The Linguistic Ability of Semantic Integration: Evidence from Proposition Entailment in Chinese Williams Syndrome

Jean C.-F. Hsu
Ovid J.-L. Tzeng
Daisy L. Hung
National Yang Ming University
Laboratory for Cognitive Neuropsychology

Abstract

This study investigates the hypothesis of selective impairment in meaning relative to form in the language processing of 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 has shown that individuals with WS have preserved normal verbal working memory and such intact verbal ability has been 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 a deficit in cognitive ability, individuals with WS are hypothesized to rely heavily on verbal working memory in learning language. This may explain the finding that the 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 selective dissociation of form from meaning in linguistic ability.

In order to examine this issue, a Bransford & Franks' recognition paradigm (1972) 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 recognition judgments about each sentence as to whether it had been heard before. They were asked to assign a recognition confidence rating value for each judgment. Two sets of 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 selective impairment hypothesis, it was expected that individuals with WS would show a high rejection rate

USTWPL **1**: 59-118, 2005.

for all new sentences and high hit rate on 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 have a high rejection rate for all new sentences and scrambled sentences, whereas the unimpaired would show a high false positive rate to all new sentences and high rejection rate for scrambled sentences. Again, the results showed that individuals with WS performed similarly to the unimpaired. In conclusion, individuals with WS showed spontaneous proposition integration in semantics like the unimpaired.

1. 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 demonstrated a double dissociation on short term storage for phonological and for visual-spatial information compared with individuals with Down syndrome. Jarrold, Baddeley, and 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 held a confound. The dissociation, they argued, might 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 those of Wang and Bellugi. Individuals with WS performed better on digit span tasks than on spatial tasks. On the contrary, individuals with DS showed the reverse pattern. Vicari, Brizzolara, Carlesimo, Pezzini, & Volterra (1996) also confirmed this dissociation. They further pursued the issue to see if there is a selective impairment within verbal working memory. They used immediate recall tasks testing individuals with WS and normal mental-age matched controls. They reasoned that if the phonological competence of individuals with WS is intact while semantic competence is deficient, then spared short-term and impaired long-term systems in working memory is to be expected. The results confirmed their predictions. When compared with normal controls, individuals with WS showed a significant difference in recall rate for primacy effect items, 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.

In a 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 participants with WS and mental age matched controls. Two of the word lists were composed of disyllabic words with high frequency or low frequency; another two were composed of four-syllable words with high or low frequency; still another two were composed of acoustically similar or dissimilar words. Participants were asked to repeat these words after presentation by the experimenter. Vicari and colleagues found that participants 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, participants with WS showed less of a frequency effect (i.e. difference in accuracy in repeating high vs. low frequency words) than normal controls. They concluded that the contribution of the phonological loop towards word span effects in participants with WS and normal controls was the same (pp921). The frequency effect was interpreted as a hyper-phonological strategy used by participants with WS relative to normal controls. We hypothesized that an impaired contribution of long-term memory to short-term memory caused this effect in participants with WS. In other words, normal controls used both phonological recoding and semantic information from long-term memory to recall words; however, participants 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, a finding which is comparable to the results discussed so far.

Karmiloff-Smith, Grant, Berthoud, Davies, Howlin, & Udwin (1997) conducted an experiment on morphosyntactic agreement in French 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 WS advantage in repeating novel words 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 also were confused in repeating the nonce words. They often asked the meaning of the nonce words, but WS individuals did not. The authors concluded that people with WS have a deficit in morphosyntactic knowledge; they further claimed a different learning path for WS children compared to that of normal controls. In other words, WS individuals seem to rely on working memory in learning language as do second language learners.

This hypothesis of second-language learning seemed to partially explain the observation of a longitudinal comparison of WS and DS populations (Singer-Harris, Bellugi, Bates, Jones, and Rossen, 1997). Since WS and DS individuals have genetic disorder and developmentally delayed, at what time 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 taking the MacArthur Communicative Development Inventory (CDI), a questionnaire for parents to report on various aspects of their children's verbal development. The results showed that there is no significant difference reported in the onset of first words in participants with WS and DS. After the second year, however, these two populations showed very different patterns in language development. WS children began to learn a great deal of vocabulary relative to their DS counterparts. WS parents also reported that their children could produce many words without understanding the meanings of those words. On the other hand, DS parents reported that their children often had good comprehension but had difficulty in producing words. Singer-Harris et al. concluded that language ability of WS and DS children diverges with the development of grammar. As grammar emerges, the WS population has relatively good productive language ability compared to DS individuals. Based on the findings of Karmiloff-Smith et al. and Singer-Harris et al., a possible explanation of the WS population's showing of a surprising linguistic ability in development in the face of their mental retardation is because of their good verbal short-term memory.

Even though a reduced frequency effect has been demonstrated in participants 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 participants with WS. They used forward digit span, backward digit span, and nonword repetition as verbal working memory indices, and two inventories for grammatical ability: the PPVT-R (Peabody Picture Vocabulary Test-Revised), which tests receptive vocabulary, and the TROG (Test for the Reception of Grammar), which measures grammatical comprehension. All these measures were tested on participants with WS and mental-age matched children with normal development. 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 children whose development was normal. On the contrary, all these measures were significantly correlated with the raw scores of the TROG on participants with WS. From these results, two conclusions were drawn: (1) due to the stronger relation between working memory and grammatical ability, participants with WS seem to rely on working memory in learning language more than unimpaired children, suggesting the possibility of 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 by participants with WS. These conclusions are comparable with the observations that verbal IQ of participants with WS is often higher than mental age matched children while their grammatical ability is much more delayed than that of 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 (1996) study as well as the results from the longitudinal study on the development of first words on participants 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 in participants 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 participants with WS, it is possible that they dissociate form and meaning on lexical items to a 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 illustrates the difficulty in lexical selection found in people with WS. Is it possible that individuals with WS dissociate grammatical knowledge and meaning on the sentential level?

There is another observation about syntax-semantics mismatch on sentential level in individuals with WS. Bellugi et al. (2000) tested WS children on counterfactual questions, a complex grammatical structure involving logical inference. The experimenter asked participants counterfactual questions and analyzed their responses in terms of both grammatical structure 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 showed full syntactic and semantic understanding of counterfactuals. 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*. Notably, they used the subjunctive mood appropriately. In contrast, 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*. However, though individuals with WS performed much better than their DS counterparts and were similar to the unimpaired controls, their semantics was not as good as their syntactic performance when compared to normal controls.

While numerous studies provide evidence of a structure-meaning dissociation in WS, there have not been any studies that have directly and clearly tested this dissociation. In the following research project reported here the major theme is the investigation of the hypothesis of selective impairment in meaning relative to form by using different types of stimuli in different modalities in Chinese.

2. The Paradox of Form and Meaning on Participants with Williams Syndrome

The hypothesis of the selective dissociation between form and meaning on sentential comprehension in participants with WS also comes from several relative clause studies (Zukowski, 2001; Grant, Valian, & Karmiloff-Smith, 2002). Though knowledge of grammatical structures was relatively spared, participants with WS were found to have difficulty in understanding relative clauses. In Zukowski's study, there were two types of relative clause gaps: subject gap (SG) and object gap (OG). An example of a subject relative is 'the woman who drove the red car walked into a supermarket'; a comparable object relative is 'the woman who the caterpillar fell on___ was eating a hamburger'. The experimenter read a scenario consisting of a sentence fragment or noun phrase with a relative clause using question intonation to elicit responses from children. Subject and object relatives were embedded in the fragments. For example, a sentence fragment with subject gap relative clause was asked to children "Which cow turned red?" while a picture with a boy and a girl who is pointing to a cow accordingly was displayed on the computer screen. Thus, the response of "The cow that the girl is pointing to__" was expected. Similarly, a sentence fragment with object gap relative clause was asked to children "Which cow is Max looking at?" while a picture with a bird (its name is Max) and a mouse that is looking at a boy and a girl respectively was shown on the screen. Participants with WS were expected to respond like "Max is looking at the cow that the girl is pointing to___". The results showed that participants with WS had high accuracy (77%) in subject relatives similar to normal children (82%). However, participants with WS showed a difficulty in producing object relatives (11%) compared to normal children (51%). Though accuracy was low for participants with WS in object gap relatives, Zukowski argued that at least 90% of all the participants with WS produced one object relative clause. From these results, it was concluded that participants with WS have near normal ability in relative clause production.

Further analysis of the production of relative clauses in Zukowski indicated that participants with WS had difficulty in understanding sentences because of mapping errors. Mapping errors, in which the subject of a relative clause is taken as the subject of a sentence fragment, were observed very frequently in the response patterns of this study. And the same error patterns were also observed on normal children. For example, when the children were asked the question "***which truck*** turned red?", 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 is "***which car*** is Max looking at?" to which 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 were observed to have these error patterns, participants with WS made more errors than normal children. Therefore, it seems that individuals with WS have good ability in building up surface structures, but find semantics problematic.

Grant, Valian, & Karmiloff-Smith (2002) conducted another study of relative clauses on English-speaking participants with WS. In their study, four groups of participants were recruited: 5-year-old, 6-year-old, 7-year-old normal children and participants with WS (mean chorological age, 17 years old; mean mental age, 8 years old). Four types of relative clauses were employed: (1) 'the boy chasing the horse is fat', which was identified as a subject-subject (SS) stimulus item; (2) 'the cat the cow chases is black', a subject-object (SO) item; (3) 'the dog chases the horse that is brown', an object-subject (OS) item; (4) 'the dog is chasing the cow the boy sees', an object-object (OO) item. Children were asked to repeat each sentence after the experimenter. According to the degree of structural complexity, it was predicted that SO stimuli would be the hardest of the stimulus types because two noun phrases were adjacent and therefore would not be easy to process in comprehension; in succeeding order of simplicity, the stimuli were ranked SO > OS > OO > SS. The results were consistent with the predictions. All participants, including participants with WS, showed the same pattern. However, the participants with WS, who had a mental age over 8-years of age, 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, participants with WS showed a delayed development in grammatical competence, or, alternatively, an impaired ability in assigning correct interpretations to structures like relative clauses. This selective impairment hypothesis between form and meaning in memory will be further investigated in the present study. However, before presenting our investigation, 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.

3. The Role of a Syntactic Form in Memory

A series of experiments by Bransford and colleagues in the early 1970s investigated how people integrate 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 contained 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 integrative process(es) 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 of meaning in conversation, or say, in memory? What is the unit of integration? Is it the exact wordings expressed of sentences and/or the linguistic ideas embedded in structures remembered in memory? What is the role of a sentence? Is the sentence 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 numbers of propositions. The sets were formed by starting from a 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 into 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, the four complex idea sets were included in and distributed across to two phases: learning and recognition. Six sentences (two ONES, two TWOS, two THREES) were selected from each idea set as learning stimuli. 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 judge whether a particular sentence had been presented in learning section. After the judgment, participants were asked to assign recognition confidence ratings on 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 greater the number of propositions in a sentence contained, the more likely it would be for participants to misrecognize the sentences as having been 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 the learning section and the other twenty four sentences were included as recognition stimuli. However, instead of mixing old sentences in the recognition stimuli, in this study six scrambled sentences which were combined by selecting propositions from different idea sets were included. That is, these scrambled sentences were really new to the participants because the events reproduced were completely unfamiliar to them. There were two types of scrambled sentences in this study: free combination and grammatical relation violation. For example, the free combination sentence, "*The old man who was smoking his pipe climbed the steep hill,*" would be derived 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*"; the grammatical relation violation sentence, "*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 with decreasing recognition confidence ratings from FOURS to ONES on new sentences and participants correctly rejected scrambled sentences as never heard stimuli. Thus these results confirmed the finding that recognition confidence rating was a function of sentence complexity or 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 integrated the 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 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 of meaning 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 comprehension.

The experimental hypothesis of the present study is that meaning is selectively impaired relative to form at the sentential level to some degree in participants with WS. If so, by using Bransford and Franks'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 participants 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 comprehension. As for old sentences, we predict that both normal people and participants 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 to be assigned. It is hypothesized that this effect of proposition integration will be observed in normal people, but not in participants with WS.

4. 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 in 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 four simple declaratives, yielding a total of twelve simple declarative propositions. These propositions had semantic relations with each other. These twelve simple declaratives differed in combination in the 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 experiment had two phases, training and recognition. Six sentences derived 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 derived from each superset sentence are listed in Table 1 as Supersets A-C. The average length for Superset A of Chinese stimuli is 13.67 characters (cf. English stimuli is 9.33 words if translated); The average length for Superset B of Chinese stimuli is 11 characters (cf. English stimuli is 7.42 words if translated); The average length for Superset C of Chinese stimuli is 11.5 characters (cf. English stimuli is 9.75 words if translated). Another nine sentences containing different propositions from another three new superset sentences were displayed as practice stimuli, which are given as Supersets D-F in Table 2. Four sentences excerpted from the training section were mixed in with the recognition materials 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 the recognition phase. All the supersets 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, the 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. 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. 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 Appendices 1, 2 and 3 for test sentences from supersets A, B, and C, respectively. In addition, Appendices 4, 5, and 6 are given for practice supersets D, E, and F, respectively.

Table 1 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 2　Practice Stimuli as Supersets 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 stages: training and recognition. All participants had to complete both sections. They were not told prior to the training session that a recognition phase would follow. All sentence stimuli were presented auditorily. During training, a fixation point displayed on a computer screen for 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 related to the content of the sentence which they were required to answer. After the training phase, there was a break for 3 to 4 minutes. Participants were then told that there was another task awaiting for them: a recognition test.

In the recognition section, participants were presented with the other half of sentences from each superset which had not been 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 each sentence had been presented in the first stage of the experiment. Meanwhile, participants were instructed that the sentences which would be presented in this section might all be 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 to indicate how confident they felt about their decision on a 5-point scale from the most confident scaling 5 down to the least confident scaling 1. Once all these requirements were fulfilled, the next recognition stimulus would be presented. All the participants were tested in a quiet room in National Tsing Hua University or in University of Maryland at College Park. Before real experiment starts, another nine practice trials were given.

Prediction

It was hypothesized that normal adults show a pattern similar 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 had not been presented previously. Participants should show that the recognition confidence rating is a function of propositional complexity, suggesting spontaneous proposition integration is taking place in sentence processing. That is, participants may think that they have heard these sentences before and show high recognition confidence ratings. Therefore, the more propositions the sentence contains, the higher recognition confidence ratings should be. In this scenario, participants were inferred to build up mental models according to the entailment relations among the propositions presented. For the old sentences, which actually were displayed in the training phase, participants should also show a high hit rate which would be reflected with high recognition confidence ratings as in Bransford and Franks. This rating was expected to be a function of number of proposition, too.

Results and Discussion

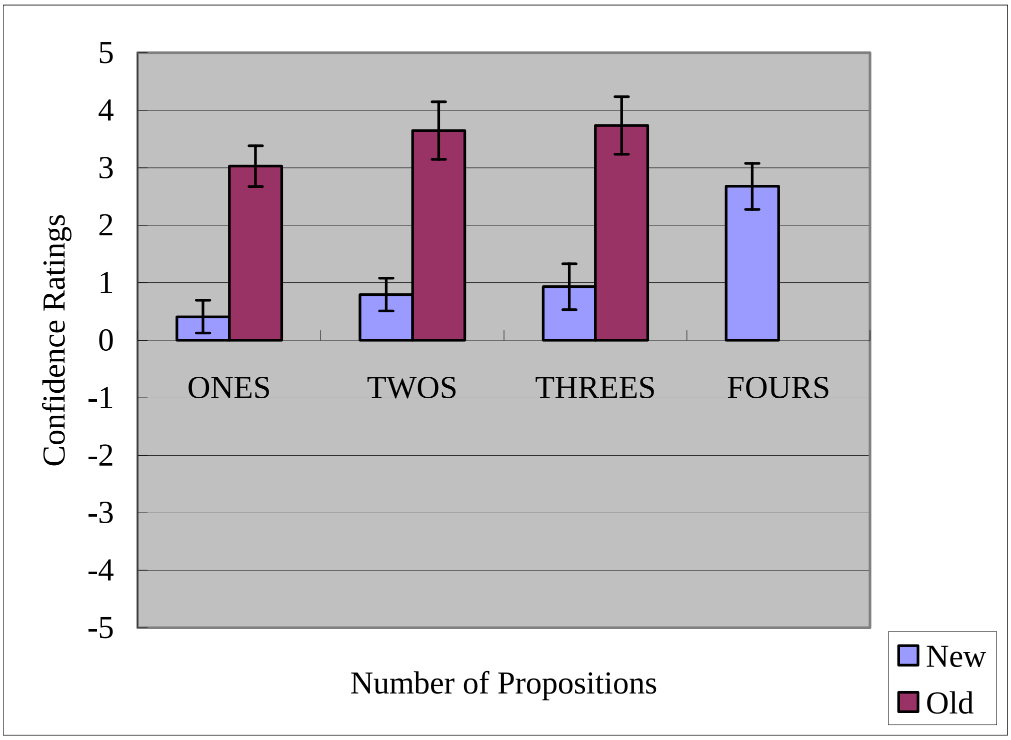
Recognition ratings were computed for each type of sentence: ONES, TWOS, THREES, FOURS, and Old. Participants' ratings were converted into numerical values. A "yes" response received a plus while a "no" response received a minus. A very high confidence rating received a 5, a high confidence rating received 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 the 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 semantically related sentences to a certain degree and built up a mental model based on sentential propositions. A one-way ANOVA showed that the propositional 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, $p = .0104$)[1].

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, $F(2, 100) = 0.88$, $p = .418$, suggesting that normal adults treated all the old sentences as highly familiar stimuli and made yes/no judgments based on the mental model built up during the training phase. Moreover, a two-way ANOVA for new and old sentences showed that the main effect of number of propositions was not significant, $F(2, 607) = 1.14$, $p = .320$. 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, $F(2, 607) = 0.03$, $p = .968$. A plotted graph based on recognition confidence ratings is given as Figure 1.

Figure 1 Comparison of New and Old Sentences for Normal People

[1] 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.



The false positive recognition rates were very high, and followed the same pattern as confidence ratings. The percent of false positives for new sentences and the percent of hit rates for old sentences are given in Table 3. A one-way ANOVA showed a significant difference between these four sentence types within the new sentences, $F(3, 562) = 37.89$, $p < .0001$. A proc mixed model with a post hoc test with the Tukey method showed that ONES received significantly lower false positives and FOURS received significantly higher 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, $p = .107$.

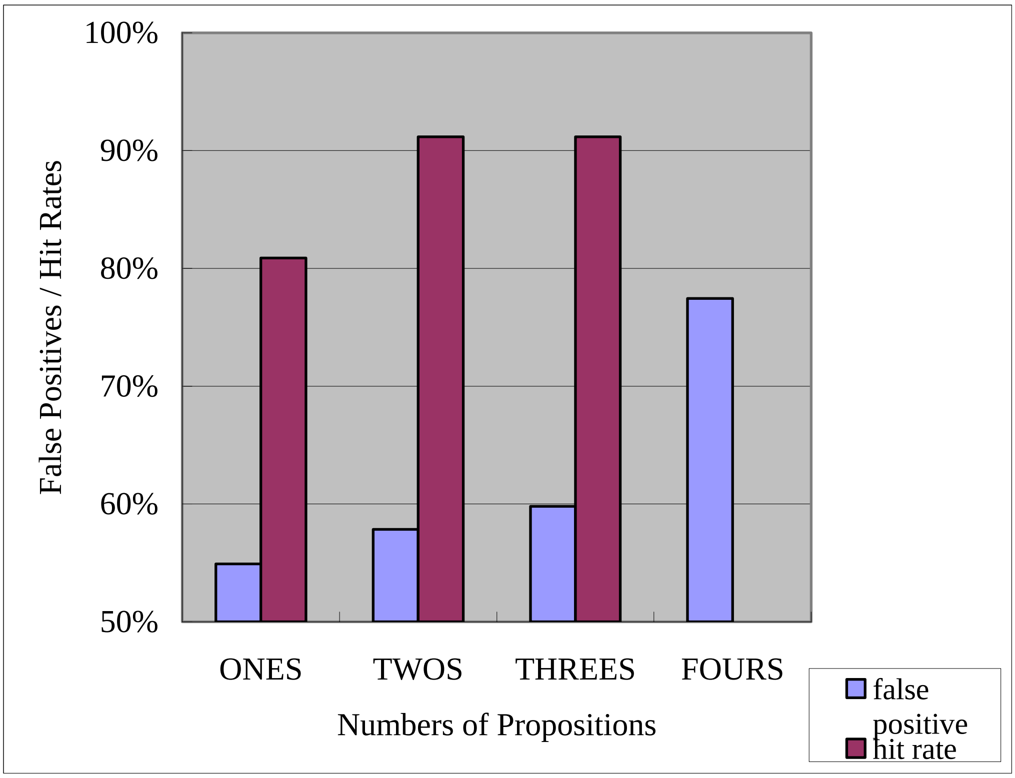
The hit rates were also very high, and showed the same pattern as confidence ratings. A one-way ANOVA showed a significant difference between these three sentence types within old sentences, $F(2, 90) = 18.26$, $p < .0001$. A proc mixed model with a post hoc test with the Tukey method showed that ONES 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 with the 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 = .047, 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 a significant difference, p = .01. A plotted graph based on false positives and hit rate is given in Figure 2.

In sum, 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 turned to the central issue of our inquiry, the extent to which verbal working memory in WS patients is different from that of unimpaired subjects. Experiment 2 used the same procedure as Experiment 1 but tested WS children.

Table 3　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
False Positives	54.90%	57.84%	59.80%	77.45%
	4.11 (0.87)	4.20 (0.81)	4.07 (0.95)	4.53 (0.68)
Hit Rates	80.88%	91.18%	91.18%	----
	4.51(0.66)	4.45(0.77)	4.55(0.68)	----

Figure 2　Comparison of Percent of False Positives and Percent of Hit Rates for Normal People



5. Experiment II: Comparison of New and Old Sentences with Chinese Participants with Williams Syndrome

Participants

Five young adults with Williams syndrome participated in this study (mean age =17yr and 9 months, range from 12yr and 8 month to 21yr and 3 month; 4 males and 1 female). Each participant was diagnosed as having this syndrome with *Fluorescent in situ hybridization* (FISH) test in a hospital or a laboratory. Individual background information is provided in Table 4.

Table 4 General Information of WS Participants

	LMH	JYL	GJH	SXY	ZHP
Chronological age	17;06	12;08	19;07	21;03	18;04
Gender	Male	Male	Male	Male	Female
FIQ (WISC III)	72	---	48	---	---
VIQ	84	---	54	---	---
PIQ	66	---	50	---	---

Design and Materials

Twenty two sentences were included in this study. 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 that in Experiment I. A training stage was required before recognition. During training, participants with WS were presented sentences auditorily and asked to name colors displayed one at a time on the computer screen. After color naming, a comprehension question associated with the training stimulus was auditorily presented. The children were required to answer the question orally right away 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.

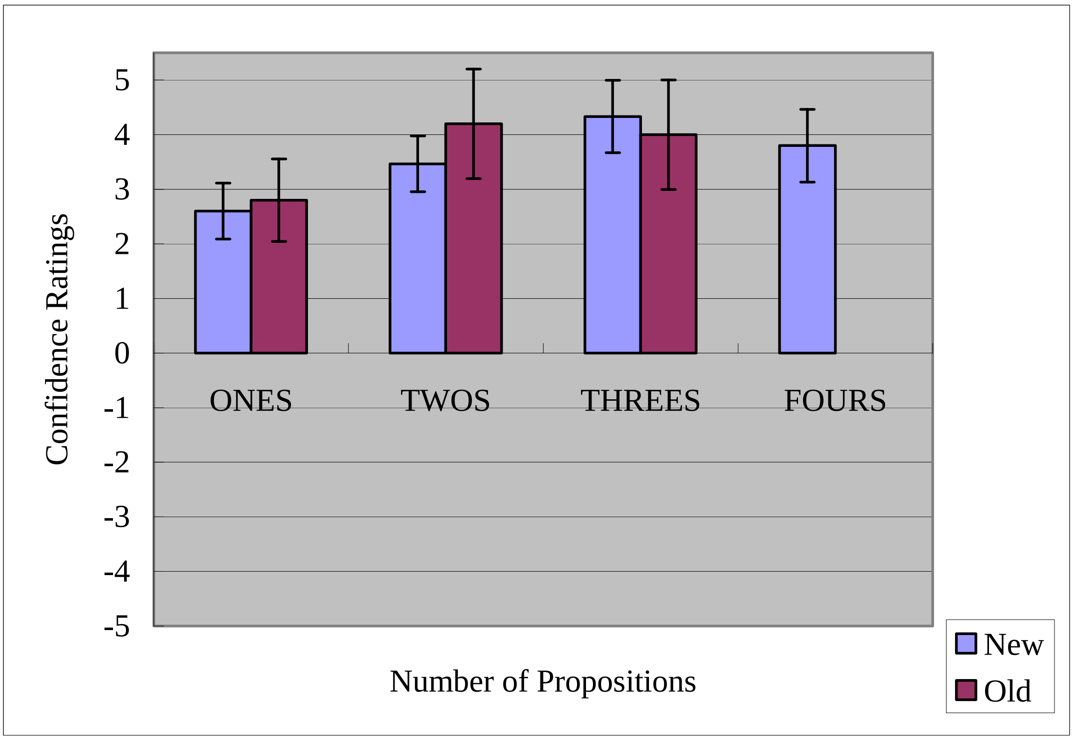
Prediction

The performance of participants with WS was predicted to be different from that of the unimpaired subjects previously tested. Participants with WS were expected to show low false positive recognition to all new sentences no matter how many propositions were embedded. That is, participants with WS should be able to correctly reject new sentences if they have superior verbal working memory and much rely on spared linguistic knowledge of grammatical structures. In other words, they are not good at building mental models involving semantic integration, so they can distinguish the sentences easily. Also, if they have an impaired ability in understanding sentences, they may not be able to use a propositional 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 participants with WS may have difficulty in integrating semantically related propositions and are limited in their capacity of 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 < .1017 although a post-hoc test with the 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 < .397. A two-way ANOVA also showed that the difference between old and new sentences was not significant, F (6, 103) = 1.192, p < .316, suggesting that participants with WS assigned, in general, high confidence values to new sentences and old sentences. The main effects of sentences with different numbers 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 = .7364, respectively. Meanwhile, the interaction between new and old sentences was also not significant, F (2, 85) = 0.23, p = .7960. A plotted graph based on confidence ratings was in Figure 3.

Figure 3 Comparison of New and Old Sentences for Participants with Williams Syndrome

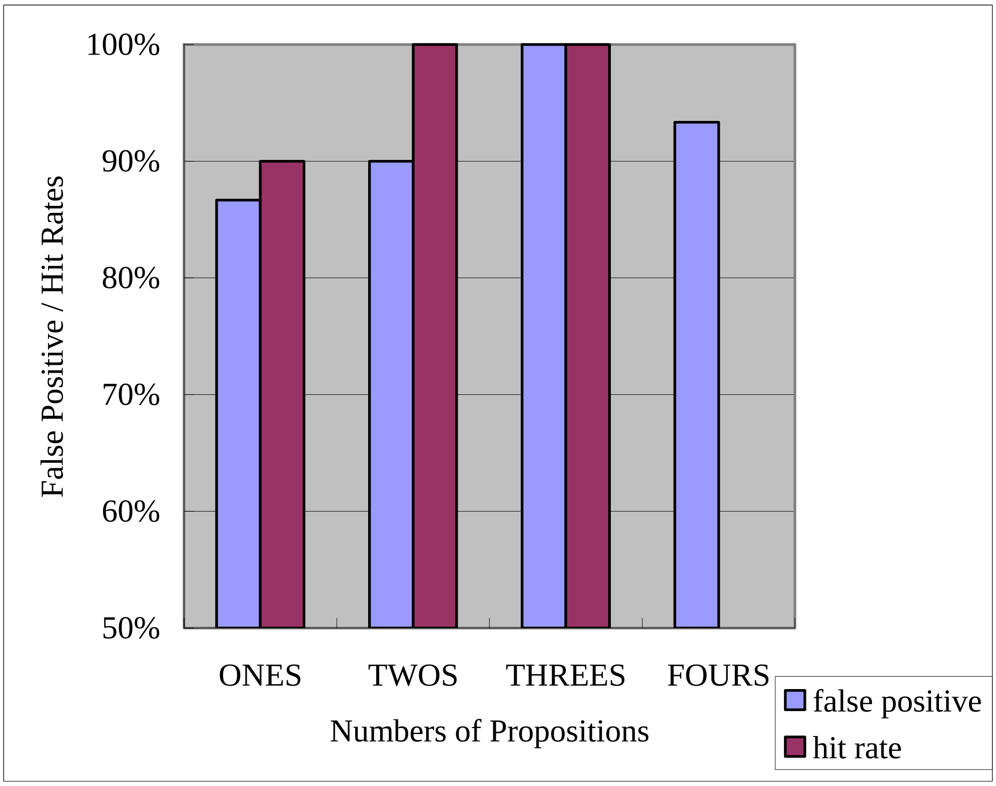


Contrary to our predictions, participants with WS showed very high false positives to all new sentences, similar to the pattern observed in normal people. The percent of false positives for new sentences and the percent of hit rates for old sentences are detailed in Table 5. A one-way ANOVA showed a significant difference between the four sentence types within new sentences, F (3, 82) = 8.30, p < .0001. A proc mixed model with a post hoc test with the Tukey method showed that there was no difference between ONES and TWOS, $p = .1420$. Both ONES and TWOS were significantly different from THREES, $p < .0001$ and $p = .0005$, respectively. But, only ONES showed a difference with FOURS, $p = .0177$. 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 with respect to integrating entailment relations comparable to normal people. Two comparisons did not reach significance, ONES-TWOS ($p = .142$) and TWOS-FOURS ($p = .229$). The hit rates were also very high, which showed the same pattern as the confidence ratings. A one-way ANOVA did not show a significant difference between these three sentence types within old sentences, $F(2, 13) = 1.87$, $p < .1925$. 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, $F(2, 85) = 1.67$, $p = .195$. A proc mixed model with a post hoc test using the Tukey method showed that the difference in TWOS between false positives and hit rates reached significance, $p = .011$. A plotted graph, based on false positives and hit rates, is given in Figure 4.

Table 5 Percent of False Positives and Mean (SD) for New Sentences and Percent of Hit Rates and Mean (SD) for Old Sentences on Participants with Williams Syndrome

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

Figure 4 Comparison of Percent of False Positives and Percent of Hit Rates for Participants with Williams Syndrome



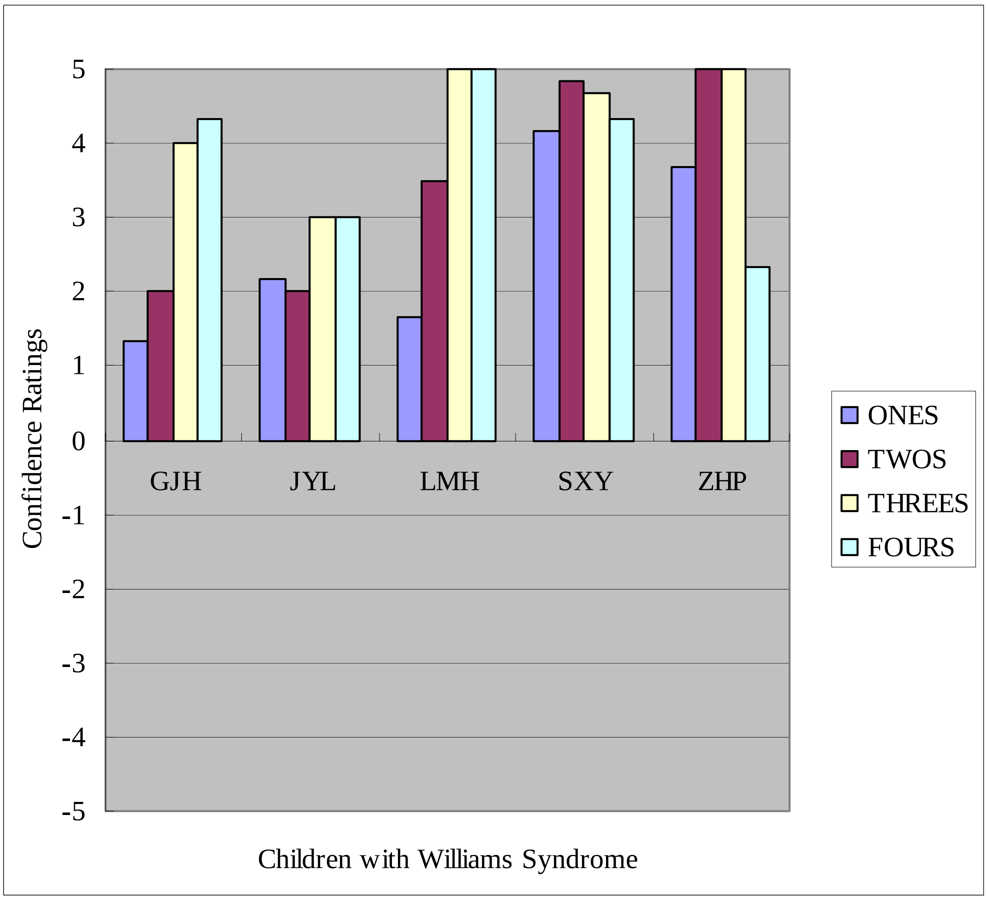
An average mean recognition confidence rating for each sentence for each WS participant is listed in Table 6. Two of the participants (GJH & LHM) showed the increasing recognition confidence ratings in accord with number of propositions in new sentences as normal people. Three of them (GJH, JYL, & LMH) showed lower confidence ratings to sentences with fewer propositions, ONES and TWOS, and higher confidence ratings to sentences with more propositions, THREES and FOURS. Meanwhile, one of participants (SXY) generally showed high recognition confidence ratings to each sentence stimulus. And a slightly different pattern was observed for another participant (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 is given in Figure 5.

Table 6 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

Figure 5 Individual Ordering for Participants with Williams Syndrome on New Sentence Conditions



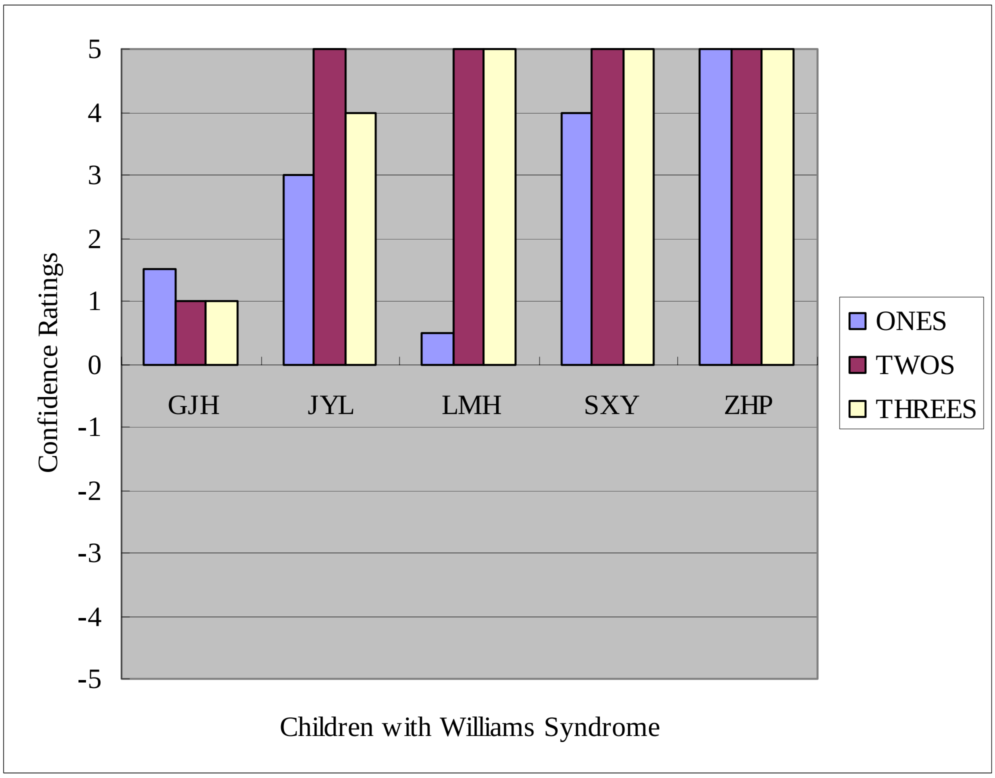
For old sentences, which were actually presented in the training phase, participants with WS showed, in general, high positive recognition ratings for all sentences, regardless of the number of propositions. A nonparametric statistics with the Kruskal-Wallis Test for old sentences with different numbers 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 $p > .607$ for overall comparisons and all $p > .667$ for paired comparisons). Detailed raw scores for each sentence condition with different numbers of propositions are listed in Table 7. A plotted graph on old sentences for each sentence condition is given in Figure 6 (Old-ONES and Old-TWOS were averaged).

Table 7　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 A, Old-ONES-2 was from superset B, Old-TWOS was from superset C.

Figure 6　Individual Ordering for Participants with Williams Syndrome on Old Sentence Conditions



General Discussion

Combining the data from Experiment I and II reveals several interesting results. First, both groups of participants, normal people and participants with WS, showed high recognition confidence ratings to all new sentences, but participants 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 participants 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 = .015. However, the effect of number of propositions was not significant across both experiments, F (2, 696) = 1.74, p = .176, the interaction between sentence types and number of propositions also was not significant, F (2, 696) = .09, p = .918, and the interaction between groups and number of propositions was not significant, F (2, 696) = .27, p = .763, either. Finally, the interaction of these three factors was not significant, F (2, 696) = .07, p = .933. In order to make a closer comparison of the performance of normal people and participants 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 participants with WS in general assigned higher confidence ratings to new sentences than the unimpaired subjects.

Second, both groups of participants showed a high percent of false positives on new sentences. Proportions of false positives, means and standard deviations on the four sentence conditions for two groups are given in Table 8 and a plotted graph is given in Figure 7. 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 = .273, and the interaction between groups and number of propositions was also not significant, F (2, 674) = .22, p = .798. Finally, the interaction of these three factors was also not significant, F (2, 674) = .32, p = .723. For the planned comparisons of normal people and participants with WS on new sentences, significant differences were 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 participants with WS in general had a higher number of 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 the three sentence conditions of the two groups are given in Table 9 and plotted in Figure 8. In our comparison of normal people and participants with WS using a three-way ANOVA, no significant difference was reflected on sentences with different numbers of propositions: ONES (p = .896), TWOS (p = .728), THREES (p

= .853). Similar pattern was observed for recognition confidence ratings. In our comparison of normal people and participants with WS, no significant difference was reflected on sentences with different numbers of propositions: ONES (p = .125), TWOS (p = .619), THREES (p = .583). These results showed that the groups did not differ in recognizing old sentences whether considered in terms of confidence ratings or hit rates.

Table 8　Percent and Mean (SD) Raw Scores for Recognition False Positive 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)

Figure 7　Percent of False Positives in Recognition for Normal people and Participants with Williams Syndrome on New Sentences

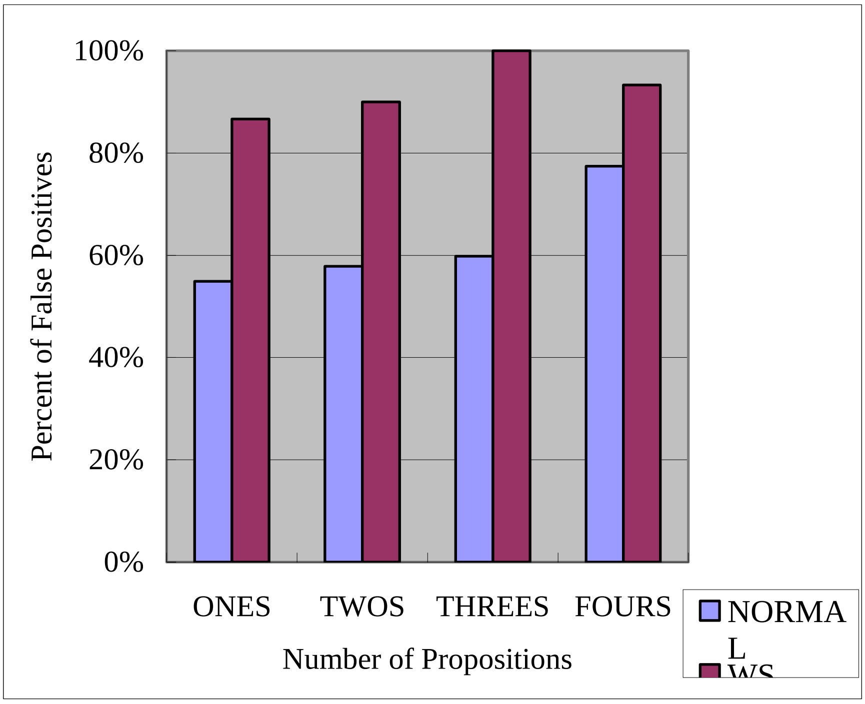
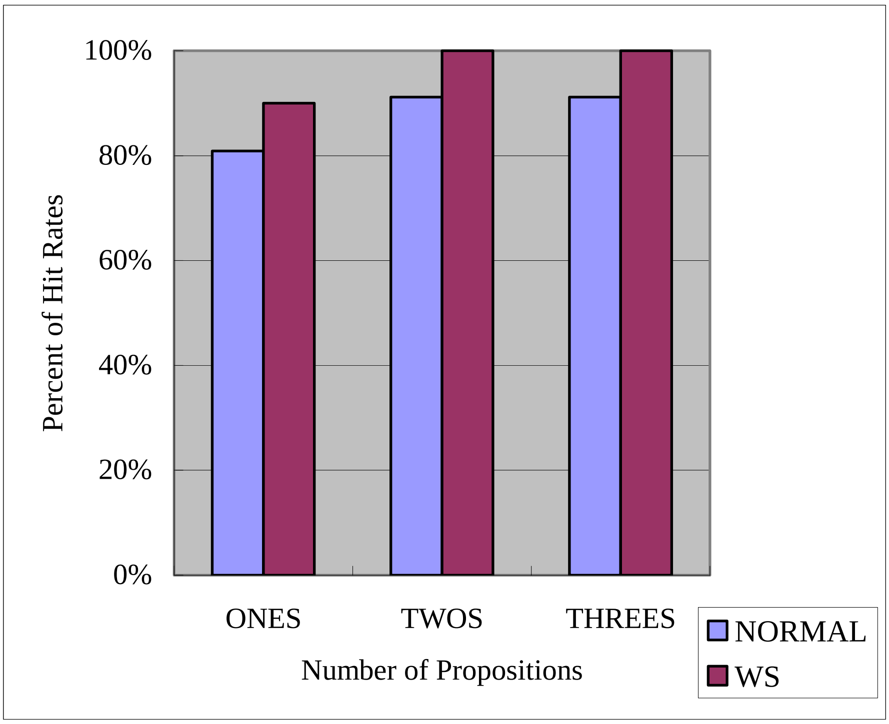


Table 9　Percent and Mean (SD) Raw Scores for Recognition Hit Rates on Old Sentences in Experiment I and II

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)

Figure 8 Percent of Hit Rates in Recognition for Normal People and Participants with Williams Syndrome on Old Sentences



Fourth, the results showed that normal people and participants 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 store them as a wholistic idea in memory. Thus, it seemed that the maximum number of propositions which could be well maintained in memory by normal people is three. It is concluded that if someone could not maintain propositions in memory, then judging coherence would be compromised. Participants with WS assigned high confidence ratings in general and also high false positives for all new sentences such that their ratings were not distinguished by number of propositions. From the data on false positives, for participants with WS, it seems that they can maintain a maximum two propositions in memory. As long as the number of propositions is over two, say three, they cannot maintain them and thus assign significantly higher recognition confidence ratings on them. In other words, the breakdown points for normal people and participants with WS were different. Based on these results, it can be inferred that participants with WS have linguistic ability in integrating propositions which have entailment relations. They can build up mental models based on the events/scenarios presented in discourse, though the ability of maintaining propositions in memory is different from the unimpaired.

It is, however, hard to conclude that participants with WS do not have difficulty in integrating semantically related sentences because there might be a confounding factor. Since recognition confidence rating to new sentences was not observed as a function of propositional complexity in the performance of participants with WS, one might argue that the general high false recognition 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.

6. 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 phase, 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? How they make inferences through listening to fragments of sentences? More relevant to the previous two experiments was a second set of questions: (2) Can participants with WS distinguish scrambled sentences from original 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, and none of them were reported as having medical problems.

Design and Materials

The same superset sentences 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 4). Practice stimuli were also presented before the test experiment (as the stimuli listed in Table 5). In this study, six sentences which contained different constituents from superset sentences in the training phase 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 A and B in the training stage: “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 are listed in Table 10. 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 characters (cf. English stimuli is 13 words if translated).

Table 10 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 & II. All participants were tested in a quiet room in National Tsing Hua University.

Prediction

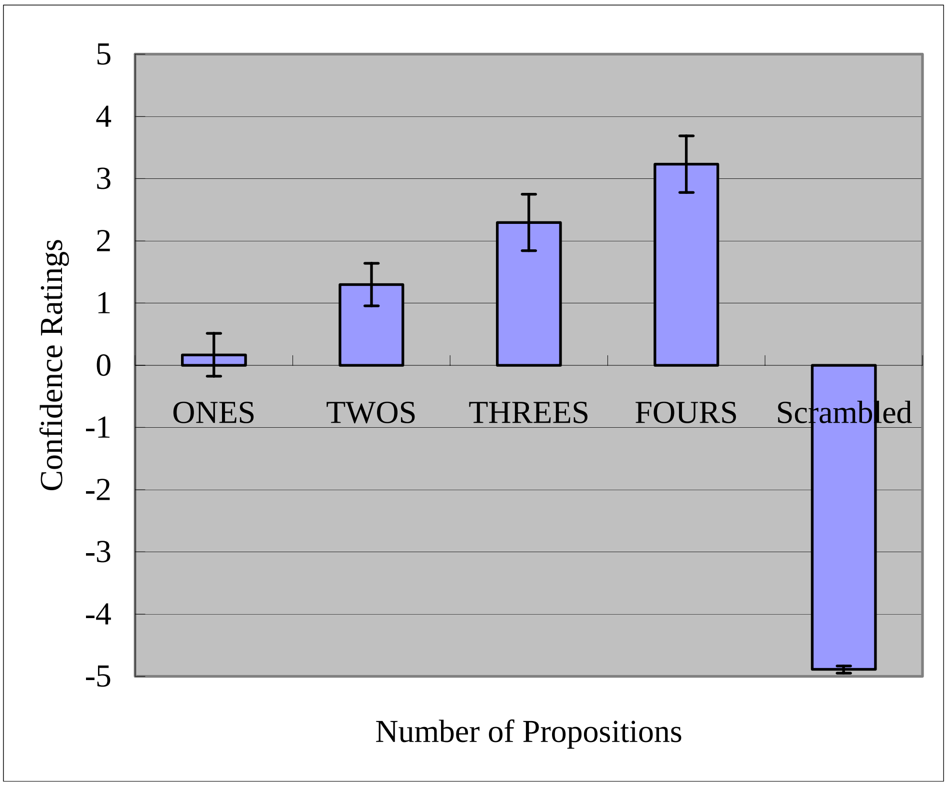
In this study, we hypothesized that participants would return high recognition confidence ratings to all new sentences as in Experiment I. Further, the recognition confidence ratings should be a function of the numbers of propositions in the new sentences. The same pattern was also expected for false positives. Therefore, the greater number of propositions a new sentence contains, the likely 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 of 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 should 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 0.166 (ONES), 1.295 (TWOS), 2.295 (THREES), and 3.231 (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, recognition confidence ratings were 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 the 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, $p < .0368$; 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, $p = .051$ and the difference between THREES and FOURS was not significant, $p = .1149$. 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 9.

Figure 9 Comparison of New and Scrambled Sentences for Normal People



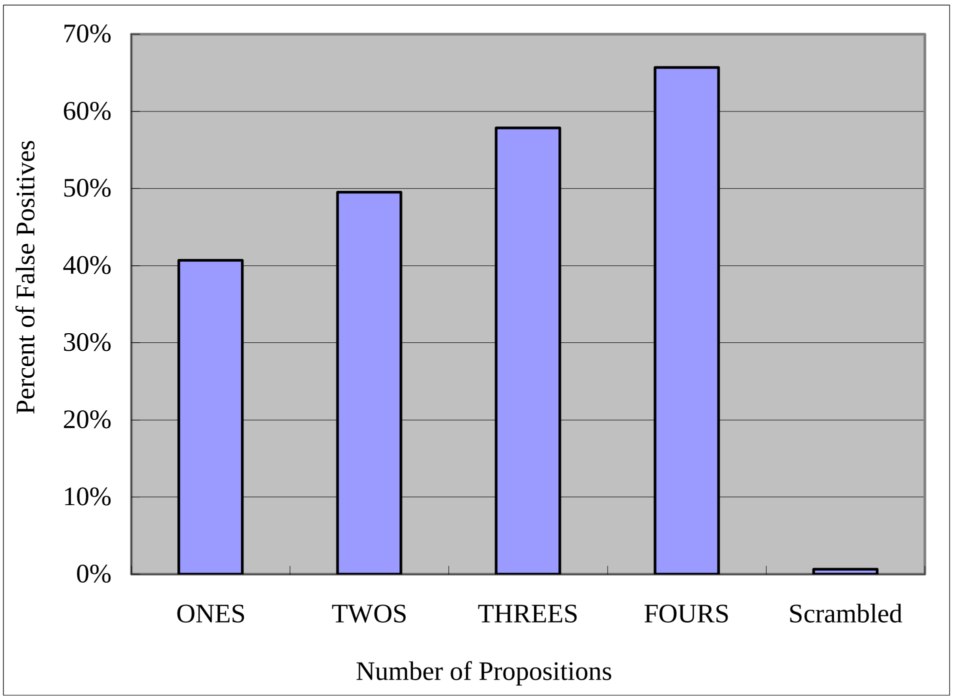
The same pattern was observed for false positives. Proportions of false positives for new sentences and for scrambled sentences are given in Table 11. A one-way ANOVA showed a significant difference between these sentence types both within New sentences and scrambled sentences, $F(4, 281) = 98.10$, $p < .0001$. A post hoc test using the Tukey method showed that scrambled sentences received significantly lower false positive rating across all conditions (all $p < .0001$ in comparisons), and FOURS received a significantly higher false positive rating than all other conditions, almost all $p < .0001$ in comparisons except with THREES ($p < .0015$). The difference between TWOS and THREES was in between. Thus, a clear function of propositional complexity was demonstrated. A plotted graph based on false positives is given in Figure 10.

Table 11 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]

a: There is only one false positive rating shown on scrambled sentences, thus no standard deviation was obtained.

Figure 10 Comparison of Percent of False Positives 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 prior training phase, as they gave very high negative recognition confidence ratings to scrambled sentences in this condition. These results clearly indicated that normal people learned the precise meanings and grammatical relations of the 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 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 easily detect the incoherence between propositions found in scrambled sentences. It is easy for normal people to learn the entailment relations among propositions and to make inferences on them. The next experiment examines the performance of participants with WS on the same materials.

7. Experiment IV: Comparison of New and Scrambled Propositions wit Chinese Participants with Williams Syndrome

Participants

The same study was conducted on participants 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 newly recruited (mean age =17 years 1 months, range from 12yr 3 months to 21yr 8 months; six males). Each participant was diagnosed as having this syndrome with *Fluorescent in situ hybridization* (FISH) test in the hospital or in laboratory prior to the experiment. Those who attended both experiments were tested with a two week interval between experimental sessions. The individual background information is listed in Table 12 below.

Table 12 General Information of WS Participants

	LMH	JYL	GJH	SXY	TSJ	CYJ
Chronological age	17:06	12:08	19:07	21:03	13	18:04
Gender	Male	Male	Male	Male	Male	Male
FIQ (WISC III)	72	---	48	---	---	48
VIQ	84	---	54	---	---	50
PIQ	66	---	50	---	---	53

Design and Materials

The same superset sentences in Experiment III tested on unimpaired individuals were included in this study. All aspects of the design were parallel.

Procedure

The procedure was the same as Experiment III. No verbal cues were given to participants during the experimental session. Nine practice trials were given before the experiment.

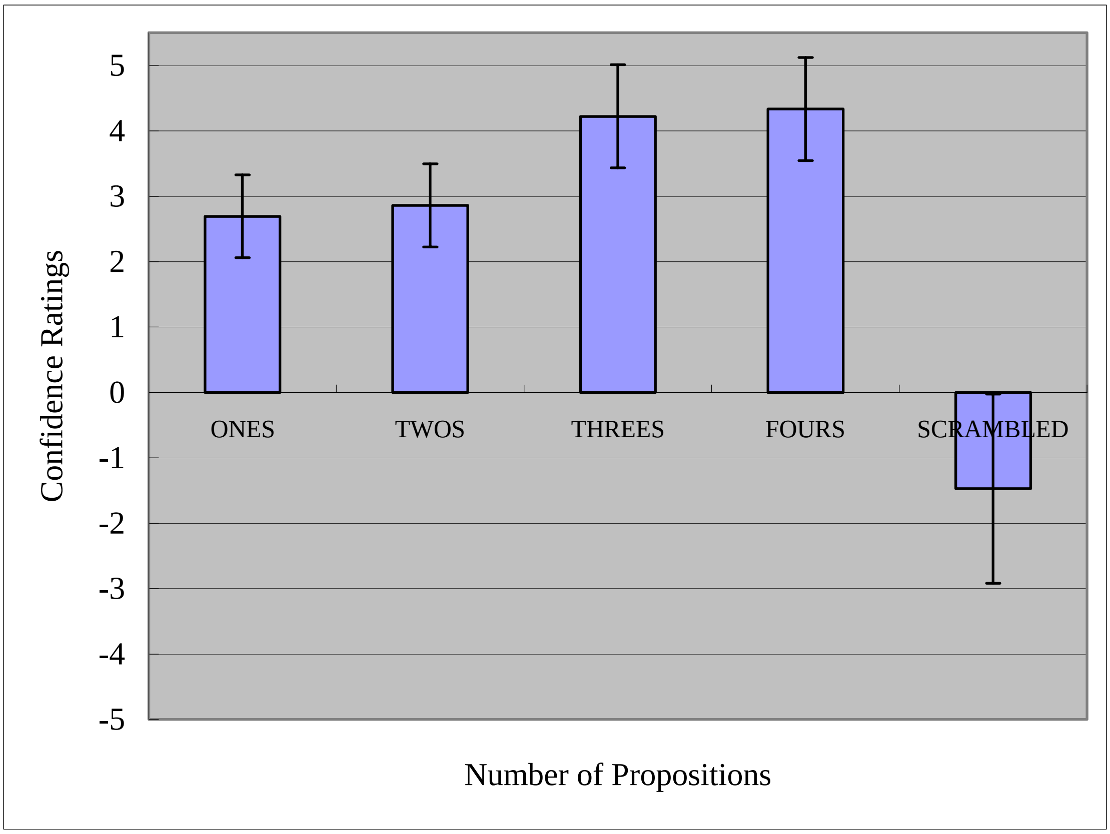
Prediction

Participants 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 participants 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 participants with WS can detect the 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 to seeing whether participants with WS tended to respond yes to all conditions without paying attention to grammatical relations between propositions in sentences. Thus, this experiment can determine whether the high recognition confidence ratings in Experiment II (i.e. comparison between new and old sentences) resulted from the ability to integrate propositions from semantically related sentences or from a yes-bias tendency.

Results and Discussion

The same method was used to convert WS participants' confidence ratings as in the earlier experiments. The averages for new sentences were 2.69 (ONES), 2.86 (TWOS), 4.22 (THREES), and 4.33 (FOURS). A one-way ANOVA showed that the difference between sentences with different number of propositions were marginally significant, $F (3, 99) = 2.33$, $p < .078$. This marginal result was analyzed based on six participants with WS and it was possible that the results would be significantly different if there were more participants with WS participated. As for the comparison of scrambled sentences (-1.47) and new sentences, the difference was significant, $F (4, 134) = 20.37$, $p < .0001$. 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 statistical analysis with the Kruskal-Wallis Test and the Mann-Whitney Test showed the same results. The negative confidence ratings to scrambled sentences reflected spared ability in detecting semantic incoherence for participants with WS. A plotted graph based on confidence ratings is given in Figure 11.

Figure 11 Comparison of New and Scrambled Sentences for Participants with Williams Syndrome

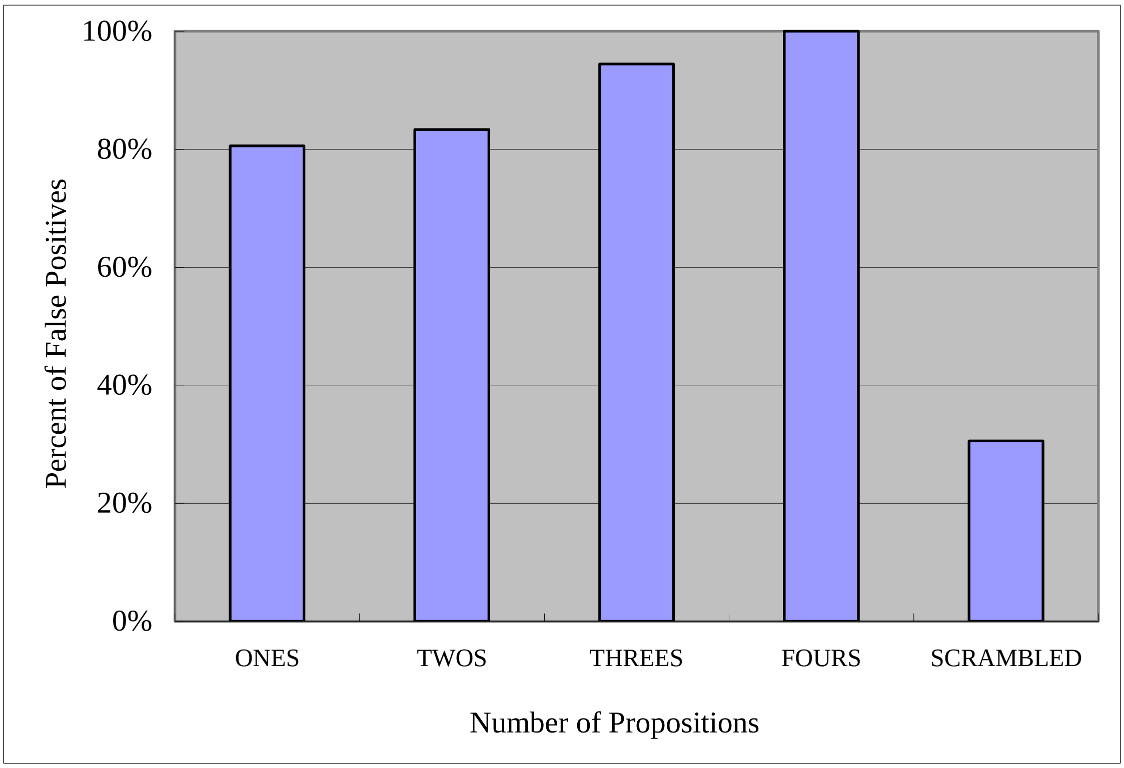


A similar pattern was observed for false positives. The percent of false positives for new sentences and for scrambled sentences is given in Table 13. A one-way ANOVA showed a significant difference between these four sentence types within new sentences, $F(3, 99) = 13.08$, $p < .0001$. A post hoc test using the Tukey method showed that ONES and TWOS received a significantly lower percent of false positives than THREES and FOURS. There was no difference between ONES-TWOS and THREES-FOURS. A clear breakdown point between these two groups was therefore demonstrated. A plotted graph based on false positives is shown in Figure 12.

Table 13 Percent of False Positives and Mean (SD) for New Sentences and for Scrambled Sentences on Participants 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)

Figure 12 Comparison of Percent of False Positives for New Sentences and Scrambled Sentences for Participants with Williams Syndrome

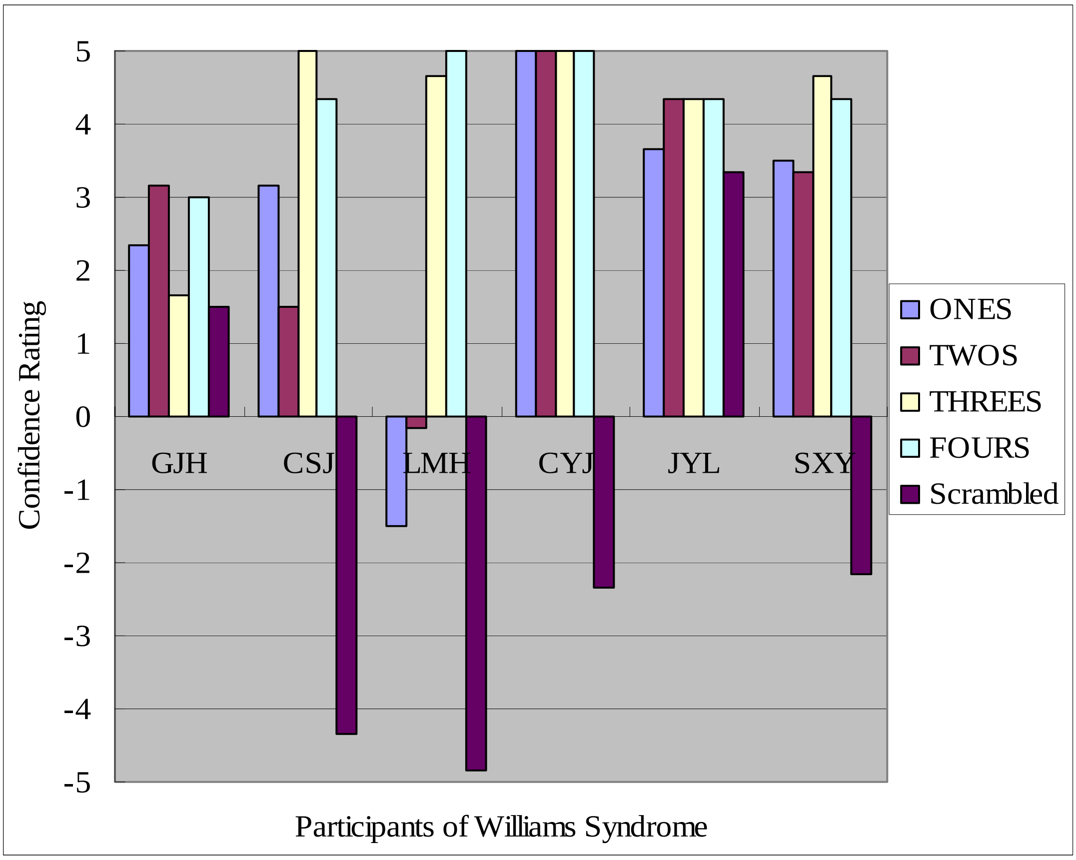


A mean recognition confidence rating for each sentence condition is listed in Table 14. Three participants with WS (CSJ, LMH, and SXY) showed higher confidence ratings on sentences with three propositions (THREES) and four propositions (FOURS) than on sentences with two propositions (TWOS) or one proposition (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 numbers of propositions and a Mann-Whitney Test for paired conditions were employed individually and, in general, the results showed no significant difference (except for 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 participants with WS (all except JYL) showed a significant difference based on the Kruskal-Wallis Test. Among these five children, all the new sentences with different numbers 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 13.

Table 14　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

Figure 13　Individual Ordering for Participants with Williams Syndrome on New and Scrambled Sentences

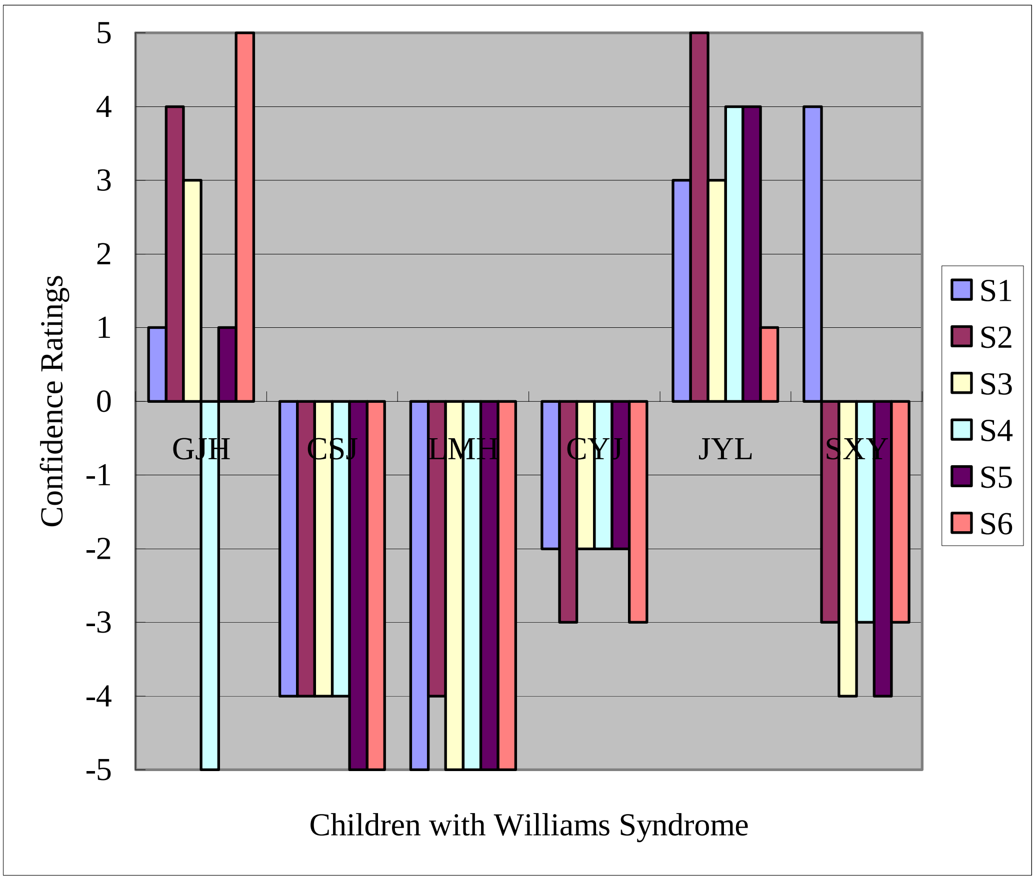


The detailed mean of recognition confidence ratings for each scrambled sentence are given in Table 15. The individual differences were very big for between-participant comparisons. Three participants (CSJ, LMH, and CYJ) assigned negative values to all scrambled sentences; however, one participant (JYL) 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) in contrast to other scrambled sentences with positive ratings. Contrastively, another person with WS (SXY) gave high negative recognition ratings to most of scrambled sentences, but did give positive recognition rating to one scrambled sentence (S1). The graph plotted individually is given in Figure 14.

Table 15 Mean Recognition Confidence Ratings of Each Sentence Stimuli

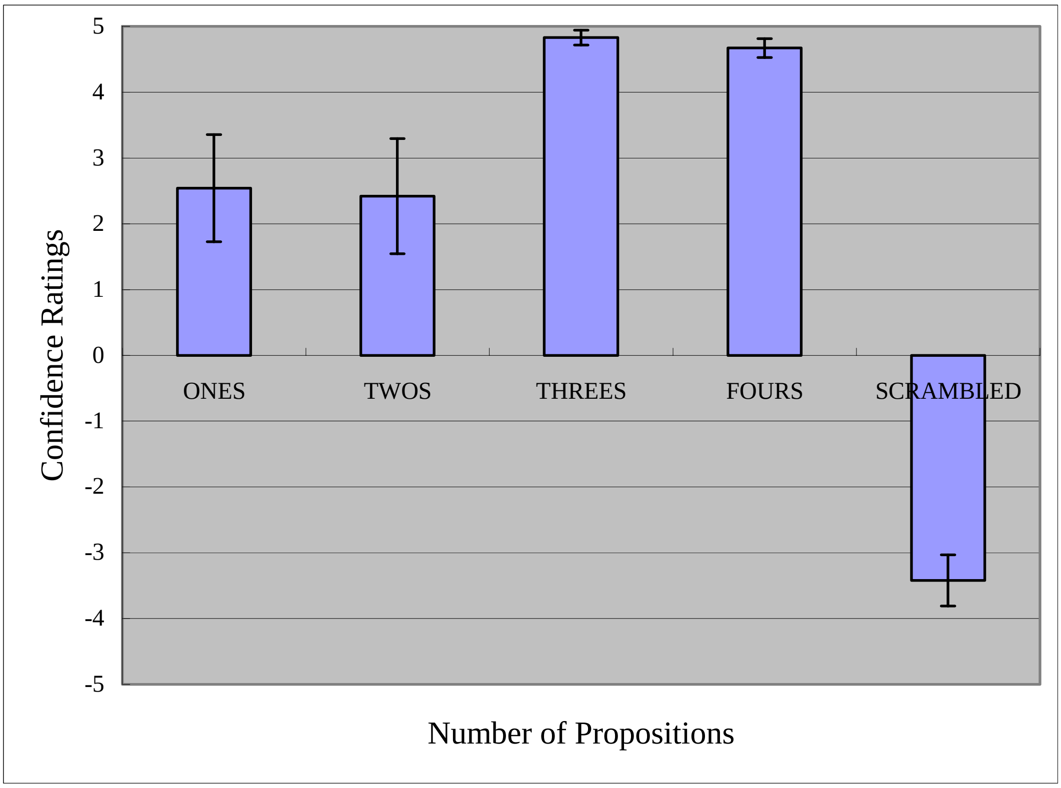
	S1	S2	S3	S4	S5	S6
GJH	1	4	3	-5	1	5
CSJ	-4	-4	-4	-4	-5	-5
LMH	-5	-4	-5	-5	-5	-5
CYJ	-2	-3	-2	-2	-2	-3
JYL	3	5	3	4	4	1
SXY	4	-3	-4	-3	-4	-3

Figure 14 Individual Ordering for Participants with Williams Syndrome on Scrambled Sentences



Since two participants 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 11) 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, $p = .2881$). A one-way ANOVA showed that the difference between new sentences and scrambled sentences showed a significant difference, $F(4, 91) = 22.791$, $p < .000$. The difference resulted mainly from the comparison between scrambled sentences and ONES, TWOS, THREES, and FOURS. Again, these results demonstrated spared linguistic ability on participants 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 in participants with WS. Thus, it can be concluded that participants with WS spontaneously integrated semantic propositions given the contexts in discourse just as normal people do.

Figure 15 Comparison of New and Scrambled Sentences for Participants with Williams Syndrome (without GJH & JYL)



General Discussion

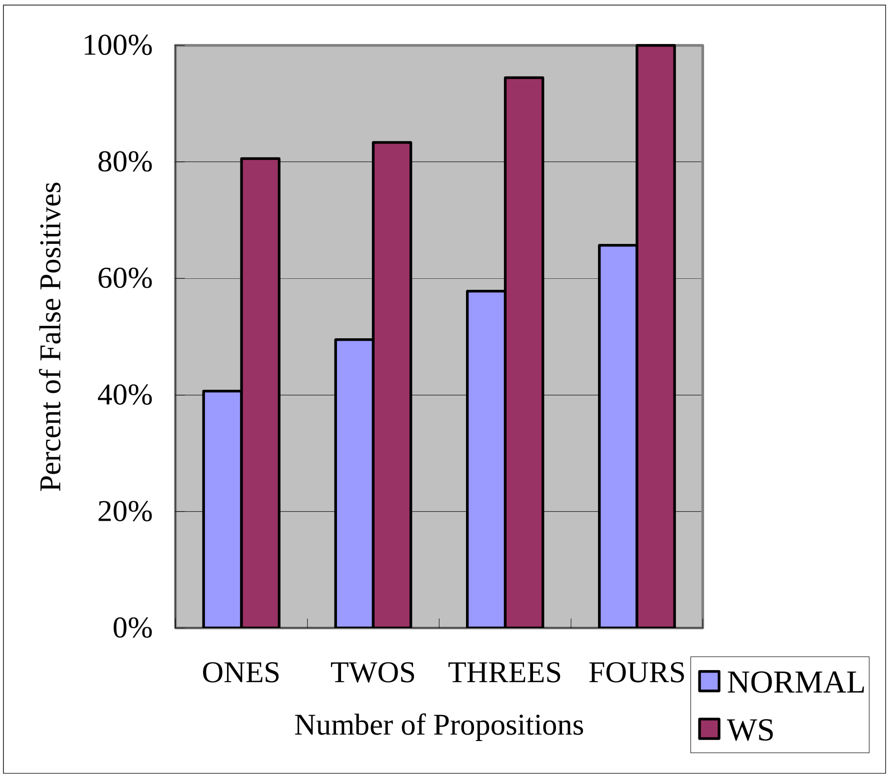
Participants with WS showed a higher proportion of false positive recognition ratings to all new sentences and scrambled sentences than normal people in this study (two-way ANOVA, $F(4, 562) = 14.79$, $p < .0001$). A clear difference between the percentage of false positives in recognition for normal people and participants with WS was quite obvious, and is represented in Table 16. 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, $p = 0.013$, between normal people and participants with WS. Further, sentences combined freely with propositions from different supersets 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 greater numbers of propositions contained in a sentence, the higher percent of false positives assigned; participants 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 propositional complexity, and successful detection of incoherence in scrambled sentences. Thus, we conclude that participants with WS do process grammatical relations of propositions of sentences in discourse. Finally, the significantly lower confidence ratings and false positives for scrambled sentences in the performance of participants with WS showed that their pattern of responses did not result from a yes-bias.

Table 16 Percent and Mean (SD) Raw Scores for Recognition False Positive 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 obtained.

Figure 16 Percent of False Positives in Recognition for Participants with Williams Syndrome on New Sentences



8. General Discussion

Where does encoding of semantic integration break down? The breakdown point of semantic integration was different for normal young and adolescents with WS in both experiments. In Experiment I, when new sentences were compared with old sentences, normal participants showed a clear new-old effect, suggesting that they could distinguish sentences which were actually presented to a certain degree. However, these same subjects 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 had not been presented in the training phase with highly negative recognition confidence ratings. As to new sentences, normal participants could not distinguish sentences with more than three propositions and misrecognized them as having been previously heard. Thus, it could be concluded from this experiment that the unimpaired can maintain at most nearly three propositions rather than four.

Why did the boundary shift for normal people to new sentences in experiments? 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 phase before and recognized them with high confidence ratings or hit rates. Under this circumstance, compared with old sentences which had actually been presented, normal people did not show as 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 mental model. Under this circumstance, compared to scrambled sentences which were combined from different supersets, 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 boundary shift in participants with WS. In Experiment II, when old sentences were lumped together as lure in recognition, participants with WS assigned in general high recognition confidence ratings to all new sentences without showing any distinguished recognition point between them (though there does show significance different on false positive). In other words, putting old sentences in recognition biased participants to assign positive values across the board. In contrast, in Experiment IV, when scrambled sentences were included in recognition, participants with WS showed sensitivity to new sentences and assigned lower recognition confidence ratings. That is, lumping scrambled sentences in recognition resulted in the change in sensitivity. Thus, participants with WS showed a distinguishing recognition boundary in the comparison of new and scrambled sentences in both recognition rates and false positives.

Did participants with WS integrate propositions semantically given the contexts in mental model? The answer to this question is positive. The supporting evidence comes from two observations. In the comparison of new and old sentences (Experiment II), participants with WS performed different false positives to new sentences having a distinguishing recognition point between TWOS and THREES. However, in recognition confidence ratings, there is no difference to all new sentences. It seems that participants with WS are prone to respond positively to all new sentences (i.e. having the yes-bias tendency). Against this yes-bias tendency, in the comparison of new and scrambled sentences, the results show that participants with WS correctly reject the scrambled sentences. This correct rejection observation suggests that participants with WS do make confidence ratings based on mental models built from sentence fragments. In other words, they infer the entailment relations embedded in propositions successfully. But, since the breakdown points for participants with WS are highly consistent (i.e. between TWOS and THREES) in experiment II and IV, it seems that they could only maintain two propositions in memory rather than three propositions as the unimpaired. It suggests that participants with WS do integrate propositions semantically in the mental models during the presentation of sentence fragments with more limited capacity in memory.

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 sentence fragments. It could be argued that the particular propositions were learned. Since each superset sentence was combined with four simple propositions (i.e. ONES), which expressed an event, participants learned each proposition rather than integrating the related propositions as a holistic semantic idea in the mental model. Thus, the more propositions a sentence includes, the easier it could be for participants to recognize them. This would be reflected in recognition confidence ratings as a function of number of propositions. That is, participants kept propositions in memory analytically (i.e. the analytic feature hypothesis). We argue against this alternative because all the sentences in recognition, including old sentences and scrambled sentences, were composed of the same propositions as in new sentence. We observed that participants show new-old effect and perform correct rejection to scrambled sentences. Thus, we conclude that participants with WS and the unimpaired build mental models based on the sentence fragments to form a wholistic representation in semantics (i.e. the wholistic representation hypothesis).

Bransford and Franks (1972) conducted a study with constrained sentences and unconstrained sentences to investigate the possibility of the analytic feature hypothesis. The atomic propositions composing these two types of sentences, constrained or unconstrained, were the same. There were eight totally different propositions used as 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 propositional relations could be found in sentences. However, unconstrained sentences were sentences with irregular relations between propositions, which meant that there was no fixed scenario as the propositions could be combined freely. The same recognition paradigm was carried out. Under the analytic feature hypothesis, it is 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 violations. On the other hand, under the wholistic 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 for both constrained and unconstrained sentences. The same finding was replicated when scrambled sentences were lumped together in recognition. Thus, it could be concluded that interrelation between propositions in sentences was learned and represented in memory instead of feature (i.e. particular proposition) 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 phase. 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 expressing a particular linguistic idea. One of the sentences in each block actually occurred in one of the paragraphs in the training phase. Participants were asked to recognize which particular sentences were presented in the paragraphs they had heard before. An effect of the agent was not predicted if sentence length was not the factor in recognition, because the additional piece of information would matter only if people would spontaneously integrate semantically related propositions. That is, short passives without agents would be recognized equally to short passives, which agents were expressed in other sentences. The results showed that these two types of short passives without agents and short passives with additional sentence expressing the agent had different recognition rates. Short passives without agents received higher recognition rates than short passives with an additional sentence following. Thus, it looked like that people spontaneously integrated the additional propositions into the short passives as a wholistic 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 solely. The factor which really mattered and was demonstrated in the present study was the number of propositions, or pieces of information, contained in a sentence. Thus, we believe that the ordering of the results (ONES < TWOS< THREES < FOURS) observed in accord with number of propositions is really a function of sentence complexity.

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 ‘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. To sum up, abstractness and concreteness do not influence participants’ representation of sentences in memory.

9. Conclusion

The aim of this study was to investigate whether individuals with WS have a selective impairment in memory for sentence meaning relative to sentence form. We hypothesized that participants 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 et al., 1997), invented objects naming (Karmiloff-Smith et al., 1997) and also studies of grammatical structures like relative clauses (Zukowski, 2001; Grant, Valian, & Karmiloff-Smith, 2002). In these previous studies, participants with WS showed 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 verbal working memory on participants with WS (Wang and Bellugi, 1994; Vicari, 1996; Jarrold et al., 1999; Robinson et al., 2003). Based on these findings, the ability of people with WS to integrate entailment relations across sentences was investigated. The performance of WS individuals was partially similar to the pattern shown on normal college students, thus suggesting that the gap of the knowledge of phrase structures and the comprehension of sentences might not exist on this genetic deficit population. In other words, participants with WS do show certain degree of proposition integration in semantics. That is, WS individuals also build mental models as the unimpaired controls. In accord with these partially similar neuropsychological patterns of WS individuals to normal controls, suggesting that WS individuals are developmental delay, but not deviant in nature.

APPENDIX 1 COMPREHENSION QUESTIONS FOR EXPERIMENTAL STIMULI OF SUPERSET A

Superset 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?
Superset 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?
Superset 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?
Superset 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?
Superset 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?
Superset 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 2 COMPREHENSION QUESTIONS FOR EXPERIMENTAL STIMULI OF SUPERSET B

Superset 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?
Superset 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?
Superset 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?
Superset 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?
Superset 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?
Superset 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 3 COMPREHENSION QUESTIONS FOR EXPERIMENTAL STIMULI OF SUPERSET C

Superset 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?
Superset 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?
Superset 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?
Superset 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?
Superset 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?
Superset 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 4 COMPREHENSION QUESTIONS FOR EXPERIMENTAL STIMULI OF SUPERSET D

Superset 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?
Superset 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?
Superset 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 5 COMPREHENSION QUESTIONS FOR PRACTICE STIMULI OF SUPERSET E

Superset 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?
Superset 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?
Superset 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 6 COMPREHENSION QUESTIONS FOR PRACTICE STIMULI OF SUPERSET F

Superset 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?
Superset 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?
Superset 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?

REFERENCES

Arwood, E. 1991. Semantics and Pragmatics Language Disorders. Maryland: An Aspern Publication.

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.

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.

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.

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.

Gernsbacher, M. 1990. Language Comprehension as Structure Building. UK: Lawrence Erlbaum Associates, Publishers.

Hulme, C., and Mackenzie, S. 1992. Working Memory and Severe Learning Difficulties. UK: Lawrence Erlbaum Associates, Publishers.

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.

Johnson-Laird, P. 1983. Mental Models. Cambridge: Harvard University Press.

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.

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.

Kintsch, W. 1998. Comprehension: A Paradigm for Cognition. Cambridge University 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.*

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 Participants with Williams Syndrome. Cortex 39.255-271.

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.

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.

Philips, C. 2004. Linguistics and Linking Problems. pp. 1-41. In Developmental 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 Participants 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.

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 Participants 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 Participants with Williams Syndrome. Cortex 32.503-514.

Vicari, S., Bellucci, S., and Carlesimo, G.A. 2001. Procedural Learning Deficit in Participants with Williams Syndrome. Neuropsychologia 39.665-677.

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.

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.

Zukowski, A. Uncovering Grammatical Competence in Participants with Williams Syndrome. 2001. Boston University, Graduate School of Arts and Sciences. Ref Type: Thesis/Dissertation.