, it could be concluded that his logical reasoning of linguistic ability is spared.

As to another WS child CYJ, for counterfactuals, his results in response latency and error rates are conflict. In response latency, CYJ showed FA < TN < FN < TA ordering, which matched the basic ordering of predictions one and two (i.e. TN < TA) based on grammatical structures of counterfactuals. This ordering cannot be found in any of individual the sixth patterns. But, in error rates, CYJ showed the results of prediction three which is represented based on sentential meaning.

For factuals, both results in response latency and error rates are the same in basic pattern (i.e. TA < TN). Further, the error rates pattern of CYJ is the same as the pattern showed of LMH, suggesting false responses did cause bigger effect in processing than true responses. If error rate is considered as a more reliable index, it is concluded that CYJ also showed a normal pattern as predicted.

In conclude, the result of CYJ, his logical reasoning ability is sort of spared as LMH.

Table 26 Individual Patterns of Counterfactual Target Clauses in The Sixth Graders

PATTERN	ORDERING	NUMBER
(1)	TA<FA<TN<FN	8
(2)	TA<TN<FA<FN	4
(3)	TA<FA<FN<TN	3
(4)	TA<FN<FA<TN	2
(5)	TA<FN<TN<FA	2
(6)	TA<TN<FN<FA	2
(7)	TN<TA<FN<FA	2
(8)	TN<FA<TA<FN	1
(9)	TN<FN<FA<TA	1
(10)	FA<TA<FN<TN	1
(11)	FN<TA<TN<FA	2
(12)	FN<TN<FA<TA	1

I General Discussion for Counterfactual Conditionals with Negation Experiments

In this study Carpenter and Just's (1973) constituent comparison model was employed. Based on CCM, number of mental operations was calculated in verification. We hypothesized that mental representations would be formed once input through different modalities was received. Comparisons of constituents were possible only when the representations were formed in working memory. Based on the representations, there were four stages hypothesized in comparison, and the comparison was from innermost proposition to outermost proposition. Propositions should be compared exhaustively before making a response. Thus, number of mental operations could be calculated for successful comprehension. For factual and counterfactual target clauses, different orderings of test conditions were ranked according to the calculated number of mental operations. From the results observed in three age groups at 0-SOA and 5-second SOA, test condition with predicate congruency received faster response latency than test condition with negation congruency in true responses (i.e. TA < TN). That is, the embedded marker (i.e. predicate) was processed prior to the embedding marker (i.e. negation). As for test conditions with false responses which were completely congruency or in-congruency, these test conditions were not significantly different and their ordering was in between. This pattern is generally observed both on factual and counterfactual target clauses. In other words, for unimpaired populations, the results confirmed the hypothesis of CCM, which hypothesizes stages of mental operations in processing complicated sentences with verification, in this case, counterfactuals. Meanwhile, it is found that at 0-SOA, when target clause and test sentence were displayed at the same time, participants did not form a representation based on grammatical structures like English, but form a representation based on exact meaning. This result indicate that for Chinese, counterfactual representation in working memory is based on meaning, which makes content and function operation possible in processing.

In these tasks, age effect is clearly demonstrated. At 0-SOA, on both factual and counterfactual target clauses, the sixth graders performed differently from both the eighth graders and college students while the eighth graders performed similar to college students. The response latency and error rates of the eighth graders did not differ with the ones of college students on factual and counterfactual target clauses. But, the sixth graders showed the slowest response latency and higher error rates than other age groups on both target clauses. At 5-second SOA, similarly, three age groups differed with one another in which the sixth graders performed the slowest in response latency and lowest accuracy among, college students performed the fastest response latency and erred least while the eighth graders performed in between. These results indicate that counterfactual reasoning is developmentally different from childhood into adulthood. Meanwhile, task effect is also demonstrated. The response latency was significantly longer when SOA was zero, suggesting a heavier processing load under time limitation. Further, the task effect was also observed on target clauses, suggesting counterfactuals are more difficult in processing than factuals. The same pattern was also observed in error rates.

For individuals with WS, they were hypothesized to form representation based on grammatical structures rather than on meaning because of dissociation between form knowledge and meaning understanding. This study on counterfactuals, which is mismatched on syntax and semantics, serves as a good testing on them. The results showed that individuals with WS performed similarly to the predictions of unimpaired populations and the patterns were like the sixth graders. Thus, form and meaning dissociation hypothesis should be reconsidered because it does not find support through this study. In sum, individuals with WS have spared linguistic ability of logical reasoning. They are not deviant, but developmental delay.

TSING HUA UNIVERSITY

Fig. 1. Response latencies of factual target clauses for three groups of participants in simultaneous task with negation.

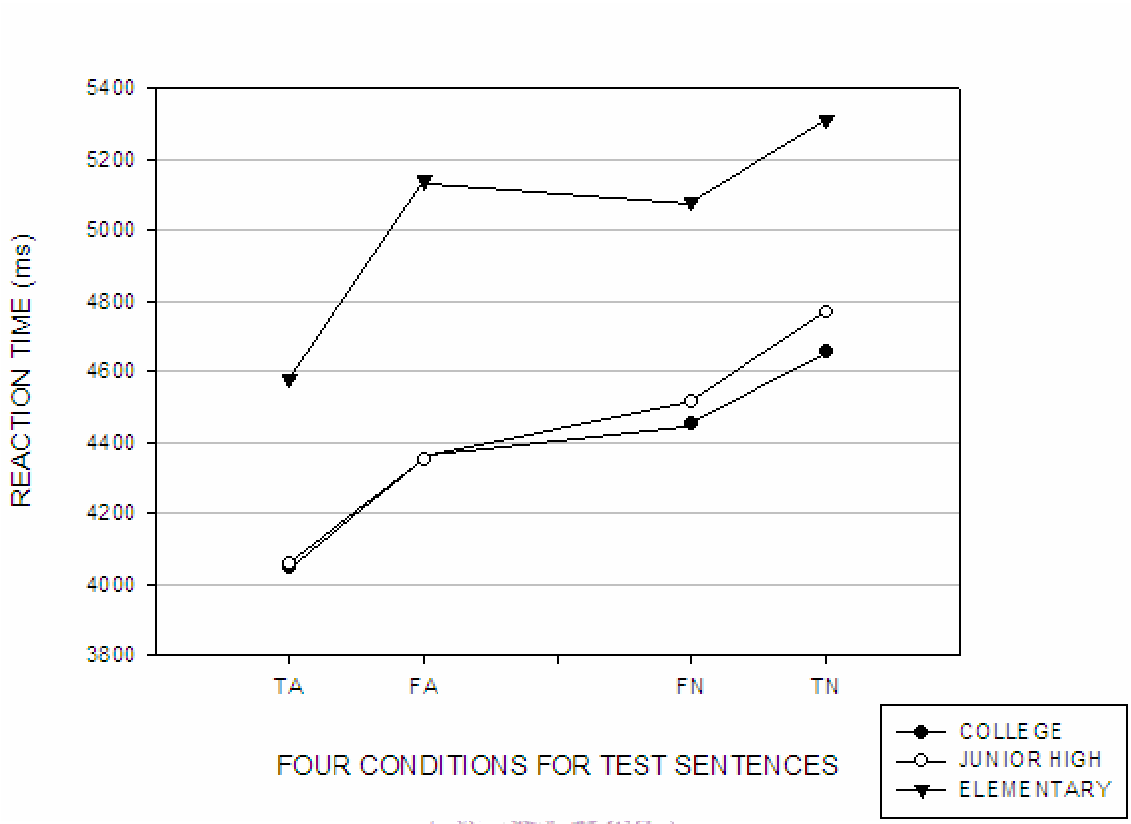


Fig. 2. Response latencies of counterfactual target clauses for three groups of participants in simultaneous task with negation.

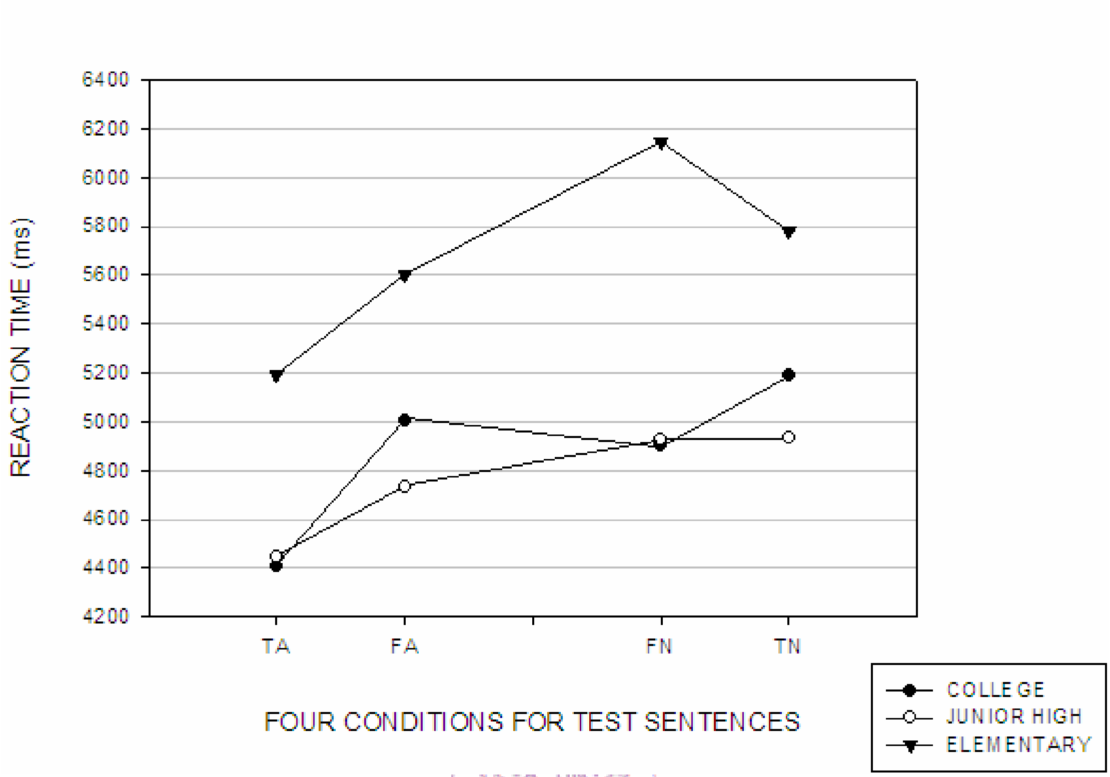


Fig. 3. Comparisons of response latencies of factual and counterfactual target clauses for three groups of participants in simultaneous task with negation.

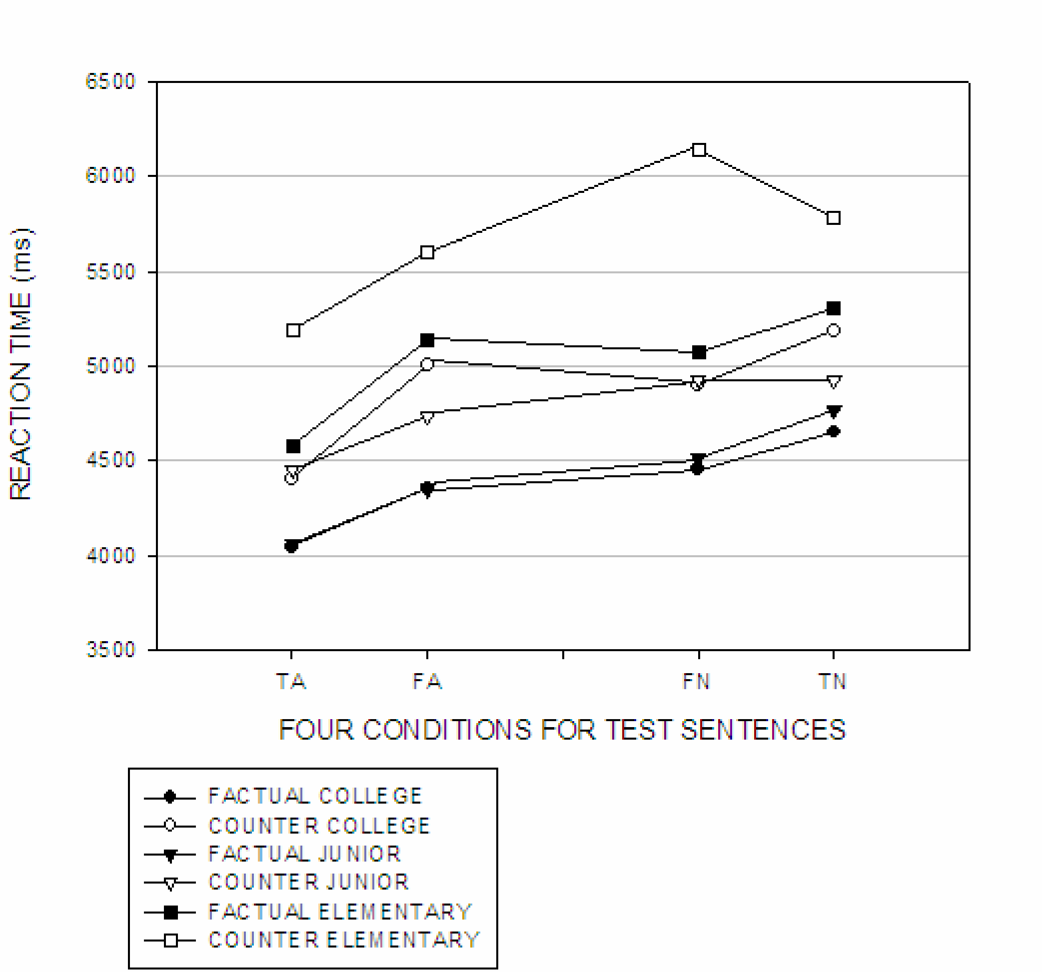


Fig. 4. Response latencies of factual target clauses for three groups of participants in delayed task with negation.

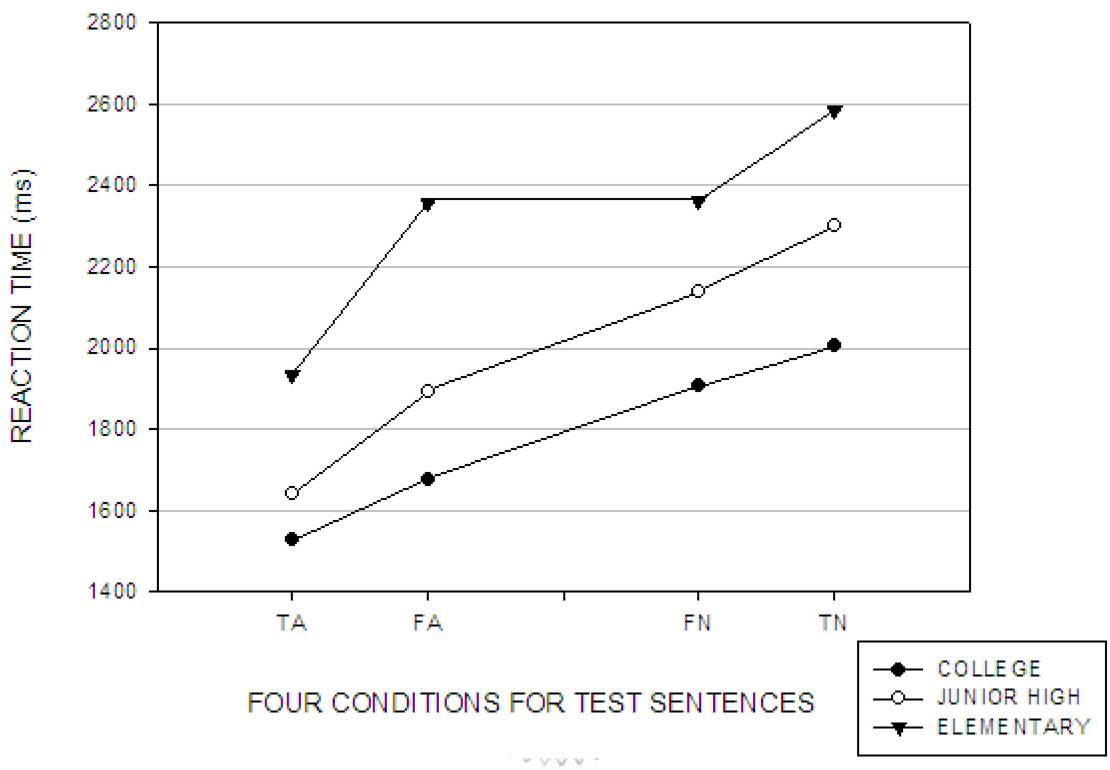


Fig. 5. Response latencies of counterfactual target clauses for three groups of participants in delayed task with negation.

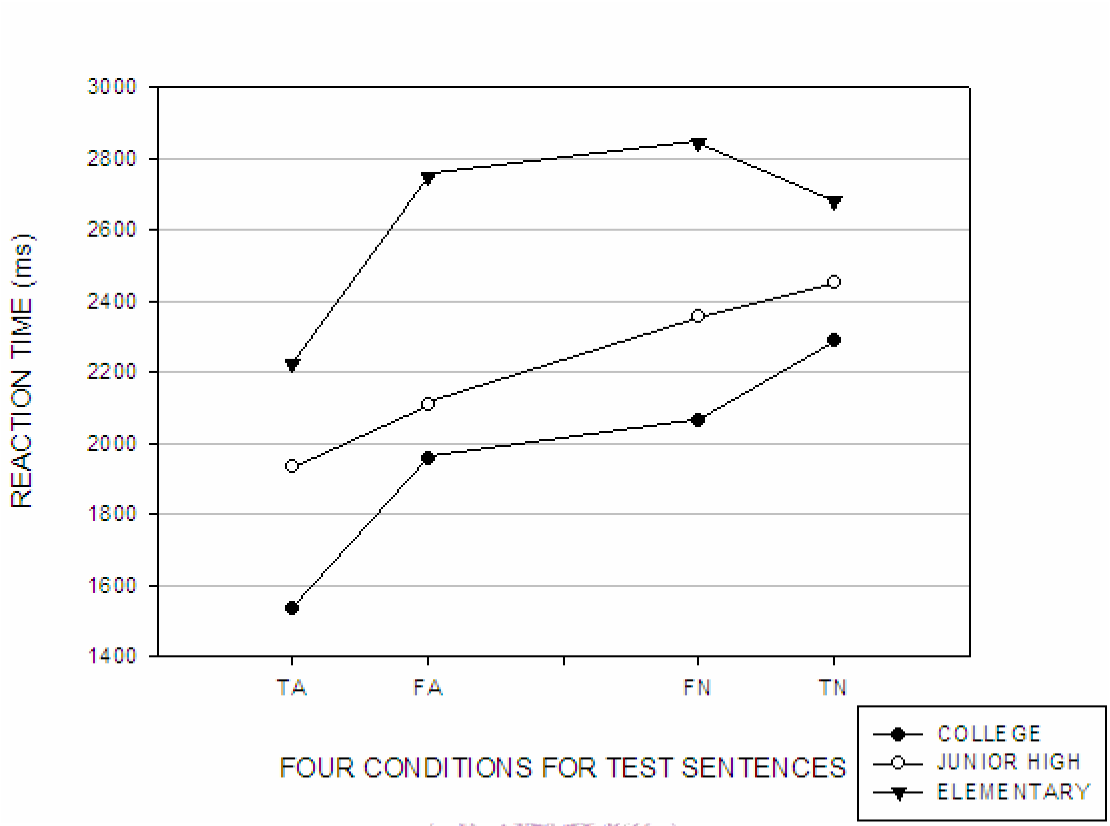


Fig. 6. Comparisons of response latencies of factual and counterfactual target clauses for three groups of participants in delayed task with negation.

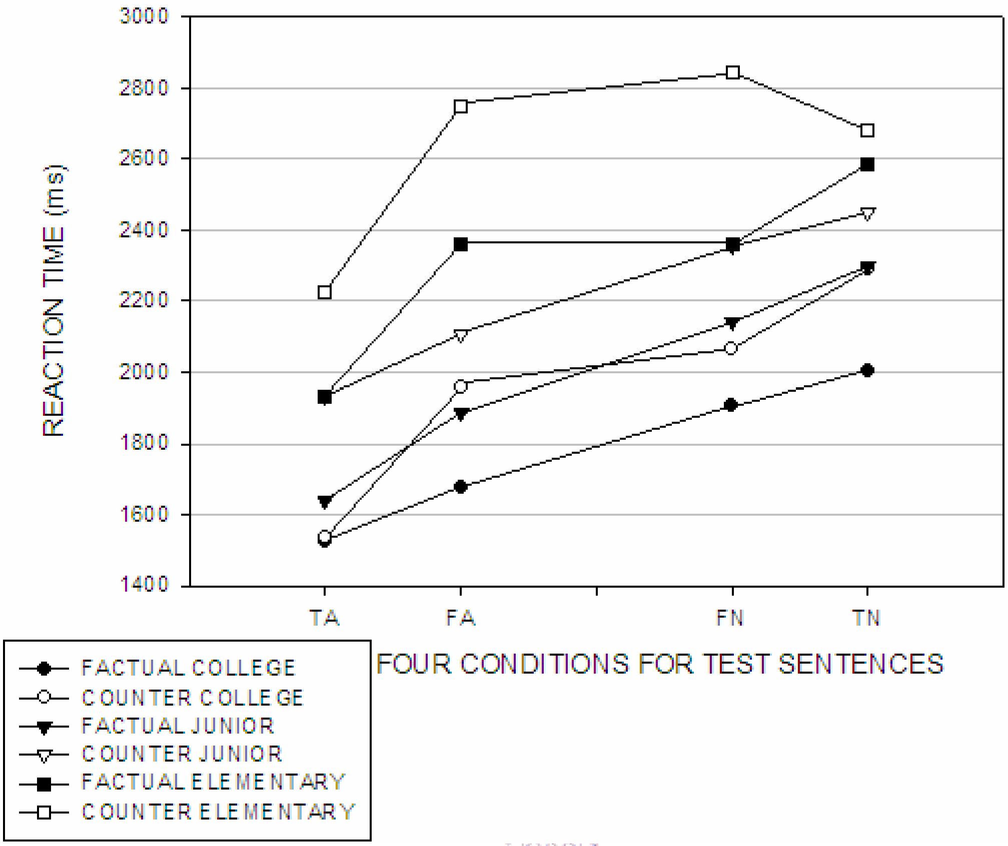
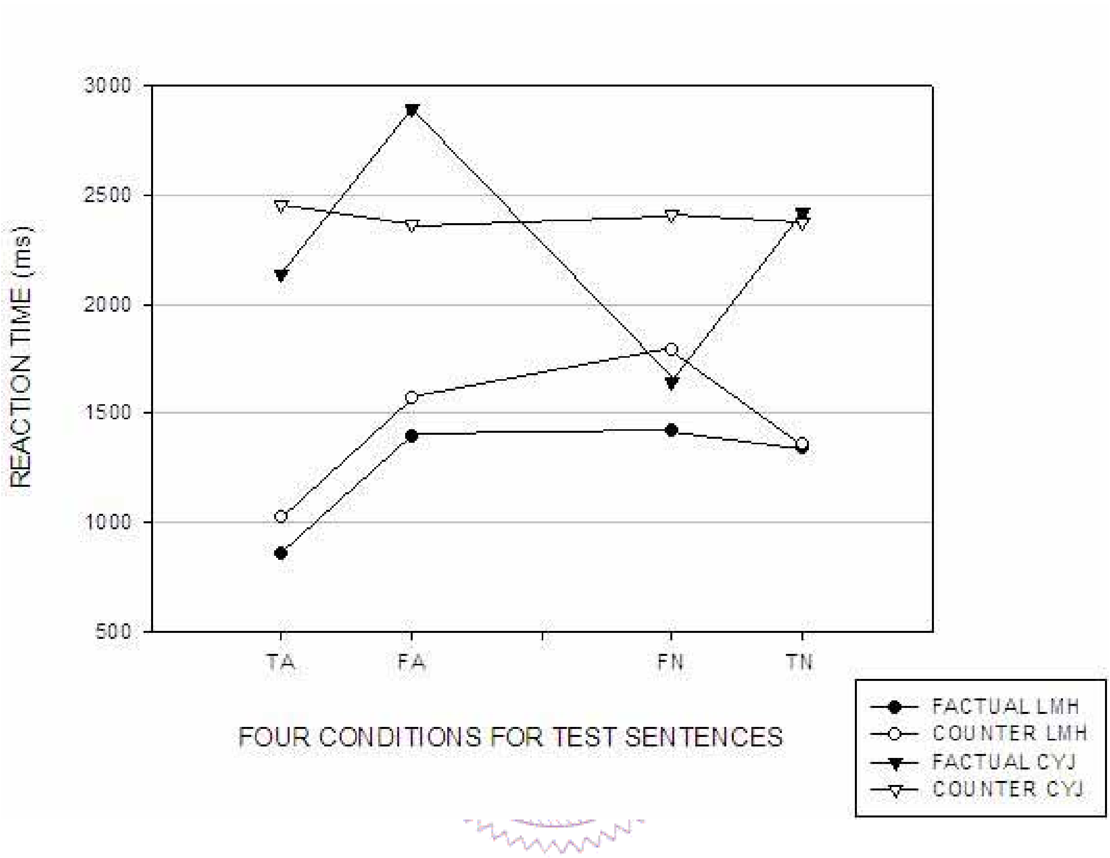


Fig. 7. Comparisons of response latencies of factual and counterfactual target clauses for individuals with Williams Syndrome in delayed task with negation.



CHAPTER III

MANUSCRIPT: YAOBUSHI P, THEN Q: THE LINGUISTIC ABILITY OF PROPOSITION REASONING IN CHINESE CHILDREN WITH WILLIAMS SYNDROME

A Abstract

This study is investigating a form-meaning dissociation hypothesis on individuals with Williams Syndrome. It is known that individuals with WS have spared grammatical knowledge accompanying with mental retardation in average of 55 low IQ. Research shows that individuals with WS have quite preserved verbal working memory at the same time, which is the possible reason to cause the emergence of language (Wang & Bellugi, 1994; Jerrold, Baddely, & Hewes, 1999; Vicari, Brizzolara, Carlesimo, Pezzini, & Volterra, 1996; Vicari, Carlesimo, Brizzolara, and Pezzini, 1996; Karmiloff-Smith, Grant, Berthoud, Davies, Howlin, & Udwin, 1997; Robinson, Mervis, & Robinson, 2003; Laing, E., Grant, J., Thomas, M. S .C. & Karmiloff-Smith, A., in press). Based on this verbal working memory advantage, it is hypothesized that individuals with WS learn their language mainly by rote memorization. In other words, though grammatical knowledge of WS individuals is strong, their semantic understanding might be weak. That is, they might have dissociation on form and meaning.

In order to answer this question, counterfactuals with *yaobushi* in Chinese are used as probes in testing the possible hypothesis. Counterfactuals are well known in the mismatch between syntax and semantics, which means that it may have two different mental representations built according to the decay of time (Carpenter, 1973). Conditional marker yaobushi is a contrary-to-fact conjunction, which denotes a

counterfactual realm of thinking unambiguously. It is interested to see whether the negation embedded in yaobushi takes a role in representation or not. If yes, then it can be concluded that participants build a form-based representation; if not, then the representation is meaning-based mentally. Carpenter's verification paradigm is employed in this study. Two experiments with different stimuli-of-asynchrony (SOA) are tested on the unimpaired controls, and the same paradigm with no limitation of SOA is tested on individuals with WS.

The four conditions of test sentences with manipulation of truth values and polarities are included as probes to target sentences of factuals and counterfactuals. It is predicted that the unimpaired would show different representations in accordance with SOA variation, in which in shorter SOA the representation of counterfactuals would be form-based and in longer SOA the representation of counterfactuals would be like the one of factuals. However, on the contrary, individuals with WS would show form-based representation even without SOA limitation because of their preserved grammatical knowledge with relatively deficient semantic understanding. Under this scenario, test sentences with negation would be responded faster and erred fewer than test sentences without negation. The results showed that individuals with WS built a meaning-based representation as the unimpaired. Meanwhile, a clear age effect is also demonstrated on the unimpaired, which shows a gradual development in logical reasoning from childhood into adulthood.

B Why *Yaobushi*?

In order to argue against that some might still criticize counterfactual conditionals with negation are hypothetical conditionals rather than contrary-to-fact conditionals, in this study, a clear counterfactual conditional marker *yaobushi* is included as counterfactual conditionals. By doing so, it can disambiguate the issue of hypothetical or counterfactual conditionals in Chinese. In Chapter III, counterfactual conditionals are coerced from a negation marker *meiyou* in if-clause and an aspect marker *le* which denotes the described event is completed in sentence final position. However, until the last grammatical element *le* is encountered, the sentence is ambiguous whether it is hypothetical or counterfactual. This ambiguity comes from the lack of tense inflections on verbs in Chinese. Thus, a counterfactual conditional marker *yaobushi* in sentence initial position helps to clarify this ambiguity. Further, it is still interested to know which mental representation of counterfactuals will be formed under time limitation like 0-SOA task in truth value judgment (Carpenter, 1973). Since there is no negation marker in the predicate of if-clause in yaobushi conditionals and yaobushi can also be equal to if this were not the case, there have several possible representations in working memory. Through this study, it is interested to investigate what representation is the one Chinese will form in processing the content and function of counterfactual conditionals.

C Rationale of Form and Meaning Representations in *Yaobushi* Experiments

A counterfactual conditional sentence with *yaobushi* might have three different representations, and the scenarios are totally different: one with negation, one without negation and another one with flip predicate. These three representations are discussed accordingly. The one with negation is literally translated as *if it were not the case*, the one without negation is processed as a whole unit called Counterfactual Marker (i.e. CF in the paper), and the one with flip predicate which changes the predicate into the one having opposite meaning (i.e. from *be late* to *be on time*). Based on constituent comparison model, the stages involved in mental operations are different for these three representations. If counterfactual marker yaobushi is taken as the representation with negation, namely, *yaobushi* equals to *if* + *not*, the prediction of the ordering of the four test conditions would be FN < TN < TA < FA. The detailed operations of affirmative conditions are listed in Table 27 and negative conditions are in Table 28 below. In terms of hypothetical operation number K, it is K, K+1, K+2, and K+3 (K equals to 6 in this case). This ordering is resulted from the mismatch of negation and counterfactual markers *if* between target sentences and test sentences in affirmative conditions, which in turn, is the result of the match between these two sentences in negative conditions.

Table 27 Representations with Negation and Hypothetical Mental Operations for the Affirmative Conditions of Counterfactual Sentences in *Yaobushi* Tasks

Stimulus and representation	**True Affirmative (TA)**	**False Affirmative (FA)**
Target sentence	Yaobushi I had been late	Yaobushi I had been late.
Test sentence	I was late	I was on time.
Target sentence rep.	[IF, NEG (late, I)]	[IF, NEG (late, I]
Test sentence rep.	(late, I)]	(on time, I)
	– + ☒	– ☒
index = false ☒	– + + ☑	– + ☑
index = true ☑	+ + +	– + + ☒
		+ + +
	response = true	response = false
	K + 2 comparisons	K + 3 comparisons
	要不是我遲到了,......... 我遲到了,..................	要不是我遲到了,.......... 我準時到,..................

Table 28 Representations with Negation and Hypothetical Mental Operations for the Negative Conditions of Counterfactual Sentences in *Yaobushi* Tasks

Stimulus and representation	**True Negative (TN)**	**False Negative (FN)**
Target sentence	Yaobushi I had been late.	Yaobushi I had been late.
Test sentence	I was not on time.	I was not late.
Target sentence rep.	[IF, NEG (late, I)]	[IF, NEG (late, I)]
Test sentence rep.	[NEG (on time, I)]	[NEG (late, I)]
	– ☒	– + + ☒
	– + + ☑	+ + +
index = false ☒	+ + +	
index = true ☑		
	response = true	response = false
	K + 1 comparisons	K comparisons (K = 6)
	要不是我遲到了,………	要不是我遲到了,…….
	我沒有準時到,…………	我沒有遲到,…………..

However, if the representation without negation is formed mentally, namely, *yaobushi* is processed as a whole unit as CF, which functions as a marker denoting a counterfactual realm for comprehenders. In this case, the ordering of the four test conditions is like the following: TA < FA < FN < TN. The detailed operations of affirmatives and negative test conditions are given below as Table 29 and Table 30. In other words, the representation without negation is the real meaning of counterfactuals. People would enter a possible world which is built in the past when the counterfactual conditional marker *yaobushi* is encountered, thus a meaning-based mental representation may be formed directly.

Table 29　Representations without Negation and Hypothetical Mental Operations for the Affirmative Conditions of Counterfactual Sentences in *Yaobushi* Tasks

Stimulus and representation	**True Affirmative (TA)**	**False Affirmative (FA)**
Target sentence	Yaobushi I had been late	Yaobushi I had been late
Test sentence	I was late	I was on time
Target sentence rep.	(late, I)	(late, I)
Test sentence rep.	(late, I)]	(on time, I)
	+	– ☒
index = false ☒		+
index = true ☑		
	response = true	response = false
	K comparisons (K = 1)	K + 2 comparisons
	要不是我遲到了,………	要不是我遲到了,……….
	我遲到了,………………	我準時到,……………….

Table 30 Representations without Negation and Hypothetical Mental Operations for the Negative Conditions of Counterfactual Sentences in *Yaobushi* Tasks

Stimulus and representation	**True Negative (TN)**	**False Negative (FN)**
Target sentence	Yaobushi I had been late.	Yaobushi I had been late.
Test sentence	I was not on time.	I was not late.
Target sentence rep.	(late, I)	(late, I)
Test sentence rep.	[NEG, (on time, I)]	[NEG (late, I)]
	– ☒	– + ☒
	– + ☑	+ + ☑
index = false ☒	+ +	
index = true ☑		
	response = true	response = false
	K + 4 comparisons	K + 3 comparisons
	要不是我遲到了,………	要不是我遲到了,…….
	我沒有準時到,…………	我沒有遲到,…………..

Once more, there is another possible alternative which might be formed mentally, that is, the flip of predicates. The negation will be preserved in the representation form because the real meaning of counterfactuals is the opposite. In this case, the predicted ordering of the four test conditions is like the following: TN < FN < FA < TA. The detailed mental operations are listed in Table 31 and Table 32 below.

Table 31 Hypothetical Mental Operations for Representations with Flip Predicates of Affirmative Conditions in Counterfactuals with *Yaobushi*

Stimulus and representation	**True Affirmative (TA)**	**False Affirmative (FA)**
Target sentence	Yaobushi I had been late	Yaobushi I had been late
Test sentence	I was late	I was on time
Target sentence rep.	[NEG (on time, I)]	[NEG (on time, I)
Test sentence rep.	(late, I)]	(on time, I)
	– ☒	– + ☒
index = false ☒	– + ☑	+ +
index = true ☑	+ +	
	response = true	response = false
	K + 3 comparisons	K + 2 comparisons
	要不是我遲到了,………	要不是我遲到了,……….
	我遲到了,………………	我準時到,……………….

Table 32 Hypothetical Mental Operations for Representations with Flip Predicates of Negative Conditions in Counterfactuals with *Yaobushi*

Stimulus and representation	**True Negative (TN)**	**False Negative (FN)**
Target sentence	Yaobushi I had been late.	Yaobushi I had been late.
Test sentence	I was not on time.	I was not late.
Target sentence rep.	[NEG, (on time, I)]	[NEG, (on time, I)]
Test sentence rep.	[NEG, (on time, I)]	[NEG, (late, I)]
	+ +	– ☒
		+ +
index = false ☒		
index = true ☑	response = true	response = false
	K comparisons (K = 2)	K + 1 comparisons
	要不是我遲到了,………	要不是我遲到了,…….
	我沒有準時到,…………	我沒有遲到,…………..

As for the consequence clause (i.e. the car would not have been driven away), there are also three possible representations as the if-clause: one with negation, one without negation, and another one with flip predicate. Based on each representation, the ordering of the four test conditions is very different. The representation with negation has the ordering like the following: FN < TN < TA < FA, which is similar to the one with flip predicate in the ordering like TN < FN < FA < TA. These two representations are called form-based representations because the conditions with negation are predicted to be responded faster than the ones without negation. The last one without negation in the representation is predicted to have the following ordering:

TA < FA < FN < TN. It is called meaning-based representation because the negation of the test conditions does not play a role in truth value judgment. Thus, the conditions without negation (i.e. affirmatives) are responded faster than the ones with negation. In other words, it is clearly to see what kind of representation is formed according to the ordering of the test conditions calculated from the CCM. It is interested to investigate whether a form-based representation would be formed in counterfactual conditionals with yaobushi, a clear counterfactual marker in Chinese. For individuals with WS, if form and meaning dissociation hypothesis is correct, they are expected to perform a form-based representation rather than meaning-based.

As for factual target clauses, since there is no difference between form-based and meaning-based representations, the ordering from the easiest to the hardest would follow the prediction of CCM in the following order: TA < FA < FN < TN. Further, this ordering should be the same in long and short SOA. Meanwhile, the prediction for counterfactual conditional clauses in longer SOA is the same as the one for factual target clauses. In other words, a meaning-based representation would be formed after certain amount of exposure.

D Language and Thought Experiment III: Simultaneous Task of Counterfactual Conditionals with *Yaobushi*

Participant: College Students

Twenty four college students from National Tsing Hua University were included (mean age = 19.13, range from 18 to 21, 22 females and 2 males). All participants were participated for course credit of Introduction to Linguistics. They were right-handed users and none of them were reported having medical problems.

Participant: The eighth graders

Twenty one the eighth graders in Fu He The eighth School participated in this study (mean age = 14, range from 13 to 15, 13 females and 8 males). They were rewarded a present after finishing the study. All participants were right-handed users and none of them were reported having medical problems.

Participant: The sixth graders

Twenty four the sixth graders in Qing Jiang The sixth School participated in this study (mean age = 12.3, range from 12 to 13, 13 females and 11 males). They were rewarded a present after finishing the study. All participants were right-handed users and none of them were reported having medical problems.

Design

There were two independent variables designed for target sentences: sentence type (i.e. factual or counterfactual) and sequence of sentence type (i.e. if-clause or consequence clause). Thus, there were four types of target sentences in this study: (1) factual-factual target sentence (FF); (2) factual-counterfactual target sentence (FC); (3) counterfactual-factual target sentence (CF); and (4) counterfactual-counterfactual target sentence (CC). There were four experimental sentences, which were under different scenario and were all common situations in daily life. Take an example of CC, a sentence like *Yaobushi I had been late, the car would not have been driven away* was designed as a target sentence, which was related to time schedule and being late. There were another two independent variables designed for test sentences: truth values (i.e. true or false) and polarity (i.e. affirmative or negative). Thus, there were four types of test sentences: (1) true affirmative (TA) like *I was late*; (2) false affirmative (FA) like *I was on time*; (3) true negative (TN) like *I was not on time*; (4)

false negative (FN) like *I was not late*. For each target clause, there were four test sentences probing the truth values of it. Thus, for each target sentence, there were eight test sentences testing its truth value. For each target sentence type, there were thirty two trials. There were four experimental scenarios designed in this study. In the end, there were 128 experimental trials. Also, 64 filler sentences were included. For each filler sentence, there were also eight test sentences for it as experimental sentences (four for if-clause and another four for consequence clause). Participants were required to judge whether the displayed test sentence was true or false based on the truth condition presupposed of the target clause.

Materials

In this study, a counterfactual conditional marker yaobushi is added in sentence initial position to avoid the possible confound of ambiguity in processing. That is, it is interested to see whether participants would have a form-based representation according to the target sentences. However, it is noteworthy that once yaobushi is put in sentence initial position, the negation in the if-clause of target sentences would be taken away. However, the negation in consequence clause would still exist. Take an example of target sentence with CC, the target sentence would be like: *Yaobushi I had been late, the car would not have been driven away*. Therefore, it is interested to see whether the counterfactual marker yaobushi is represented as *if* + *not*, or a whole unit CF. All the target sentences were listed in Appendix 10, 12, 14, and 16. Besides, the corresponding test sentences for each target sentence were listed in Appendix 11, 13, 15, and 17. The filler sentences were the same as tasks with negation, which could be referred in Appendix 9.

Procedure

The procedure was parallel to simultaneous task with negation in section E. In the beginning, a fixation point was displayed on the computer screen for 500ms. After this arousal to make participants pay attention, a target sentence and a test sentence were presented on the screen at the same time. The target sentence was above the test sentence. After 2 seconds, the target sentence disappeared and only the test sentence remained presenting on the screen. Participants were required to make judgment of the truth value of the test sentence based on the meaning of the target clause. If the truth condition between the target clause and the test sentence matched, participants were instructed to press the left button of a mouse as soon as possible; however, if any mismatches were found, they should press the right button right away. There were four blocks in this study. Each block contained 48 experimental trials and a break was expected between two blocks. Thus, there were three blocks in this study. All participants were given 8 practice trials to make sure their understanding of this task before running experimental trials.

Prediction

Counterfactual conditionals with yaobushi may have different results because of the different representations. First, yaobushi is treated as a counterfactual (CF) marker without decomposing its parts. That is, there is no negation in representation of if-clause in a complex form. For example, for a counterfactual target clause *Yaobushi I had been late, the car would not have been driven away*, the complex form would be [CF (late, I)]. According to congruent mapping strategy hypothesized in time limitation, test sentences with true affirmatives (i.e. the representation is (late, I)) would receive fastest response latency, then test sentences with false affirmatives (i.e. the representation is (on time, I)) would have longer response latency, and then test sentences with false negatives (i.e. [false (late, I)] would be the second longest response latency condition, which in turn is faster than test sentences with true negatives (i.e. the representation is [false (on time, I)]. So, the ordering of the four conditions tested is predicted to be like this: TA < FA < FN < TN. If numbers of mental operations are transformed in to hypothetical letter K, the ordering is like: K, K +1, K +5, K + 6 (K equals to 6 in this case). In other words, if a simpler form is represented (i.e. the representation without conditional marker and negation marker), counterfactual conditionals with yaobushi would have exactly the same prediction as the one in counterfactual conditionals with negation in delayed task. Second, yaobushi is decomposed as a conditional marker (ruguo, *if*) accompanying with a negation (meiyou, *not*), which means that there would have a negation formed in the representation. Thus, a more complex representation would be formed mentally. The prediction would be like the one in simultaneous task with negation. The ordering of the four test conditions would be like the following: FN < TN < TA < FA.

Results: College students

The latencies and error rates to respond to factual and counterfactual clauses were shown in Table 33 below.

Counterfactual vs. factual target clauses

Participants responded to counterfactual target clauses and factual target clauses were 5230ms and 4663ms, respectively. These two response latencies were significantly different ($F(1, 2717) = 72.66$, $p < .0001$). Meanwhile, participants made more errors on test sentences probing counterfactual target clauses than test sentences probing factual target clauses ($F(1, 2740) = 145.37$, $p < .0001$). For factual target clauses, there was an interaction between truth values and polarities and this interaction reached significant ($F(1, 1358) = 6.32$, $p = .01$). This interaction of error rates was not found (F (1, 1381) = 0.32, p = .57). For counterfactual target clauses, there was also a significant interaction between truth values and polarities (F (1, 1330) = 15.74, p < .0001). This interaction of error rates was not found (F (1, 1353) = 0.00, p = .9866).

The interaction of the four conditions on counterfactual and factual target clauses was not significant (F (3, 2711) = 1.40, p = .24). The main effects of clause types (i.e. counterfactual or factual) and the four conditions of test sentences were significant, p < .0001. All the comparisons of the difference of each condition between factual target clauses and counterfactual target clauses was highly significant at p = .0001 level. Meanwhile, the interaction of the four conditions on error rates on both target clauses was not significant (F (3, 2734) = 0.16, p = .92). The main effect of error rates on target clauses was significant (F (1, 2734) = 145.63, p < .0001), but the main effect of error rates on the four conditions was not significant difference (F (3, 2734) = 0.52, p = .66). The difference of each condition on error rates between factual and counterfactual target clauses reached highly significance (p < .0001), suggesting that college students erred more on counterfactual target clauses than on factual target clauses.

Counterfactual vs. factual clause position

Response latency to counterfactual target clauses in if-clause was 4959ms and in consequence clause was 5494ms. The difference of their latency did not reach significance (F (1, 1354) = 19.34, p < .0001), suggesting clause positions for counterfactual targets made difference in processing. The same pattern was not found on their error rates (F (1, 1355) = 0.00, p = .9), suggesting that participants did not make more errors because of the clause positions. Moreover, response latency to factual target clauses in if-clause was 4472ms and in consequence clause was 4806ms.

The difference of their latency reached significance (F (1, 1382) = 11.36, p = .0008), suggesting clause positions for factual targets did have influence in processing. Their error rates did not show any difference (F (1, 1382) = 0.03, p = .8), suggesting that participants did not make more errors because of the clause positions. For college students, clause positions caused processing difference no matter on factual or counterfactual target clauses.

Counterfactual vs. factual sentence type

Response latencies for counterfactual target clauses in four experimental sentence types were 5314ms, 5420ms, 4985ms, and 5203ms, respectively. A one-way ANOVA results did not show significant difference among them (F (3, 1352) = 2.46, p = .06), suggesting that none of the counterfactual target clauses in these four experimental sentences was more difficult among them. That is, they were all similar in degree of difficulty. The same pattern was found on their error rates in one-way ANOVA (F (3, 1353) = 0.08, p = .9). Response latencies for factual target clauses in four experimental sentences were 4605ms, 4737ms, 4536ms, and 4665ms, respectively. A one-way ANOVA results did not show significant difference among them (F (3, 1380) = 0.73, p = .5), suggesting that none of the factual target clauses in these four experimental sentences was more difficult among them. In other words, they were all similar in degree of difficulty. The same pattern was found on their error rates in one-way ANOVA (F (3, 1381) = 0.11, p = .9).

Test sentences in counterfactual targets

For counterfactual target sentences with yaobushi at zero SOA, the condition ordering was like the following: TA < TN < FA < FN, which was different from the prediction based on CCM. The condition ordering was expected to be like: TA < FA <

FN < TN. Participants responded fastest to test sentences with true affirmatives (4683ms), next was to test sentences with false affirmatives (5383ms), the third was to test sentences with false negatives (5389ms), and the last was to test sentences with true negatives (5522ms). A one-way ANOVA showed that the difference between these the four conditions was significant ($F (3, 1330) = 15.16$, $p < .0001$). A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. The major difference was attributed to the comparison of TA and other groups in $p < .0001$ level. Parallel to Carpenter's findings on counterfactual clauses and modified by Chinese stimuli, as predicted, test sentences with matched representations in predicates were responded significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA ($p < .0001$). Error rates for factual target clauses in the four conditions did not reach significant difference to one another ($F (3, 1381) = 0.46$, $p = .7$) but the difference of error rates for counterfactual target clauses in the four conditions reached significance ($F (3, 1353) = 0.18$, $p = .9$).

Test sentences in factual targets

As predicted, for factual target clauses no matter how short was the display, the condition ordering from the easiest to the hardest was like the following: TA < FA < FN < TN. That is, college students responded fastest to test sentences with true affirmatives (4285ms), next was to test sentences with false affirmatives (4596ms), the third was to test sentences with false negatives (4797ms), and the last was to test sentences with true negatives (4889ms). A one-way ANOVA showed that the difference between these the four conditions was significant ($F (3, 1358) = 10.40$, $p < .0001$). A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. The results showed almost all the comparisons were significant, but two comparison were not (FA vs. FN, $p = .1037$

and FN vs. TN, p = .4). Parallel to Carpenter's findings on factual clauses, test sentences with matched representations in predicates were responded significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA (p = .004).

Affirmative vs. negative test sentences

Test sentences with affirmatives were responded faster than test sentences with negatives for factual target clauses (4440ms and 4840ms, respectively) (F (1, 1382) = 16.69, p < .0001). The difference in their error rates was not significant (F (1, 1383) = 0.94, p = .3). Similarly, test sentences with affirmatives were responded faster than test sentences with negatives for counterfactual target clauses (5015ms and 5455ms, respectively) (F (1, 1354) = 13.27, p = .0003). The difference in their error rates was not significant (F (1, 1355) = 0.44, p = .5).

TSING HUA UNIVERSITY

Table 33 Response Latency (in ms) and Error Rates in Simultaneous Task with Yaobushi for College Students

Type of Clause		TA	FA	FN	TN	Total
Factual	RT	4285.80	4596.61	4797.55	4889.22	4663.82
	Errors	9.76%	9.71%	9.61%	9.40%	9.62%
Counterfactual	RT	4683.20	5383.16	5389.59	5522.52	5230.96
	Errors	11.59%	11.52%	11.39%	11.45%	11.49%

Results: The Eighth Graders

The latencies and error rates to respond to factual and counterfactual clauses were shown in Table 34 below.

Counterfactual vs. factual target clauses

Participants responded to counterfactual target clauses and factual target clauses were 4692ms and 4336ms, respectively. These two response latencies were significantly different ($F(1, 2454) = 41.91$, $p < .0001$). Meanwhile, participants made more errors on test sentences probing counterfactual target clauses than test sentences probing factual target clauses ($F(1, 2476) = 25.65$, $p < .0001$). For factual target clauses, there was an interaction between truth values and polarities and this interaction did not reach significant difference ($F(1, 1227) = 3.41$, $p = .0649$). The same pattern was found on error rates ($F(1, 1249) = 0.03$, $p = .8517$). For counterfactual target clauses, there was also a significant interaction between truth values and polarities ($F(1, 1199) = 1.85$, $p = .1$). The same pattern was found on error rates ($F(1, 1221) = 0.19$, $p = .6$). The interaction of the four conditions on factual and counterfactual target clauses was not significant ($F(3, 2448) = 1.20$, $p = .3$), suggesting in general response latency on factual target clauses was not faster than response latency on counterfactual target clauses. But, the main effects of target clauses and test conditions were significant at .0001. The difference of each condition between factual target clauses and counterfactual target clauses was highly significant. Meanwhile, the interaction of the four conditions on error rates on both target clauses was not significant ($F(3, 2448) = 0.16$, $p = .9$). But, the main effects of error rates on target clauses and test conditions were both significantly different at .0001. The difference of each condition on error rates between factual and counterfactual target clauses reached highly significance ($p < .0001$), suggesting that college students erred more on test sentences for counterfactual target clauses than on test sentences for factual target clauses.

Counterfactual vs. factual clause position

Response latency to counterfactual target clauses in if-clause was 4581ms and in consequence clause was 4829ms. The difference of their latency reached significance (F (1, 1222) = 8.50, p = .003), suggesting clause positions for counterfactual targets made difference in processing. The same pattern was not found on their error rates (F (1, 1223) = 0.03, p = .8), suggesting that participants did not make more errors because of the clause positions. Moreover, response latency to factual target clauses in if-clause was 4217ms and in consequence clause was 4479ms. The difference of their latency reached significance (F (1, 1250) = 11.09, p = .0009), suggesting clause positions for factual targets did have influence in processing. Their error rates did not show any difference (F (1, 1251) = 0.18, p = .6), suggesting that participants did not make more errors because of the clause positions. For the eighth graders, clause positions caused processing difference no matter on factual or counterfactual target clauses.

Counterfactual vs. factual sentence type

Response latencies for counterfactual target clauses in four experimental sentence types were 4682ms, 4761ms, 4569ms, and 4805ms, respectively. A one-way ANOVA results did not show significant difference among them (F (3, 1220) = 1.47, p = .2), suggesting that none of the counterfactual target clauses in these four experimental sentences caused more difficulty than others. They were all similar in degree of difficulty. The same pattern was found on their error rates in one-way ANOVA (F (3, 1221) = 0.13, p = .9).

Response latencies for factual target clauses in four experimental sentence types were 4281ms, 4478ms, 4359ms, and 4269ms, respectively. A one-way ANOVA results did not show significant difference among them ($F(3, 1248) = 1.50$, $p = .2$), suggesting that none of the factual target clauses in these four experimental sentences caused more difficulty than others. They were all similar in degree of difficulty or easiness. The same pattern was found on their error rates in one-way ANOVA ($F(3, 1249) = 0.75$, $p = .5$).

Test sentences in counterfactual targets

For counterfactual target sentences, the condition ordering was like the following: TA < FA < FN < TN. This ordering result seemed to show that the eighth graders treated yaobushi as a counterfactual conditional marker rather than decomposed it as a conditional marker with a negation because the numbers of mental operations exactly parallel to the predictions of factual target clauses. Participants responded fastest to test sentences with true affirmatives (4393ms), next was to test sentences with false affirmatives (4421ms), the third was to test sentences with false negatives (4916ms), and the last was to test sentences with true negatives (5111ms). A one-way ANOVA showed that the difference between these the four conditions was significant, $F(3, 1199) = 18.90$, $p < .0001$. A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. The major difference was attributed to the comparison of TA and other groups in $p < .0001$ level. There were two comparisons which did not reach significance (i.e. TA vs. FA, $p = .7806$; FN vs. TN, $p = .1$). Again, contrary to Carpenter's findings on counterfactual clauses and modified by Chinese stimuli, test sentences with matched representations in predicates were not responded significantly faster than test sentences with mismatched representations in predicates, TA < FA ($p < .0001$). Error rates for factual target clauses in the four conditions did not reach significant difference to one another

($F (3, 1249) = 1.59$, $p = .1$), and also error rates for counterfactual target clauses in the four conditions was not significant ($F (3, 1221) = 0.74$, $p = .5$).

Test sentences in factual targets

As predicted, for factual target clauses, the condition ordering from the easiest to the hardest for the eighth graders was like the following: TA < FA < FN < TN. That is, participants responded fastest to test sentences with true affirmatives (4106ms), next was to test sentences with false affirmatives (4208ms), the third was to test sentences with false negatives (4472ms), and the last was to test sentences with true negatives (4631ms). A one-way ANOVA showed that the difference between these the four conditions was significant ($F (3, 1227) = 9.77$, $p < .0001$). A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. The results showed almost all the comparisons were significant, but comparisons between affirmatives and negatives were not significant (i.e. TA vs. FA, $p = .3013$; FN vs. TN, $p = .1$). Contrary to Carpenter's findings on factual target clauses, test sentences with matched representations in predicates were not responded significantly faster than test sentences with mismatched representations in predicates. Though response latency was faster for matched condition than mismatched condition in 101ms, they were not different to each other (i.e. TA < FA, $p = .3$).

Affirmative vs. negative test sentences

Test sentences with affirmatives were responded faster than test sentences with negatives for factual target clauses (4154ms and 4550ms, respectively) ($F (1, 1250) = 25.01$, $p < .0001$). The difference in their error rates was not significant ($F (1, 1251) = 0.06$, $p = .8$). Similarly, test sentences with affirmatives were responded faster than test sentences with negatives for counterfactual target clauses (4406ms and 5013ms, respectively) (F (1, 1222) = 50.66, p < .0001). The difference in their error rates was also significant (F (1, 1223) = 0.00, p = .9).

Table 34 Response Latency (in ms) and Error Rates in Simultaneous Task with Yaobushi for The eighth graders

Type of Clause		TA	FA	FN	TN	Total
Factual	RT	4106.12	4208.04	4472.32	4631.36	4336.75
	Errors	14.70%	13.85%	14.03%	14.74%	14.33%
Counterfactual	RT	4393.88	4421.35	4916.57	5111.58	4692.38
	Errors	16.30%	15.39%	15.64%	16.12%	15.88%

Results: The Sixth Graders

The latencies and error rates to respond to factual and counterfactual clauses were shown in Table 35 below.

Counterfactual vs. factual target clauses

Participants responded to counterfactual target clauses and factual target clauses were 4774ms and 4442ms, respectively. These two response latencies were significantly different (F (1, 2558) = 51.13, p < .0001). Meanwhile, participants made more errors on test sentences probing counterfactual target clauses than test sentences probing factual target clauses (F (1, 2581) = 190.01, p < .0001).

For counterfactual target clauses, there was also a significant interaction between truth values and polarities (F (1, 1225) = 11.82, p = .0006). The same pattern was not found on error rates (F (1, 1248) = 1.09, p = .2). The interaction of the four conditions on counterfactual and factual target clauses was not significant ($F (3, 2552) = 0.61$, $p = .6$), suggesting response latency on factual target clauses in general was not faster than response latency on counterfactual target clauses. But, the main effects of clause types and test conditions were all significant at .0001.

For factual target clauses, there was an interaction between truth values and polarities and this interaction reached significant difference ($F (1, 1304) = 16.50$, $p < .0001$). The same pattern was not found on error rates ($F (1, 1327) = 0.29$, $p = .5$).

The difference of each condition between factual target clauses and counterfactual target clauses was highly significant. Meanwhile, the interaction of the four conditions on error rates on both target clauses was not significant ($F (3, 2552) = 0.17$, $p = .9$). The main effect of error rates on target clauses was significant ($F (1, 2552) = 1378.93$, $p < .0001$), but the main effect of error rates on test conditions was not significant difference ($F (3, 2552) = 0.12$, $p = .9$). The difference of each condition on error rates between factual and counterfactual target clauses reached highly significance, suggesting that the sixth graders erred more on test sentences for counterfactual target clauses than on test sentences for factual target clauses like other groups.

Counterfactual vs. factual clause position

Response latency to counterfactual target clauses in if-clause was 4671ms and in consequence clause was 4844ms. The difference of their latency also reached significance ($F (1, 1249) = 5.43$, $p = .01$), suggesting clause positions for counterfactual targets made difference in processing. However, the difference of error rates was not significance ($F (1, 1250) = 1.80$, $p = .1$), suggesting that participants did not make more errors because of the clause positions.

Response latency to factual target clauses in if-clause was 4329ms and in consequence clause was 4536ms. The difference of their latency reached significance (F (1, 1328) = 9.39, p = .002), suggesting clause positions for factual targets did have influence in processing. Their error rates did not show any difference (F (1, 1329) = 0.58, p = .4), suggesting that participants did not make more errors because of the clause positions. For the sixth graders, clause positions caused processing difference no matter on factual or counterfactual target clauses just like the patterns observed on college students and the eighth graders.

Counterfactual vs. factual sentence type

Response latencies for counterfactual target clauses in four experimental sentence types were 4676ms, 4845ms, 4674ms, and 4831ms, respectively. A one-way ANOVA results did not show significant difference among them (F (3, 1247) = 1.53, p = .2), suggesting that none of the counterfactual target clauses in these four experimental sentences caused more difficulty than one another. They were all similar in degree of difficulty. The same pattern was found on their error rates in one-way ANOVA (F (3, 1248) = 0.03, p = .9).

Response latencies for factual target clauses in four experimental sentence types were 4399ms, 4541ms, 4295ms, and 4487ms, respectively. A one-way ANOVA results did not show significant difference among them (F (3, 1326) = 2.50, p = .05), suggesting that none of the factual target clauses in these four experimental sentences caused more difficulty than one another. They were all similar in degree of difficulty. The same pattern was found on their error rates in one-way ANOVA F((3, 1327) = 0.21, p = .8).

Test sentences in counterfactual targets

For counterfactual target sentences, the condition ordering was like the following:

TA < FA < FN < TN, which was different from the ordering predicted based on decomposing hypothesis. For the sixth graders, it seemed that they also processed yaobushi as a whole counterfactual conditional and thus formed a simpler representation. Participants responded fastest to test sentences with true affirmatives (4400ms), next was to test sentences with false affirmatives (4779ms), the third was to test sentences with false negatives (4924ms), and the last was to test sentences with true negatives (4981ms). A one-way ANOVA showed that the difference between these the four conditions was significant ($F (3, 1225) = 17.91$, $p < .0001$). A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. The results showed almost all the comparisons were significant, but comparisons between false responses and negatives were not significant (FA vs. FN, $p = .1445$; FN vs. TN, $p = .5$). Parallel to Carpenter's findings on counterfactual clauses and modified by Chinese stimuli, as predicted, test sentences with matched representations in predicates were responded significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA ($p < .0001$). Error rates for factual target clauses in the four conditions did not reach significant difference to one another ($F (3, 1327) = 1.32$, $p = .2$), but the difference of error rates for counterfactual target clauses in the four conditions reached significance ($F (3, 1248) = 0.84$, $p = .4$).

Test sentences in factual targets

As predicted, for factual target clauses, the condition ordering from the easiest to the hardest was like the following: TA < FA < FN < TN. Participants responded fastest to test sentences with true affirmatives (4064ms), next was to test sentences with false affirmatives (4379ms), the third was to test sentences with false negatives (4576ms), and the last was to test sentences with true negatives (4759ms). A one-way ANOVA showed that the difference between these the four conditions was significant ($F(3, 1304) = 24.17$, $p < .0001$). A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. The results showed almost all the comparisons were significant, but comparisons between false responses and negatives were not significant (FN vs. TN, $p = .05$). Parallel to Carpenter's findings on factual clauses, test sentences with matched representations in predicates were responded significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA ($p = .0001$).

Affirmative vs. negative test sentences

Test sentences with affirmatives were responded faster than test sentences with negatives for counterfactual target clauses (4571ms and 4953ms, respectively) ($F(1, 1249) = 28.29$, $p < .0001$). The difference in their error rates was also not significant ($F(1, 1250) = 0.03$, $p = .8$).

Test sentences with affirmatives were responded faster than test sentences with negatives for factual target clauses (4212ms and 4665ms, respectively) ($F(1, 1328) = 45.21$, $p < .0001$). The difference in their error rates was not significant ($F(1, 1329) = 2.36$, $p = .1$).

Table 35 Response Latency (in ms) and Error Rates in Simultaneous Task with Yaobushi for The sixth graders

Type of Clause		TA	FA	FN	TN	Total
Factual	RT	4064.30	4379.19	4576.45	4759.89	4442.20
	Errors	13.41%	12.85%	12.53%	12.73%	12.89%

Counterfactual	RT	4400.67	4779.00	4924.95	4981.89	4774.71
	Errors	17.84%	16.64%	17.18%	17.23%	17.25%

Summary

Table 36 summarized the findings of experiment III in three age groups. Basically, these three groups followed the same pattern. The orderings of factuals target clauses matched the predictions. As for the counterfactual target clauses, the pattern of college students was different from other groups. Though the difference reached significance, condition TN was responded faster than conditions with false responses. Meanwhile, the difference for factuals and counterfactuals from other two groups both reached significance ($F(11, 3874) = 11.81$, $p < .0001$ for factuals, and $F(11, 3734) = 13.12$, $p < .0001$ for counterfactuals). Meanwhile, all the error rates were not significant for both target clauses. All three age groups performed consistently shorter response latency and fewer errors on factuals than on counterfactuals. Interesting, clause effects on both target clauses were significant, suggesting different clause positions caused different processing load in simultaneous task. These results were different from the ones in Experiment I, which participants in different ages did not show clause effect on counterfactuals. Interactions of truth value and polarity in response latency and error rates were clearly observed on factual target clauses and on counterfactual target clauses, except for the eighth graders. Further, as predicted, there is no sentence type effect cross the board, indicates sentences have similar degree in difficulty. The polarity effect is robust in all three groups, suggesting a faster response in affirmatives than the one in negatives.

Table 36 Summary Findings of Three Groups in Delayed Experiment of Counterfactual Conditionals with *Yaobushi*

		COLLEGE	THE EIGHTH	THE SIXTH
Factual Ordering		TA < FA < FN < TN	TA < FA < FN < TN	TA < FA < FN < TN
Factual p-value	(RT)	p < .0001	p < .0001	p < .0001
	(Error Rates)	p = .7110	p = .1911	p = .2647
Counterfactual Ordering		TA < TN < FA < FN	TA < FA < FN < TN	TA < FA < FN < TN
Counterfactual p-value	(RT)	p < .0001	p < .0001	p < .0001
	(Error Rates)	p = .9	p = .5	p = .4
Factual vs. Counterfactual p-value	(RT)	p < .0001	p < .0001	p < .0001
	(Error Rates)	p < .0001	p < .0001	p < .0001
Factual interaction of truth values and polarities	(RT)	p = .01	p = .06	p < .0001
	(Error Rates)	p = .5	p = .8	p = .5
Counterfactual interaction of truth values and polarities	(RT)	p < .0001	p = .1741	p = .0006
	(Error Rates)	p = .9	p = .6	p = .2
Interaction of the	(RT)	p = .2	p = .3	p = .6

four conditions on factual and counterfactual	(Error Rates)	p = .9	p = .9	p = .9
Factual clause effect	(RT)	p = .0008	p = .0009	p = .002
	(Error Rates)	p = .8	p = .6	p = .4
Counterfactual clause effect	(RT)	p < .0001	p = .003	p = .01
	(Error Rates)	p = .9	p = .8	p = .1
Factual sentence type effect	(RT)	p = .5	p = .2	p = .05
	(Error Rates)	p = .9561	p = .5	p = .8864
Counterfactual sentence type effect	(RT)	p = .06	p = .211	p = .2043
	(Error Rates)	p = .9	p = .9445	p = .9
Factual polarity effect	(RT)	p < .0001	p < .0001	p < .0001
	(Error Rates)	p = .3	p = .8	p = .1
Counterfactual polarity effect	(RT)	p = .0003	p < .0001	p < .0001
	(Error Rates)	p = .5	p = .9	p = .8

General Discussion of Age Effect

Counterfactual vs. factual target clauses in different ages

A response latency of counterfactual and factual target clauses showed interaction among three age groups ($F(2, 7694) = 5.86$, $p = .002$). The main effect of age group was not significant ($F(2, 7694) = 2.74$, $p = .06$) and the main effect of target clause was significant ($F(2, 7694) = 167.56$, $p < .0001$). The main difference came from the comparisons of college students and other two groups on counterfactual target clauses (i.e. college vs. the eighth graders, $p = .006$; college vs. the sixth graders, $p = .04$). The differences of response latency on factual and counterfactual target clauses did not show any significance. However, the interaction of response latency of factual and counterfactual target clauses on the four conditions for three age groups was significant ($F(23, 7676) = 19.98$, $p < .0001$).

A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed, showing mainly the significance came from counterfactual target clauses comparisons like FA, FN, TN conditions between college students and the eighth graders; FA, TN conditions between college and the sixth graders. Meanwhile, only one comparison on factual target clauses was significant (i.e. FN) between college students and the eighth graders. These results indicated that for younger participants like the eighth and the sixth graders, all conditions of test sentences for factual and counterfactual target clauses were equally easy or difficult for them in processing, but for the oldest participants, college students, only counterfactuals differed from other groups.

Test sentences in counterfactual targets in different ages

The interaction between truth values and polarity on counterfactual target clauses for three groups reached significant difference ($F(11, 3734) = 13.12$, $p < .0001$), so did the interaction on factual target clauses ($p < .0001$). The interaction between clause positions and different age groups on counterfactual target clauses was significant ($F (5, 3740) = 12.85, p < .0001$). Clause effect was clearly demonstrated on comparison of college students and other two groups, and the effect was only observed on consequence clauses. None of the comparisons between the eighth graders and the sixth graders on clause positions reached significance, suggesting that these two groups did not show any difference in processing counterfactual target clauses in terms of clause positions.

Test sentences in factual targets in different ages

The interaction between clause positions and different age groups on factual target clauses was not significant ($F (2, 3880) = 0.48, p = .6$). Meanwhile, the main effect of age group was not significant ($F (2, 3880) = 2.06, p = .1$), and the main effect of clause positions was significant ($F (1, 3880) = 44.73, p < .0001$). Almost all the comparisons between if-clauses and consequence clauses were not significant for factual target clauses (except clause effect on consequence clause between college students and the eighth graders), showing overall easiness in processing for these three groups.

Counterfactual vs. factual sentence type in different ages

The interaction between age group and different experimental sentences was not significant on counterfactuals ($F (6, 3734) = 1.55, p = .1579$), suggesting that there was no specific experimental sentence caused more difficult processing. The main effect of age group was significant ($F (2, 3734) = 3.08, p = .0462$) and main affect of experimental sentence was also significant ($p = .0005$). The major difference came from experimental sentences 1 and 2 between college students and other two groups. The interaction between age group and different target sentences was not significant on factual target clauses (F (6, 3874) = 0.56, p = .7), suggesting that there was no specific experimental sentence caused more difficult processing. The main effect of age group was not significant (F (2, 3874) = 2.05, p = .1) and main affect of experimental sentence was significant (p = .02). All the comparisons between two clauses on factual target clauses were not significant (except experimental sentence 1 between college students and the eighth graders).

Generally speaking, none of the comparisons between the eighth graders and the sixth graders reached significant difference. However, the comparisons which reached significant difference between college students and other two groups were mainly on experimental sentences 1 and 2 of counterfactual target clauses.

Affirmative vs. negative test sentences in factual sentences in different ages

The interaction between age group and polarity on counterfactual target clauses was not significant (F (2, 3740) = 0.19, p = .8). The main effect of age group was significant, p = .0467 and the main effect of polarity was also significant, p < .0001. Only the comparisons between college students and the eighth graders were significant, suggesting a general pattern of faster response latency on test sentences with affirmatives and negatives for college students than for the eighth graders. Other comparisons of other groups did not reach significance. An interaction between age group and polarity on factual target clauses was also not significant (F (2, 3880) = 0.70, p = .4). The main effect of age group was not significant (p = .1) and the main effect of polarity was significant (p < .0001). All the comparisons between any two groups were not significant (except test sentences with negatives between college and the eighth graders).

These results indicated that the younger groups like the sixth graders and the eighth graders did not show difference from each other in response latency on both test sentences with affirmatives and negatives, so did the comparisons between college students and the sixth graders. However, the difference in response latency between college students and the eighth graders reached significance on test sentences with both polarities (except test sentences with affirmatives for factual target clauses).

Table 37 Response Latency (in ms) and Error Rates in Simultaneous Task with Yaobushi on Factual and Counterfactual Target Clauses for Three Age Groups

Type of Clause	Group	TA	FA	FN	TN	Total
Factual	College	4285.80	4596.61	4797.55	4889.22	4663.82
		9.76%	9.71%	9.61%	9.40%	9.62%
	The Eighth	4106.12	4208.04	4472.32	4631.36	4336.75
		14.70%	13.85%	14.03%	14.74%	14.33%
	Sixth	4064.30	4379.19	4576.45	4759.89	4442.20
		13.41%	12.85%	12.53%	12.73%	12.89%
Counterfactual	College	4683.20	5383.16	5389.59	5522.52	5230.96
		11.59%	11.52%	11.39%	11.45%	11.49%
	The Eighth	4393.88	4421.35	4916.57	5111.58	4692.38
		16.30%	15.39%	15.64%	16.12%	15.88%
	The Sixth	4400.67	4779.00	4924.95	4981.89	4774.71
		17.84%	16.64%	17.18%	17.23%	17.25%

E Language and Thought Experiment IV:

Delayed Task of Counterfactual Conditionals with *Yaobushi*

Participant: College Students

Twenty three participants were recruited in this study, including nineteen college students from National Tsing Hua University and four graduate students from University of Maryland at College Park (mean age = 21.26, range from 19 to 30, 17 females and 6 males). All students from Tsing Hua University were participated for course credit of Introduction to Linguistics and students from UMCP were rewarded a present after participating this experiment. They were right-handed users and none of them were reported having medical problems. Participants were tested in a quiet room in laboratory of Cognitive Neuropsychology in National Yang Ming University. All of them were tested one at a time.

Participant: The Eighth Graders

Twenty nine the eighth graders in Fu He The eighth School participated in this study (mean age = 14.41, range from 13 to 15, 16 females and 13 males). They were rewarded a present after finishing the study. All participants were right-handed users and none of them were reported having medical problems. All junior students were tested in a computer room of their school and all were tested at a time.

Participant: The Sixth Graders

Twenty four the sixth graders in Qing Jiang The sixth School participated in this study (mean age = 12.43, range from 11 to 13, 14 females and 10 males). They were rewarded a present after finishing the study. All participants were right-handed users and none of them were reported having medical problems. All the sixth graders were tested in a computer room of their own school and all were tested at a time.

Design and Materials

A verification task from Carpenter (1973) was employed in this study. Two independent variables were designed in target sentences (i.e. clause types and clause positions) and also another two independent variables were included in test sentences (i.e. truth values and polarities). Each experimental sentence received eight test sentences which were categorized as the four conditions (i.e. TA, FA, FN, and TN). These eight test sentences were used to probe if-clause or consequence clause. Four test sentences for each experimental sentence. The same experimental sentences were used as in simultaneous task with yaobushi, which could be referred in Appendix 10, 12, 14, 16, and test sentences could be referred in Appendix 11, 13, 15, 17. Practice materials could be referred in Appendix 9. The only difference of this study with simultaneous study was stimuli-of-asynchrony (SOA). In this study, the presented time between target clauses and test sentences was 5 seconds.

Procedure

A fixation point was displayed on the computer screen for 500ms. After this arousal of participants' attention, a target clause was displayed. After 5 seconds, a test sentence appeared instead. Participants were required to make judgment whether the test sentence was congruent with the target clause in its truth value. They were instructed to click buttons of the mouse as quickly and accurate as possible (i.e. left for correct response and right for incorrect response). Before experimental trials, practice trials were given to participants to make sure their understanding of this task. There were four blocks in whole experiment and it lasted about 30 minutes. Between two blocks, there was a break. In this case, there were three.

Prediction

Counterfactual conditionals are predicted to be processed as factuals after 5-second SOA. That is, the real meaning of counterfactuals is interpreted and represented mentally. In this case, a counterfactual conditional like Yaobushi I had been late, the car would not have been driven away is presented to participants in 5 seconds, it is predicted that only the simper representation would be formed (i.e. (late, I)). If so, the ordering of the four test sentences would be like this: TA < FA < FN < TN.

Results: College Students Data

The latencies and error rates to respond to factual and counterfactual clauses were shown in Table 38 below.

Counterfactual vs. factual target clauses

Participants responded to counterfactual target clauses and factual target clauses were 1919ms and 1827ms, respectively. These two response latencies were significantly different ($F(1, 2662) = 6.19$, $p = .01$). Meanwhile, participants made more errors on test sentences probing counterfactual target clauses than test sentences probing factual target clauses ($F(1, 2684) = 13.22$, $p = .0003$).

For counterfactual target clauses, there was also a significant interaction between truth values and polarities ($F(1, 1312) = 21.88$, $p < .0001$). The same pattern was not found on error rates ($F(1, 1334) = 0.04$, $p = .8$). The interaction of the four conditions on factual and counterfactual target clauses was not significant ($F(3, 2656) = 1.90$, $p = .1$). The main effects of target clauses and conditions of test sentences were significant. The differences of conditions with affirmatives (i.e. TA and FA) between factual target clauses and counterfactual target clauses were highly significant, but the differences of conditions with negatives (i.e. FN and TN) between factual and counterfactual target clauses were not significant. The results seemed to indicate that for college students test sentences with negatives were not easy to them no matter target clauses were factual or counterfactual.

For factual target clauses, there was an interaction between truth values and polarities and this interaction reached significant difference (F (1, 1322) = 29.79, $p < .0001$). The same pattern was not found on error rates (F (1, 1344) = 0.46, $p = .4$).

The interaction of the four conditions on error rates on both target clauses was not significant (F (3, 2656) = 0.04, $p = .9$). The main effect of error rates on target clauses was significant (F (1, 2656) = 30.18, $p < .0001$), but the main effect of error rates on conditions of test sentences was not significant (F (3, 2656) = 0.09, $p = .9$). The difference of each condition on error rates between factual and counterfactual target clauses reached significance, suggesting that college students erred more on counterfactual target clauses than on factual target clauses.

Counterfactual vs. factual clause position

Response latency to counterfactual target clauses in if-clause was 1836ms and in consequence clause was 1987ms. The difference of their latency also reached significance (F (1, 1335) = 6.20, $p = .01$), suggesting clause positions for counterfactual targets made difference in processing. The same pattern was found on their error rates (F (1, 1336) = 0.28, $p = .5$), suggesting that participants did not make more errors because of the clause positions.

Response latency to factual target clauses in if-clause was 1712ms and in consequence clause was 1922ms. The difference of their latency reached significance (F (1, 1345) = 14.12, $p = .0002$), suggesting clause positions for factual targets did have influence in processing. Their error rates did not show any difference (F (1, 1346) = 0.68, $p = .4$), suggesting that participants did not make more errors because of the clause positions.

For college students, clause positions caused processing difference no matter on factual or counterfactual target clauses.

Counterfactual vs. factual sentence type

Response latencies for counterfactual target clauses in four experimental sentence types were 1789ms, 1913ms, 2022ms, and 1931ms, respectively. A one-way ANOVA results did not show significant difference among them ($F(3, 1333) = 2.60$, $p = .05$), suggesting that none of the counterfactual target clauses in these four experimental sentences caused more difficulty than others. They were all similar in degree of difficulty. The same pattern was found on their error rates in one-way ANOVA ($F(3, 1334) = 0.14$, $p = .9$).

Response latencies for factual target clauses in four experimental sentence types were 1766ms, 1813ms, 1827ms, and 1851ms, respectively. A one-way ANOVA results did not show significant difference among them ($F(3, 1343) = 0.44$, $p = .7$), suggesting that none of the factual target clauses in these four experimental sentences caused more difficulty than others. They were all similar in degree of difficulty. The same pattern was found on their error rates in one-way ANOVA ($F(3, 1344) = 0.25$, $p = .8$).

For college students, clause positions did not cause any processing difference to them no matter the target clause was factual or counterfactual.

Test sentences in counterfactual targets

For counterfactual target sentences at 5-second SOA, the condition ordering was like the following: TA < FA < FN < TN, which matched with the prediction based on CCM. Participants responded fastest to test sentences with true affirmatives (1578ms), next was to test sentences with false affirmatives (1990ms), the third was to test sentences with false negatives (2033ms), and the last was to test sentences with true negatives (2091ms). A one-way ANOVA showed that the difference between these the four conditions was significant ($F(3, 1312) = 21.64$, $p < .0001$). A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. The major difference was attributed to the comparisons of TA and other groups at $p < .0001$ level. Parallel to Carpenter's findings on counterfactual clauses and modified by Chinese stimuli, as predicted test sentences with matched representations in predicates were responded significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA ($p < .0001$). Error rates for factual target clauses in the four conditions did not reach significant difference to one another ($F(3, 1344) = 0.20$, $p = .8$), but the difference of error rates for counterfactual target clauses in the four conditions reached significance ($F(3, 1334) = 0.14$, $p = .9$).

Test sentences in factual targets

As expected, for factual target clauses at 5-second SOA, the condition ordering was like the ordering observed in simultaneous task: TA < FA < FN < TN. That is, participants responded fastest to test sentences with true affirmatives (1432ms), next was to test sentences with false affirmatives (1795ms), the third was to test sentences with false negatives (1952ms), and the last was to test sentences with true negatives (2111ms). A one-way ANOVA showed that the difference between these the four conditions was significant ($F(3, 1322) = 36.32$, $p < .0001$).

A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. The results showed all comparisons were significant. Parallel to Carpenter's findings on factual clauses, test sentences with matched representations in predicates were responded significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA (p < .0001).

Affirmative vs. negative test sentences

For counterfactual target clauses, test sentences with affirmatives were responded faster than test sentences with negatives (1778ms and 2061ms, respectively) (F (1, 1335) = 23.15, p < .0001). The difference in their error rates was also not significant (F (1, 1336) = 0.23, p = .6).

For factual target clauses, test sentences with affirmatives were responded faster than test sentences with negatives (1611ms and 2030ms, respectively) (F (1, 1335) = 23.15, p < .0001). The difference in their error rates was not significant (F (1, 1346) = 0.10, p = .7)

.

Table 38 Response Latency (in ms) and Error Rates in Delayed Task with Yaobushi for College Students

Type of Clause		TA	FA	FN	TN	Total
Factual	RT	1432.06	1795.76	1952.34	2111.41	1827.35
	Errors	8.30%	8.20%	8.28%	8.07%	8.21%
Counterfactual	RT	1578.70	1990.78	2033.16	2091.92	1919.15
	Errors	8.98%	8.83%	8.76%	8.80%	8.85%

Results: The Eighth Graders Data

The latencies and error rates to respond to factual and counterfactual clauses were shown in Table 39 below.

Counterfactual vs. factual target clauses

For counterfactual target clauses, there was also a significant interaction between truth values and polarities ($F (1, 1593) = 30.66$, $p < .0001$). The same pattern was found on error rates ($F (1, 1261) = 0.14$, $p = .7$). The interaction of the four conditions on factual and counterfactual target clauses was significant ($F (3, 3242) = 2.77$, $p = .0402$), suggesting in general faster response latency on factual target clauses than on counterfactual target clauses. The differences of conditions with false responses between factual target clauses and counterfactual target clauses were highly significant, but the differences of conditions with true responses between these two targe clauses were not.

The interaction of the four conditions on error rates on both target clauses was not significant ($F (3, 3242) = 0.00$, $p = .9$). The main effect of error rates on target clauses was significant ($F (1, 3242) = 252.45$, $p < .0001$), but the main effect of error rates on the four conditions was not significant difference ($F (3, 3242) = 0.41$, $p = .7$). The difference of each condition on error rates between factual and counterfactual target clauses reached highly significance ($p < .0001$), suggesting that the eighth graders erred more on counterfactual target clauses than on factual target clauses as college students.

Participants responded to factual target clauses and counterfactual target clauses were 2004ms and 2185ms, respectively. These two response latencies were significantly different ($F (1, 3248) = 17.17$, $p < .0001$). Meanwhile, participants made more errors on test sentences probing counterfactual target clauses than test sentences probing factual target clauses ($F (1, 3276) = 40.23$, $p < .0001$). For factual target clauses, there was an interaction between truth values and polarities and this interaction reached significant difference (F (1, 1621) = 17.61, p < .0001). The same pattern was not found on error rates (F (1, 1649) = 0.51, p = .4).

Counterfactual vs. factual clause position

Response latency to counterfactual target clauses in if-clause was 2109ms and in consequence clause was 2236ms. The difference of their latency did not reach significance (F (1, 1622) = 3.42, p = .06), suggesting clause positions for counterfactual targets did not make difference in processing. The same pattern was found on their error rates (F (1, 1623) = 0.05, p = .8), suggesting that participants did not make more errors because of the clause positions.

Response latency to factual target clauses in if-clause was 1805ms and in consequence clause was 2171ms. The difference of their latency reached significance (F (1, 1650) = 37.00, p < .0001), suggesting clause positions for factual targets did have influence in processing. Their error rates did not show any difference (F (1, 1651) = 0.01, p = .9), suggesting that participants did not make more errors because of the clause positions.

For the eighth graders, clause position effect was only found on factual target clauses.

Counterfactual vs. factual sentence type

Response latencies for counterfactual target clauses in four experimental sentence types were 1993ms, 2224ms, 2348ms, and 2118ms, respectively. A one-way ANOVA results did not show significant difference among them (F (3, 1620) = 4.65, p = .003). Experimental sentence 1 was significantly different from sentence 2 and 3. Meanwhile experimental sentence 3 was significantly different from sentence 4. The same pattern was found on their error rates in one-way ANOVA (F (3, 1621) = 0.10, p

= .9).

Response latencies for factual target clauses in four experimental sentence types were 1833ms, 2066ms, 2033ms, and 2014ms, respectively. A one-way ANOVA results showed significant difference among them ($F (3, 1648) = 2.91$, $p = .03$). Experimental sentence 1 was significantly responded faster than sentence 2, 3, and 4, but not other comparisons. The same pattern was found on their error rates in one-way ANOVA ($F (3, 1649) = 0.12$, $p = .9$).

Test sentences in counterfactual targets

For counterfactual target sentences at 5-second SOA, the condition ordering was like the following: TA < FA < FN < TN. Participants responded fastest to test sentences with true affirmatives (1711ms), next was to test sentences with false affirmatives (2198ms), the third was to test sentences with false negatives (2288ms), and the last was to test sentences with true negatives (2520ms). A one-way ANOVA showed that the difference between these the four conditions was significant ($F (3, 1593) = 28.95$, $p < .0001$). A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. The results showed that almost all the comparisons were significant, but one wasn't (i.e. FA and FN, $p = .3$). Parallel to Carpenter's findings on counterfactual clauses and modified by Chinese stimuli, as predicted, test sentences with matched representations in predicates were responded significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA ($p < .0001$). Error rates for factual target clauses in the four conditions did not reach significant difference to one another, $F (3, 1649) = 0.60$, $p = .6$, and also the difference of error rates for counterfactual target clauses in the four conditions was not significant ($F (3, 1621) = 0.45$, $p = .7$).

Test sentences in factual targets

As predicted, for factual target clauses at 5-second SOA, the condition ordering from the easiest to the hardest was like the following: TA < FA < FN < TN. That is, participants responded fastest to test sentences with true affirmatives (1695ms), next was to test sentences with false affirmatives (1863ms), the third was to test sentences with false negatives (2059ms), and the last was to test sentences with true negatives (2360ms). A one-way ANOVA showed that the difference between these the four conditions was significant ($F (3, 1621) = 25.04$, $p < .0001$).

A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. The results showed all comparisons were significant. Parallel to Carpenter's findings on factual clauses, test sentences with matched representations in predicates were responded significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA ($p = .0258$).

Affirmative vs. negative test sentences

For counterfactual target clauses, test sentences with affirmatives were responded faster than test sentences with negatives (1950ms and 2406ms, respectively) ($F (1, 1622) = 42.09$, $p < .0001$). The difference in their error rates was also not significant ($F (1, 1623) = 0.53$, $p = .4$).

For factual target clauses, test sentences with affirmatives were responded faster than test sentences with negatives (1777ms and 2207ms, respectively) ($F (1, 1650) = 50.88$, $p < .0001$). The difference in their error rates was not significant ($F (1, 1651) = 0.00$, $p = .9$).

Table 39 Response Latency (in ms) and Error Rates in Delayed Task with *Yaobushi* for The Eighth Graders

Type of Clause		TA	FA	FN	TN	Total
Factual	RT	1695.15	1863.17	2059.05	2360.36	2004.30
	Errors	10.74%	10.19%	10.39%	10.51%	10.46%
Counterfactual	RT	1711.00	2198.58	2288.13	2520.97	2185.39
	Errors	12.17%	11.79%	11.67%	11.82%	11.87%

Results: The Sixth Graders Data

The latencies and error rates to respond to factual and counterfactual clauses were shown in Table 40 below.

NATIONAL TSING HUA UNIVERSITY

Counterfactual vs. factual target clauses

Participants responded to counterfactual target clauses and factual target clauses were 2154ms and 1978ms, respectively. These two response latencies were significantly different ($F(1, 2613) = 16.18$, $p < .0001$). Meanwhile, participants made more errors on test sentences probing counterfactual target clauses than test sentences probing factual target clauses ($F(1, 2636) = 108.34$, $p < .0001$).

For counterfactual target clauses, there was also a significant interaction between truth values and polarities ($F(1, 1269) = 8.64$, $p = .0034$). The same pattern was found on error rates ($F(1, 1292) = 0.51$, $p = .4738$). The interaction of the four conditions on factual and counterfactual target clauses was not significant ($F(3, 2607) = 1.79$, $p = .1$). The main effects of clause types and conditions of test sentences were significant. The difference of each condition between factual target clauses and counterfactual target clauses was highly significant (except TN condition, $p = .9$).

For factual target clauses, there was an interaction between truth values and polarities and this interaction reached significant difference ($F (1, 1315) = 34.73$, $p < .0001$). The same pattern was not found on error rates ($F (1, 1338) = 0.27$, $p = .6$).

The interaction of the four conditions on error rates on both target clauses was not significant ($F (3, 2607) = 0.24$, $p = .8$). The main effect of error rates on target clauses was significant ($F (1, 2607) = 425.26$, $p < .0001$), but the main effect of error rates on the four conditions was not significant difference ($F (3, 2607) = 0.26$, $p = .8$). The difference of each condition on error rates between factual and counterfactual target clauses reached highly significance ($p < .0001$), suggesting that the sixth graders erred more on counterfactual target clauses than on factual target clauses as other groups.

Counterfactual vs. factual clause position

Response latency to counterfactual target clauses in if-clause was 2073ms and in consequence clause was 2223ms. The difference of their latency reached significance ($F (1, 1293) = 4.73$, $p = .02$), suggesting clause positions for counterfactual targets made difference in processing. The same pattern was found on their error rates ($F (1, 1294) = 0.30$, $p = .5$), suggesting that participants did not make more errors because of the clause positions.

Response latency to factual target clauses in if-clause was 1826ms and in consequence clause was 2126ms. The difference of their latency reached significance ($F (1, 1339) = 22.63$, $p < .0001$), suggesting clause positions for factual targets did have influence in processing. Their error rates did not show any difference ($F (1, 1340) = 0.00$, $p = .9$), suggesting that participants did not make more errors because of the clause positions.

For the sixth graders, clause positions caused processing difference no matter on factual or counterfactual target clauses.

Counterfactual vs. factual sentence type

Response latencies for counterfactual target clauses in four experimental sentence types were 2142ms, 2084ms, 2310ms, and 2040ms, respectively. A one-way ANOVA results did not show significant difference among them ($F(3, 1291) = 2.84$, $p = .03$). The differences between experimental sentence 2 and sentence 3 and the difference between experimental sentence 3 and sentence 4 were also significant. The same pattern was not observed on their error rates in one-way ANOVA ($F(3, 1292) = 0.37$, $p = .7$).

Response latencies for factual target clauses in four experimental sentence types were 1897ms, 2016ms, 2030ms, and 1962ms, respectively. A one-way ANOVA results did not show significant difference among them ($F(3, 1337) = 1.04$, $p = .3$), suggesting that none of the factual target clauses in these four experimental sentences caused more difficulty than others. They were all similar in degree of difficulty. The same pattern was found on their error rates in one-way ANOVA ($F(3, 1338) = 0.31$, $p = .8$).

Test sentences in counterfactual targets

For counterfactual target sentences at 5-second SOA, the condition ordering was like the following: TA < FA < FN < TN. Participants responded fastest to test sentences with true affirmatives (1850ms), next was to test sentences with false affirmatives (2175ms), the third was to test sentences with false negatives (2286ms), and the last was to test sentences with true negatives (2302ms). A one-way ANOVA showed that the difference between these the four conditions was significant ($F(3, 1269) = 12.27$, $p < .0001$). A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. The major difference was attributed to the comparison of TA and other groups in $p < .0001$ level. Parallel to Carpenter’s findings on counterfactual clauses and modified by Chinese stimuli, as predicted, test sentences with matched representations in predicates were responded significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA ($p < .0001$). Error rates for factual target clauses in the four conditions did not reach significant difference to one another ($F(3, 1338) = 0.40$, $p = .7$), and also the difference of error rates for counterfactual target clauses in the four conditions did not reach significance ($F(3, 1292) = 0.25$, $p = .8$).

Test sentences in factual targets

As predicted, for factual target clauses at 5-second SOA, the condition ordering from the easiest to the hardest was like the following: TA < FA < FN < TN. That is, participants responded fastest to test sentences with true affirmatives (1575ms), next was to test sentences with false affirmatives (2015ms), the third was to test sentences with false negatives (2073ms), and the last was to test sentences with true negatives (2295ms). A one-way ANOVA showed that the difference between these the four conditions was significant ($F(3, 1315) = 28.74$, $p < .0001$).

A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. The results showed almost all the comparisons were significant, but one comparison wasn’t (FA vs. FN, $p = .7$). Parallel to Carpenter’s findings on factual clauses, test sentences with matched representations in predicates were responded significantly faster than test sentences with mismatched representations in predicates, namely, TA < FA ($p < .0001$).

Affirmative vs. negative test sentences

For counterfactual target clauses, test sentences with affirmatives were responded faster than test sentences with negatives (2007ms and 2294ms, respectively) ($F(1, 1293) = 16.18$, $p < .0001$). The difference in their error rates was also not significant ($F(1, 1294) = 0.02$, $p = .8$).

For factual target clauses, test sentences with affirmatives were responded faster than test sentences with negatives (1788ms and 2179ms, respectively) ($F(1, 1339) = 39.14$, $p < .0001$). The difference in their error rates was not significant ($F(1, 1340) = 0.91$, $p = .3$).

Table 40 Response Latency (in ms) and Error Rates in Delayed Task with *Yaobushi* for The Sixth Graders

Type of Clause		TA	FA	FN	TN	Total
Factual	RT	1575.86	2015.52	2073.00	2295.33	1978.39
	Errors	12.53%	12.37%	12.25%	12.12%	12.33%
Counterfactual	RT	1850.70	2175.21	2286.82	2302.35	2154.14
	Errors	15.19%	14.70%	14.95%	14.84%	14.93%

Summary

Table 41 summarized the findings of experiment IV in three age groups. The orderings of factual and counterfactual target clauses matched the predictions. Since SOA was 5 seconds, both factual and counterfactual orderings are the same. Affirmative with true response (TA) condition has the shortest response latency, negative with true response (TN) condition has the longest, and false responses in affirmative or negative are in between. All the orderings on both factual and counterfactual across three age groups are significant difference in reaction times, but not in error rates. Counterfactuals are responded slower in general than factuals.

The interaction of factual target clauses and counterfactual target clauses in the four conditions is significant only in the eighth group, suggesting factual target clauses generally received shorter response latency and lower error rates than counterfactual target clauses on all conditions. This pattern is not observed on college group and the sixth group. It seems that for the oldest and the youngest groups, factual and counterfactual target clauses are not different in long SOA.

Interactions of truth value and polarity in response latency are clearly observed both on factual target clauses and on counterfactual target clauses. Clause effect is observed on factual target clauses and also on counterfactual target clauses (except for the eighth group). Thus, it makes difference whether the target clause is in the first clause (i.e. if-clause) or in the second clause (i.e. consequence clause). However, experimental sentence types do cause processing differences on both factual and counterfactual target clauses in the eighth group. They showed difference response latency to different sentences. The same effect is also observed on counterfactual target clauses in the sixth group.

Meanwhile, as predicted, polarity effect is very robust on both factual and counterfactual target clauses in response latency across all three age groups. But, polarity effect does not show on error rates.

Table 41 Summary Findings of Three Groups in Delayed Experiment of Counterfactual Conditionals with *Yaobushi*

	COLLEGE	THE EIGHTH	THE SIXTH
Factual Ordering	TA < FA < FN	TA < FA < FN	TA < FA < FN

		< TN	< TN	< TN
Factual p-value	(RT)	p < .0001	p < .0001	p < .0001
	(Error Rates)	p = .8	p = .6	p = .7
Counterfactual Ordering		TA < FA < FN < TN	TA < FA < FN < TN	TA < FA < FN < TN
Counterfactual p-value	(RT)	p < .0001	p < .0001	p < .0001
	(Error Rates)	p = .9352	p = .7156	p = .8618
Factual vs. Counterfactual p-value	(RT)	p = .01	p < .0001	p < .0001
	(Error Rates)	p = .0003	p < .0001	p < .0001
Factual interaction of truth values and polarities	(RT)	p < .0001	p < .0001	p < .0001
	(Error Rates)	p = .4	p = .4	p = .6
Counterfactual interaction of truth values and polarities	(RT)	p < .0001	p < .0001	p = .003
	(Error Rates)	p = .8	p = .7	p = .4
Interaction of the four conditions on factual and counterfactual	(RT)	p = .1	p = .04	p = .1
	(Error Rates)	p = .9	p = .9	p = .8
Factual clause effect	(RT)	p = .0002	p < .0001	p < .0001

	(Error Rates)	p = .4	p = .9	p = .9
Counterfactual clause effect	(RT)	p = .01	p = .06	p = .02
	(Error Rates)	p = .5	p = .8	p = .5
Factual sentence type effect	(RT)	p = .7	p = .03	p = .3
	(Error Rates)	p = .8	p = .9	p = .8
Counterfactual sentence type effect	(RT)	p = .05	p = .003	p = .03
	(Error Rates)	p = .9	p = .9	p = .7
Factual polarity effect	(RT)	p < .0001	p < .0001	p < .0001
	(Error Rates)	p = .7486	p = .9553	p = .3408
Counterfactual polarity effect	(RT)	p < .0001	p < .0001	p < .0001
	(Error Rates)	p = .6343	p = .4654	p = .8894

General Discussion of Age Effect

Counterfactual vs. factual target clauses in different ages

A response latency of counterfactual and factual target clauses showed interaction among three age groups ($F (2, 8523) = 1.38$, $p = .2$). The main effect of age group was not significant ($F (2, 8523) = 1.34$, $p = .2$) and the main effect of target clause was significant ($F (1, 8523) = 37.58$, $p < .0001$). None of the differences of response latency on factual and counterfactual target clauses reached significance. However, the interaction of response latency of factual and counterfactual target clauses on the four conditions for three age groups was significant ($F(23, 8505) = 21.30$, $p < .0001$). A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed, showing almost all comparisons between each condition in three age groups were not significant (except TN condition for the comparison of college and the eighth graders).

Test sentences in counterfactual targets in different ages

The interaction between truth values and polarity on counterfactual target clauses for three groups reached significant difference ($F(11, 4174) = 17.63$, $p < .0001$), so did the interaction on factual target clauses ($p < .0001$). The interaction between clause positions and different age groups on counterfactual target clauses was significant ($F(5, 4180) = 4.14$, $p = .0009$). All the comparisons of three groups did not reach significance, suggesting in general similar response latency for both target clauses while SOA was 5 seconds.

Test sentences in factual targets in different ages

The interaction between clause positions and different age groups on factual target clauses was not significant ($F(2, 4264) = 1.57$, $p = .2085$). Meanwhile, the main effect of age group was not significant ($F(2, 4264) = 1.02$, $p = .3615$), and the main effect of clause positions was significant ($F(1, 4264) = 77.19$, $p < .0001$). All the comparisons between if-clauses and consequence clauses were not significant for factual target clauses, showing overall easiness in processing for these three groups.

Counterfactual vs. factual sentence type in different ages

The interaction between age group and different experimental sentences was not significant on counterfactuals ($F (6, 4174) = 1.32$, $p = .2$), suggesting that there was no specific experimental sentence caused more difficult processing. The main effect of age group was not significant ($F (2, 4174) = 1.48$, $p = .2$) and main affect of experimental sentence was significant ($p < .0001$). None of the comparisons was significant. The interaction between age group and different target sentences was not significant on factual target clauses ($F (6, 4258) = 0.70$, $p = .6$), suggesting that there was no specific experimental sentence caused more difficult processing. The main effect of age group was not significant ($F (2, 4258) = 1.04$, $p = .3$) and main affect of experimental sentence was significant ($p = .01$). All the comparisons between two clauses on factual target clauses were not significant.

Affirmative vs. negative test sentences in factual sentences in different ages

The interaction between age group and polarity on counterfactual target clauses was not significant ($F (2, 4180) = 2.54$, $p = .07$). The main effect of age group was not significant ($p = .2$) and the main effect of polarity was also significant ($p < .0001$). Only one comparison was significant (i.e. test sentences with negatives between college students and the eighth graders). Other comparisons of other groups did not reach significance.

An interaction between age group and polarity on factual target clauses was also not significant ($F (2, 4264) = 0.09$, $p = .9$). The main effect of age group was not significant ($p = .3$) and the main effect of polarity was significant ($p < .0001$). All the comparisons between any two groups were not significant.

These results indicated that the younger groups like the sixth graders and the eighth graders did not show difference from each other in response latency on both test sentences with affirmatives and negatives, so did the comparisons between college students and the sixth graders. However, there was one difference in response latency between college students and the eighth graders reached significance, (i.e. test sentences with negatives).

Table 42 Response Latency (in ms) and Error Rates in Delayed Task with *Yaobushi* on Factual and Counterfactual Target Clauses for Three Age Groups

Type of Clause	Group	TA	FA	FN	TN	Total
Factual	College	1432.06	1795.76	1952.34	2111.41	1827.35
		8.30%	8.20%	8.28%	8.07%	8.21%
	The Eighth	1695.15	1863.17	2059.05	2360.36	2004.30
		10.74%	10.19%	10.39%	10.51%	10.46%
	The Sixth	1575.86	2015.52	2073.00	2295.33	1978.39
		12.53%	12.37%	12.25%	12.12%	12.33%
Counterfactual	College	1578.70	1990.78	2033.16	2091.92	1919.15
		8.98%	8.83%	8.76%	8.80%	8.85%
	The Eighth	1711.00	2198.58	2288.13	2520.97	2185.39
		12.17%	11.79%	11.67%	11.82%	11.87%
	The Sixth	1850.70	2175.21	2286.82	2302.35	2154.14
		15.19%	14.70%	14.95%	14.84%	14.93%

General Discussion of Task Effect in Counterfactuals with *Yaobushi*

A response latency of factual and counterfactual target clauses in experiments with different SOA showed a significant interaction among three age groups ($F(7, 16E3) = 40.48$, $p < .0001$). The main effect of age group was not significant (p

= .3546) while the main effects of tasks with different SOA and clause types both were significant at .0001. All the comparisons of target clauses in three different age groups in different experiments reached significance, suggesting a clear task effect on each age group.

The interaction of response latency of factual and counterfactual target clauses on the four conditions for three age groups in different experiments was also significant ($F (47, 16E3) = 349.53$, $p < .0001$). A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed, showing all the comparisons reached significance. Task difference caused processing difference in each group.

The interaction between truth values and polarity on counterfactual target clauses for three groups in different experiments reached significant difference ($F (23, 7973) = 350.00$, $p < .0001$), so did the interaction on factual target clauses ($F (23, 8197) = 362.70$, $p < .0001$).

The interaction between clause positions and different age groups on counterfactual target clauses in different experiments was significant ($F (11, 7985) = 695.48$, $p < .0001$), suggesting that clause positions received different processing in different experiments.

The interaction between clause positions and different age groups on factual target clauses was also significant ($F (7, 8209) = 11.05$, $p < .0001$). The pattern was exactly parallel to the findings on counterfactual target clauses.

The interaction between age group and different experimental sentences was significant on counterfactuals in different experiments ($F (17, 7973) = 12.11$, $p < .0001$).

The interaction between age group and polarity on factual target clauses was significant ($F (18, 8197) = 430.14$, $p < .0001$). All the comparisons were significant within each group. An interaction between age group and polarity on counterfactual target clauses was also significant ($F(7, 7985) = 23.28$, $p < .0001$).

These results indicated that different tasks with different SOA may cause processing difference, which was reflected clearly on each age group.

F Williams Syndrome Study

Participants

Three Williams Syndrome individuals were recruited to join in this study. Each participant was diagnosed to be one of the members having this syndrome with *Fluorescent in situ hybridization* (FISH) test in hospital or in laboratory before the experiment. After finishing the experiment, each of them rewarded a present. The mean age of these three participants was 14yr and 8m.

Design and Materials

The design is exactly the same as the one tested on normal participants in Experiment IV. The only different point is that there is no time limit for individuals with WS to make judgment. That is, there is no limitation of SOA. They can read or comprehend target sentences as long as they need. The stimuli of target sentences are listed in Appendix 10, 12, 14, and 16. Meanwhile, the stimuli of test sentences are listed in Appendix 11, 13, 15, and 17. Practice Stimuli is listed in Appendix 9.

Procedure

This task is pretended to be teacher-student game, which means that individuals with WS play a role of teaching the computer to say correct meaning (i.e. the test sentences) of the scenario (i.e. the target sentences). They were instructed to reward or punish the computer by answering yes or no via a mouse. If a yes was pressed, the computer would be encouraged and happy; if a no was responded, the computer would be reminded to say right thing next time. They should not be afraid to punish the computer because it might say wrong thing sometimes and not aware of it. So, participants were encouraged to be a good and responsible teacher. This pretence makes individuals with WS have fun in doing this task, thus they did not feel bored. The real procedure is like the following: a fixation point appeared on the computer screen for 500ms. Then a target sentence showed up instead without time limitation. Participants with WS were required to read and comprehend it and instructed to press the spacebar after reading. Next, a test sentence was displayed, and they should press the mouse to make judgment of it whether the test sentence matched with the target sentence in its meaning. All the participants were tested in their own houses and rewarded a present after finishing the task.

TSING HUA UNIVERSITY

Results: Individual Analysis

Since participants with Williams Syndrome may have their own characteristics in performance, their data would be presented by individual first and by group later.

Participant: LMH

The latencies and error rates to respond to factual and counterfactual clauses were shown in Table 43 below.

Counterfactual vs. factual target clauses

LMH responded to counterfactual target clauses and factual target clauses were 1203ms and 1218ms, respectively. These two response latencies were not significantly different (F (1, 110) = 0.03, p = .8). Meanwhile, LMH made more errors on test sentences probing counterfactual target clauses than test sentences probing factual target clauses (F (1, 110) = 6.31E17, p < .0001).

For counterfactual target clauses, the difference was also not significant between truth values and polarities (F (1, 47) = 2.22, p = .1). The main effect of clauses were not significant (p = .6), but the main effect of the four conditions was significant (p = .003).

For factual target clauses, there was an interaction between truth values and polarities and this interaction was not significantly different (F (1, 57) = 0.14, p = .7).

But, the interaction of the four conditions on factual and counterfactual target clauses was not significant (F (3, 104) = 1.56, p = .2), suggesting that response latency did not differ on certain condition of both target clauses. Meanwhile, none of the difference between each condition of factual and counterfactual target clauses was significant.

Counterfactual vs. factual clause position

Response latency to counterfactual target clauses in if-clause was 1193ms and in consequence clause was 1212ms. The difference of their latency also did not reach significance (F (1, 49) = 0.02, p = .8), suggesting clause positions for counterfactual targets also did not make any difference in processing.

Response latency to factual target clauses in if-clause was 1216ms and in consequence clause was 1221ms. The difference of their latency did not reach significance (F (1, 59) = 0.00, p = .9), suggesting clause positions for factual targets did not cause difference in processing.

For LMH with Williams Syndrome, clause positions did not cause processing difference no matter on factual or counterfactual target clauses.

Counterfactual vs. factual sentence type

Response latencies for counterfactual target clauses in four experimental sentence types were 1086ms, 1405ms, 1312ms, and 1066ms, respectively. A one-way ANOVA results did not show significant difference among them ($F(3, 47) = 1.86$, $p = .14$), suggesting that none of the counterfactual target clauses in these four experimental sentence types caused more difficulty than others. They were all similar in degree of difficulty.

Response latencies for factual target clauses in four experimental sentence types were 1033ms, 1385ms, 1259ms, and 1190ms, respectively. A one-way ANOVA results did not show significant difference among them ($F(3, 57) = 1.69$, $p = .1$), suggesting that none of the factual target clauses in these four experimental sentences caused more difficulty than others. They were all similar in degree of difficulty.

Test sentences in counterfactual targets

For counterfactual target clauses, the condition ordering was like the following: TA < FA < FN < TN, which did not match any predictions. It seemed that truth value was a more important factor to this child with WS. LMH responded fastest to test sentences with true affirmatives (969ms), next was to test sentences with false affirmatives (1055ms), the third was to test sentences with false negatives (1238ms), and the last was to test sentences with true negatives (1493ms). A one-way ANOVA showed that the difference between these the four conditions was significant ($F(3, 47) = 5.02$, $p = .004$). A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. The major difference was attributed to the comparisons of TN with TA ($p = .0007$) and TN with FA ($p = .009$).

Test sentences in factual targets

For factual target clauses, the condition ordering from the easiest to the hardest was like the following: TA < FA < TN < FN. That is, participants responded fastest to test sentences with true affirmatives (1016ms), next was to test sentences with false affirmatives (1244ms), the third was to test sentences with true negatives (1251ms), and the last was to test sentences with false negatives (1395ms). A one-way ANOVA showed that the difference between these the four conditions was significant ($F(3, 57) = 1.86$, $p = .1$), suggesting no difference between each test condition. Thus, the congruency principle proposed by Carpenter was not confirmed on this participant, TA vs. FA ($p = .1$).

Affirmative vs. negative test sentences

For counterfactual target clauses, test sentences with affirmatives were responded faster than test sentences with negatives (1003ms and 1395ms, respectively) ($F(1, 49) = 11.63$, $p = .001$).

For factual target clauses, test sentences with affirmatives were responded faster than test sentences with negatives (1130ms and 1316ms, respectively) ($F(1, 59) = 2.56$, $p = .1$), suggesting test sentences with affirmatives were easier than test sentences with negatives for factual target clauses to LMH.

To sum up, for LMH, the effect of polarity did make difference no matter on factual or counterfactual target clauses as unimpaired participants.

Table 43 Response Latency (in ms) and Error Rates in Delayed Task with *Yaobushi* for Individuals with Williams Syndrome (LMH)

Type of Clause	Variables	TA	FA	FN	TN	Total
Factual	RT	1016.78	1244.84	1395.78	1251.67	1218.98
	Error Rates	0%	0%	18.75%	0%	4.69%

Counterfactual	RT	969.07	1055.41	1238.60	1493.56	1203.39
	Error Rates	6.25%	37.50%	37.50%	0%	20.31%

Participant: ZYL

The latencies and error rates to respond to factual and counterfactual clauses were shown in Table 44 below.

Counterfactual vs. factual target clauses

ZYL responded to counterfactual target clauses and factual target clauses were 8253ms and 8344ms. These two response latencies were not significantly different ($F (1, 63) = 0.00$, $p = .9$.) Meanwhile, ZYL made more errors on test sentences probing counterfactual target clauses than test sentences probing factual target clauses ($F (1, 63) = 2.86E14$, $p < .0001$). For factual target clauses, there was an interaction between truth values and polarities and this interaction was not significantly different ($F (1, 28) = 0.58$, $p = .4$).

For counterfactual target clauses, the difference was not significant between the interaction of truth values and polarities ($F (1, 29) = 1.94$, $p = .1$). The main effects of clauses and the four conditions were both not significant ($p = .9$ for clauses; $p = .3$ for the four conditions). The interaction of the four conditions on factual and counterfactual target clauses was not significant ($F (3, 57) = 0.10$, $p = .9$), suggesting that response latency did not differ on certain condition of both target clauses. Meanwhile, none of the difference between each condition of factual and counterfactual target clauses was significant.

Counterfactual vs. factual clause position

Response latency to counterfactual target clauses in if-clause was 7932ms and in consequence clause was 8593ms. The difference of their latency also did not reach significance (F (1, 31) = 0.10, p = .7), suggesting clause positions for counterfactual targets also did not make any difference in processing.

Response latency to factual target clauses in if-clause was 8282ms and in consequence clause was 8398ms. The difference of their latency did not reach significance (F (1, 30) = 0.00, p = .9), suggesting clause positions for factual targets did not have influence in processing for ZYL.

For ZYL with Williams Syndrome, clause positions did not cause processing difference no matter on factual or counterfactual target clauses

Counterfactual vs. factual sentence type

Response latencies for counterfactual target clauses in four experimental sentence types were 10731ms, 7282ms, 7289ms, and 6430ms, respectively. A one-way ANOVA results did not show significant difference among them (F (3, 29) = 1.09, p = .3), suggesting that none of the counterfactual target clauses in these four experimental sentence types caused more difficulty than others. They were all similar in degree of difficulty.

Response latencies for factual target clauses in four experimental sentence types were 8892ms, 7300ms, 10416ms, and 6690ms, respectively. A one-way ANOVA results did not show significant difference among them (F (3, 28) = 0.60, p = .6), suggesting that none of the factual target clauses in these four experimental sentences caused more difficulty than others. They were all similar in degree of difficulty.

Test sentences in counterfactual targets

For counterfactual target clauses, the condition ordering was like the following: TA < FN < FA < TN, which did not match any predictions. It seemed that truth value was a more important factor to this child with WS. ZYL responded fastest to test sentences with true affirmatives (6187ms), next was to test sentences with false negatives (6926ms), the third was to test sentences with false affirmatives (9215ms), and the last was to test sentences with true negatives (10271ms). A one-way ANOVA showed that the difference between these the four conditions was significant ($F(3, 29) = 1.21$, $p = .3$).

A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. None of the difference between each condition was significant. Again, Carpenter’s congruency principle was not confirmed, TA vs. FA ($p = .35$).

Test sentences in factual targets

For factual target clauses, the condition ordering from the easiest to the hardest was like the following: TA < FN < FA < TN. That is, participants responded fastest to test sentences with true affirmatives (7303ms), next was to test sentences with false negatives (7676ms), the third was to test sentences with false affirmatives (8578ms), and the last was to test sentences with false negatives (9644ms). A one-way ANOVA showed that the difference between these the four conditions was significant ($F(3, 28) = 0.27$, $p = .8$), suggesting no difference between each test condition. Thus, the congruency principle proposed by Carpenter was not confirmed on this participant, TA vs. FA ($p = .6$).

Affirmative vs. negative test sentences

For counterfactual target clauses, test sentences with affirmatives were responded faster than test sentences with negatives (6944ms and 9484ms, respectively) ($F(1, 31) = 1.58$, $p = .2$).

For factual target clauses, test sentences with affirmatives were responded faster than test sentences with negatives (7781ms and 8906ms, respectively) (F (1, 30) = 0.27, p = .6), suggesting test sentences with affirmatives were easier than test sentences with negatives for factual target clauses to ZYL.

To sum up, for ZYL, the effect of polarity did make difference no matter on factual or counterfactual target clauses as unimpaired participants.

Table 44 Response Latency (in ms) and Error Rates in Delayed Task with *Yaobushi* for Individuals with Williams Syndrome (ZYL)

Type of Clause	Variables	TA	FA	FN	TN	Total
Factual	RT	7303.46	8578.65	7676.72	9644.43	8344.10
	Error Rates	37.50%	62.50%	62.50%	37.50%	50%
Counterfactual	RT	6187.83	9215.70	6926.55	10271.35	8253.05
	Error Rates	25%	75%	75%	18.75%	48.44%

Participant: TSJ

The latencies and error rates to respond to factual and counterfactual clauses were shown in Table 45 below.

Counterfactual vs. factual target clauses

TSJ responded to counterfactual target clauses and factual target clauses were 6078ms and 6581ms, respectively. These two response latencies were not significantly different (F (1, 65) = 0.21, p = .6). Meanwhile, TSJ made more errors on test sentences probing counterfactual target clauses than test sentences probing factual target clauses (F (1, 65) = 1.04E16, p < .0001).

For factual target clauses, there was a significant interaction between truth values and polarities (F (1, 32) = 4.66, p = .03).

For counterfactual target clauses, there was also a significant interaction between truth values and polarities (F (1, 27) = 6.36, p = .01). The main effect of clause type was not significant (p = .7), but the main effect of the four conditions was significant (p = .003).

The interaction of the four conditions on counterfactual and factual target clauses was not significant (F (3, 59) = 0.51, p = .6), suggesting that response latency did not differ on certain condition of both target clauses. Meanwhile, none of the difference between each condition of factual and counterfactual target clauses was significant.

Counterfactual vs. factual clause position

Response latency to counterfactual target clauses in if-clause was 6019ms and in consequence clause was 6150ms. The difference of their latency also did not reach significance (F (1, 29) = 0.01, p = .9), suggesting clause positions for counterfactual targets also did not make any difference in processing.

Response latency to factual target clauses in if-clause was 6663ms and in consequence clause was 6489ms. The difference of their latency did not reach significance (F (1, 34) = 0.01, p = .9), suggesting clause positions for factual targets did not have influence in processing for TSJ.

For TSJ with Williams Syndrome, clause positions did not cause processing difference no matter on factual or counterfactual target clauses.

Counterfactual vs. factual sentence type

Response latencies for counterfactual target clauses in four experimental sentence types were 6898ms, 5126ms, 6412ms, and 5594ms, respectively. A one-way ANOVA results did not show significant difference among them (F (3, 27) = 0.27, p

= .8), suggesting that none of the counterfactual target clauses in these four experimental sentence types caused more difficulty than others. They were all similar in degree of difficulty.

Response latencies for factual target clauses in four experimental sentence types were 5025ms, 9106ms, 7085ms, and 4692ms, respectively. A one-way ANOVA results did not show significant difference among them (F (3, 32) = 2.30, p = .09), suggesting that none of the factual target clauses in these four experimental sentences caused more difficulty than others. They were all similar in degree of difficulty.

Test sentences in counterfactual targets

For counterfactual target clauses, the condition ordering was like the following: FA < FN < TA < TN, which did not match any predictions. It seemed that truth value was a more important factor to this child with WS. TSJ responded fastest to test sentences with false affirmatives (5210ms), next was to test sentences with false negatives (5442ms), the third was to test sentences with true affirmatives (5849ms), and the last was to test sentences with true negatives (14596ms). A one-way ANOVA showed that the difference between these the four conditions was significant (F (3, 27) = 3.79, p = .02).

A proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. The major difference was attributed to the comparisons of TN with other groups (TN vs. TA, p = .0060; TN vs. FA, p = .003; TN vs. FN, p = .003). This time Carpenter's congruency principle was also not confirmed, TA vs. FA (p = .7).

Test sentences in factual targets

For factual target clauses, the condition ordering from the easiest to the hardest was like the following: TA < FN < FA < TN. That is, participants responded fastest to test sentences with true affirmatives (5008ms), next was to test sentences with false negatives (6163ms), the third was to test sentences with false affirmatives (6656ms), and the last was to test sentences with false negatives (11739ms). A one-way ANOVA showed that the difference between these the four conditions was significant ($F(3, 32) = 1.87$, $p = .1$), suggesting no difference between each test condition. Thus, the congruency principle proposed by Carpenter was not confirmed on this participant, TA vs. FA ($p = .4$).

Affirmative vs. negative test sentences

For counterfactual target clauses, test sentences with affirmatives were responded faster than test sentences with negatives (5529ms and 6663ms, respectively) ($F(1, 29) = 0.54$, $p = .4$).

Test sentences with affirmatives were responded faster than test sentences with negatives for factual target clauses (6079ms and 7208ms, respectively) ($F(1, 34) = 0.56$, $p = .4$), suggesting test sentences with affirmatives were easier than test sentences with negatives for factual target clauses to TSJ.

To sum up, for TSJ, the effect of polarity did make difference no matter on factual or counterfactual target clauses as unimpaired participants.

Table 45 Response Latency (in ms) and Error Rates in Delayed Task with *Yaobushi* for Individuals with Williams Syndrome (TSJ)

Type of Clause	Variables	TA	FA	FN	TN	Total
Factual	RT	5008.56	6656.68	6163.14	11739.70	6581.57
	Error Rates	56.25%	18.75%	18.75%	81.25%	43.75%

Counterfactual	RT	5849.05	5210.63	5442.90	14596.60	6078.33
	Error Rates	50%	50%	18.75%	87.50%	51.56%

Group Data Analysis

The average latencies and error rates to respond to factual and counterfactual clauses of three Williams Syndrome children were listed in Table 46.

Counterfactual vs. factual target clauses

Participants responded to counterfactual target clauses and factual target clauses were 4482ms and 4540ms, respectively. These two response latencies were not significantly different ($F (1, 240) = 0.11$, $p = .7$). Meanwhile, participants made more errors on test sentences probing counterfactual target clauses than test sentences probing factual target clauses ($F (1, 242) = 17.91$, $p < .0001$).

For factual target clauses, there was an interaction between truth values and polarities and this interaction did not reach significant difference ($F (1, 123) = 2.20$, $p = .1$). The same pattern was not found on error rates ($F (1, 125) = 0.36$, $p = .5$).

For counterfactual target clauses, the interaction between truth values and polarities was also not significant ($F (1, 109) = 3.72$, $p = .05$). The same pattern was found on error rates ($F (1, 111) = 1.06$, $p = .3$). Thus, none of the conditions was significant difference between factual and counterfactual target clauses.

The interaction of the four conditions on counterfactual and factual target clauses was not significant ($F (3, 234) = 0.22$, $p = .8$), suggesting the response latency did not differ on certain condition of target clauses. The main effect of clause type was not significant ($p = .6$), but the main effect of four test conditions was significant ($p = .01$). None of the difference between factual and counterfactual target clauses was significant.

Counterfactual vs. factual clause position

Response latency to counterfactual target clauses in if-clause was 4583ms and in consequence clause was 4496ms. The difference of their latency did not reach significance ($F(1, 112) = 0.11$, $p = .7$), suggesting clause positions for counterfactual targets did not make difference in processing. The same pattern was found on their error rates ($F(1, 113) = 0.44$, $p = .5$), suggesting that participants did not make more errors because of the clause positions.

Response latency to factual target clauses in if-clause was 4489ms and in consequence clause was 4476ms. The difference of their latency did not reach significance ($F(1, 126) = 0.01$, $p = .9$), suggesting clause positions for factual targets did not have influence in processing. Their error rates did not show any difference ($F(1, 127) = 0.00$, $p = .9$), suggesting that participants did not make more errors because of the clause position.

For children with Williams Syndrome, clause position did not cause processing difference no matter on factual or counterfactual target clauses.

Counterfactual vs. factual sentence type

Response latencies for counterfactual target clauses in four experimental sentence types were 5612ms, 4329ms, 4261ms, and 3631ms, respectively. A one-way ANOVA results did not show significant difference among them ($F(3, 110) = 1.10$, $p = .3$), suggesting that none of the counterfactual target clauses in these four experimental sentence types caused more difficulty than one another. They were all similar in degree of difficulty. The same pattern was found on their error rates in one-way ANOVA ($F(3, 111) = 0.10$, $p = .9$).

Response latencies for factual target clauses in four experimental sentence types were 4060ms, 5112ms, 5314ms, and 3418ms, respectively. A one-way ANOVA results showed significant difference among them (F (3, 124) = 1.28, p = .2). The same pattern was found on their error rates in one-way ANOVA (F (3, 125) = 0.06, p = .9).

Test sentences in counterfactual targets

For counterfactual target sentences, the condition ordering was like the following: TA < FA < FN < TN. This result seemed to indicate that truth value factor was a more important factor for children with Williams Syndrome to make judgment. Participants responded fastest to test sentences with true affirmatives (3873ms), next was to test sentences with false affirmatives (4050ms), the third was to test sentences with false negatives (4105ms), and the last was to test sentences with true negatives (6019ms). A one-way ANOVA showed that the difference between these the four conditions was not significant (F (3, 109) = 2.60, p = .05). The same finding was observed on error rates of counterfactuals (F (3, 111) = 0.66, p = .5). Once more, contrary to Carpenter's findings on counterfactual clauses, test sentences with matched representations in predicates were not responded significantly faster than test sentences with mismatched representations in predicates, namely, TA vs. FA (p = .6).

Test sentences in factual targets

For factual target clauses, the condition ordering for children with Williams Syndrome from the easiest to the hardest was like the following: TA < FN < FA < TN. That is, participants responded fastest to test sentences with true affirmatives (3768ms), next was to test sentences with false negatives (4510ms), the third was to test sentences with false affirmatives (4512ms), and the last was to test sentences with true negatives (5230ms). A one-way ANOVA showed that the difference between these the four conditions was not significant (F (3, 123) = 1.08, p = .3). The same finding was also observed on error rates of factuals (F (3, 125) = 0.25, p = .8). Contrary to Carpenter's findings on factual clauses, test sentences with mismatched representations in predicates were responded significantly faster than test sentences with matched representations in predicates, namely, TA vs. FA (p = .4).

Affirmative vs. negative test sentences

For counterfactual target clauses, test sentences with affirmatives were responded faster than test sentences with negatives (3941ms and 5128ms, respectively) (F (1, 112) = 2.70, p = .1). The difference in their error rates was also not significant (F (1, 113) = 0.02, p = .9).

Test sentences with affirmatives were responded faster than test sentences with negatives for factual target clauses (4151ms and 4852ms, respectively) (F (1, 126) = 1.14, p = .2). The difference in their error rates was not significant (F (1, 127) = 0.00, p = .9).

Table 46 Response Latency (in ms) and Error Rates in Delayed Task with *Yaobushi* for Individuals with Williams Syndrome

Type of Clause	Variables	TA	FA	FN	TN	Total
Factual	RT	3768.58	4512.17	4510.20	5230.69	4482.99
	Error Rates	31.25%	27.08%	33.33%	39.58%	32.81%
Counterfactual	RT	3873.78	4050.09	4105.55	6019.93	4540.45
	Error Rates	27.08%	54.17%	43.75%	35.42%	40.10%

Table 47 Response Latency (in ms) and Error Rates in Delayed Task with *Yaobushi* for Individuals with Williams Syndrome

Type of Clause	Participant	TA	FA	FN	TN	Total
Factual	LHM	1016.78	1244.84	1395.78	1251.67	1218.98
		0%	0%	18.75%	0%	4.69%
	ZYL	7303.46	8578.65	7676.72	9644.43	8344.10
		37.50%	62.50%	62.50%	37.50%	50%
	TSJ	5008.56	6656.68	6163.14	11739.70	6581.57
		56.25%	18.75%	18.75%	81.25%	43.75%
Counterfactual	LMH	969.07	1055.41	1238.60	1493.56	1203.39
		6.25%	37.50%	37.50%	0%	20.31%
	ZYL	6187.83	9215.70	6926.55	10271.35	8253.05
		25%	75%	75%	18.75%	48.44%
	TSJ	5849.05	5210.63	5442.90	14596.60	6078.33
		50%	50%	18.75%	87.50%	51.56%

Summary

In this study, three participants of WS performed non-form-based representations. Participant LMH showed a pattern which is similar to the one observed from unimpaired participants, namely, TA < FA < TN < FN. Though the ordering is slightly different from the prediction (i.e. TA < FA < FN < TN), basically the pattern followed the predicted ordering of meaning-based representation of CCM. That is, affirmatives were responded faster than negatives and this pattern was consistent on both factual and counterfactual target clauses. Further, only the main effect of counterfactuals on response latency of four test conditions reached significance.

For ZYL and TSJ, they also performed a pattern which is similar to the prediction of meaning-based representation ordering. Though the ordering of false responses is different from the exact prediction (i.e. FN < FA vs. FA < FN), basically the pattern is the same because TA is the fastest condition, TN is the slowest and FN/FA are in between.

Participant TSJ also showed significant difference of test conditions on counterfactuals, but participant ZYL did not. Besides, TSJ also showed a pattern which was observed in unimpaired participants: the interaction of truth values and polarities of both factuals and counterfactuals. These interactions mean that the effects caused on conditions were different. They did not show clause effect both on factuals and counterfactuals, suggesting different clause positions did not cause any processing difference on them. This observation is very different from the unimpaired which showed the significant clause position differences on both target clauses. The sentence type effect is not significant, which is similar to the pattern of the unimpaired. Moreover, all the difference between affirmatives and negatives was not significantly different except the one of counterfactuals in response latency of participant LHM.

Table 48 Summary Findings of Two Children with Williams Syndrome in Delayed Experiment of Counterfactual Conditionals

	LMH	ZYL	TSJ
Factual Ordering	TA < FA < TN < FN	TA < FN < FA < TN	TA < FN < FA < TN
Factual p-value (RT)	$p = .1$	$p = .8$	$p = .1$
Counterfactual Ordering	TA < FA < FN < TN	TA < FN < FA < TN	FA < FN < TA < TN

Counterfactual p-value	(RT)	p = .004	p = .3	p = .02
Factual vs. Counterfactual p-value	(RT)	p = .8	p = .9	p = .6
Factual interaction of truth values and polarities	(RT)	p = .7	p = .4	p = .03
Counterfactual interaction of truth values and polarities	(RT)	p = .1	p = .1	p = .01
Interaction of the four conditions on factual and counterfactual	(RT)	p = .2	p = .9	p = .6
Factual clause effect	(RT)	p = .9	p = .9	p = .9
Counterfactual clause effect	(RT)	p = .8	p = .7	p = .9
Factual sentence type effect	(RT)	p = .1	p = .6	p = .09

Counterfactual sentence type effect	(RT)	p = .1	p = .3	p = .8
Factual polarity effect	(RT)	p = .1	p = .6	p = .4
Counterfactual polarity effect	(RT)	p = .001	p = .2	p = .4

G Comparison of Individual of the Sixth Graders and Children with Williams Syndrome

From the patterns of counterfactuals observed on the sixth graders, there are totally ten types of orderings can be summarized. There are four participants performed the same pattern as participant LMH (i.e. pattern 1). Meanwhile, pattern 2, which is observed on participant ZYL, is also found on four participants. Participant TSJ performed a similar pattern as pattern (10) with a slightly different ordering of true responses. However, the basic picture is the same because false responses triumph true responses. That is, the patterns individuals of WS showed are not deviant from the patterns observed on the unimpaired. In other words, WS individuals performed in the normal range.

Table 49 Individual Patterns of Counterfactual Target Clauses with *Yaobushi* in The Sixth Graders

PATTERN	ORDERING	NUMBER
(1)	TA < FA < FN < TN	4
(2)	TA < FN < FA < TN	4

(3)	TA < FN < TN < FA	5
(4)	FA < TA < FN < TN	2
(5)	TA < FA < TN < FN	1
(6)	FA < TA < TN < FN	2
(7)	TN < FN < FA < TA	1
(8)	TA < TN < FA < FN	2
(9)	TA < TN < FN < FA	1
(10)	FA < FN < TN < TA	1

Fig. 8. Comparisons of response latencies of factual target clauses for three groups of participants in simultaneous task with *yaobushi*.

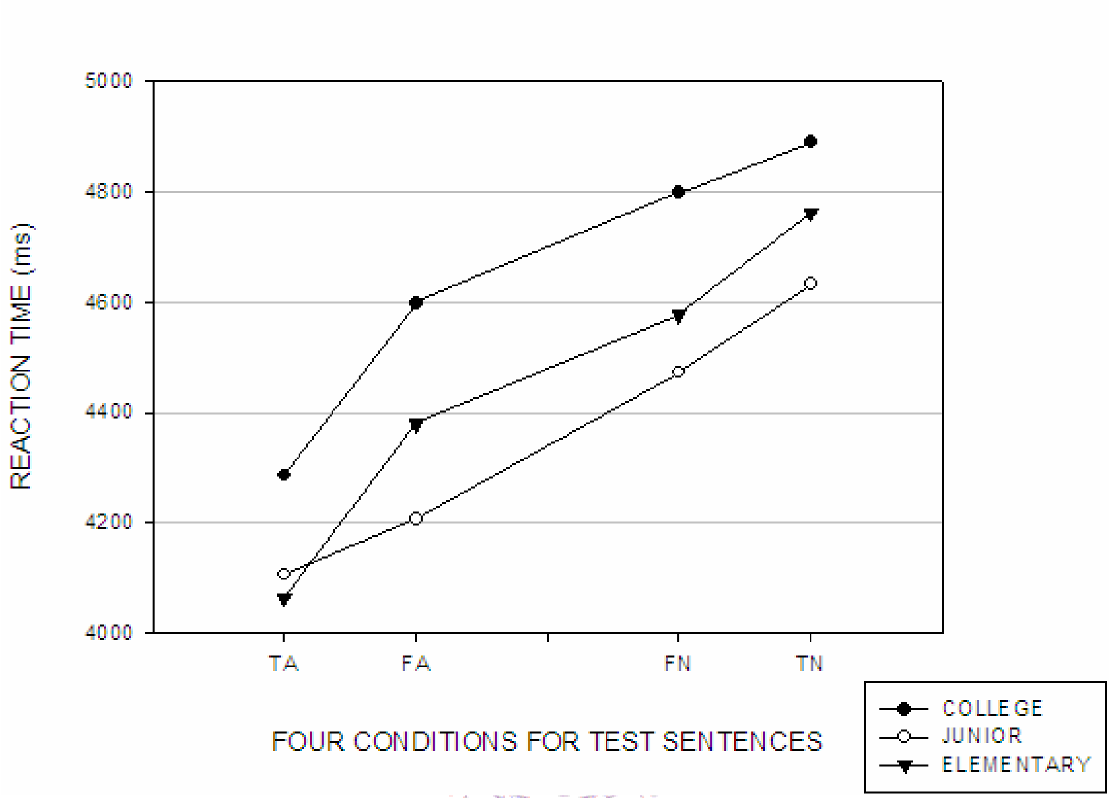


Fig. 9. Comparisons of response latencies of counterfactual target clauses for three groups of participants in simultaneous task with *yaobushi*.

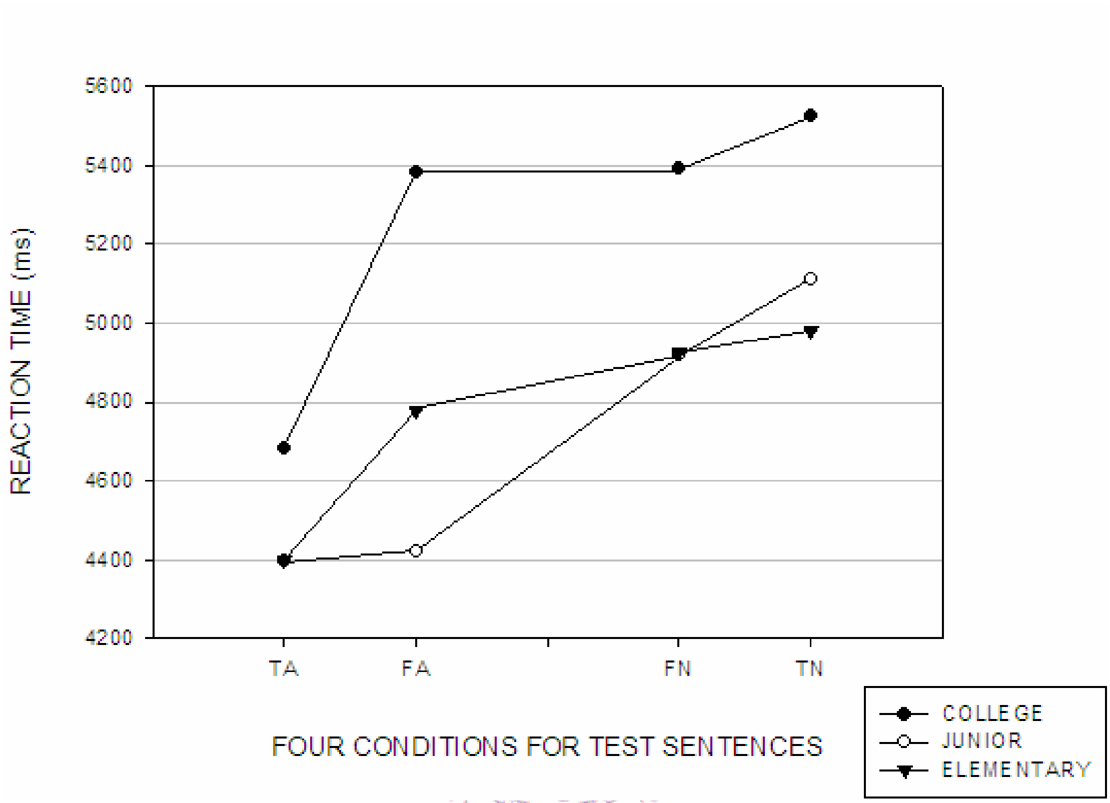


Fig. 10. Comparisons of response latencies of factual and counterfactual target clauses for three groups of participants in simultaneous task with *yaobushi*.

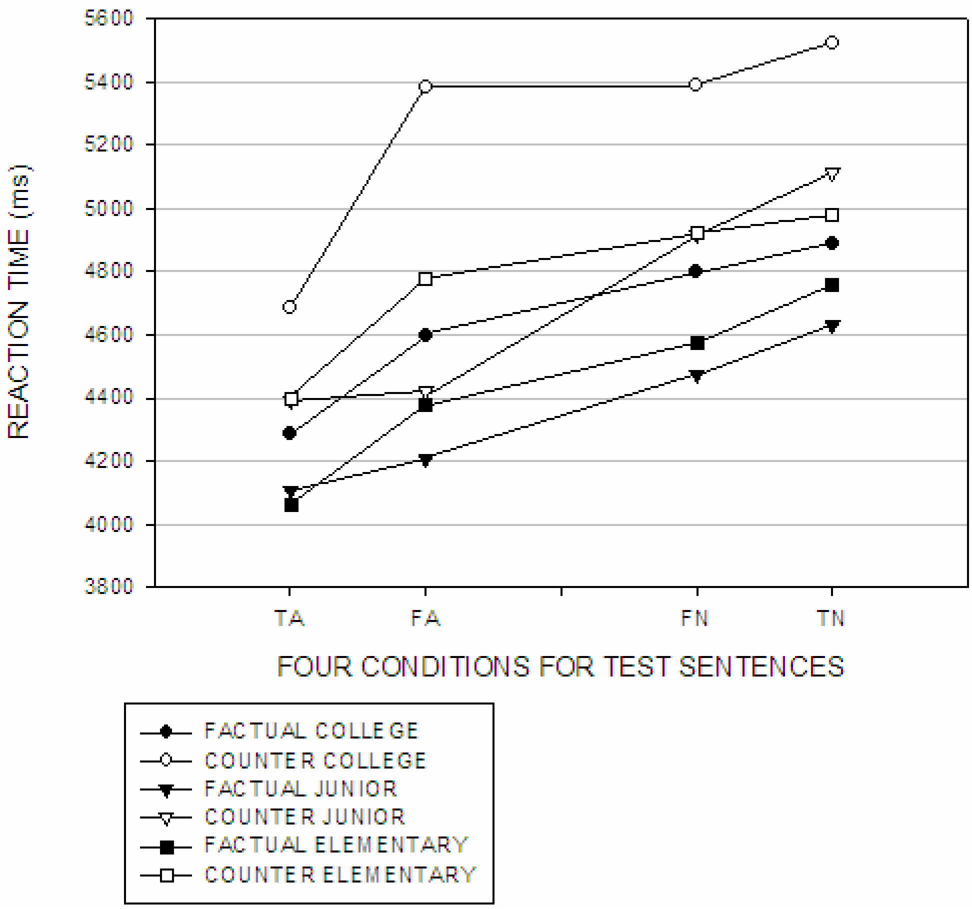


Fig. 11. Comparisons of response latencies of factual target clauses for three groups of participants in delayed task with *yaobushi.*

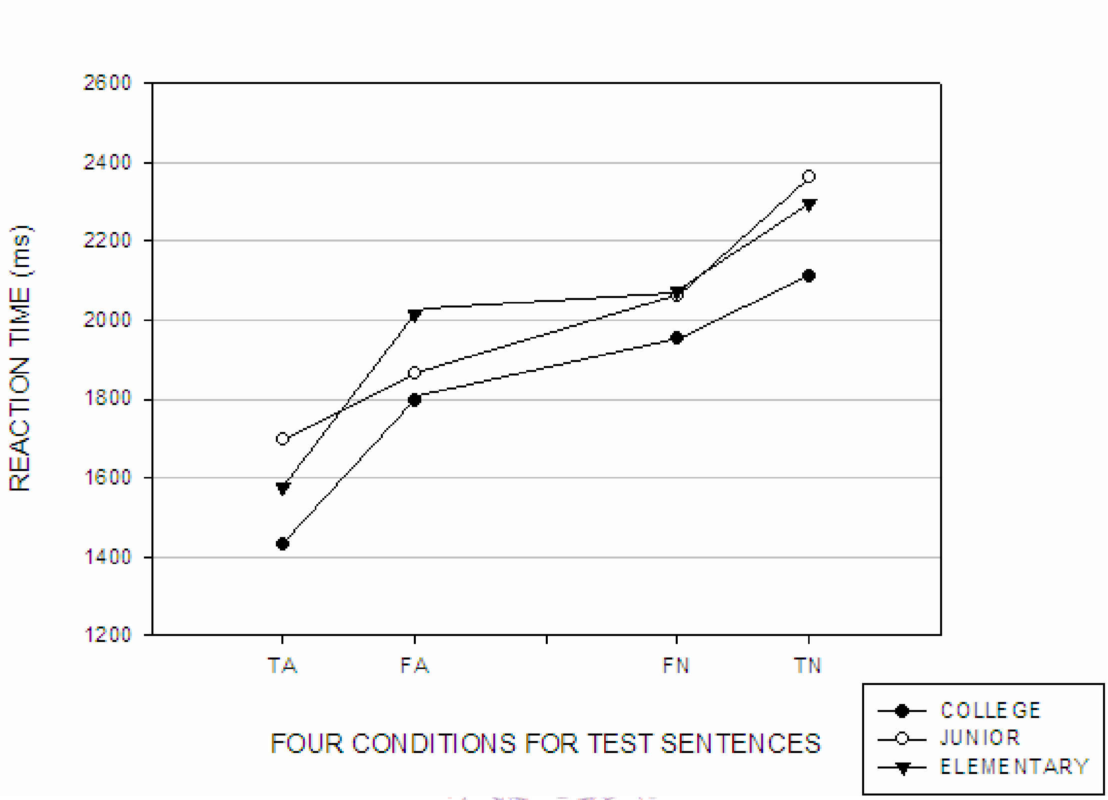


Fig. 12. Comparisons of response latencies of counterfactual target clauses for three groups of participants in delayed task with *yaobushi*.

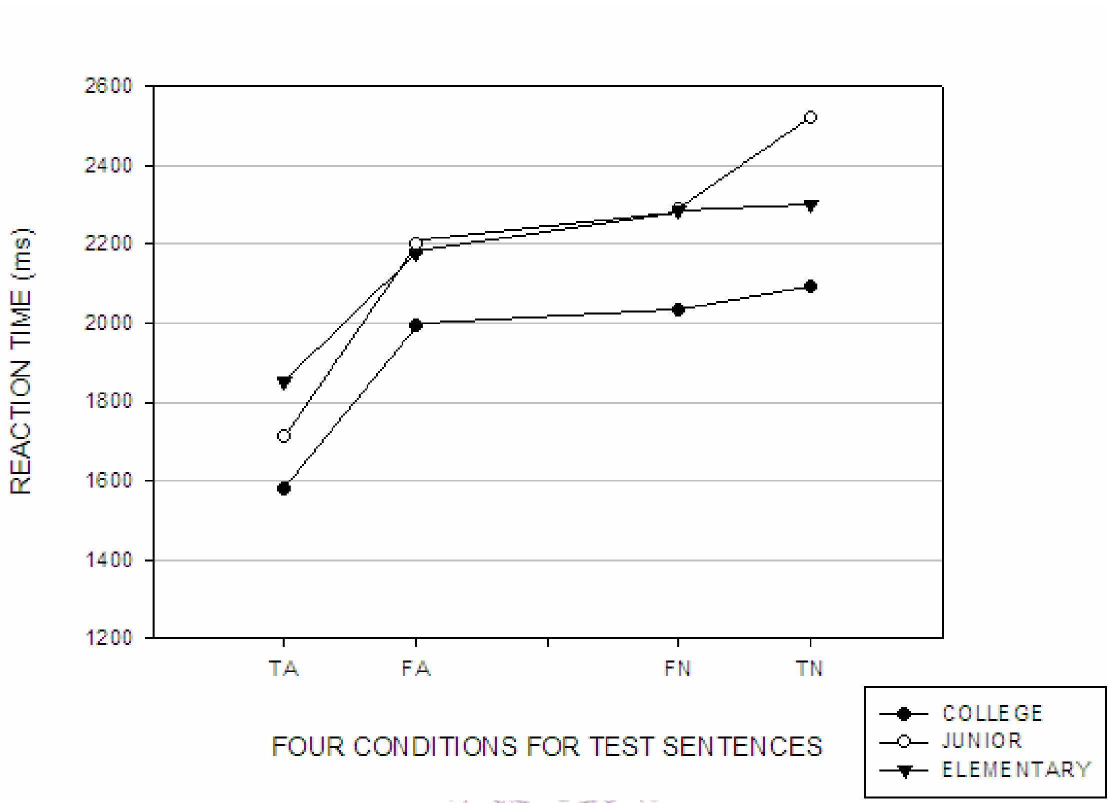


Fig. 13. Comparisons of response latencies of factual and counterfactual target clauses for three groups of participants in delayed task with *yaobushi*.

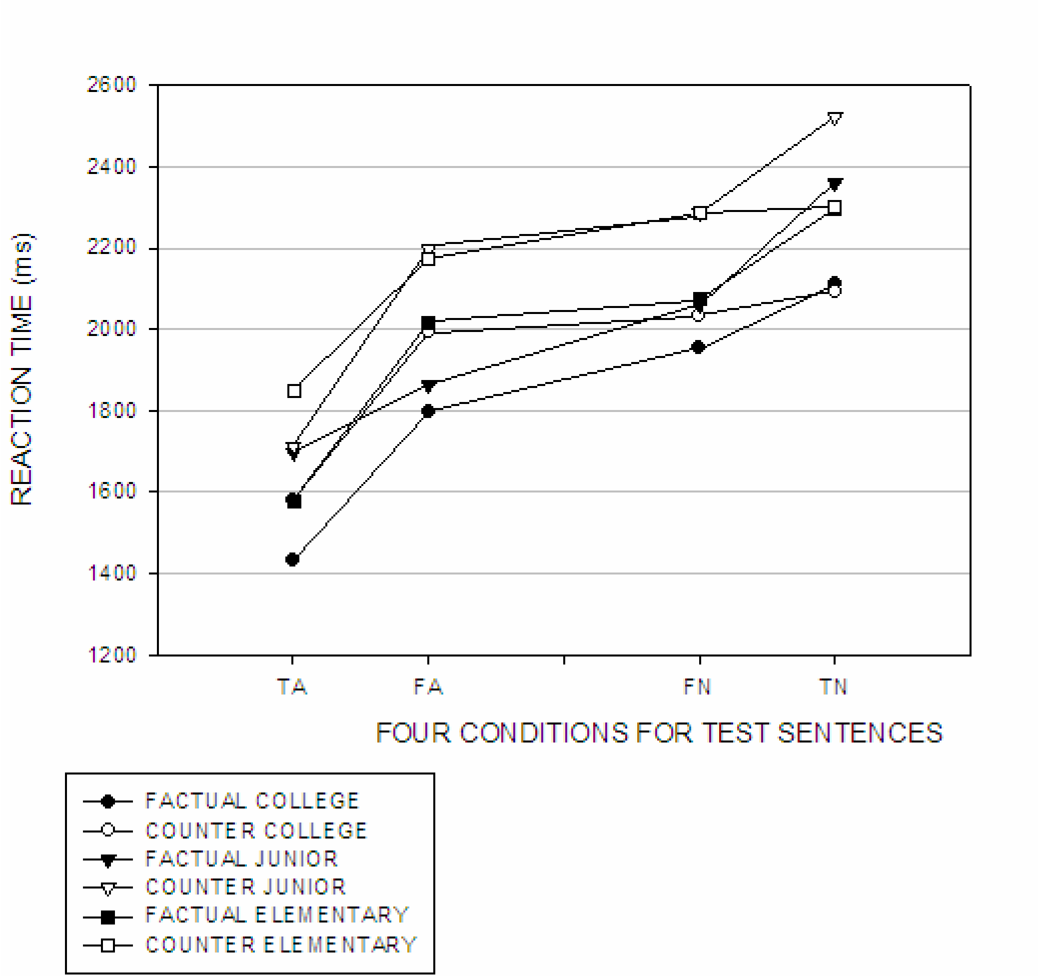


Fig. 14. Comparisons of response latencies of factual target clauses for individuals with Williams Syndrome in delayed task with *yaobushi*.

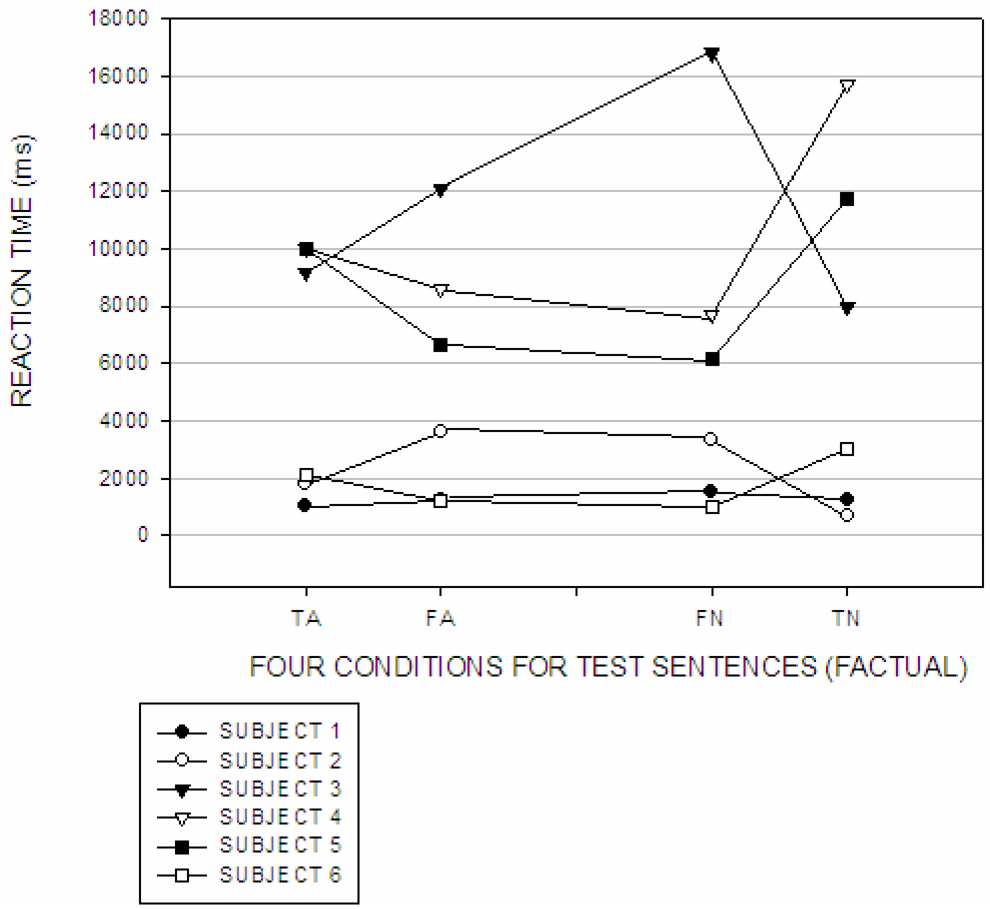
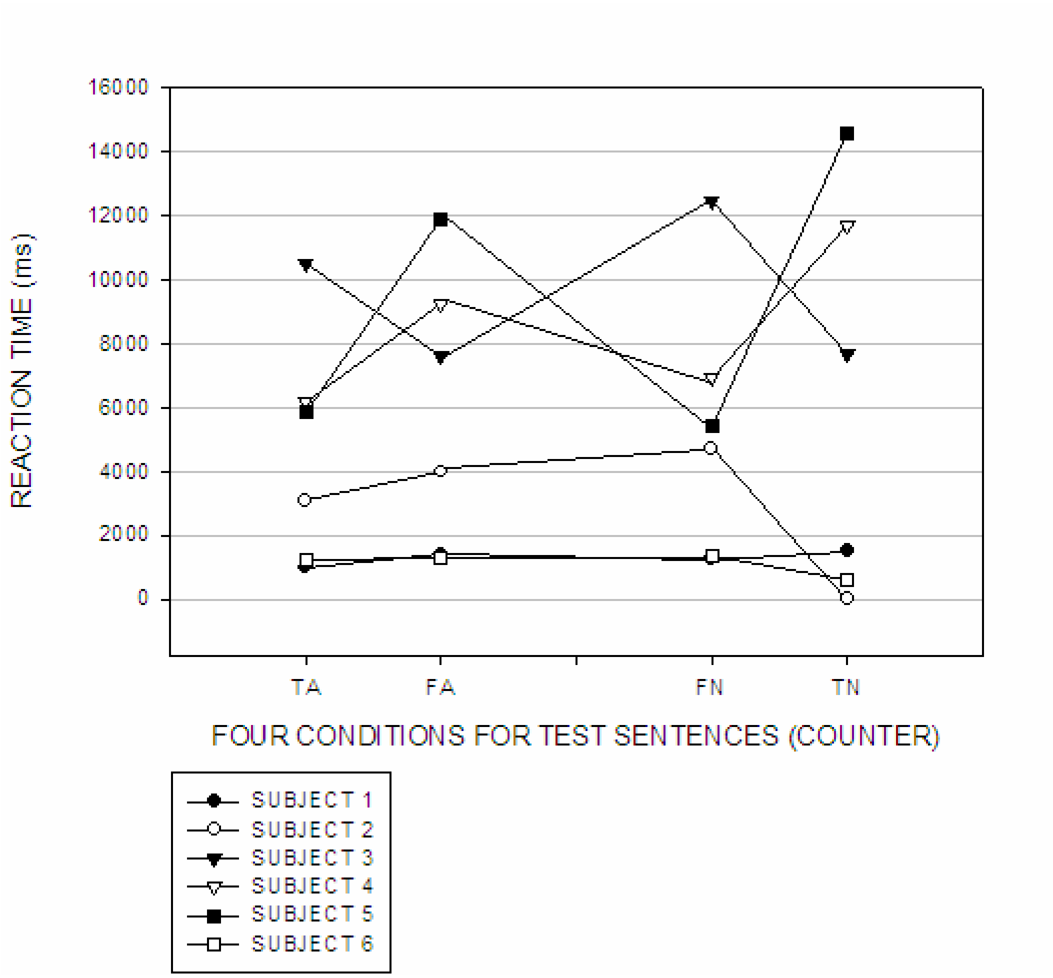


Fig. 15. Comparisons of response latencies of counterfactual target clauses for individuals with Williams Syndrome in delayed task with *yaobushi*.



CHAPTER IV

MANUSCRIPT: THE LINGUISTIC ABILITY OF SEMANTIC INTEGRATION: EVIDENCE FROM PROPOSITION ENTAILMENT IN CHINESE CHILDREN WITH WILLIAMS SYNDROME

A Abstract

This study investigated the hypothesis of selective impairment on form and meaning in language processing on individuals with Williams Syndrome. It has been known that individuals with WS have spared grammatical knowledge even with mental retardation (average IQ of 55 or below) and poor cognition. Past research also showed that individuals with WS preserved normal verbal working memory and such intact verbal ability was thought to be responsible for their relatively good language performance (Wang & Bellugi, 1994; Jerrold, Baddely, & Hewes, 1999; Vicari, Brizzolara, Carlesimo, Pezzini, & Volterra, 1996; Vicari, Carlesimo, Brizzolara, and Pezzini, 1996; Karmiloff-Smith, Grant, Berthoud, Davies, Howlin, & Udwin, 1997; Robinson, Mervis, & Robinson, 2003; Laing, E., Grant, J., Thomas, M. S .C. & Karmiloff-Smith, A., in press). With a good verbal memory but deficit cognitive ability, individuals with WS are hypothesized to rely heavily on verbal working memory in learning their language. This may explain the finding that grammatical knowledge of WS individuals is strong while their semantic understanding might be weak (Zukowski, 2001; Grant, Valian, and Karmiloff-Smith, 2002). In other words, these individuals might have dissociation on form and meaning in linguistic ability.

In order to examine this issue, a Bransford & Franks' paradigm (1972) in recognition was employed. Participants were trained to implicitly learn a series of

sentences which were composed of different events from various superset sentences, and later to make judgments about each sentence in recognition as to whether it had been heard before. They were asked to assign a recognition confidence rating value for each judgment. Two experiments were conducted. The first set involved experiments in which recognition score and confidence ratings of new and old sentences were compared across the unimpaired and individuals with WS (Experiments I & II). According to form and meaning dissociation hypothesis, it was expected that individuals with WS would perform a high rejection rate to all new sentences and high hit rate to all old sentences. On the other hand, the unimpaired would show a high false positive rate to all new sentences and high hit rate to all old sentences. The results showed that individuals with WS performed similarly to the unimpaired, i.e. their chronological age-matched normal controls. In the second set of experiments, new and scrambled sentences were compared (Experiments III & IV). It was expected that individuals with WS would perform a high rejection rate to all new sentences and scrambled sentences, whereas the unimpaired would show a high false positive rate to all new sentences and high rejection rate to scrambled sentences. Again, the results showed that individuals with WS performed similarly to the unimpaired. In conclusion, individuals with WS showed spontaneous semantic integration like the unimpaired.

B The Paradox of Form and Meaning on Children with Williams Syndrome

The hypothesis of the dissociation between form and meaning on sentential comprehension on children with WS also comes from several observations of relative clause studies (Zukowski, 2001; Grant, Valian, & Karmiloff-Smith, 2002). Though knowledge of grammatical structures was relatively spared, children with WS were found to have difficulty in understanding relative clauses. In Zukowski's study, there were two types of relative clause gaps: participant gap (SG) and object gap (OG). For example, a participant gap relative clause is like 'the woman who drove the red car walked into a supermarket'; an object gap relative clause is like 'the woman who the caterpillar fell on was eating a hamburger'. The experimenter read a scenario to each child and asked a question to elicit responses of relative clauses. These two types of gap relative clauses were embedded in two different kinds of sentences: full sentence and noun phrase. For example, a sentence with object gap relative clause was asked to children like "The cow that the girl is pointing to__" which was expected to elicit a response with a noun phrase like "Max is looking at the cow that the girl is pointing to". The results showed that children with WS had high accuracy (77%) in participant gap relative clauses similar to normal children (82%). However, children with WS showed a difficulty in producing object gap relative clauses (11%) compared to normal children (51%). Though accuracy was low for children with WS, at least 9 out of 10 children with WS produced one object gap sentence. From these results, we concluded that children with WS have near normal ability in producing relative clauses.

Further analysis in the production of relative clauses showed that children with WS had difficulty in understanding sentences because of mapping errors. Mapping errors, which mistook the participant of a relative clause as the participant of a matrix clause in a full sentence elicitation or a head noun of a noun phrase, were observed very frequently. And the same error patterns were also observed on normal children. For example, when a question "***which truck*** turned red?" was asked to children, many of them replied that "***the girl*** that is jumping over the truck turned red" instead of replying to the target "***the truck*** that the girl is jumping over turned red". Another example like "***which car*** is Max (a bird's name) looking at?" was asked, they replied that "***the pigeon*** that is flying over the car" instead of replying the target "***the car*** that the pigeon is flying over". Though both groups of participants were observed to have these error patterns, children with WS erred more than normal children. Thus, we concluded that children with WS seem to have normal linguistic knowledge for producing grammatical structures, but are selectively impaired in understanding sentential meanings.

Grant, Valian, & Karmiloff-Smith (2002) conducted another study in relative clauses on English-speaking children with WS in Britain. In their study, four groups of participants were recruited: 5-year-old, 6-year-old, and 7-year-old normal children and children with WS (mean chorological age was 17 years old; mean mental age was 8 years old). Four types of relative clauses were designed: (1) 'the boy chasing the horse is fat', which was identified as participant-participant sentence stimuli (SS); (2) 'the cat the cow chases is black', which was participant-object sentence stimuli (SO); (3) 'the dog chases the horse that is brown', which was object-participant sentence stimuli (OS); (4) 'the dog is chasing the cow the boy sees', which was object-object sentence stimuli (OO). Children were asked to repeat each sentence after the experimenter spoke to them. According to the degree of structure complexity, we predicted that SO stimuli was the hardest condition because two noun phrases were in order and were not easy to process in comprehension, which in turn was easier than OS stimuli, which in turn was easier than OO stimuli, which in turn was easier than SS stimuli. The results matched with the predictions. All participants, including children with WS, showed the same pattern. However, children with WS, given their mental age was over 8-year-old, did not show better performance than 6-year-old and 7-year-old normal children. They showed the same level of performance as 5-year-old normal children. In other words, children with WS showed a delayed development on grammatical structures, or say, impaired ability in understanding sentential meanings like relative clauses. This dissociation between form and meaning in memory will be further investigated in the present study. However, before presenting our investigation of these questions on children with WS, we will briefly discuss experiments on normal adults examining the role of syntactic form in memory, which will suggest a new method of investigation of language in WS.

C The Role of a Syntactic Form in Memory

A series of experiments by Bransford and colleagues in the early 1970s investigated how people integrated information from various sentences, which expressed partial meanings in communication (Bransford and Franks, 1971, 1972; Bransford, Barclay, and Franks, 1972; Franks and Bransford, 1972, 1974; Singer and Rosenberg, 1973; Franks and Bransford, 1974). Sentences contain what are known as propositions, parts of the meaning of the utterance in which it is expressed (Lyons, 1995:118). Everyday conversation is made possible only through the ability to quickly extract these propositions from the speaker's sentences. Some form of integration process allows people to form a holistic representation from these pieces of information. The main question Bransford and colleagues were interested in is: what is the unit in conversation, or say, in memory? What is the unit of integration? Is it the exact wordings expressed on sentences or the linguistic ideas embedded in structures remembered in memory? What is the role of a sentence? Is it a unit in memory or a unit in communication carrying information?

In order to answer these questions, Bransford and Franks (1971) conducted a series of comprehension experiments. Participants were presented with sentences containing different number of propositions. The sets were formed by beginning with each sentence containing four propositions, and then breaking them down into different declarative statements based on the free combination of different number of propositions. For example, a declarative sentence could be a sentence with four propositions (FOURS): “*The ants in the kitchen ate the sweet jelly which was on the table*”; three propositions (THREES): “*The ants ate the sweet jelly which was on the table*”, “*The ants in the kitchen ate the jelly which was on the table*” and “*The ants in the kitchen ate the sweet jelly*”; two propositions (TWOS): ”*The ants in the kitchen ate the jelly*”, “*The ants ate the sweet jelly*”, “*The sweet jelly was on the table*” and “*The ants ate the jelly which was on the table*”; or only one proposition (ONES): “*The ants were in the kitchen*”, “*The jelly was on the table*”, “*The jelly was sweet*” and “*The ants ate the jelly*”. Therefore, a four-proposition sentence could be decomposed to a number of sentences including one FOURS, three THREES, four TWOS, and four ONES. All these declarative sentences formed a complex idea set, in some sense equal to the sentence with four propositions.

In Bransford and Franks’s studies, there four complex idea sets were included and distributed to two phases: learning and recognition. Six sentences (two ONES, two TWOS, two THREES) were selected from each idea sets as learning stimuli first. In the learning phase, participants were required to listen to these sentences auditorily, and later responded to an elliptical question. Another twenty-four sentences, which were selected from four complex idea sets (two ONES, two TWOS, one THREES, one FOURS), were presented as recognition stimuli, along with six sentences which were actually presented in learning section. In recognition, participants were asked to make judgment whether the particular sentence had been presented in learning section before. After the judgment, participants were asked to assign recognition confidence ratings in a 5-point scale. The experimenter coded ‘yes’ responses a positive value, suggesting that participants felt that they had heard the sentences before, and ‘no’ responses were assigned a negative value.

The hypothesis behind the experiment was that people might maintain sentence meaning without memorizing syntactic structures. That is, people might form a holistic semantic representation rather than a particular sentence representation in memory. If this hypothesis was correct, we predicted that participants might think that they recognized the sentences which were not presented in learning section because of spontaneous semantic integration processing. Furthermore, we predicted that recognition confidence ratings would be a function of sentence complexity, which was defined based on different number of propositions embedded in a sentence. The more propositions a sentence contained, the easier participants would misrecognize the sentences as heard before. The results confirmed the predictions. Recognition confidence ratings followed this pattern: FOURS > THREES > TWOS > ONES. Bransford and colleagues concluded that participants integrated linguistic information from successive and nonconsecutive sentences spontaneously, and that the more propositions the sentences contained, the higher the recognition ratings would be assigned. This finding was replicated in using different sentence types.

The same procedure was applied with new idea sets. The design was the same, twenty four sentences were designed as stimuli in learning section and other twenty four sentences were included as recognition stimuli. However, instead of mixing old sentences, in this study six scrambling sentences which were combined from selecting propositions of different idea sets were included in recognition. That is, these scrambling sentences were really new to the participants because the events reproduced were completely unfamiliar to them. There were two types of scrambling sentences in this study: free combination and grammatical relation violation. For example, a free combination scrambling sentence like "*The old man who was smoking his pipe climbed the steep hill*" may come from two complex linguistic ideas: "*The old car pulling the trailer climbed the steep hill*" and "*The tall tree in the front yard shaded the man who was smoking his pipe*"; a grammatical relation violation scrambling sentence like "*The scared cat ran from the barking dog which jumped on* *the table*" (i.e. it is the dog which jumped on the table, not the cat) was created from sentence with exact wordings: "*The scared cat running from the barking dog jumped on the table*" (i.e. it is the cat which jumped on the table, not the dog). The results demonstrated that participants performed decreasing recognition confidence ratings from FOURS to ONES in new sentences and participants correctly rejected scrambling sentences as never heard stimuli. Thus these results confirmed the finding that recognition confidence rating was a function of sentence complexity, or say, number of propositions.

Based on the results of these studies, Bransford and colleagues claimed that a holistic semantic idea was learned rather than particular sentences. Participants did not learn the particular sentence structures, but rather the integrated semantic information expressed in sentences. Due to this spontaneous integration, participants almost always recognized the sentences as presented before. Moreover, participants learned the precise meaning of propositions and grammatical relations between them in sentences, but through integrating semantic information which was derivable from presented sentences instead of focusing on the exact wordings. Bransford and Franks claimed, then, that the sentence is not a unit in memory, but a unit in communicating linguistic ideas. In other words, a syntactic form is not represented in memory, but the meaning will be retained. Further, extracting semantic information by integrating propositions conveyed in sentences is a spontaneous process in language.

The experimental hypothesis of the present study is that form and meaning is dissociated on sentential level to a certain degree in children with WS. If so, by using Bransford and Frank's recognition paradigm, it is predicted that for normal people the confidence ratings or false positives will be very high for all new sentences no matter which types of sentences are lumped together (i.e. old or scrambled). However, we predict that children with WS, will show low confidence ratings and low false positives for all new sentences due to the possible discrepancy between form memorization and meaning understanding. As for old sentences, we predict that both normal people and children with WS will give high ratings. Finally, due to the manipulation of number of propositions, an ordering of recognition confidence ratings is expected to be like the following: ONES < TWOS < THREES < FOURS, suggesting normal ability in semantic integration. That is, the more propositions a sentence contains, the higher confidence ratings in recognition would be assigned. It is hypothesized that this effect of proposition integration will be observed in normal people, but not in children with WS.

D Experiment I: Comparison of New and Old Sentences

Participants

Thirty four participants were tested in this study. Twenty three were undergraduates from National Tsing Hua University participating for course credit of Introduction to Linguistics; eleven were graduates from University of Maryland at College Park participating for reimbursement (mean age=21.6, range from 18 to 32, 21 females and 13 males). All participants were right-handed and none of them were reported as having medical problems. All were native speakers of Mandarin Chinese.

Design and Materials

Three sentences were used for the basic supersets of ideas. Each superset sentence contained four propositions and was broken down into twelve simple declaratives. These propositions had semantic relations with each other. These twelve simple declaratives differed in combination of propositions: (1) Four-proposition sentences (FOURS), which exhaustively listed all the propositions in a sentence. For example, 森林裡的大野狼抓到了正在草叢裡吃紅蘿蔔的小白兔 (A wild wolf in the forest caught a rabbit which was eating carrots in brushwood); (2) Three-proposition sentences (THREES), which combined any three propositions. For example, 森林裡的大野狼抓到了草叢裡的小白兔 (A wild wolf in the forest caught a rabbit which was in brushwood); (3) Two-proposition sentences (TWOS), which combined any two propositions. For example, 小白兔正在草叢裡吃紅蘿蔔 (A rabbit was eating carrots in brushwood); (4) One-proposition sentences (ONES), which only contained one proposition. For example, 小白兔在草叢裡 (A rabbit was in brushwood). Basically, propositions were defined as locations (e.g. mice were in the kitchen; kids were in the classroom), properties (e.g. cakes were made of strawberry; kids were cute), and events (e.g. mice were eating cakes; the wolf caught a rabbit). The more propositions contained in a sentence, the more complex a sentence structure was.

The study was composed of two sections, training and recognition. Six sentences from each superset sentence were selected as stimuli in the training section (two ONES, two TWOS, two THREES) and another six sentences from each superset were left as recognition stimuli (two ONES, two TWOS, one THREES, one FOURS). Those sentences were actually new sentences to all participants, as they were never presented in the training section. The sentence stimuli of each superset sentence were listed in Table 50 as Superset A to C. The average length for Superset A of Chinese stimuli is 13.67 (cf. English stimuli is 9.33 if translated); The average length for Superset B of Chinese stimuli is 11 (cf. English stimuli is 7.42 if translated); The average length for Superset C of Chinese stimuli is 11.5 (cf. English stimuli is 9.75 if translated). Another nine sentences containing different propositions from another three new superset sentences were displayed as practice stimuli, which were given as Superset Idea D to F in Table 51. Four sentences excerpted in the training section were mixed in recognition as Old sentence stimuli (two ONES, one TWOS, one

THREES), which are marked with an asterisk in Table 1. Thus, twenty two test trials were included as stimuli in recognition. All the superset ideas used in this study were vivid events, including concrete objects (e.g. cakes, carrots), familiar cartoon characters (e.g. Mickey Mouse, Snoopy, rabbits), imaginable activities (e.g. playing games, eating), and highly frequent settings for children (e.g. kindergarten, aquarium). All sentence stimuli were recorded as mono sound waves in 44100 frequency by a female voice and presented using Praat software. All the sentences in the training section were presented randomly and no sentences selected from the same superset idea were presented consecutively. Four random lists were compiled for distribution across participants.

After listening to each sentence in the training section, participants were required to name colors one at a time displayed on the computer screen. This color naming was designed to interrupt the phonological loop in working memory so that participants could not use subvocal rehearsal to memorize the sentence just heard. After color naming, in order to make sure that participants did understand the sentences and implicitly learned the presented sentences, a comprehension question was presented to each sentence. These comprehension questions were also recorded as mono sound waves. For example, after presentation of a training sentence like “Koalas were on the trees“, a comprehension question like “Where were the koalas?” was asked immediately to participants. Participants had to answer the comprehension question to complete a trial. Once these procedures were fulfilled, the training section was completed. The comprehension questions paired with presented sentences are listed in Appendix 1, 2 and 3 for test sentences from superset idea A, B, and C, respectively. In addition, Appendix 4, 5, and 6 are given for practice superset idea D, E, and F, accordingly.

Table 50 Experimental Stimuli as Supersets A – C

Number of propositions	Sentences
Superset A---森林裡的大野狼抓到了正在草叢裡吃紅蘿蔔的小白兔 (23) A wild wolf in the forest caught a rabbit which was eating carrots in brushwood. (15)	
FOURS	森林裡的大野狼抓到了正在草叢裡吃紅蘿蔔的小白兔 (23) A wild wolf in the forest caught a rabbit which was eating carrots in brushwood. (15)
THREES	大野狼抓到了正在草叢裡吃紅蘿蔔的小白兔 (19) A wild wolf caught a rabbit which was eating carrots in brushwood. (12)
	★森林裡的大野狼抓到了草叢裡的小白兔 (17) A wild wolf in the forest caught a rabbit which was in brushwood. (13)
	森林裡的大野狼抓到了正在吃紅蘿蔔的小白兔 (20) A wild wolf in the forest caught a rabbit which was eating carrots. (13)
TWOS	森林裡的大野狼抓到了小白兔 (13) A wild wolf in the forest caught a rabbit. (9)
	大野狼抓到了正在吃紅蘿蔔的小白兔 (16) A wild wolf caught a rabbit which was eating carrots. (10)
	小白兔正在草叢裡吃紅蘿蔔 (12) A rabbit was eating carrots in brushwood. (7)

	大野狼抓到了在草叢裡的小白兔 (14) A wild wolf caught a rabbit which was in brushwood. (10)
ONES	★大野狼在森林裡 (7) A wild wolf was in the forest. (7)
	小白兔在草叢裡 (7) A rabbit was in brushwood. (5)
	大野狼抓到小白兔 (8) A wild wolf caught a rabbit. (6)
	小白兔在吃紅蘿蔔 (8) A rabbit was eating carrots. (5)
Superset B---廚房裡的老鼠正在偷吃桌子上的草莓蛋糕 (18)	The mice in the kitchen were eating strawberry cakes on the table. (12)
FOURS	廚房裡的老鼠正在偷吃桌子上的草莓蛋糕 (18) The mice in the kitchen were eating strawberry cakes on the table. (12)
THREES	老鼠正在偷吃桌子上的草莓蛋糕 (14) The mice were eating strawberry cakes on the table. (9)
	廚房裡的老鼠正在偷吃桌子上的蛋糕 (16) The mice in the kitchen were eating cakes on the table. (11)
	廚房裡的老鼠正在偷吃草莓蛋糕 (14) The mice in the kitchen were eating strawberry cakes. (9)
TWOS	廚房裡的老鼠正在偷吃蛋糕 (12) The mice in the kitchen were eating cakes. (8)
	老鼠正在偷吃草莓蛋糕 (10) The mice were eating strawberry cakes. (6)

	草莓蛋糕在桌子上 (8) Strawberry cakes were on the table. (6)
	老鼠正在偷吃桌子上的蛋糕 (12) The mice were eating cakes on the table. (8)
ONES	老鼠在廚房裡 (6) The mice were in the kitchen. (6)
	★蛋糕在桌子上 (6) Cakes were on the table. (5)
	老鼠正在偷吃蛋糕 (8) The mice were eating cakes. (5)
	蛋糕是草莓口味的 (8) Those were strawberry cakes. (4)
Superset C---幼稚園裡可愛的小朋友正在教室裡玩遊戲 (18) Cute kindergarten kids were playing games in the classroom. (9)	
FOURS	幼稚園裡可愛的小朋友正在教室裡玩遊戲 (18) Cute kindergarten kids were playing games in the classroom. (9)
THREES	可愛的小朋友正在教室裡玩遊戲 (14) Cute kids were playing games in the classroom. (8)
	幼稚園裡的小朋友正在教室裡玩遊戲 (16) Kindergarten kids were playing games in the classroom. (8)
	幼稚園裡可愛的小朋友正在玩遊戲 (15) Cute kindergarten kids were playing games. (6)
TWOS	幼稚園裡的小朋友正在玩遊戲 (13) Kindergarten kids were playing games. (5)

	★可愛的小朋友正在玩遊戲 (11) Cute kids were playing games. (5)
	小朋友正在教室裡玩遊戲 (11) Kids were playing games in the classroom. (7)
	可愛的小朋友正在教室裡 (11) Cute kids were in the classroom. (6)
ONES	小朋友在幼稚園裡 (8) Kids were in the kindergarten. (5)
	小朋友很可愛 (6) Kids were very cute. (4)
	小朋友正在玩遊戲 (8) Kids were playing games. (4)
	小朋友在教室裡 (7) Kids were in the classroom. (5)

Table 51 Practice Stimuli as Idea set D-F

Superset D---動物園裡的無尾熊正在高高的樹上吃油加利葉 Koalas in the zoo were eating leaves on tall trees.	
TWOS	動物園裡的無尾熊正在樹上 Koalas in the zoo were on tall trees.
ONES	無尾熊正在樹上 Koalas were on the trees.
	無尾熊在吃油加利葉 Koalas were eating leaves.
Superset E---米老鼠和史努比正在公園裡玩蹺蹺板	

Mickey Mouse and Snoopy were playing seesaw in the park.	
THREES	米老鼠正在公園裡玩蹺蹺板 Mickey Mouse was playing seesaw in the park.
ONES	史努比在公園裡 Snoopy was in the park.
	米老鼠在玩蹺蹺板 Mickey Mouse was playing seesaw.
Superset F---水族箱裡的魚和螃蟹正在吃飼料 Fish and crabs were eating feeding stuffs in the aquarium.	
FOURS	水族箱裡的魚和螃蟹正在吃飼料 Fish and crabs were eating feeding stuffs in the aquarium.
TWOS	魚和螃蟹正在吃飼料 Fish and crabs were eating feeding stuffs.
	水族箱裡有魚和螃蟹 Fish and crabs were in the aquarium.

Procedure

The experimental task consisted of two sections: training and recognition. All participants had to complete both sections. They were not told prior to the training session that a recognition section would follow, nor that they might be integrating information from related sentences. All sentence stimuli were presented auditorily. During training, a fixation point displayed on the computer screen 500ms alerted participants to the beginning of each trial. A test sentence followed the fixation point. After presentation of the test sentence, color naming was required. Four colors were presented one at a time: yellow, blue, red, and green. Participants were asked to name the colors accordingly during two-second exposure. The same color could appear twice and all colors were displayed randomly. After color naming, participants would hear a comprehension question, which was related to the content of the sentence just heard. Then participants were required to answer the comprehension question. After the training section, there was a break for 3 to 4 minutes. Now participants were told that there was another task waiting for them: a recognition test.

In the recognition section, participants were presented with the other half of sentences from each superset idea which were not previously presented (i.e. New sentences) and another four sentences which were actually presented in training section (i.e. Old sentences). In this section, participants were required to indicate whether the particular sentence was presented in the training section before. Meanwhile, participants were instructed that the sentences which would be presented in this section might be all new or all old to them, or any distribution in between. If a sentence was recognized as heard before, participants responded by clicking the left button of a mouse; if a sentence was recognized as never heard before, they responded by clicking the right button of the mouse. After this yes/no judgment, participants were required to make a recognition confidence rating using the keyboard for each response they just made to indicate how confident they felt about their decision in 5-point scale from the most confident scaling 5 down to the least confident scaling 1. Once all these requirements were fulfilled, the next trial would begin. Nine practice trials were given before the actual experiment. All the participants were tested in a quiet room in National Tsing Hua University or in University of Maryland at College Park.

Prediction

Normal adults were hypothesized to demonstrate a similar pattern to the results reported in Bransford and Franks (1971, 1972, 1973, and 1974). In other words, they should show high recognition confidence rating and also high false positive recognition for the new sentences with shared propositions, although those sentences were never presented before. Participants should show that the recognition confidence rating is a function of proposition complexity, suggesting a spontaneous proposition integration taking place in sentence processing. That is, participants may think that they had heard these sentences before and show high recognition confidence ratings. Therefore, the more propositions the sentence contained, the higher recognition confidence ratings should be. In this scenario, participants were inferred to build up mental models according to the entailment relation of propositions presented. For the old sentences, which actually were displayed in the training section, participants should also show a high hit rate which would be reflected with high recognition confidence ratings as Bransford and Franks.

TSING HUA UNIVERSITY

Results and Discussion

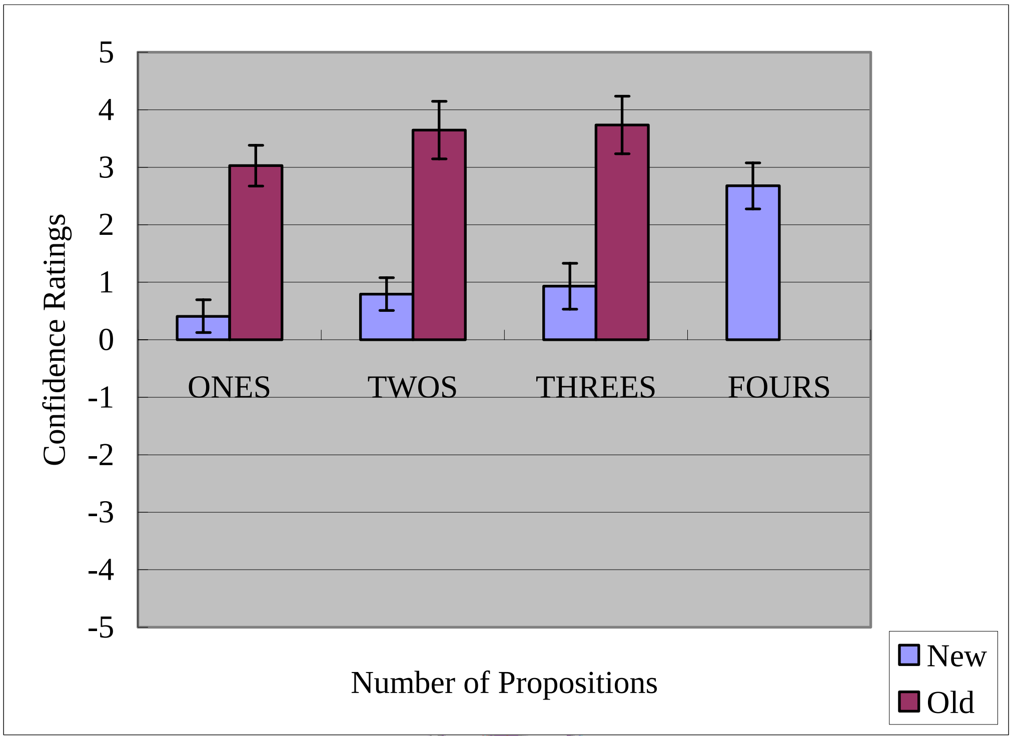
Recognition ratings were computed for each type of sentence: ONES, TWOS, THREES, FOURS, and Old. Participants' ratings were converted into numerical values. A "yes" response received a plus while a "no" response received a minus. A very high confidence rating received a 5, a high confidence rating received a 4, a middle confidence rating received a 3, a low confidence rating received a 2, and a no confidence rating received a 1. In this way, a 10-point rating scale emerged, ranging from plus 5 to minus 5. Zero was excluded.

Due to an uneven number of trials across four experimental conditions, a proc mixed model with a post hoc test of least significance means (LSMEANS) using the Tukey method was employed. A clear ordering was apparent according to the number of propositions: the mean ratings for New sentences were 0.408 (ONES), 0.794

(TWOS), 0.931 (THREES), and 2.676 (FOURS). Normal people in general showed high recognition confidence ratings to all sentence types, suggesting that they integrated semantic related sentences to a certain degree and built up a mental model based on sentential propositions. A one-way ANOVA showed that the proposition complexity effect was significant, $F(3, 574) = 7.63$, $p < .0001$, suggesting that participants performed high recognition confidence ratings to the more complex superset sentences and lower recognition confidence ratings to those sentences with fewer propositions. The major difference was seen in the comparison of FOURS to other conditions (FOURS vs. ONES, $p < .0001$; FOURS vs. TWOS, $p = .0007$; FOURS vs. THREES, $p = .01$)[7]. For Old sentences, the recognition confidence ratings were very high across the board: the mean ratings for old sentences were ONES (3.02), TWOS (3.64), and THREES (3.73). The difference between Old sentences was not significant, $F(2, 100) = 0.88$, $p = .4$, suggesting that normal adults treated all the Old sentences as highly familiar stimuli and made yes/no judgment based on the built up mental model built during the training section. Moreover, a two-way ANOVA for New and Old sentences comparison showed that The main effect of number of propositions was not significant, $F(2, 607) = 1.14$, $p = .3$. The main effect of new-old sentences was significant, $F(1, 607) = 49.14$, $p < .0001$, for New ONES and Old ONES ($p < .0001$), for New TWOS and Old TWOS ($p < .0001$), and for New THREES and Old THREES ($p < .0003$). However, the interaction between number of propositions and sentence type was not significant, $F(2, 607) = 0.03$, $p = .9$. A plotted graph based on recognition confidence ratings is given below as Figure 16.

[7] A nonparametric statistics with Kruskal-Wallis Test and Mann-Whitney Test for two independent samples were employed at the same time. The results were similar, but the results generated by using proc mixed model with least significance difference were more conservative. Thus, we used the results generated from proc mixed model and made conclusions from these results for all the data sets in this paper.

Fig. 16. Comparison of New and Old Sentences for Normal People.



Meanwhile, the false positive recognition rates were very high, and followed the same pattern as confidence ratings. Percent of false positives for new sentences and percent of hit rates for old sentences were given in Table 52. A one-way ANOVA showed a significant difference between these four sentence types within New sentences, F (3, 562) = 37.89, p < .0001. A proc mixed model with a post hoc test of Tukey method showed that ONES received significantly lowest false positives and FOURS received significantly highest false positives (i.e. ONES vs. TWOS, p =.0048; ONES vs. THREES, p = .0001; ONES vs. FOURS, p < .0001; FOURS vs. TWOS, p < .0001; FOURS vs. THREES, p < .0001). The difference between TWOS and THREES did not reach significance, p = .1. The hit rates were also very high, and showed the same pattern as confidence ratings. A one-way ANOVA showed a significant difference between these three sentence types within Old sentences, F (2, 90) = 18.26, p < .0001. A proc mixed model with a post hoc test of Tukey method showed that ONES hit rates were significantly lower than the other conditions (i.e. ONES vs, TWOS, $p < .0001$; ONES vs. THREES, $p < .0001$). But, the difference between TWOS and THREES did not reach significance, $p = 1.000$. Another proc mixed model with a post hoc test of Tukey method showed The main effect of number of propositions was significant, $F (2, 585) = 21.23$, $p < .0001$ and The main effect of sentence types was also significant, $F (1, 585) = 425.95$, $p < .0001$. Meanwhile, the interaction between these two factors was also significant, $F (2, 585) = 3.06$, $p = .04$, for New ONES and Old ONES ($p < .0001$), for New TWOS and Old TWOS ($p < .0001$), for New THREES and Old THREES ($p < .0001$). A two-tailed t-test was employed for the difference on means between New and Old sentences and the results showed that only New ONES (4.11) and Old ONES (4.51) reached significant difference, $p = .01$. A plotted graph based on false positives and hit rates is given below as Figure 17.

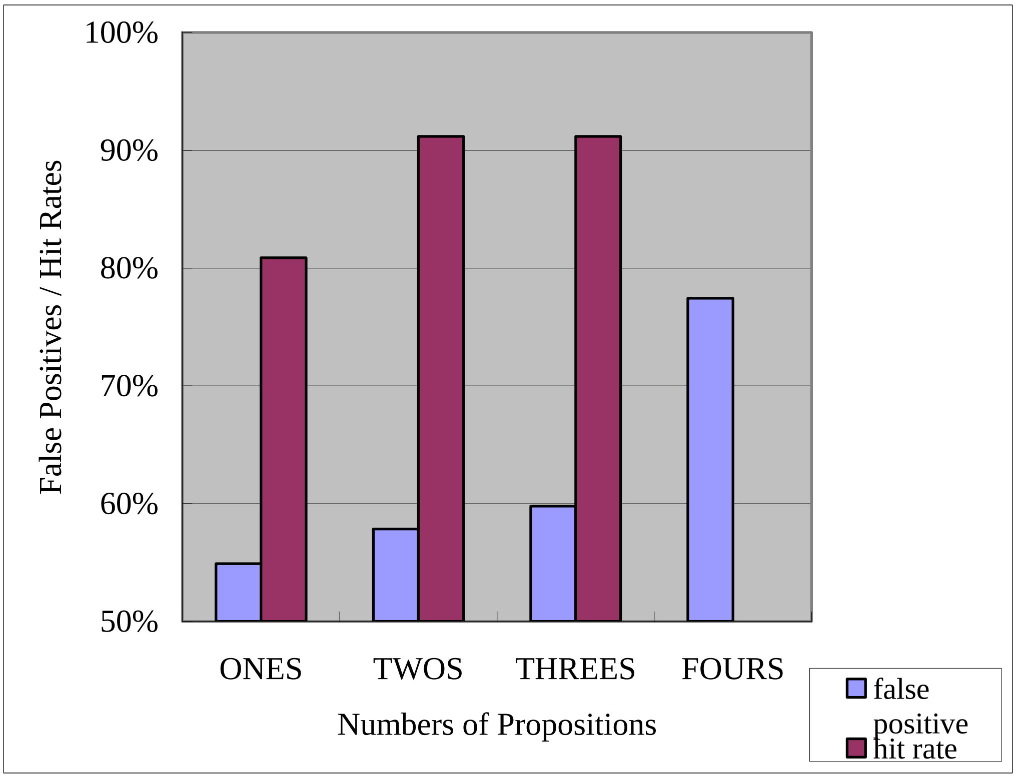
In summary, we found that normal people showed spontaneous integration of propositions which had entailment relations. Having replicated that the basic effects reported by Bransford and Franks (1971), we continued on to the participant of our inquiry, the extent to which verbal working memory in WS patients is different than in normals. Experiment 2 used the same procedure as Experiment 1 but tested WS children.

Table 52 Percent of False Positives and Mean (SD) for New Sentences and Percent of Hit Rates and Mean (SD) for Old Sentences on Normal People

Normal	ONES	TWOS	THREES	FOURS
New	54.90%	57.84%	59.80%	77.45%
	4.11 (0.87)	4.20 (0.81)	4.07 (0.95)	4.53 (0.68)

Old	80.88%	91.18%	91.18%	----
	4.51(0.66)	4.45(0.77)	4.55(0.68)	----

Fig. 17. Comparison of Percent of False Positives and Percent of Hit Rates for Normal People.



E Experiment II: Comparison of New and Old Sentences with Chinese Children with Williams Syndrome

Participants

Five young adults with Williams Syndrome participated in this study (mean age =17;9, range from 12yr and 8m to 21yr and 3m; 4 males and 1 female). Each participant was diagnosed to be one of the members having this syndrome with *Fluorescent in situ hybridization* (FISH) test in hospital or in laboratory.

Design and Materials

Twenty two sentences were included in this study. Six out of twelve from each superset sentence were presented to children with Williams Syndrome in the training section. The other half of the sentences from each superset sentence which were not seen in the training section were mixed with four sentences chosen from training section as recognition stimuli. All the sentence stimuli, including practice and test trials, were the same as Experiment I, which can be referred to in Table 1 and Table 2.

Procedure

The procedure was parallel to Experiment I. A training section was required before recognition. During training, children with Williams Syndrome were presented sentences auditorily and asked to name colors displayed one at a time on the computer screen. After color naming, a comprehension question to each sentence was displayed auditorily to each child. They were required to answer the question right away after the comprehension question was presented and instructed to give the answer based on the sentence they just heard. No verbal cue was given while test trials were presented. Nine practice trials were presented before the experiment.

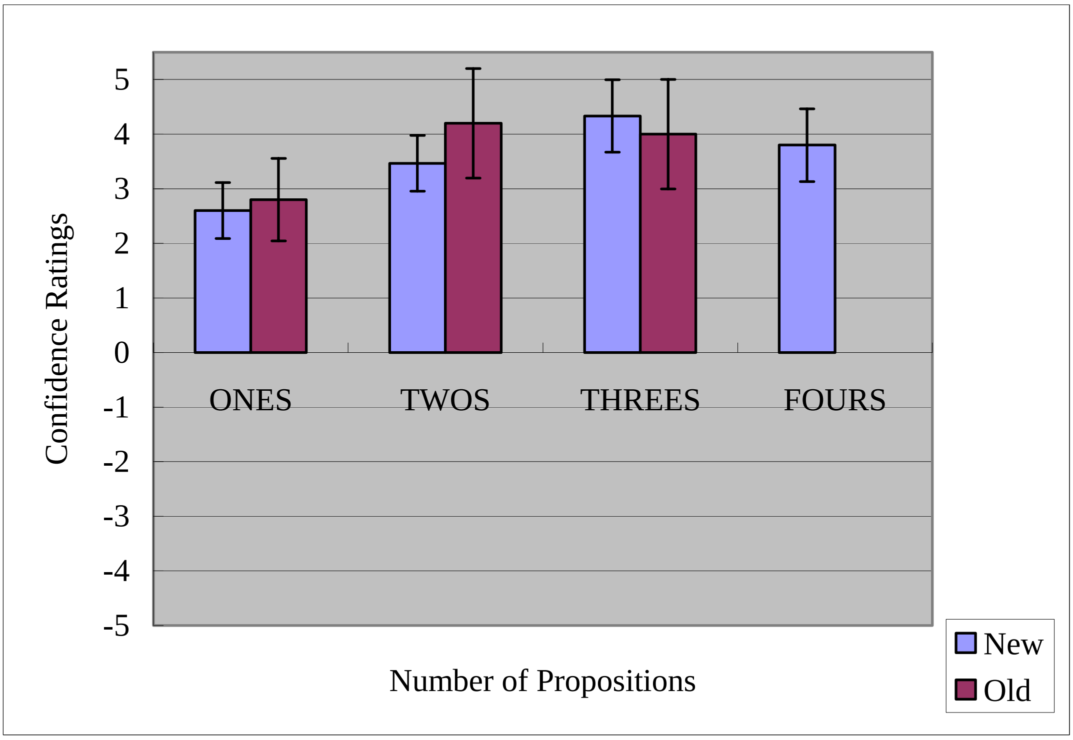
Prediction

The performance of children with Williams Syndrome was predicted to be different from normals. Children with WS were expected to show low false positive recognition to all new sentences no matter how many number of propositions were embedded. That is, children with WS should be able to correctly reject those new sentences if they have superior verbal working memory and spared linguistic knowledge of grammatical structures. In other words, they are not good at building mental models, so they can distinguish the sentences easily. Also, if they have an impaired ability in understanding sentences given the provided events in discourse, they may not be able to use a proposition integration strategy, leading to a lower false positive rate. Thus, recognition should not be a function of the number of propositions in the sentence. If this is true, it can be inferred that children with WS may have difficulty in integrating semantic related propositions and are bad at building mental models from contexts in discourse. Consistent with this prediction, they should show high hit rates to old sentences because they can correctly recognize the particular forms.

Results and Discussion

The same conversion of recognition confidence ratings as in Experiment 1 was calculated and averaged for each condition. The mean ratings for New sentences from ONES to FOURS were 2.6 (ONES), 3.46 (TWOS), 4.33 (THREES), and 3.8 (FOURS), displaying a generally high recognition confidence ratings for all new sentences. A one-way ANOVA showed that the proposition complexity effect was not significant, $F(3, 82) = 2.14$, $p < .1$ although a post-hoc test with Tukey method showed that the difference between ONES and THREES reached significance ($p = .02$). For Old sentences, the difference in confidence values between sentences with different number of propositions was not significant, $F(2, 13) = 0.99$, $p < .3$. A two-way ANOVA also showed that the difference between Old and New sentences was not significant, $F(6, 103) = 1.192$, $p < .3$, suggesting that children with WS assigned in general high confidence values to New sentences and Old sentences. The main effects of sentences with different number of propositions and new/old sentences in recognition were not significantly different, $F(2, 85) = 2.86$, $p = .0746$ and $F(2, 85) = 0.11$, $p = .7$, respectively. Meanwhile, the interaction between new and old sentences was also not significant, F (2, 85) = 0.23, p = .7. A plotted graph based on confidence ratings was given below as Figure 18.

Fig. 18. Comparison of New and Old Sentences for Children with Williams Syndrome.

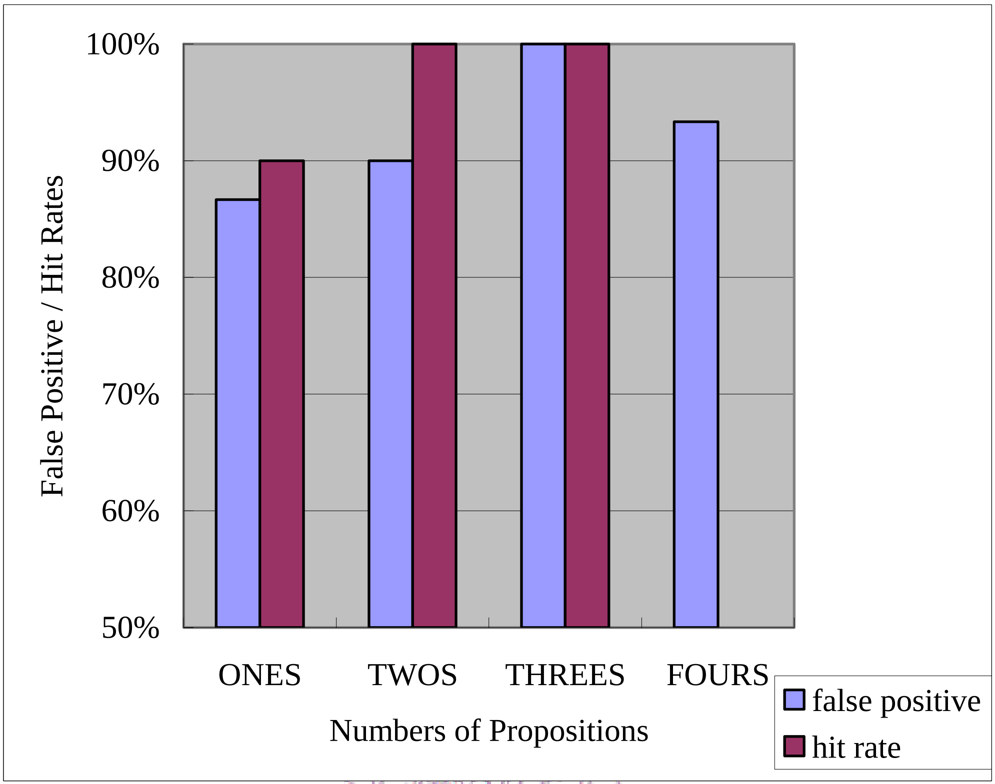


Contrary to our predictions, children with WS showed very high false positives to all new sentences, similar to the pattern observed in normal people. Percent of false positives for new sentences and percent of hit rates for old sentences are listed and detailed in Table 53. A one-way ANOVA showed a significant difference between these four sentence types within New sentences, F (3, 82) = 8.30, p < .0001. A proc mixed model with a post hoc test of Tukey method showed that there was no difference between ONES and TWOS, p = .14. Both ONES and TWOS were significantly different from THREES, p < .0001 and p = .0005, respectively. But, only ONES showed difference with FOURS, p = .01. In other words, WS children showed higher misrecognition rates on sentences with more propositions and lower misrecognition rates on sentences with fewer propositions. These results indicated that WS children showed spared linguistic ability of entailment relations like normal people. Two comparisons did not reach significance, ONES-TWOS ($p = .1$) and TWOS-FOURS ($p = .2$). The hit rates were also very high, which showed the same pattern as confidence ratings. A one-way ANOVA did not show a significant difference between these three sentence types within Old sentences, $F(2, 13) = 1.87$, $p < .19$. A two-way ANOVA showed the main effect of number of propositions was significant, $F(2, 85) = 11.26$, $p < .0001$ and the main effect of sentence types was also significant, $F(1, 85) = 4.34$, $p < .04$. However, the interaction between these two factors was not significant, $F(2, 85) = 1.67$, $p = .1$. A proc mixed model with a post hoc test of Tukey method showed that New TWOS and Old TWOS reached significant difference, $p = .01$. A plotted graph based on false positives and hit rates is given below as Figure 19.

Table 53 Percent of False Positives and Mean (SD) for New Sentences and Percent of Hit Rates and Mean (SD) for Old Sentences on Children with Williams Syndrome

WS	ONES	TWOS	THREES	FOURS
New	86.67%	90%	100%	93.33%
	3.54 (1.56)	4.22 (1.16)	4.33 (0.98)	4.29 (1.27)
Old	90%	100%	100%	----
	3.56(1.74)	4.20(1.6)	4.00(1.55)	----

Fig. 19. Comparison of Percent of False Positives and Percent of Hit Rates for Children with Williams Syndrome.



An average mean recognition confidence rating of each sentence for each WS participant is listed in Table 54 below. Two of the participants (GJH & LHM) showed the increasing recognition confidence ratings in accord with number of propositions in New sentences as normal people. Three of them (GJH, JYL, & LMH) showed lower confidence ratings to sentences with fewer propositions like ONES and TWOS and higher confidence ratings to sentences with more propositions like THREES and FOURS. Meanwhile, one of them (SXY) generally showed high recognition confidence ratings to each sentence stimuli. And a slightly different pattern was observed on one of them (ZHP), who gave the lowest recognition confidence rating to sentences with four propositions, contrary to the prediction. A nonparametric statistics with Kruskal-Wallis Test for New sentences with different number of propositions and

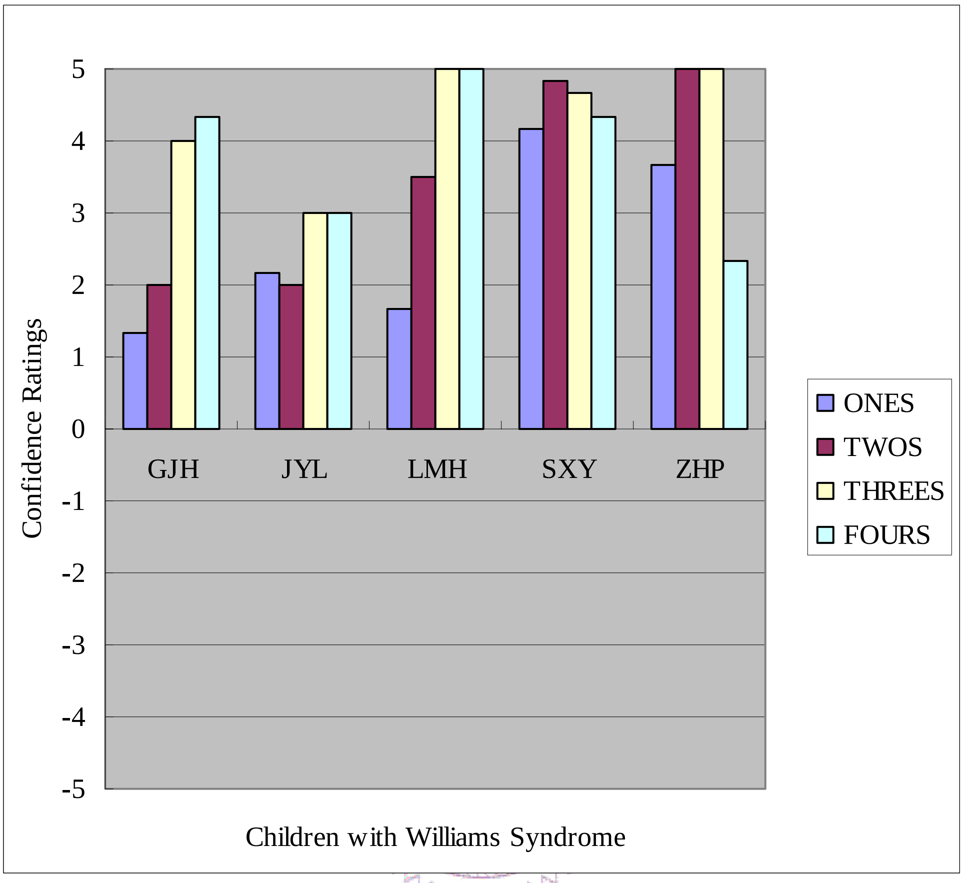
Mann-Whitney Test for paired conditions was employed individually and in general the results showed no significant difference except one WS participant (GJH). The p value for the comparison of New sentences was .035 and the p values both for the comparison of ONES-THREES and ONES-FOURS were .048. The graph plotted individually was given in Figure 20 below.

Table 54 Mean Recognition Confidence Ratings of Each Sentence Condition

	ONES	TWOS	THREES	FOURS
GJH	1.33	2	4	4.33
JYL	2.17	2	3	3
LMH	1.67	3.5	5	5
SXY	4.17	4.83	4.67	4.33
ZHP	3.67	5	5	2.33

TSING HUA UNIVERSITY

Fig. 20. Individual Ordering for Children with Williams Syndrome on New Sentence Conditions.



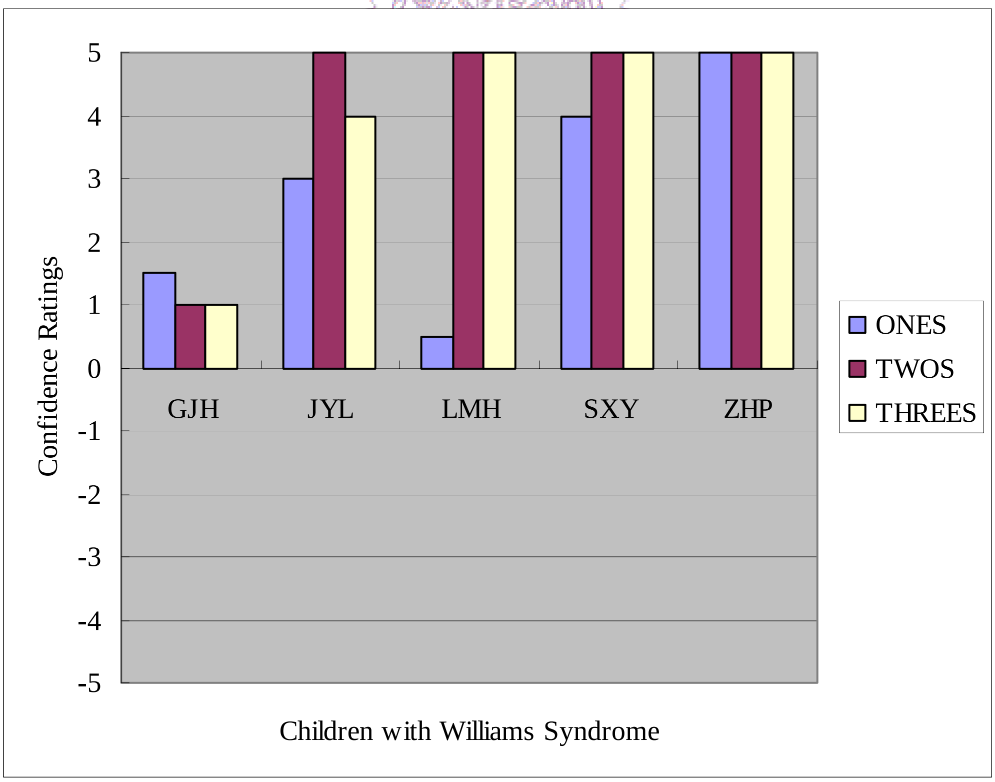
For Old sentences, which were actually displayed in the training section, children with Williams Syndrome showed in general high positive recognition ratings to all sentences, regardless of number of propositions. A nonparametric statistics with Kruskal-Wallis Test for Old sentences with different number of propositions and Mann-Whitney Test for paired conditions were employed individually and in general the results showed no significant difference for any pair compared (all $p > .60$ for overall comparisons and all $p > .66$ for paired comparisons). Detailed raw scores of each sentence condition with different number of propositions were listed in Table 57 below. A plotted graph on old sentences for each sentence condition was given in Figure 21 below (Old-ONES and Old-TWOS were averaged).

Table 55 Detailed Recognition Confidence Ratings of Each Old Sentence

	Old-ONES-1	Old-ONES-2	Old-TWOS	Old-THREES
GJH	2	1	1	1
JYL	5	1	5	4
LMH	5	-4	5	5
SXY	4	4	5	5
ZHP	5	5	5	5

Note: Old-ONES-1 and Old-THREES were selected from superset idea A, Old-ONES-2 was from superset idea B, Old-TWOS was from superset idea C.

Fig. 21. Individual Ordering for Children with Williams Syndrome on Old Sentence Conditions.



General Discussion

Combining the data from Experiments 1 and 2 reveals several interesting results. First, both groups of participants, normal people and children with WS, showed high recognition confidence ratings to all new sentences, but children with WS assigned higher confidence ratings across the board than normal people. A three-way ANOVA showed that the main effect of sentence type was significantly different, $F(1, 696) = 7.93$, $p = .005$, suggesting that new sentences received higher recognition confidence ratings than old sentences. There was also the main effect of groups, $F(1, 696) = 7.93$, $p = .005$, suggesting that children with WS in general assigned higher positive values than normal people. The interaction between sentence types and groups was significant, $F(1, 696) = 5.93$, $p = .01$. However, the effect of number of propositions was not significant across both experiments, $F(2, 696) = 1.74$, $p = .1$, and the interaction between sentence types and number of propositions was also not significant, $F(2, 696) = .09$, $p = .9$, and the interaction between groups and number of propositions was also not significant, $F(2, 696) = .27$, $p = .7$. The interaction of these three factors was also not significant, $F(2, 696) = .07$, $p = .9$. In order to make a closer comparison of the performance of normal people and children with WS on new sentences, planned comparisons between normal and WS groups were made on new sentences for each proposition condition. All three comparisons came out significantly different: ONES ($p = .002$), TWOS ($p = .0002$), THREES ($p = .0009$). These results confirmed that children with WS in general assigned higher confidence ratings on new sentences than normal people.

Second, both groups of participants showed a high percent of false positives on new sentences. Detailed proportions of false positives, means and standard deviations on four sentence conditions for two groups are given in Table 56 and a plotted graph accordingly is given as Figure 22. A three-way ANOVA showed that the main effects of three factors (i.e. sentence type, groups, and number of propositions) were all significantly different, $p < .0001$. The interaction between sentence types and groups was significant, $F(1, 674) = 47.60$, $p < .0001$. But, the interaction between sentence types and number of propositions was not significant, $F(2, 674) = 1.30$, $p = .2$, and the interaction between groups and number of propositions was also not significant, $F(2, 674) = .22$, $p = .7$. The interaction of these three factors was also not significant, $F(2, 674) = .32$, $p = .7$. For the planned comparisons of normal people and children with WS on new sentences, significant difference was reflected on all three proposition type conditions: ONES ($p < .0001$), TWOS ($p < .0001$), THREES ($p = .0001$). These results followed the same pattern as confidence ratings, showing that children with WS in general performed higher false positives on new sentences than normal people.

Third, for old sentences, hit rates were also very high for both groups. Detailed hit rates for three sentence conditions on two groups were given in Table 57 and a plotted graph accordingly was given as Figure 23. In our comparison of normal people and children with WS using a three-way ANOVA, no significant difference was reflected on sentences with different number of propositions: ONES ($p = .8$), TWOS ($p = .7$), THREES ($p = .8$). Similar pattern was observed for recognition confidence ratings. In our comparison of normal people and children with WS, no significant difference was reflected on sentences with different number of propositions: ONES ($p = .1$), TWOS ($p = .6$), THREES ($p = .5$). These results showed that both groups did not differ in recognizing old sentences whether looking at confidence ratings or hit rates.

Table 56 Percent and Mean (SD) Raw Scores for Recognition False Positive (FP) Errors on New Sentences in Experiment I and II

Group	ONES	TWOS	THREES	FOURS
Normal	54.90%	57.84%	59.80%	77.45%
	4.11 (0.87)	4.20 (0.81)	4.07 (0.95)	4.53 (0.68)
WS	86.67%	90%	100%	93.33%
	3.54 (1.56)	4.22 (1.16)	4.33 (0.98)	4.29 (1.27)

Fig. 22. Percent of False Positives in Recognition for Normal people and Children with Williams Syndrome on New Sentences.

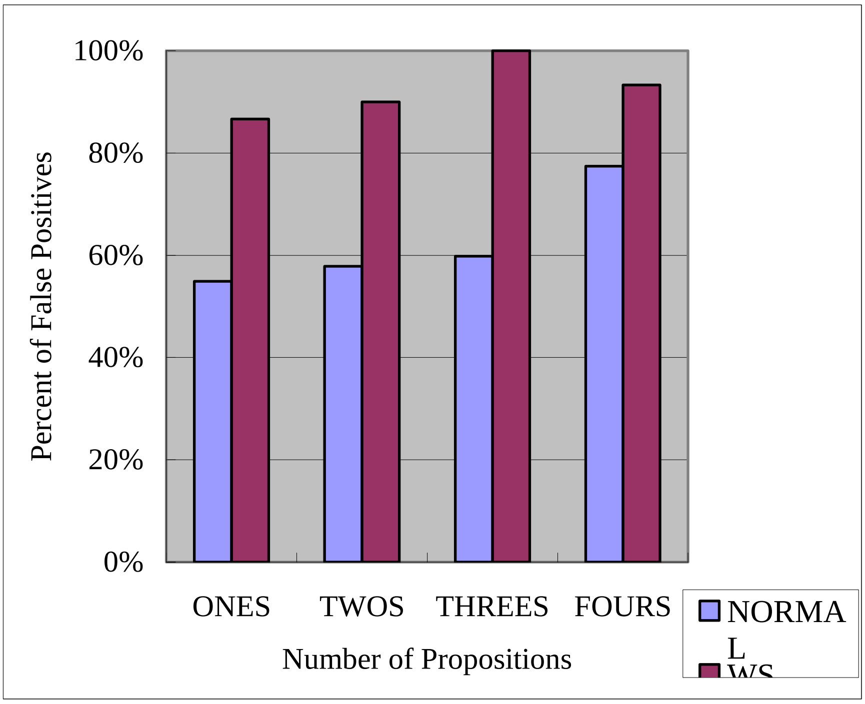
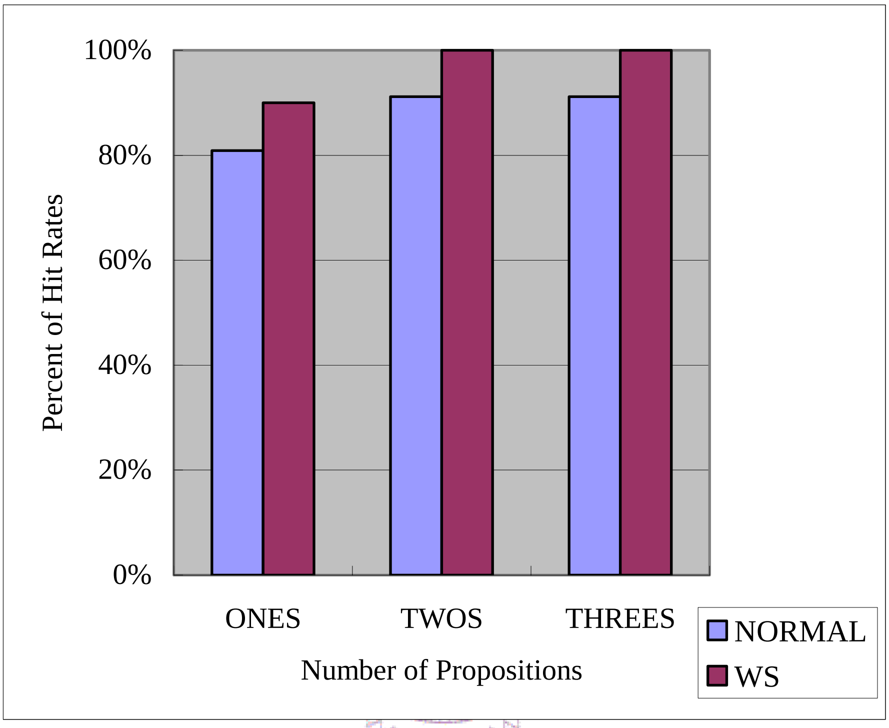


Table 57 Percent and Mean (SD) Raw Scores for Recognition Hit Rates on Old Sentences in Experiment I and II

Group	ONES	TWOS	THREES
Normal	80.88%	91.18%	91.18%
	4.51(0.66)	4.45(0.77)	4.55(0.68)
WS	90%	100%	100%

	3.56(1.74)	4.20(1.6)	4.00(1.55)

Fig. 23. Percent of Hit Rates in Recognition for Normal People and Children with Williams Syndrome on Old Sentences.



Fourth, the results showed that normal people and children with WS patterned differently on new sentences. For normal people, when sentences contained up to three propositions, they tended to recognize those sentences as not-so-familiar sentences, suggesting that they can maintain three propositions in memory. However, when there were more than three propositions in a sentence, normal people could not discriminate those sentences as new and assigned significantly higher positive values to them in recognition. This finding was consistent with the hypothesis that people would spontaneously integrate partial meanings from non-consecutively presented sentences and stored them as a holistic semantic idea in memory. Thus, it seemed that the maximum number of propositions which could be maintained for normal people in memory was three. It is concluded that if someone could not maintain propositions in memory, then judging coherence was compromised. For children with WS, they assigned high confidence ratings in general and also high false positives on all new sentences such that their ratings were not distinguished by number of propositions. From the data showed on false positives, for children with WS, it seems that they could contain maximum two propositions in memory. As long as the number of propositions was over two like three, they cannot maintain them and thus assigned significantly higher recognition confidence ratings on them. In other words, the breakdown points for normal people and children with WS were different. Based on these results, it can be inferred that children with WS have linguistic ability in integrating propositions which have entailment relations. They can build up mental models according to the events/scenarios presented in discourse.

Fifth, differential performance on new and old sentences in the recognition task was expressed to a different extent in normal people and children with WS. In confidence ratings, normal people showed a significant difference in recognition between new and old sentences on ONES, TWOS, THREES, but children with WS did not show any evidence of distinguishing these two kinds of sentences. The same findings were confirmed in the comparison of false positives and hit rates for normal people and children with WS (except one difference on TWOS). We concluded that normal people had difficulty in attempting to maintain three propositions while working on the fourth; however, children with WS could not maintain any as seen by the high confidence ratings generally. Thus, normal people's results matched the prediction that their confidence ratings/false positives would be very high and that they would be a function of proposition complexity. However, children with WS's results did not match the predictions that they showed high confidence ratings/false positives to all new sentences, but their ratings in false positives showed different relation to proposition complexity as normal people. At the same time, it is hard to infer that children with WS do not have difficulty in integrating semantically related sentences because there might be a confounding factor. Since recognition confidence rating was not observed as a function of proposition complexity on the performance of children with WS, one might argue that the general high false recognitions came from their yes-bias tendency (i.e. they are prone to say yes to all conditions). Experiment III and IV were conducted to control for this confound.

F Experiment III: Comparison of New and Scrambled Propositions

Parallel to the second experiment in Bransford and Franks (1971), the third and fourth experiments in our study compared recognition between new sentences and scrambled sentences. Instead of mixing old and new sentences in the recognition stage, scrambled sentences which combined propositions from different superset sentences were included. In doing this study we were interested in several questions: (1) what do comprehenders retain from sentences? Do participants memorize the particular propositions or the grammatical relations between them? Do they build mental models analytically or holistically based on the given contexts in discourse? More relevant to the previous two experiments was the second set of questions: (2) Can children with WS distinguish scrambled sentences from old sentences? Or would they again respond yes to all conditions, suggesting the influence of a yes-bias? We expected that this second set of experiments would help us answer these questions.

Participants

Twenty six undergraduates from National Tsing Hua University were recruited (mean age = 19.6, ranging from 18 to 21, 24 females and 2 males), participating for course credit in Introductory Linguistics. They were right-handed users and none of them were reported as having medical problems.

Design and Materials

The same superset sentence ideas from Experiment I and II were included in this study. The sentence stimuli were all parallel. Eighteen sentences from these superset ideas were displayed as New sentences (as the stimuli listed in Table 50). Practice stimuli were also presented before the test experiment (as the stimuli listed in Table 51). In this study, six sentences which contained different constituents from superset sentences in the training section were mixed together as stimuli. For example, a scrambled sentence like “草叢裡的大野狼正在抓吃紅蘿蔔的老鼠 (A wild wolf in brushwood was catching mice that were eating carrots) came from two different superset ideas A and B in the training section: 森林裡的大野狼抓到了正在草叢裡吃紅蘿蔔的小白兔 (A wild wolf in the forest caught a rabbit which was eating carrots in brushwood) and 廚房裡的老鼠正在偷吃桌子上的草莓蛋糕 (The mice in the kitchen were eating strawberry cakes on the table). These scrambled sentences were listed in Table 58 below. Almost all the scrambled sentences contained four propositions (except the first sentence which contained three propositions) and the mean length of the Chinese stimuli was 17.67 (cf. English stimuli is 13 if translated).

Table 58　Scrambled Stimuli

Scrambled1	草叢裡的大野狼正在抓吃紅蘿蔔的老鼠 (17) A wild wolf in brushwood was catching mice that were eating carrots. (12)
Scrambled2	幼稚園裡可愛的小朋友抓到了廚房裡的老鼠 (19) Cute kids in the kindergarten caught the mice that were in the

	kitchen. (13)
Scrambled3	動物園裡的無尾熊正在玩遊戲吃草莓蛋糕 (18) Koalas in the zoo were playing games and eating strawberry cakes. (11)
Scrambled4	可愛的小白兔正在教室裡吃桌上的紅蘿蔔 (18) Cute rabbits were eating carrots which were on the table in the classroom. (13)
Scrambled5	廚房裡的小朋友正在吃桌上的紅蘿蔔和草莓 (19) Kids were eating carrots and strawberries which were on the table in the kitchen. (14)
Scrambled6	可愛的老鼠正在教室裡吃草莓蛋糕 (15) Cute mice were eating strawberry cakes in the classroom. (9)

Procedure

The procedure was also parallel to Experiment I. All participants were tested in a quiet room in National Tsing Hua University.

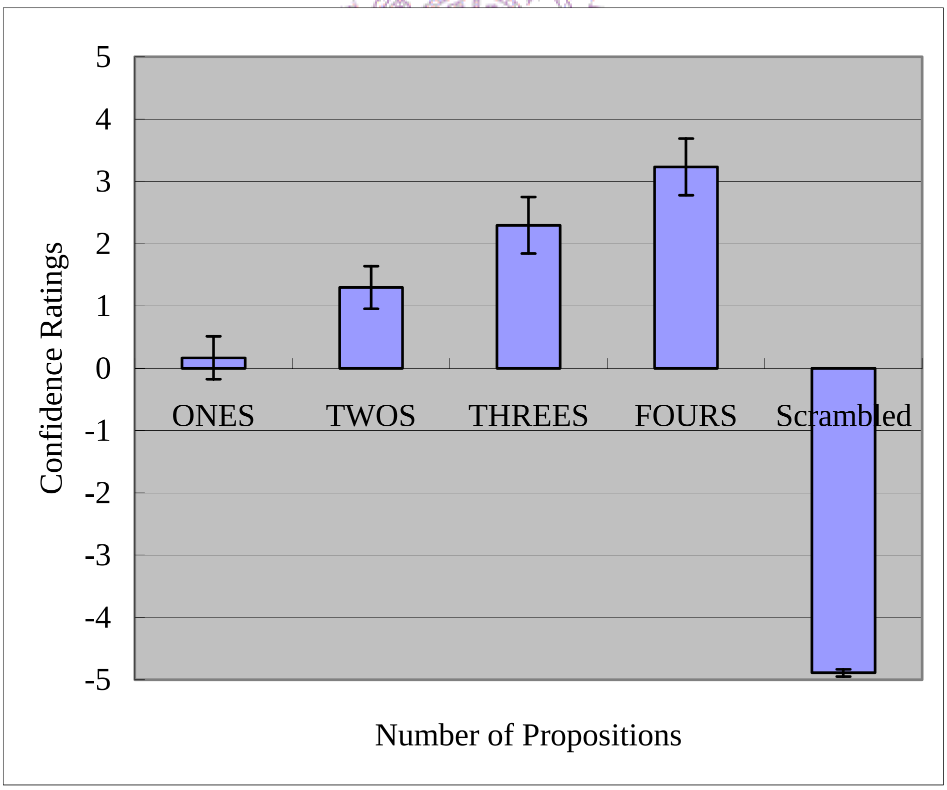
Prediction

In this study, we hypothesized that participants would return high recognition confidence ratings to all new sentences as in Experiment I. Further, the recognition confidence ratings should be a function of the number of propositions in the new sentences. The same pattern was also expected for false positives. Therefore, the more propositions a new sentence contained, the easier it would be for participants to misrecognize its familiarity. The reason for this hypothesis was the evidence from Experiments I and II for spontaneous integration on semantic propositions, meaning that normal people can spontaneously put pieces of information together from contexts in discourse. Moreover, participants were predicted to learn the grammatical relations from the presented sentences in the training section, rather than memorizing the exact wordings or particular propositions. If this prediction was correct, scrambled sentences would be recognized as new.

Results and Discussion

Participants' confidence ratings were converted into numerical values in the same way as in Experiments I and II. The average rating for New sentences accordingly was 0.16 (ONES), 1.29 (TWOS), 2.29 (THREES), and 3.23 (FOURS). A one-way ANOVA showed a significant difference for number of propositions, $F(3, 439) = 13.71$, $p < .0001$, suggesting that participants gave high recognition confidence ratings to those sentences with more propositions and lower recognition confidence ratings to those sentences with fewer propositions. Thus, the recognition confidence rating was a function of number of propositions. Due to an uneven number of trials across four experimental conditions, a proc mixed model with a post hoc test of least significance means (LSMEANS) by using Tukey method was employed. The results showed that the difference between numbers of propositions was significant. The main difference resulted from comparisons between ONES to other sentences (e.g. to TWOS, $p < .03$; to THREES, $p < .0002$; to FOURS, $p < .0001$), and the difference of TWOS to FOURS was also significant ($p < .001$). Meanwhile, the difference between TWOS and THREES approached significance, $p = .05$ and the difference between THREES and FOURS was not significant, $p = .1$. These results indicated that normal people were able to maintain particular sentences up to two propositions and could not give accurate recognition judgments for sentences with more than three propositions. As for the comparison of scrambled sentences (-4.89) and New sentences, this difference was also significant, $F(4, 594) = 124.37$, $p < .0001$. In this study, scrambled sentences were consistently recognized as never heard before, as reflected in the highly negative rating scores, and this indicated an accurate encoding of semantic content and grammatical relations between propositions. A post hoc test with the Tukey method showed a clear difference ($p < .0001$) between scrambled and ONES, TWOS, THREES, FOURS accordingly. It can be inferred that the breakdown point in integration of entailment relations was up to two propositions and the identification of new grammatical/semantic relations between propositions in sentences was spontaneous for normal people. A plotted graph based on recognition confidence ratings is given below as Figure 24.

Fig. 24. Comparison of New and Scrambled Sentences for Normal People.

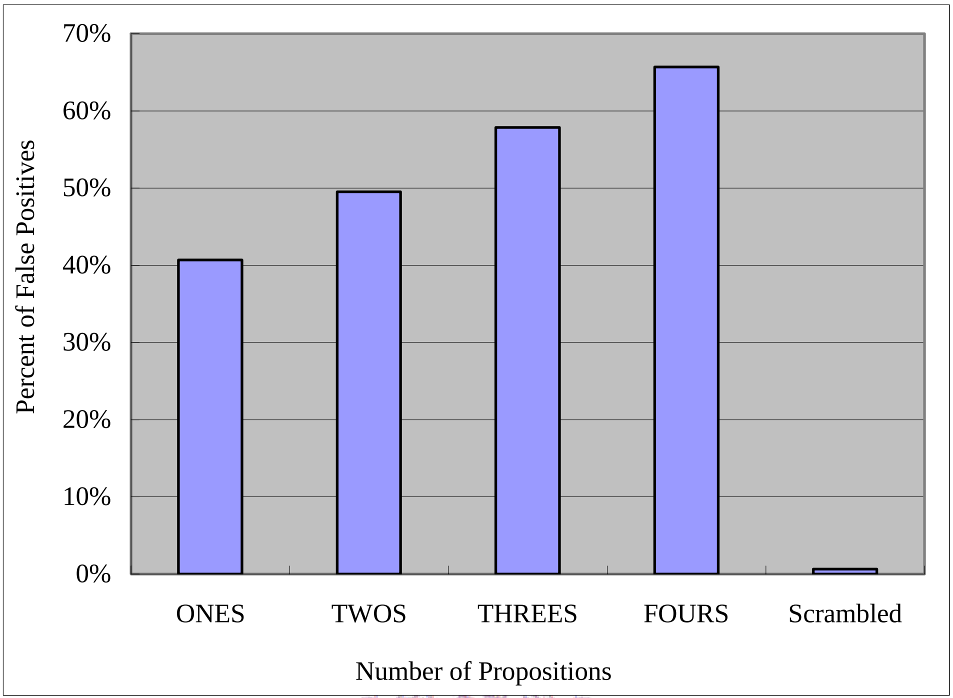


The same pattern was observed for false positives. Proportions of false positives for new sentences and for scrambled sentences are given in Table 59. A one-way ANOVA showed a significant difference between these sentence types both within New sentences and scrambled sentences, $F (4, 281) = 98.10$, $p < .0001$. A post hoc test of Tukey method showed that scrambled sentences received the significantly lowest false positive rating among all conditions (all $p < .0001$ in comparisons), and FOURS received the significantly highest false positive rating than all other conditions, almost all $p < .0001$ in comparisons except with THREES ($p < .001$). The difference between TWOS and THREES was in between. Thus, a clear function of proposition complexity was demonstrated. A plotted graph based on false positives is given below as Figure 25.

Table 59 Percent of False Positives and Mean (SD) for New Sentences and for Scrambled Sentences on Normal People

Normal	ONES	TWOS	THREES	FOURS	Scrambled
Percent of FP	40.69%	49.51%	57.84%	65.69%	0.64%
Mean (SD)	3.95 (0.90)	4.15 (0.91)	4.22 (0.74)	4.39 (0.80)	3[a]

Fig. 25. Comparison of Percent of False Positives (FP) for New Sentences and Scrambled Sentences for Normal People.



General Discussion

In this study, normal people were highly confident that they had never heard the scrambled sentences in the training section before, as they gave very high negative recognition confidence ratings to sentences in this condition. These results clearly indicated that normal people learned the precise meanings and grammatical relations of propositions, which were derivable from semantically related and non-consecutively presented sentences. That is, participants built up the mental models based on the whole events rather than memorizing particular propositions. In agreement with the earlier experiments, participants were very confident that they had heard the new (non-scrambled) sentences with more than two propositions, which actually were not displayed before. Again, recognition confidence ratings were found to covary with the number of semantic propositions embedded in sentences. The ordering of these the four conditions was ONES < TWOS < THREES < FOURS. The same ordering was also observed in false positive ratings. The results matched the predictions. To sum up, normal people did spontaneously integrate semantic information from propositions expressed in sentences and detect the incoherence between propositions like scrambled sentences easily. It is easy for normal people to learn the entailment relations between propositions. The next experiment examines the performance of children with WS on the same materials.

G Experiment IV: Comparison of New and Scrambled Propositions with Chinese Children with Williams Syndrome

Participants

The same study was conducted on children with WS. Six young adults with WS participated in this study: four of them participated in Experiment II, and another two participants (TSJ and CYJ) with Williams Syndrome were recruited. Each participant was diagnosed to be one of the members having this syndrome with *Fluorescent in situ hybridization* (FISH) test in hospital or in laboratory before the experiment. Those who attended both experiments were tested with two weeks difference (mean age =17yr and 1m, range from 12yr and 8m to 21yr and 3m; six males).

Design and Materials

The same superset sentence ideas in Experiment III tested on normal people were included in this study. There were twenty four test sentences in recognition. Six out of twelve were included from each superset idea. Therefore, eighteen sentence stimuli were presented in the training section. The other half sentences of each superset idea were mixed with another six scrambled sentences, which were composed of propositions across different superset idea, and were presented as recognition stimuli.

All the sentences were presented randomly and no two sentences were presented consecutively from the same superset idea.

Procedure

The procedure was the same as Experiment III. No verbal cue was given to participants during the experiment session. Nine practice trials were given before the experiment.

Prediction

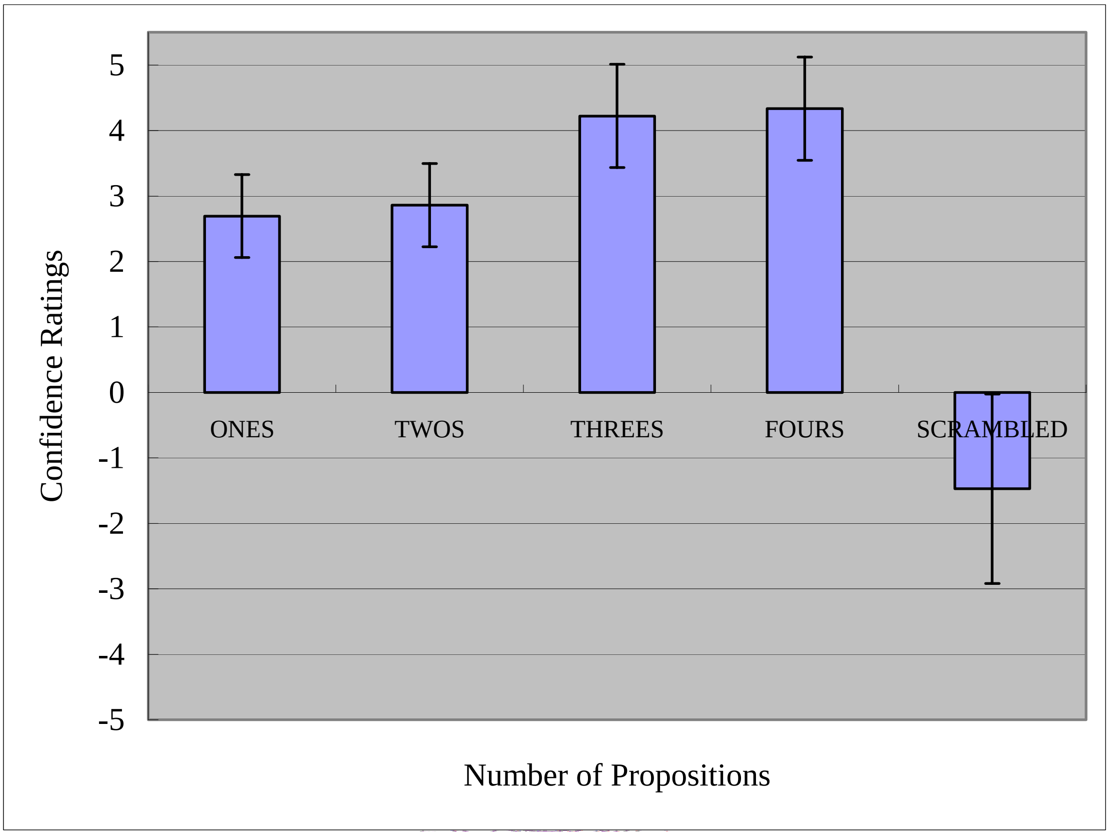
Children with WS were predicted to show high recognition confidence ratings to all new sentences as in Experiment II. The same pattern should apply to false positives. For scrambled sentences, there were two possible predictions. If children with WS were prone to assign positive values to all sentences blindly without differentiating the real events presented, we predicted that they would show high positive recognition confidence ratings to all scrambled sentences. On the contrary, if children with WS can detect incoherence which scrambled sentences showed, they were predicted to assign negative confidence ratings, suggesting that they had never heard these sentences before. In other words, scrambled sentences were the key point to see whether children with WS are tended to respond yes to all conditions without paying attention to the grammatical relations between propositions in sentences. Thus, this experiment can determine whether the high recognition confidence ratings in Experiment II resulted from the ability to integrate propositions from semantically related sentences or from their yes-bias tendency.

Results and Discussion

The same method was used to convert WS participants' confidence ratings as in the earlier experiments. The average for New sentences was 2.69 (ONES), 2.86 (TWOS), 4.22 (THREES), 4.33 (FOURS). A one-way ANOVA showed that the difference between sentences with different number of propositions were marginally significant, $F (3, 99) = 2.33$, $p < .07$. This marginal result was analyzed based on six children with WS and it was possible that the results would be significantly different if there were more children with WS participated. As for the comparison of scrambled sentences (-1.47) and New sentences, the difference was significant, $F (4, 134) = 20.3$, $p < .0001$. A post hoc test with the Tukey method showed a significant difference ($p < .0001$) between scrambled sentences and ONES, TWOS, THREES, FOURS accordingly. A nonparametric statistics with Kruskal-Wallis Test and Mann-Whitney Test showed the same results. The negative confidence ratings to scrambled sentences reflected spared ability in detecting semantic incoherence for children with WS. A plotted graph based on confidence ratings is given below as Figure 26.

TSING HUA UNIVERSITY

Fig. 26. Comparison of New and Scrambled Sentences for Children with Williams Syndrome.

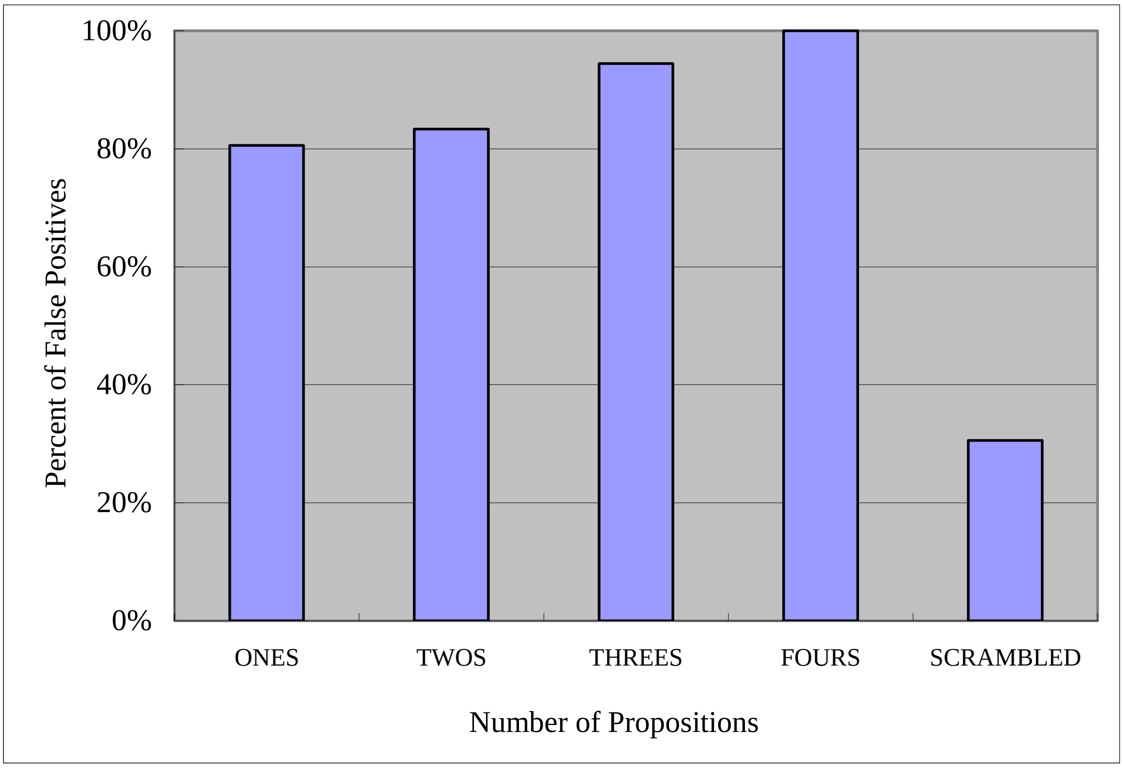


A similar pattern was observed for false positives. The percent of false positives for new sentences and for scrambled sentences is given in Table 60. A one-way ANOVA showed a significant difference between these four sentence types within New sentences, $F (3, 99) = 13.08, p < .0001$. A post hoc test of Tukey method showed that ONES and TWOS received significantly lower percent of false positives than THREES and FOURS. There was no difference between ONES-TWOS and THREES-FOURS. A clear breakpoint between these two groups was therefore demonstrated. A plotted graph based on false positives is given below as Figure 27.

Table 60 Percent of False Positives (FP) and Mean (SD) for New Sentences and for Scrambled Sentences on Children with Williams Syndrome

WS	ONES	TWOS	THREES	FOURS	Scrambled
Percent of FP	80.56%	83.33%	94.44%	100%	30.56%
Mean (SD)	4.28 (1.19)	4.40 (0.97)	4.71 (0.47)	4.33 (0.91)	3.17(1.47)

Fig. 27. Comparison of Percent of False Positives for New Sentences and Scrambled Sentences for Children with Williams Syndrome.



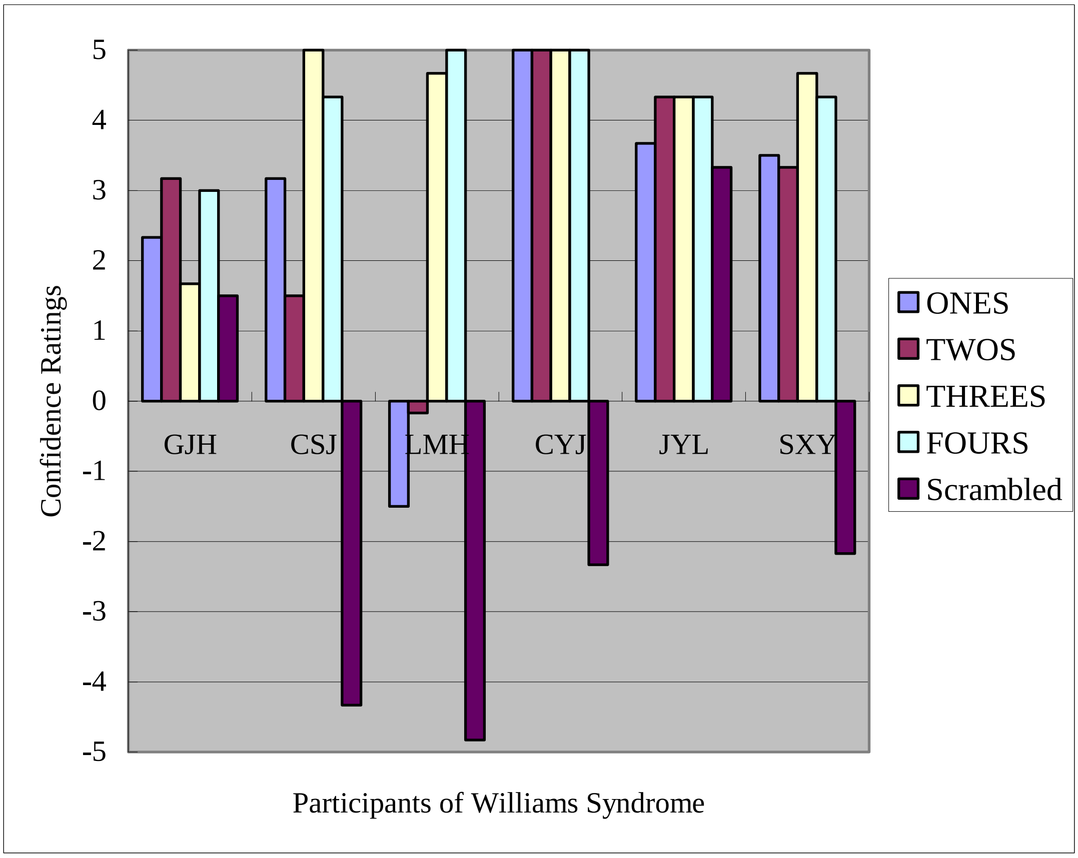
A mean recognition confidence rating for each sentence condition is listed in Table 61 below. Three children with WS (CSJ, LMH, and SXY) showed higher confidence ratings on sentences with three propositions (THREES) and four propositions (FOURS) than on sentences with two propositions (TWOS) and one proposition (ONES). Moreover, among all stimuli, scrambled sentences received the lowest recognition confidence ratings across all participants, though not all of them were judged as negative. Two participants (CSJ and LMH), had similar performance to normal people, showing almost ceiling negative recognition confidence ratings to scrambled sentences. Another two participants (CYJ and SXY) assigned negative values on scrambled sentences while giving very high positive values to all new sentences. However, two participants (GJH and JYL) assigned positive values to scrambled sentences, which in turn had the lowest values among all sentence stimuli. A nonparametric Kruskal-Wallis Test for New sentences with different number of propositions and Mann-Whitney Test for paired conditions was employed individually and in general the results showed no significant difference (except LMH, the p value for the comparison of ONES-FOURS was .048). In the comparison of New sentences and scrambled sentences, five out of six children with WS (except JYL) showed significant difference based on the Kruskal-Wallis Test. Among these five children, all the new sentences with different number of propositions were rated differently from scrambled sentences (except GJH showing a non-significant difference between THREES vs. scrambled and CSJ between TWOS vs. scrambled). The graph plotted individually is given in Figure 28 below.

Table 61 Mean Recognition Confidence Ratings of Each Sentence Condition

	ONES	TWOS	THREES	FOURS	Scrambled
GJH	2.33	3.17	1.67	3	1.5
CSJ	3.17	1.5	5	4.33	-4.33
LMH	-1.5	-0.17	4.67	5	-4.83
CYJ	5	5	5	5	-2.33
JYL	3.67	4.33	4.33	4.33	3.33

SXY	3.5	3.33	4.67	4.33	-2.17

Fig. 28. Individual Ordering for Children with Williams Syndrome on New and Scrambled Sentences.



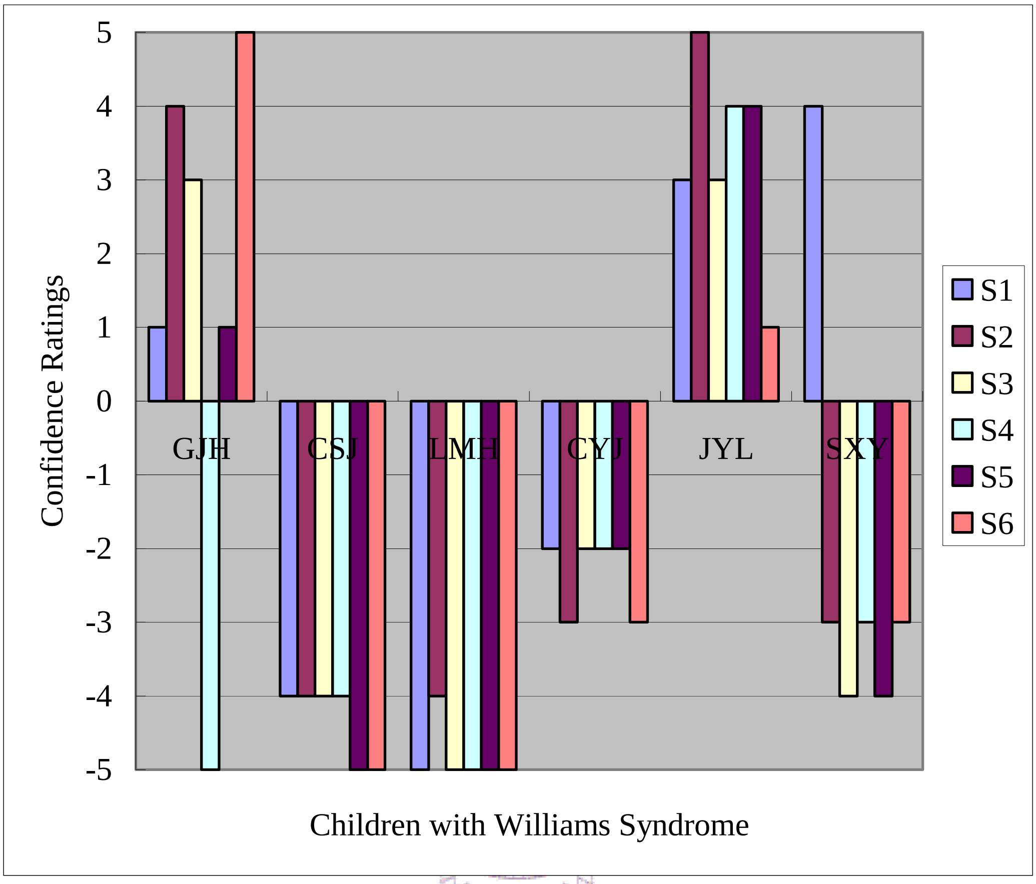
A detailed mean of recognition confidence rating of each scrambled sentence is given in Table 62 below. The individual differences were very big for between-participant comparisons. Three participants (CSJ, LMH, and CYJ) assigned negative values to all scrambled sentences; however, one participant (JYL) assigned positive values to all scrambled sentences. Furthermore, the individual differences were also very big for within-participant comparisons. One participant (GJH) gave very high negative recognition rating to one scrambled sentence (S4) contrary to other scrambled sentences with positive ratings. In contrastive, there was another person with Williams Syndrome (SXY) gave high negative recognition ratings to most of scrambled sentences, but gave positive recognition rating to only one scrambled sentence (S1). The graph plotted individually is given in Figure 29 below.

Table 62 Mean Recognition Confidence Ratings of Each Sentence Stimuli

	S1	S2	S3	S4	S5	S6
GJH	1	4	3	-5	1	5
CSJ	-4	-4	-4	-4	-5	-5
LMH	-5	-4	-5	-5	-5	-5
CYJ	-2	-3	-2	-2	-2	-3
JYL	3	5	3	4	4	1
SXY	4	-3	-4	-3	-4	-3

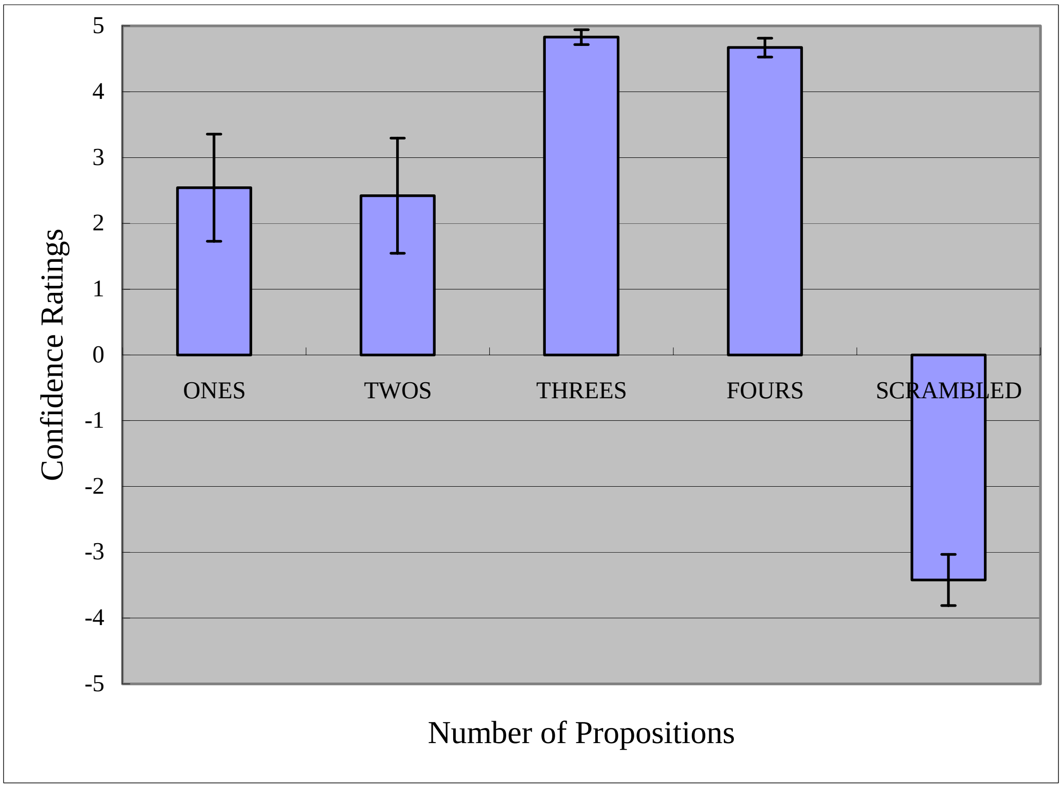
TSING HUA UNIVERSITY

Fig. 29. Individual Ordering for Children with Williams Syndrome on Scrambled Sentences.



Since two children with WS (GJH & JYL) assigned positive values to almost all scrambled sentences, they might be yes-bias children. Thus, a new averaged mean was calculated after taking away their data. The pattern was similar to the original one (cf. Figure 26) with higher confidence ratings on sentences with three and four propositions. Scrambled sentences had higher negative ratings (-3.42) than the original rating (-1.47), but the difference did not reach significance (t-test, $p = .28$). A one-way ANOVA showed that the difference between new sentences and scrambled sentences was significantly different, $F(4, 91) = 22.7$, $p < .000$. The difference mainly resulted from the comparison between scrambled sentences and ONES, TWOS, THREES, and FOURS. Again, these results demonstrated spared linguistic ability on children with WS for detecting semantic incoherence. Further, it clarified that the high recognition confidence ratings assigned to new sentences in Experiment II (comparison of new and old sentences) and in this experiment (comparison of new and scrambled sentences) did not result from a yes-bias tendency on children with WS. Thus, it can be concluded that children with WS spontaneously integrated semantic propositions given the contexts in discourse like normal people.

Fig. 30. Comparison of New and Scrambled Sentences for Children with Williams Syndrome (without GJH & JYL).



General Discussion

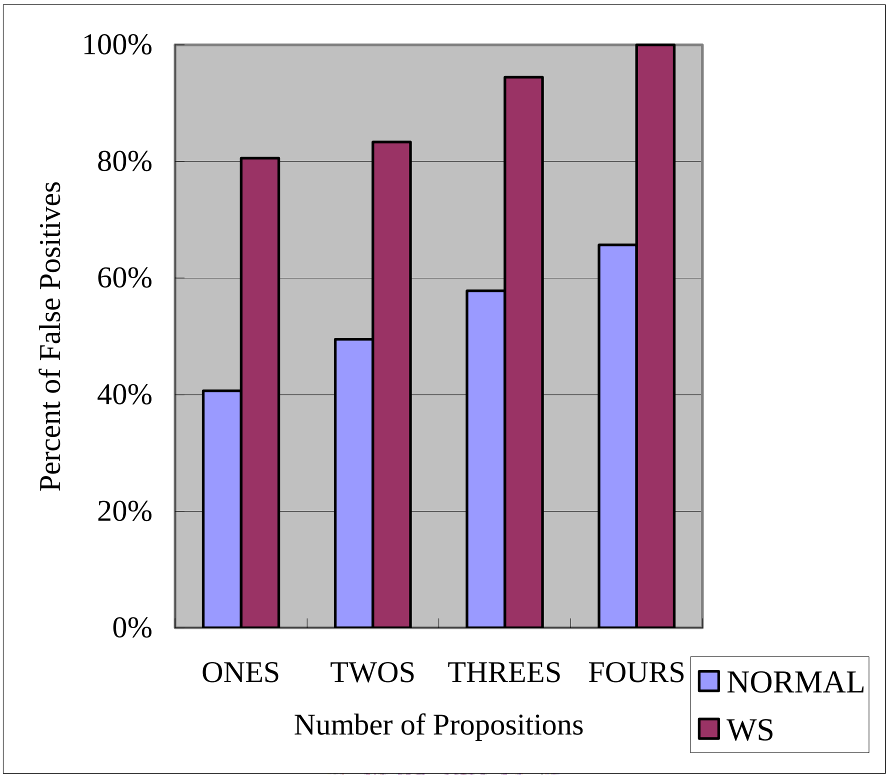
Children with WS showed higher proportion of false positive recognition ratings to all new sentences and scrambled sentences than normal people in this study (two-way ANOVA, $F(4, 562) = 14.79$, $p < .0001$). A clear difference between the percent of false positives in recognition on normal people and children with WS was very obvious, and is represented in Table 63. A paired t-test showed that the difference in mean of false positive recognition was significant for THREES (i.e. sentences with three propositions), 4.22 vs. 4.71, $p = 0.01$, between normal people and children with WS. Further, sentences combined freely with propositions from different idea sets received very negative recognition ratings for both groups, suggesting that the semantic and grammatical relations between propositions were encoded and used in recognition. In general, the more propositions contained in a sentence, the higher percent of false positives assigned. In general, children with WS showed similar pattern as normal people in Experiment III: high confidence ratings and high false positives on new sentences, as a function of proposition complexity, and successful detection of incoherence in scrambled sentences. Thus, we concluded that children with WS did integrate semantic relations within sentences in discourse. Finally, the significantly lower confidence ratings and false positives of scrambled sentences performed on children with WS showed that their pattern of responses did not result from a yes-bias.

Table 63 Percent and Mean (SD) Raw Scores for Recognition False Positive (FP) Errors on New Sentences and Scrambled Sentences in Experiment III and IV

Group	ONES	TWOS	THREES	FOURS	Scrambled
Normal	40.69%	49.51%	57.84%	65.69%	0.64%
	3.95 (0.90)	4.15 (0.91)	4.22 (0.74)	4.39 (0.80)	3[a]
WS	80.56%	83.33%	94.44%	100%	30.56%
	4.28 (1.19)	4.40 (0.97)	4.71 (0.47)	4.33 (0.91)	3.17(1.47)

[a]: Only one positive value assigned to a scrambled sentence. No standard deviation could be calculated.

Fig. 31. Percent of False Positives (FP) in Recognition for Children with Williams Syndrome on New Sentences.



H General Discussion

Where does encoding of semantic integration break down? The breakdown point of semantic integration was different for normal adults and children with WS in both experiments. In Experiment I, when new sentences were compared with old sentences, normal adults showed a clear new-old effect, suggesting that they could distinguish sentences which were actually presented to a certain degree. However, normal adults showed a failure in distinguishing sentences with four propositions, which was reflected both in high recognition confidence ratings and in a high number of false positive recognition ratings. Thus, we concluded from this experiment that normal people have difficulty in attempting to maintain three propositions while working on the fourth. However, in Experiment III, the breakdown point shifted. When scrambled sentences were lumped together with new sentences, normal adults easily distinguished the sentences which were never presented in the training section before with highly negative recognition confidence ratings. As to new sentences, normal adults could not distinguish sentences with more than three propositions and misrecognized them as heard before. Thus, it could be concluded from this experiment that normal people could maintain at most nearly two propositions rather than three.

Why did the boundary shift? It could result from the different composition of sentence types across experiments. When old sentences were mixed in recognition, normal people showed high sensitivity to them due to the exposure in the training section before and recognized them with high confidence ratings or hit rates. Under this circumstance, compared with old sentences which were actually presented, normal people did not show high sensitivity to new sentences as old ones, thus they assigned lower recognition confidence ratings, or say, false positive rates. However, when scrambled sentences were included in recognition, normal people correctly rejected them as never heard because of the incompatibility of the representation of sentences in their memory. Under this circumstance, compared to scrambled sentences which were combined from different superset ideas, normal people showed a lower sensitivity to new sentences because at least new sentences contained the exact wordings as the trained sentences. Thus, the distinguishing boundary shifted.

This effect caused an even more dramatic effect on children with WS. In Experiment II, when old sentences were lumped together with new sentences in recognition, children with WS assigned in general high recognition confidence ratings (i.e. high false positive rates) to all new sentences without showing any distinguished recognition point between them. In other words, putting old sentences in recognition biased participants to assign positive values across the board. In the contrary, in Experiment IV, when scrambled sentences were included in recognition, children with WS showed sensitivity to new sentences and assigned lower recognition confidence ratings. That is, including scrambled sentences in recognition resulted in the change in sensitivity. Thus, children with WS showed a distinguishing recognition boundary in the comparison of new and scrambled sentences, but not in the comparison of new and old sentences.

If this interpretation of the results is correct, did children with WS integrate semantic propositions given the contexts in discourse? The answer to this question is positive. From the results observed in recognition confidence ratings of Experiment II, it was hard to make this conclusion because children with WS showed in general high recognition confidence ratings to all new sentences without a distinguishing recognition point as Bransford and Franks' hypothesized. It seemed that they were prone to respond positively to all sentences (i.e. the yes-bias tendency). However, from the results of false positives analyzed in Experiment II, it seemed that children with WS showed a breakdown point between TWOS and THREES. That is, children with WS could maintain two propositions in memory rather than three as normal people. Thus, it is highly possible that children with WS can build mental model based on entailment relations embedded in propositions. This observation was parallel to the results found in Experiment IV, the comparison of new and scrambled sentences, it was clear that children with WS did integrate semantic propositions during the presentation of trained sentences.

What did participants learn? Two alternatives could account for the effect of number of propositions. It actually is not necessary to claim that participants integrate propositions from different sentences. Some could argue that the features of the sentences were learned. Since sentences were combined with four simple clauses (i.e.

ONES), which expressed an event, participants learned each clause rather than integrating them as a holistic semantic idea. Thus, the more features a sentence has, the easier participants could recognize them. This would be reflected on recognition confidence ratings as a function of number of propositions. Under this scenario, however, we would hypothesize that participants would not show old-new effect because all of them were composed of the same propositions. Bransford and Franks (1972) conducted a study with constrained sentences and unconstrained sentences to investigate the possibility of the analytic feature hypothesis. The basic propositions composing these two types of sentences which were constrained or unconstrained were the same. There were eight totally different propositions used as basic atoms in sentences, for example, ‘the man was rich’, ‘the man lived next door’, ‘the man wore a hat’, and ‘the hat was green’. Constrained sentences were designed from two fixed superset sentences, which were broken down into different sub-sentences in recognition. These superset sentences were created based on the basic propositions and no violations of proposition relations could be found in sentences. However, unconstrained sentences were sentences with irregular relations between propositions, which meant that there was no fixed scenario beforehand. Those propositions could be combined freely. The same paradigm of recognition was conducted. Under the analytic feature hypothesis, we predicted that there should be no new-old effect observed in both constrained and unconstrained sentences because no violation of proposition relations would be detected. In other words, entailment relations cannot play a role in detecting violation. On the other hand, under the holistic representation hypothesis, the new-old sentence effect should be detected in both constrained and unconstrained sentences. The results showed a very clear and strong new-old sentence effect in both types of sentence stimuli. The same finding was replicated when scrambled sentences were lumped together in recognition. Thus, we concluded that interrelation between propositions in sentences was learned and represented in memory instead of feature memorization.

Could recognition confidence rating be a length effect rather than a function of number of propositions? Bransford and Franks (1974) conducted a study with passives to reject the length effect hypothesis. In their study, four types of sentences were presented auditorily in three different paragraphs: full passives, full actives, short passives, and short passives with a generalized actor (i.e. someone) in the training section. There were two types of short passives: one was without an agent and the other was with an agent, which was expressed in a sentence following the short passive (e.g. After the harvest a huge feast was served. Mrs. Brown, who did it, was a very good cook). Later, in recognition, short passives were lumped together with other foils presented on a sheet with nine blocks. Each block contained five syntactic structures of a particular linguistic idea. One of the sentences in each block actually occurred in one of the paragraphs in the training section. Participants were asked to recognize which particular sentences were presented on the paragraphs they heard before. An effect of the agent was predicted if sentence length was not the factor in recognition, because the additional piece of information would matter only if people would spontaneously integrate semantic related propositions. That is, short passives without agents would be recognized better than short passives with agents which were expressed in another sentences. The results showed that these two types of short passives had different recognition rates. Short passives without introducing agents received higher recognition rates than short passives with an additional sentence following. Thus, it looked like people spontaneously integrated the additional propositions into the short passives as a holistic semantic representation, and that therefore the recognition rates for the short passives were low. This finding can be seen as evidence against the length effect hypothesis, which claimed that the recognition difference was resulted from the sentence length. The factor which really mattered and was demonstrated in this study was whether the number of propositions, or pieces of information, contained in a sentence. Thus, we believed that the ordering of the results (ONES < TWOS< THREES < FOURS) we observed in accord with number of propositions in this study was really a function of sentence complexity. At the same time, this is indirect evidence. It would be better to control number of propositions of a similar grammatical unit. For example, sentences with different number of propositions could be compared among property (i.e. adjectives) or location (i.e. prepositions). This would be more direct evidence regarding whether length is a confounding factor. We leave this comparison for a study in the future.

Could the concreteness of sentences make any difference to the results? Begg and Paivio (1970) hypothesized that concrete sentences might be easier for participants to store in memory as images, while abstract sentences might have to be stored in a more verbal coding of exact wordings presented. Bransford and Franks (1972) conducted an experiment with abstract sentences like 'The arrogant attitude expressed in the speech lead to immediate criticism' and 'The unrealistic goals proposed by the leader resulted in frequent disillusionment'. They demonstrated the same ordering observed as when the concrete sentences were used (Bransford and Franks, 1971), suggesting that recognition confidence ratings were a function of number of propositions. Thus, abstractness and concreteness probably doesn't influence participants' representation of sentences in memory.

I Conclusions

The aim of this study was to investigate whether memory for sentence form and meaning dissociate on children with WS. We hypothesized that children with WS have spared grammatical knowledge, but impaired semantic interpretation. This hypothesis came from the studies of lexical semantics like homonyms (Rossen, et al., 1996; Wang & Bellugi, 1993; Bellugi et al., 2000), longitudinal observation of vocabulary growth (Singer-Harris, 1997), invented objects naming (Karmiloff-Smith, 1997) and also studies of grammatical structures like relative clauses (Zukowski, 2001; Grant, Valian, & Karmiloff-Smith, 2002). From these previous studies, children with WS performed an extremely good ability in producing low frequency words, giving secondary meanings of homonyms, nonword repetition advantage, and mapping errors in relative clause elicitation. This hypothesis might result from the advantage of working memory on children with WS (Wang and Bellugi, 1994; Vicari, 1996; Jarrold et al., 1999; Robinson et al., 2003). Thus, proposition integration of entailment relations was tested on children with WS. The results did not support the form and meaning dissociation hypothesis. In Experiment II, children with WS in general assigned high positive values, including confidence ratings and false positives, to all conditions on new sentences and also high hit rates to old sentences. There was no complexity effect of number of propositions on confidence ratings of new sentences and hit rates of old sentences. But, there showed a breakdown point on false positives of new sentences. The breakdown point was between TWOS and THREES, which was different from the breakdown point on normal people (cf. between THREES and FOURS). Due to this discrepancy, it is possible that they do not have the linguistic ability to spontaneously integrate propositions, or that they were showing a simple yes-bias because the significantly higher recognition ratings and number of false positives. However, in Experiment IV, children with WS show a successful detection of sentences which were scrambled from distinct superset sentences. They showed similar patterns both on confidence ratings and false positives. Meanwhile, the confidence ratings and hit rates on old sentences for children with WS and normal people were very high without any difference. Therefore, it can be concluded that children with WS are able to integrate semantically related propositions of sentences based on given contexts. They have good ability in building up mental models based on entailment relations of sentences.

TSING HUA UNIVERSITY

APPENDIX 1　TARGET SENTENCES I IN FUOR SENTENCE TYPS IN NEGATION EXPERIMENTS

Sentence Type	Target Sentence
Factual-Factual	我遲到了，車子也開走了
	I was late, the car was driven away.
Factual-Counterfactual	我遲到了，否則車子就不會開走了
	I was late, otherwise the car wouldn’t have been driven away.
Counterfactual-Factual	如果我沒有遲到，車子也提早開走了
	If I hadn’t been late, the car was driven away earlier.
Counterfactual-Counterfactual	如果我沒有遲到，車子就不會開走了
	If I hadn’t been late, the car wouldn’t have been driven away.

APPENDIX 2　TEST SENTENCES FOR TARGET SENTENCES I IN FOUR SENTENCE TYPES IN NEGATION EXPERIMENTS

Stimuli Type	Test Sentence for Clause 1	Test Sentence for Clause 2
True-Affirmative	我遲到了 I was late	車子開走了 The car was driven away
True-Negative	我沒有準時到 I was not on time	車子不在原地 The car was not at the same place
False-Affirmative	我準時到 I was on time	車子還在原地 The car was at the same place
False-Negative	我沒有遲到 I was not late	車子沒有開走 The car wasn’t driven away

APPENDIX 3　TARGET SENTENCES II IN FUOR SENTENCE TYPS IN NEGATION EXPERIMENTS

Sentence Type	Target Sentence
Factual-Factual	病人死了，醫生也離開了 The patient died, the doctor left.
Factual-Counterfactual	病人死了，否則醫生就不會離開了 The patient died, otherwise the doctor wouldn't have left.
Counterfactual-Factual	如果病人沒有死，醫生也會先離開 If the patient hadn't died, the doctor would have left.
Counterfactual-Counterfactual	如果病人沒有死，醫生就不會離開了 If the patient hadn't died, the doctor wouldn't have left.

APPENDIX 4　TEST SENTENCES FOR TARGET SENTENCES II IN FOUR SENTENCE TYPES IN NEGATION EXPERIMENTS

Stimuli Type	Test Sentence for Clause 1	Test Sentence for Clause 2
True-Affirmative	病人死了 The patient died	醫生離開了 The doctor left
True-Negative	病人沒有活著 The patient wasn't alive	醫生沒有留下 The doctor didn't stay
False-Affirmative	病人活著 The patient was alive	醫生留下來 The doctor stayed
False-Negative	病人沒有死 The patient didn't die	醫生沒有離開 The doctor didn't leave

APPENDIX 5 TARGET SENTENCES III IN FUOR SENTENCE TYPS IN NEGATION EXPERIMENTS

Sentence Type	Target Sentence
Factual-Factual	颱風來了，機場也關閉了 The hurricane came, the airport was closed.
Factual-Counterfactual	颱風來了，否則機場就不會關閉了 The hurricane came, otherwise the airport wouldn't have been closed.
Counterfactual-Factual	如果颱風沒有來，機場也會關閉一陣子 If the hurricane hadn't come, the airport would have been closed for a while.
Counterfactual-Counterfactual	如果颱風沒有來，機場就不會關閉了 If the hurricane hadn't come, the airport wouldn't have been closed.

APPENDIX 6 TEST SENTENCES FOR TARGET SENTENCES III IN FOUR SENTENCE TYPES IN NEGATION EXPERIMENTS

Stimuli Type	Test Sentence for Clause 1	Test Sentence for Clause 2
True-Affirmative	颱風來了 The hurricane came.	機場關閉了 The airport was closed.
True-Negative	颱風沒有遠離 The hurricane wasn't far away.	機場沒有開放 The airport was not opened.
False-Affirmative	颱風遠離了 The hurricane was far away.	機場開放著 The airport was opened.
False-Negative	颱風沒有來 The hurricane didn't come.	機場沒有關閉 The airport was not closed.

APPENDIX 7 TARGET SENTENCES IV IN FUOR SENTENCE TYPS IN NEGATION EXPERIMENTS

Sentence Type	Target Sentence
Factual-Factual	貓來了，老鼠都跑走了 The cat came, the mice all run away.
Factual-Counterfactual	貓來了，否則老鼠就不會跑走了 The cat came, otherwise the mice wouldn't run away.
Counterfactual-Factual	如果貓沒有來，老鼠也都跑走了 If the cat hadn't come, the mice wouldn't have run away.
Counterfactual-Counterfactual	如果貓沒有來，老鼠就不會跑走了 If the cat hadn't come, the mice wouldn't have run away.

APPENDIX 8 TEST SENTENCES FOR TARGET SENTENCES IV IN FOUR SENTENCE TYPES IN NEGATION EXPERIMENTS

Stimuli Type	Test Sentence for Clause 1	Test Sentence for Clause 2
True-Affirmative	貓來了 The cat came.	老鼠跑走了 The mice run away.
True-Negative	貓沒有走掉 The cat didn't go away.	老鼠沒有留下來 The mice didn't stay.
False-Affirmative	貓走掉了 The cat went away.	老鼠留下來 The mice stayed.
False-Negative	貓沒有來 The cat didn't come.	老鼠沒有跑走 The mice didn't run away.

APPENDIX 9 FILLER SENTENCES IN NEGATION EXPERIMENTS

Fillers	Sentences
Filler 1	我準時到了，車子還在原地 I was on time, the car was still there.
Filler 2	我沒有遲到，車子也還在原地 I was not late, the car was still there.
Filler 3	病人活著，醫生也留下來了 The patient was alive, the doctor also stayed.
Filler 4	病人沒有死，醫生也留下來了 The patient did not die, the doctor also stayed.
Filler 5	颱風遠離了，機場也開放了 Typhoon was away, the airport also opened.
Filler 6	颱風沒有來，機場也開放了 Typhoon did not come, the airport also opened.
Filler 7	貓走掉了，老鼠又跑出來了 The cat was gone, mice showed up again.
Filler 8	貓沒有出現，老鼠又跑出來了 The cat did not show up, mice showed up again.

TSING HUA UNIVERSITY

APPENDIX 10　TARGET SENTENCES I IN FUOR SENTENCE TYPS IN YAOBUSHI EXPERIMENTS

Sentence Type	Target Sentence
Factual-Factual	我遲到了，車子也開走了 I was late, the car was driven away.
Factual-Counterfactual	我遲到了，否則車子就不會開走了 I was late, otherwise the car wouldn't have been driven away.
Counterfactual-Factual	要不是我遲到了，車子也開走了 If I hadn't been late, the car was driven away earlier.
Counterfactual-Counterfactual	要不是我遲到了，車子就不會開走了 If I hadn't been late, the car wouldn't have been driven away.

APPENDIX 11　TEST SENTENCES FOR TARGET SENTENCES I IN FOUR SENTENCE TYPES IN YAOBUSHI EXPERIMENTS

Condition	Test Sentence for Clause 1	Test Sentence for Clause 2
True-Affirmative	我遲到了 I was late	車子開走了 The car was driven away
True-Negative	我沒有準時到 I was not on time	車子不在原地 The car was not at the same place
False-Affirmative	我準時到 I was on time	車子還在原地 The car was at the same place
False-Negative	我沒有遲到 I was not late	車子沒有開走 The car wasn't driven away

APPENDIX 12 TARGET SENTENCES II IN FUOR SENTENCE TYPS IN YAOBUSHI EXPERIMENTS

Sentence Type	Target Sentence
Factual-Factual	病人死了，醫生也離開了 The patient died, the doctor left.
Factual-Counterfactual	病人死了，否則醫生就不會離開了 The patient died, otherwise the doctor wouldn’t have left.
Counterfactual-Factual	要不是病人死了，醫生也會先離開 If the patient hadn’t died, the doctor would have left.
Counterfactual-Counterfactual	要不是病人死了，醫生就不會離開了 If the patient hadn’t died, the doctor wouldn’t have left.

APPENDIX 13 TEST SENTENCES FOR TARGET SENTENCES II IN FOUR SENTENCE TYPES IN YAOBUSHI EXPERIMENTS

Condition	Test Sentence for Clause 1	Test Sentence for Clause 2
True-Affirmative	病人死了 The patient died	醫生離開了 The doctor left
True-Negative	病人沒有活著 The patient wasn’t alive	醫生沒有留下 The doctor didn’t stay
False-Affirmative	病人活著 The patient was alive	醫生留下來 The doctor stayed
False-Negative	病人沒有死 The patient didn’t die	醫生沒有離開 The doctor didn’t leave

APPENDIX 14　TARGET SENTENCES III IN FUOR SENTENCE TYPS IN YAOBUSHI EXPERIMENTS

Sentence Type	Target Sentence
Factual-Factual	颱風來了，機場也關閉了 The hurricane came, the airport was closed.
Factual-Counterfactual	颱風來了，否則機場就不會關閉了 The hurricane came, otherwise the airport wouldn't have been closed.
Counterfactual-Factual	要不是颱風來了，機場也會關閉一陣子 If the hurricane hadn't come, the airport would have been closed for a while.
Counterfactual-Counterfactual	要不是颱風來了，機場就不會關閉了 If the hurricane hadn't come, the airport wouldn't have been closed.

APPENDIX 15　TEST SENTENCES FOR TARGET SENTENCES III IN FOUR SENTENCE TYPES IN YAOBUSHI EXPERIMENTS

Condition	Test Sentence for Clause 1	Test Sentence for Clause 2
True-Affirmative	颱風來了 The hurricane came.	機場關閉了 The airport was closed.
True-Negative	颱風沒有遠離 The hurricane wasn't far away.	機場沒有開放 The airport was not opened.
False-Affirmative	颱風遠離了 The hurricane was far away.	機場開放著 The airport was opened.
False-Negative	颱風沒有來 The hurricane didn't come.	機場沒有關閉 The airport was not closed.

APPENDIX 16 TARGET SENTENCES IV IN FUOR SENTENCE TYPS IN YAOBUSHI EXPERIMENTS

Sentence Type	Target Sentence
Factual-Factual	貓來了，老鼠都跑走了 The cat came, the mice all run away.
Factual-Counterfactual	貓來了，否則老鼠就不會跑走了 The cat came, otherwise the mice wouldn't run away.
Counterfactual-Factual	要不是貓來了，老鼠也都跑走了 If the cat hadn't come, the mice wouldn't have run away.
Counterfactual-Counterfactual	要不是貓來了，老鼠就不會跑走了 If the cat hadn't come, the mice wouldn't have run away.

APPENDIX 17 TEST SENTENCES FOR TARGET SENTENCES IV IN FOUR SENTENCE TYPES IN YAOBUSHI EXPERIMENTS

Condition	Test Sentence for Clause 1	Test Sentence for Clause 2
True-Affirmative	貓來了 The cat came.	老鼠跑走了 The mice run away.
True-Negative	貓沒有走掉 The cat didn't go away.	老鼠沒有留下來 The mice didn't stay.
False-Affirmative	貓走掉了 The cat went away.	老鼠留下來 The mice stayed.
False-Negative	貓沒有來 The cat didn't come.	老鼠沒有跑走 The mice didn't run away.

APPENDIX 18 COMPREHENSION QUESTIONS FOR EXPERIMENTAL STIMULI OF IDEA SET A

Idea A learned sentence ONES #1	大野狼在森林裡 A wild wolf was in the forest.
Question	哪裡有大野狼？ Where was the wild wolf?
Idea A learned sentence ONES #2	小白兔在草叢裡 A rabbit was in brushwood.
Question	草叢裡有什麼？ What was there in brushwood?
Idea A learned sentence TWOS #1	森林裡的大野狼抓到了小白兔 A wild wolf in the forest caught a rabbit.
Question	小白兔被什麼抓到了？ What was the rabbit caught?
Idea A learned sentence TWOS #2	大野狼抓到了正在吃紅蘿蔔的小白兔 A wild wolf caught a rabbit which was eating carrots.
Question	小白兔正在吃什麼？ What was the rabbit eating?
Idea A learned sentence THREES #1	大野狼抓到了正在草叢裡吃紅蘿蔔的小白兔 A wild wolf caught a rabbit which was eating carrots in brushwood.
Question	小白兔在哪裡？ Where was the rabbit?
Idea A learned sentence THREES #2	森林裡的大野狼抓到了草叢裡的小白兔 A wild wolf in the forest caught a rabbit which was in brushwood.
Question	大野狼抓到了什麼？ What did the wild wolf catch?

APPENDIX 19 COMPREHENSION QUESTIONS FOR EXPERIMENTAL STIMULI OF IDEA SET B

Idea B learned sentence ONES #1	老鼠在廚房裡 The mice were in the kitchen.
Question	老鼠在哪裡？ Where were the mice?
Idea B learned sentence ONES #2	蛋糕在桌子上 Cakes were on the table.
Question	桌子上有什麼？ What were on the table?
Idea B learned sentence TWOS #1	廚房裡的老鼠正在偷吃蛋糕 The mice in the kitchen were eating cakes.
Question	老鼠正在做什麼？ What were the mice doing?
Idea B learned sentence TWOS #2	老鼠正在偷吃草莓蛋糕 The mice were eating strawberry cakes.
Question	誰在吃草莓蛋糕？ Who were eating strawberry cakes?
Idea B learned sentence THREES #1	老鼠正在偷吃桌子上的草莓蛋糕 The mice were eating strawberry cakes on the table.
Question	老鼠正在偷吃什麼？ What were the mice eating secretly?
Idea B learned sentence THREES #2	廚房裡的老鼠正在偷吃桌子上的蛋糕 The mice in the kitchen were eating cakes on the table.
Question	蛋糕在哪裡？ Where were the cakes?

APPENDIX 20 COMPREHENSION QUESTIONS FOR EXPERIMENTAL STIMULI OF IDEA SET C

Idea C learned sentence ONES #1	小朋友在幼稚園裡 Kids were in the kindergarten.
Question	小朋友在哪裡？ Where were the kids?
Idea C learned sentence ONES #2	小朋友很可愛 Kids were very cute.
Question	小朋友很怎麼樣？ How were those kids?
Idea C learned sentence TWOS #1	幼稚園裡的小朋友正在玩遊戲 Kindergarten kids were playing games.
Question	小朋友正在做什麼？ What were the kids doing?
Idea C learned sentence TWOS #2	可愛的小朋友正在玩遊戲 Cute kids were playing games.
Question	可愛的小朋友正在做什麼？ What were the cute kids doing?
Idea C learned sentence THREES #1	可愛的小朋友正在教室裡玩遊戲 Cute kids were playing games in the classroom.
Question	誰在玩遊戲？ Who were playing games?
Idea C learned sentence THREES #2	幼稚園裡的小朋友正在教室裡玩遊戲 Kindergarten kids were playing games in the classroom.
Question	幼稚園裡的小朋友正在哪裡玩遊戲？ Where were the kindergarten kids playing games?

APPENDIX 21 COMPREHENSION QUESTIONS FOR EXPERIMENTAL STIMULI OF IDEA SET D

Idea D learned sentence ONES #1	無尾熊正在樹上 Koalas were on the trees.
Question	無尾熊在哪裡？ Where were the koalas?
Idea D learned sentence ONES #2	無尾熊在吃油加利葉 Koalas were eating leaves.
Question	無尾熊正在做什麼？ What were the koalas doing?
Idea D learned sentence TWOS #1	動物園裡的無尾熊正在樹上 Koalas in the zoo were on tall trees.
Question	哪裡有無尾熊？ Where were the koalas?

APPENDIX 22 COMPREHENSION QUESTIONS FOR PRACTICE STIMULI OF IDEA SET E

Idea E learned sentence ONES #1	史努比在公園裡 Snoopy was in the park.
Question	誰在公園裡？ Who was in the park?
Idea E learned sentence ONES #2	米老鼠在玩蹺蹺板 Mickey Mouse was playing seesaw.
Question	誰在玩蹺蹺板？ Who was playing seesaw?
Idea E learned sentence THREES #1	米老鼠正在公園裡玩蹺蹺板 Mickey Mouse was playing seesaw in the park.
Question	米老鼠在做什麼？ What was Mickey Mouse doing?

APPENDIX 23 COMPREHENSION QUESTIONS FOR PRACTICE STIMULI OF IDEA SET F

Idea F learned sentence TWOS #1	魚和螃蟹正在吃飼料 Fish and crabs were eating feeding stuffs.
Question	魚正在做什麼？ What was the fish doing?
Idea F learned sentence TWOS #2	水族箱裡有魚和螃蟹 Fish and crabs were in the aquarium.
Question	水族箱裡有什麼？ What were in the aquarium?
Idea F learned sentence THREES #1	水族箱裡的魚和螃蟹正在吃飼料 Fish and crabs were eating feeding stuffs in the aquarium.
Question	螃蟹正在做什麼？ What were the crabs doing?

TSING HUA UNIVERSITY

REFERENCE

Arwood, E. 1991. Semantics and Pragmatics Language Disorders. Maryland: An Aspern Publication.

Au, Terry K.-F. 1983. Chinese and English Counterfactuals: The Sapir-Whorf Hypothesis Revisited. Cognition 15.155-187.

Au, Terry K.-F. 1984. Counterfactuals: In Reply to Alfred Bloom. Cognition 17.289-302.

Bellugi, U., Lichtenberger, L., Jones, W., Lai, Z., and George, M. 2000. The Neurocognitive Profile of Williams Syndrome: A Complex Pattern of Strengths and Weaknesses. Journal of Cognitive Neuroscience 12.7-29.

Bellugi, U., P. Wang, and Jernigan. 1994. Williams Syndrome: An Unusual Neuropsychological Profile, ed. by S. Broman., Grafman, 23-56. New Jersey: Lawrence Erlbaum Associates.

Bellugi, U., Mills, D., Jernigan, T. Hickok, G., and Galaburda, A. 1999. Linking Cognition, Brain Structure, and Brain Function in Williams Syndrome. pp.111-136. Neurodevelopmental Disorders. (ed.) Tager-Flusberg, H. Cambridge: The MIT Press.

Bishop, D., Hartley, J., and Weir, F. 1994. Why and When Do Some Language-Impaired Children Seem Talkative? A Study of Initiation in Conversations of Children with Semantic-Pragmatic Disorder. Journal of Autism and Developmental Disorders 24.177-97.

Bishop, D. 1997. Uncommon Understanding: Development and Disorders of Language Comprehension in Children. UK: Psychology Press Limited.

Bloom, A. 1981. The Linguistic Shaping of Thought: A Study in the Impact of Language on Thinking in China and the West. Lawrence Erlbaum Associates.

Bloom, A. H. 1984. Caution-The Words You Use May Affect What You Say: A Response to Au. Cognition 17.275-287.

Bransford, J. and Franks, J. 1971. The Abstraction of Linguistic Ideas. Cognitive Psychology 2.331-350.

Bransford, J. and Franks, J. 1972. The Abstraction of Linguistic Ideas: A Review. Cognition 1.213-48.

Bransford, J. Barclay, J. and Franks, J. 1972. Sentence Memory: a Constructive Versus Interpretive Approach. Cognitive Psychology 3,193-209.

Caplan, D. and Waters, G. 1999. Verbal Working Memory and Sentence Comprehension. Behavior and Brain Science 22.77-126.

Carpenter, P. 1973. Extracting Information From Counterfactual Clauses. Journal of Verbal Learning and Verbal Behavior 12.512-21.

Carpenter, P. and Just, M. 1975. Sentence Comprehension: A Psycholinguistic Processing Model of Verification. Psychological Review 82.45-73.

Catlin, J., and Jones, N. 1976. Verifying Affirmative and Negative Sentences. Psychological Review, 83(6), 497-501.

Chao, Y.-R. 1955. Notes on Chinese Grammar and Logic, 237-49. Anwar S. Dil: Stanford University Press.

Chao, Y.-R. 1959. How Chinese Logic Operates, ed. by Anwar S.Dil, 250-9. Stanford University Press.

Clark, H. 1976. Semantics and Comprehension. Mouton Publishers.

Condon, J. 1985. Semantics and Communication. New York, London: Macmillan Publishing Company, Collier Macmillan Publishers.

Curtiss, S., Katz, W. and Tallal, P. 1992. Delay versus Deviance in the Language Acquisition of Language-Impaired Children. Journal of Speech and Hearing Research 35.373-383.

Ernst, T. 1995. Negation in Mandarin Chinese. Natual Language and Linguistic Theory, 13, 665-707. Netherlands: Kluwer Academic Publishers.

Franks, J. and Bransford, J. 1972. The Acquisition of Abstract Ideas. Journal of Verbal Learning and Verbal Behavior 11.311-315.

Franks, J. and Bransford, J. 1974. A Brief Note on Linguistic Integration. Journal of Verbal Learning and Verbal Behavior 13.217-219.

Franks, J. and Bransford, J. 1974. Memory for Syntactic Form as a Function of Semantic Context. Journal of Experimental Psychology 103(5).1037-1039.

Franklin, M. and S. Barten. Franklin, M. Barten S. (ed.) 1988. Child Language: A Reader. New York: Oxford University Press, Inc.

Galaburda, A. and Bellugi, U. 2000. Mutli-Level Analysis of Cortical Neuroanatomy in Williams Syndrome. Journal of Cognitive Neuroscience 12.74-88.

Gathercole, S., and Baddeley, A. 1993. Working Memory and Language. UK: Lawrence Erlbalm Associates, Publishers.

Grant, J., Valian, V., and Karmiloff-Smith, A. 2002. A Study of relative clauses in Williams Syndrome. Journal of Child Language 29.403-16.

Hintikka, J. 2002. Negation in Logic and in Natural Language. Linguistics and Philosophy, 25, 585-600. Netherlands: Kluwer Academic Publishers.

Hulme, C., and Mackenzie, S. 1992. Working Memory and Severe Learning Difficulties. Hove, UK: Lawrence Erlbaum Associates.

Iatridou, S. 2000. The Grammatical Ingredients of Counterfactuality. Linguistic Inquiry, 31(2), 231-270.

Ippolito, M. 2002. On the Temporal Dimension of Counterfactuality. pp. 237-255. In Proceedings of the 32th Annual Conference on Northeast Linguistics Society (NELS).

James, F. 1986. Semantics of the English Subjunctive. Vancouver: University of British Columbia Press.

Jarrold, C., Baddeley, A.D., Hewes, A.K. 1999. Genetically Dissociated Components of Working Memory: Evidence from Down’s and Williams Syndrome. Neuropsychologia 37.637-651.

Jaszczolt, K. M. 1999. Discourse, Beliefs and Intentions: Semantic Defaults and Propositional Attitude Ascription. Elsevier Science Ltd.

Jones, W., Bellugi, U., Lai, Z., Chiles, M., Reilly, J., Lincoln, A., and Adolphs, R. 2000. Hypersociability in Williams Syndrome. Journal of Cognitive Neuroscience 12.30-46.

TSING HUA UNIVERSITY

Karmiloff-Smith, A., Grant, J., Berthoud, I., Davies, M., Howlin, P., and Udwin, O. 1997. Language and Williams Syndrome: How Intact is "Intact"? Child Development 68.246-62.

Karmiloff-Smith, A., and Thomas, M. 2003. What Can Developmental Disorders Tell Us About The Neurocomputational Constraints That Shape Development? The Case of Williams Syndrome. Development and Psychopathology, 15, 969-990.

Karmiloff-Smith, A., Brown, J., Grice, S., and Peterson, S. 2003. Dethroning the Myth: Cognitive Dissociates and Innate Modularity in Williams Syndrome. pp. 227-242. In In Developmental Neuropsychology, 23(1&2) (Ed.) Mervis, C. New Jersey: Lawrence Erlbaum Associates, Publishers.

Kazanina, N., and Philips, C. 2003. Temporal Reference Frames and the Imperfective Paradox. pp. 287-300. In Proceedings of West Cost Conference of Formal Linguistics (WCCFL). (Ed.) Garding, G., and Tsujimura, M. Somerville, MA: Cascadilla Press.

Korenberg, J., Chen, X.-N., Hirota, H., Lai, Z., Bellugi, U., Burian, D., Roe, B., and Matsuoka, R. 2000. Genome Structure and Cognitive Map of Williams Syndrome. Journal of Cognitive Neuroscience 12.89-107.

Laing, E., Grant, J., Thomas, M. S .C. & Karmiloff-Smith, A. (in press). Love is….an abstract word: The influence of phonological and semantic factors on verbal short-term memory in Williams Syndrome. *Cortex.*

Landau, B. 1996. Multiple geometric representations of objects in languages and language learners, ed. by Bloom, P. Peterson M. Nadel L. and Garrett M., 317-64. Cambridge: The MIT Press.

Lenhoff, H., Wang, P., Greenberg, F., Bellugi, U. 1997. Williams Syndrome and the Brain. Scientific American. 68-73.

Levy, Y. and Bechar, T. 2003. Cognitive Lexical and Morpho-Syntactic Profiles of Israeli Children with Williams Syndrome. Cortex 39.255-271.

Liu, Lisa G. 1985. Reasoning Counterfactuality in Chinese: Are There Any Obstacles? Cognition 21.239-270.

Majerus, S. 2004. Phonological Processing in Williams Syndrome. pp.125-142. Williams Syndrome Across Languages. (Eds.) Bartke, S., and Siegmuller, J. Philadelphia: John Benjamins Publication.

Majerus, S. Linden, M.V., Mulder, L., Meulemans, T., and Peters, F. 2004. Verbal Short-term Memory Reflects the Sublexical Organization of the Phonological Language Network: Evidence from an Incidental Phonotactic Learning Paradigm. Journal of Memory and Language 51.297-306.

Martin, R., and Lesch, M. 1996. Associations and Dissociations between Language Impairment and List Recall: Implications for models of STM. pp149-177. Models of Short-Term Memory. (ed.) Gathercole, S. Hove, UK: Psychology Press.

Mervis, C.B., Morris, C.A., Bertrand, J., and Robinson, B.F. 1999. Williams Syndrome: Findings from an Integrated Program of Research. pp.65-110. Neurodevelopmental Disorders. (ed.) Tager-Flusberg, H. Cambridge: The MIT Press.

Mervis, C. 2003. Williams Syndrome: 15 Years Of Psychological Research. pp. 1-12. In Developmental Neuropsychology, 23(1&2) (Ed.) Mervis, C. New Jersey: Lawrence Erlbaum Associates, Publishers.

Montgomery, J. 2000. Verbal Working Memory and Sentence Comprehension in Children with Specific Language Impairment. Journal of Speech, Language and Hearing Research, 43, 293-308.

Mills, D., Alvares, T., George, M., Appelbaum, L., Bellugi, U., and Neville, H. 2000. Electrophysiological Studies of Face Processing in Williams Syndrome. Journal of Cognitive Neuroscience 12.47-64.

Nadel, L. 1999. Down Syndrome in Cognitive Neuroscience Perspective. pp.197-222. Neurodevelopmental Disorders. (ed.) Tager-Flusberg, H. Cambridge: The MIT Press.

Nevins, A. 2002. Counterfactuality Without Past Tense. pp. 441-450. In Proceedings of the 32th Annual Conference on Northeast Linguistics Society (NELS).

Nichols, S., Jones, W., Roman, M., Wulfeck, B., Delis, D. Reilly, J., and Bellugi, U. 2004. Mechanisms of Verbal Memory Impairment in Four Neurodevelopmental Disorders. Brain and Language 88.180-189.

Parker, J. and P. Parker. 2002. The Official Parent's Sourcebook on Williams Syndrome: A Revised and Updated Directory for the Internet Age. San Diego: ICON Group International, Inc.

Perkins, D. 1985. Reasoning as Imagination. Interchange, 6(1), 14-26. The Ontario Institute for Studies in Education.

Philips, C. 2004. Linguistics and Linking Problems. pp. 1-41. In Developmental

TSING HUA UNIVERSITY

Language Disorders: From Phenotypes to Etiologies (Eds.) Rice, M. and Warren, S. NJ: Lawrence Erlbaum.

Prasad, P., and Arunkumar, S. 2004. From Short-Term Memory to Semantics-A Computational Model. Neural Computational & Application, 13, 157-167.

Reiss, A., Eliez, S., Schmitt, E., Straus, E., Lai, Z., Jones, W., and Bellugi, U. 2000. Neuroanatomy of Williams Syndrome: A High Resolution MRI study. Journal of Cognitive Neuroscience 12.65-73.

Robinson, B.F., Mervis, C., and Robinson, B.W. 2003. The Roles of Verbal Short-Term Memory and Working Memory in the Acquisition of Grammar by Children with Williams Syndrome. pp. 13-31. In Developmental Neuropsychology, 23(1&2) (Ed.) Mervis, C. New Jersey: Lawrence Erlbaum Associates, Publishers.

Semel, E. and S. Rosner. 2003. Understanding Williams Syndrome: Behavioral Patterns and Interventions . Lawrence Erlbaum Associates.

Sigman, M. 1999. Developmental Deficits in Children with Down Syndrome. pp.179-196. Neurodevelopmental Disorders. (ed.) Tager-Flusberg, H. Cambridge: The MIT Press.

Singer, M., and Rosenburg, S. 1973. The Role of Grammatical Relations in the Abstraction of Linguistic Ideas. Journal of Verbal Learning and Verbal Behavior 12.273-284.

Slobin, D. 2003. Language and Thought Online: Cognitive Consequences of Linguistic Relativity. pp. 157-191. In Language in Mind: Advances in the

Study of Language and Thought. Bradford Books.

Stiles-Davis, J. Kritchevsky M. Bellugi U. Stiles-Davis, J. Kritchevsky M. Bellugi U. (ed.) 1988. Spatial Cognition: Brain Bases and Development. New Jersey: Lawrence Erlbaum Association.

Takano, Y. 1989. Methodological Problems in Cross-Cultural Studies of Linguistic Relativity. Cognition 31.141-162.

Trabasso, T. Discussion of the Papers by Bransford and Johnson and Clark, Carpenter, and Just: Language and Cognition. 439-460.

Tyler, L., Karmiloff-Smith, A., Voice, J., Stevens, T., Grant, J., Udwin, O., Davies, M., and Howlin, P. 1997. Do Individuals with Williams Syndrome Have Bizarre Semantics? Evidence for Lexical Organization Using an On-line Task. Cortex 33, 515-527.

Ullman, M. 2004. Contributions of Memory Circuits to language: the declarative/procedural model. Cognition 92.231-270.

Vicari, S., Carlesimo, G., Brizzolara, D., and Pezzini, G. 1996. Short-term Memory in Children with Williams Syndrome: A Reduced Contribution of Lexical-Semantic Knowledge to Word Span. Neuropsychologia, 34(9).919-925.

Vicari, S., Brizzolara, D., Carlesimo, G., Pezzini, G. Volterra, V. 1996. Memory Abilities in Children with Williams Syndrome. Cortex 32.503-514.

Vicari, S., Bellucci, S., and Carlesimo, G.A. 2001. Procedural Learning Deficit in Children with Williams Syndrome. Neuropsychologia 39.665-677.

TSING HUA UNIVERSITY

Vicari, S., Caselli, M.C., Gagliardi, C., Tonucci, F. And Volterra, V. 2002. Language Acquisition in Special Populations: a Comparison between Down and Williams Syndrome. Neuropsychologia 40.2461-2470.

Volterra, V., Longobardi, E., Pezzini, G., Vicari, S., and Antenore, C. 1999. Visuo-Spatial and Linguistic Abilities in a Twin with Williams Syndrome. Journal of Intellectual Disability Research 43(4).294-305.

Wang, P. and Bellugi, U. 1993. Williams Syndrome, Down Syndrome, and Cognitive Neuroscience. American Journal of Diseases of Children 147.1246-51.

Wang, P. and Bellugi, U. 1994. Evidence from Two Genetic Syndromes for a Dissociation between Verbal and Visual-Spatial Short-Term Memory. Journal of Clinical and Experimental Neuropsychology. 16(2).317-322.

Wu, Z.-Y. Exploring Counterfactuals in English and Chinese. 1989. University of Massachusetts.

Wu, H.-F. 1994. "If Triangles Were Circles,..."---A Study of Counterfactuals in Chinese and in English. Taipei: The Crane Publishing Co.,Ltd.

Zukowski, A. Uncovering Grammatical Competence in Children with Williams Syndrome. 2001. Boston University, Graduate School of Arts and Sciences. Ref Type: Thesis/Dissertation.

TSING HUA UNIVERSITY