# The development of metaphorical language comprehension in typical development and in Williams syndrome

**Journal Article**

# The development of metaphorical language comprehension in typical development and in Williams syndrome

**Abstract**

The domain of figurative language comprehension was used to probe the developmental relation between language and cognition in typically developing individuals and individuals with Williams syndrome. Extending the work of Vosniadou and Ortony (1983), the emergence of non-literal similarity and category knowledge was investigated in 117 typically developing children aged between 4 and 12, 19 typically developing adults, 15 children with Williams syndrome between 5 and 12 years of age, and 8 adults with Williams syndrome. Participants were required to complete similarity and categorization statements by selecting one of two words (e.g., either "The sun is like...?" or "The sun is the same kind of thing as...?") with word pairs formed from items that were literally, perceptually, or functionally similar to the target word, or else anomalous (e.g., 'moon', 'orange', 'oven', or 'chair', respectively). Results indicated that individuals with Williams syndrome may access different, less abstract knowledge in figurative language comparisons, despite the relatively strong verbal abilities found in this disorder.

Although metaphor and analogy have traditionally been viewed as a relatively rare linguistic ornament that complement literal language, recent research suggests that metaphorical language is, in fact, commonplace in everyday communication (e.g., Graesser, Long & Mio, 1989; Pollio, Barlow, Fine & Pollio, 1977). For example, various strands of linguistic evidence suggest that metaphor is important for communicating, and perhaps reasoning, about abstract concepts (e.g. Gibbs, 1994; Lakoff & Johnson, 1980). Understanding metaphorical language necessitates certain degrees of proficiency in both cognition and language, and relies on several component abilities, such as processing capacity, metalinguistic skill, an understanding of communicative pragmatics and semantic knowledge (see Vosniadou, 1987a, 1987b).

Although definitions of what constitutes a metaphorical comparison vary considerably (e.g., Bowdle & Gentner, 2005), it seems uncontroversial that understanding verbal metaphor involves accessing some similarity between two terms, while recognizing that two terms belong to different conventional categories. This, of course, necessitates knowledge of such categories. If so, then one can take an initial step in ascertaining whether children are capable of understanding metaphor at a given stage in development by investigating their ability to understand non-literal similarity statements, that is, that items that fall in different semantic or conceptual categories can nevertheless be similar in certain ways. Vosniadou and Ortony (1983) required children between the ages of 3 and 6, as well as adults, to complete similarity statements by selecting one of two words from either (1) a metaphorical/literal pair of alternatives, (2) a literal/anomalous pair, or (3), a metaphorical/anomalous pair. For example, the experimenter would say “A river is like a...?” and the participant was required to respond with either *snake* or *lake* in (1), *lake* or *cat* in (2), and *snake* or *cat* in (3). The authors assumed that selecting metaphorical or literal responses over anomalous ones justified attributing the ability to distinguish meaningful similarity statements from anomalous ones, and that selecting a literal response over a metaphorical one indicated that the terms of the literal comparison were understood as more similar than the terms in the metaphorical comparison.

Even the youngest children were able to distinguish meaningful similarity statements from anomalous ones.

In addition to this comparison task, which involved perceiving some similarity between two terms, a further group of participants undertook a categorization task, which tested for knowledge of conventional categories. This was identical in format to the comparison task, except that the experimenter would say (e.g.) “A river is the same kind of thing as...?” and there was no metaphorical/anomalous condition. In the categorization task, literal responses were clearly correct responses and metaphorical responses incorrect (a river is the same kind of thing as a lake, but not a snake). Children aged 4-years old and above showed that they possessed the conventional categories that were assumed to be transgressed in the metaphorical juxtapositions of the comparison task.

The current study adapted and extended the above paradigm, with two main aims. The first was to investigate the relative emergence of perceptual and functional non-literal similarity in typically developing children. The second was to explore the development of requisites for metaphor comprehension in Williams syndrome. Functional similarity can be defined as ‘a correspondence based on what things do’ (e.g., the Sun is like an oven because they both produce heat). There is robust evidence that young children find metaphorical comparisons based on physical or perceptual similarity easier to understand than those based on abstract and complex relations (e.g., Billow, 1975; Gentner & Stuart, 1983), but it is not yet clear when this difference emerges, or whether this is the case for individuals with Williams syndrome.

Gentner (1988) proposed that a ‘relational shift’ occurs during development (at around 6-7 years old) whereby children interpret metaphorical comparisons first in terms of object similarity (i.e., attributional/perceptual similarity) and then in terms of relational similarity. However, Goswami (1996) has argued against the relational shift hypothesis, citing a study by Goswami and Brown (1989) in which children as young as 3-years old were able to correctly complete pictorial analogies based on familiar causal relations (though see Ratterman & Gentner, 1998). Goswami

argued that as soon as relational knowledge is acquired, it can be used by children in comprehending relational comparisons. Setting aside the nuances of this debate, there is ample evidence that semantic knowledge has an effect on metaphor comprehension (e.g., Keil, 1986). Because individuals with WS appear to use figurative language that they do not understand, a specific hypothesis that the current study will address is the possibility that the WS sample will struggle to understand non-literal similarity, even if the requisite knowledge is in place.

Williams syndrome is of particular relevance to studying figurative language because it is a developmental disorder in which aspects of language development are in advance of the general level of cognitive functioning (e.g., Bellugi, Birhle, Jernigan, Trauner, & Doherty, 1990; Brock, Jarrold, Farran, & Laws, 2006). Despite overall IQs falling with the 50-70 range, productive language has been reported to include rare or low frequency vocabulary in some individuals (see Thomas, Dockrell, Messer, Parmigiani, Ansari & Karmiloff-Smith, 2006, for a review) and also to include figurative language (e.g., Bertrand, Mervis, Armstrong & Ayers, 1994). To what extent does this output reflect proper understanding of figurative language? Given the processing of metaphorical comparisons is intimately related to conceptual development, how does this play out in a disorder showing uneven development of language and more general cognitive abilities?

One intriguing aspect of language in WS is the unusual words that are reportedly used by people with the disorder (Bellugi *et al*., 1990; Bellugi, Wang & Jernigan, 1994; Udwin & Dennis, 1995). Thomas and colleagues (Thomas *et al*., 2006) have suggested that use of low frequency words is a linguistic social engagement device, since people with WS have been described as having a "hypersocial" personality profile (Gosch & Pankau, 1997; Howlin *et al*., 1998; Jones, Bellugi, Lai, Chiles, Reilly, Lincoln & Adolphs, 2000) and have been reported to make frequent use of pragmatic conversational devices to engage speakers' attention (Reilly, Losh, Bellugi & Wulfeck, 2004). There are examples of people with WS using various parts of language, such as clichés, idioms and figurative language that can nevertheless appear somewhat inappropriate to the conversational context (Bertrand, Mervis, Armstrong & Ayers, 1994; Udwin & Yule, 1990). This contextual

inappropriateness may reflect a poor underlying knowledge of the meaning of such language. This important notion is supported by the findings that individuals with WS appear to have great difficulty in explaining the meanings of proverbs and metaphors, with attempts focusing on surface features (Bertrand *et al*., 1994), and also have great trouble distinguishing lies from jokes, tending simply to recount events when required to justify their interpretation (Sullivan, Winner & Tager-Flusberg, 2003). It should be noted, however, that it is difficult to distinguish shortcomings that relate to limited semantic and conceptual knowledge from weakness in metacognitive abilities, i.e., offering explanations and justifications.

One possibility is that the figurative language used by individuals with WS is 'frozen', that is to say that each figurative phrase is invariant and produced from memory rather than 'on-line' processes (Annaz, van Herwegen, Thomas, Fishman, Karmiloff-Smith, & Runbland, in press). If people with WS use frozen language that they do not, in fact, properly understand and which is not representative of their cognitive abilities, it may lead to problems for teachers, parents and caregivers attempting to communicate effectively with people with the disorder. Moreover, use of frozen language may lead to inappropriate judgments of language abilities in WS, in turn leading to people with WS facing language that they do not understand. It seems possible that the social difficulties experienced by children and adolescents with WS (e.g., Gosch & Pankau, 1997) may, in part, be caused or exacerbated by difficulties in using language appropriately, and by failing to understand language used by others. The current study, therefore, additionally provides a way of investigating whether people with WS are possessed of semantic knowledge proportionate to their apparent linguistic competence at a given point in development. Specifically, the current study indirectly addressed whether it might be possible for individuals with WS to have relatively advanced figurative language without the normal underlying conceptual knowledge, as a form of phrased vocabulary.

For the current study, a developmental trajectories approach was adopted (see Thomas, Annaz, Ansari, Scerif, Jarrold & Karmiloff-Smith, in press), in which functions of task performance

and age are constructed, thereby allowing developmental change to be compared across typically and atypically developing groups. Although, ideally, longitudinal methods would be used to reveal such patterns, cross-sectional studies can give an initial approximation of developmental trajectories, which can subsequently be validated by longitudinal investigations. Trajectories that link performance to measures of mental age can be used to ascertain whether any such performance difference is in line with the developmental state of other measures of cognition, that is, to reveal the developmental relations that exist within disorders that show uneven cognitive profiles.

*Method*

In addition to the perceptual metaphorical comparisons used by Vosniadou and Ortony (1983), functional metaphorical comparisons were also included (e.g., “Eyes are like a camera”, “The Sun is like an oven”), to target the differential emergence of functional and perceptual non-literal similarity and so tap a more sophisticated level of similarity to be accessed in non-literal comparisons. Although ‘metaphorical’ is the term used by Vosniadou and Ortony (1983), henceforth we use the terms ‘non-literally similar’ or ‘non-literal similarity’, to avoid confusion with the various definitions of ‘metaphor’. There are several advantages to using the Vosniadou and Ortony (1983) paradigm in the current context: first, unlike many ways of investigating figurative competence, the task has neither metalinguistic nor metacognitive aspects (i.e., understanding of language and cognition) that could prevent participants from successfully demonstrating understanding of non-literal similarity and semantic categories. This is important with respect to the participants with WS, because such individuals appear to have particular difficulty with metalinguistic and metacognitive tasks (Bertrand *et al*., 1994; Sullivan *et al*., 2003), although it is not clear whether this issue reflects pragmatic difficulty (Brock, 2007), information processing limitations (*cf*. Vosniadou, 1987b), or some other factor. Whatever the case, elimination of such task demands would also enable younger children to demonstrate competences more central to the task design.

*Participants*

There were two experimental groups: 136 typically developing (TD) individuals and 23 individuals with a clinical diagnosis of WS, confirmed by the fluorescence in situ hybridization (FISH) test. Of these 23 individuals with WS, 15 were children recruited through the Williams Syndrome Foundation UK; the other 8 participants with WS were adults approached via the Williams Syndrome Association (US). US and UK participants took part in the same experiment. The typically developing group included 117 children and 19 adults; the children were recruited through local schools, the adults were undergraduate students. The WS sample size, although relatively large compared with many other studies of WS (given the rareness of the disorder), was much smaller than the TD sample. The large TD sample served to robustly investigate developmental patterns in typical development, so that any developmental atypicality of the WS group could be seen with reference to that sample.

We took the verbal mental age of our TD participants to be the same as their chronological age. An alternative approach would have been to assess the language ability of each TD child on the standardized vocabulary test and then assign him or her a mental age. However, this would only serve to reference one sample of typically developing children (our sample of 117) to the performance of another sample of typically developing children (those used to norm the standardized vocabulary test). Given the relatively large size of our TD sample, we elected to avoid this additional transform and simply take the variation of language ability observed in our TD sample as representative of typical development.

In a preliminary test session, the WS group was assessed on the British Picture Vocabulary Scale II (BPVS; Dunn, Dunn, Whetton, & Burley, 1997), a measure of receptive vocabulary, and its US equivalent, the Peabody Picture Vocabulary Test III (PPVT; Dunn & Dunn, 1997). Raw scores on these tests were converted into mental age equivalents, to gain an estimate of verbal mental age. Individuals with WS were additionally assessed on the Pattern Construction sub-test of the British Abilities Scales (BAS; Elliot *et al*., 1996) and the US equivalent, the Pattern Construction sub-test of the Differential Ability Scales (DAS; Elliot, 1990). The general population mean for the BPVS is 100 (SD: 15) whereas the general population mean for the British Abilities Scales subtests is 50 (SD: 10).

The standard scores of the UK and US samples on vocabulary and visuospatial tests did not reliably differ (mean and [standard deviation] for vocabulary standard score: UK = 80.2 [19.1], US = 71.3 [20.4], $t(21) = 1.02$, $p = .318$; visuospatial standard score: UK = 22.5 [6.3], US = 20.1 [.4], $t(21) = 1.07$, $p = .295$). Therefore the two groups were equally representative of the WS cognitive profile. Participant details are given in Table 1.

*======== insert Table 1 about here =========*

*Materials*

The materials were based on those used by Vosniadou and Ortony (1983). Nine concrete nouns were used as target words for the TD group. Where there was any US-UK difference in materials (i.e., tap/faucet, doughnut/donut), the culturally appropriate stimulus was used. As the relatively large number of trials generated by these nine words was not viable with the WS group, five words were selected for comparison and only data for these words are reported here. Each target word was matched with a series of comparison terms, related literally (L), perceptually (P), functionally (F), or anomalously (A). See Table 2 for a list of the stimulus words used in the experiment. The stimulus words were printed onto small laminated labels (measuring approximately 4x4cm) in capital letters. Supporting pictures were in the form of color photographs, selected to represent unambiguously each of the stimulus words used in the experiment. The materials for the pre-test were a small round stone (target), a small rubber ball (perceptual match), and a stone that was not round (literal match).

*======== insert Table 2 about here =========*

*Procedure*

Prior to testing, a pre-test was administered using the materials described at the end of the above subsection, in order to ensure that participants understood the difference between comparison and categorization that was central to the task. Participants were then presented with a single practice trial of the categorization task, using the practice stimuli Globe (target), Pin (A) and Map (L). There was also a single practice trial of the comparison task using the same target, but Book (A) and Hot air balloon (P) as response options. Children that failed the pre-test would have been excluded from participating in the rest of the experiment, but none did.

There were two tasks, each of which involved a two-alternative forced choice. In the comparison task, participants were required to complete sentences of the form "X is like...?" with one of a pair of words, where A was one of the target words listed in Table 2 and the two response options were drawn from each target word's set of four related words (L, P, F and A). Participants were presented with all possible pairwise response options over the course of testing (L/A, L/P, L/F, P/A, F/A, P/F, i.e., 6 possible pairs for each of the 5 targets), such that there was a total of 30 trials (6x5). Order was blocked by target item but randomized within block; order of targets was also randomized.

The categorization task was similar, except that sentences of the form "X is the same kind of thing as...?" were to be completed. Literal matches were always one of the response options on each trial (L/A, L/P, L/F), so the fully factorial set of trials summed to 15 (3x5).

On each trial, the sentence (i.e., "X is the same kind of thing as...?") was read aloud by the experimenter and the participant chose a response from nouns printed onto laminated paper. This response selection obviated concerns about the original paradigm, where a participant might perform poorly simply because he or she has poor short-term memory (*cf*. Purser & Jarrold, in press). For the participants with WS, all trials were administered with pictorial support for each noun, in order to ensure that the terms of each trial were understood. Half of the TD group was given pictorial support; the other half was not, and this split was even across the chronological age range. The reason for this was to be able to test the effect of pictorial support on patterns of

responding, to ensure that pictures did not bias participants to choose perceptual responses. For all participants, the comparison and categorization tasks were administered in blocks, categorization first and then comparison.

During testing, participants were asked to justify their comparison and categorization responses. Roughly, a justification was requested every 5$^{th}$ or 6$^{th}$ trial, in order to collect sufficient data for analysis purposes without substantially prolonging testing time. The randomization of the comparison/categorization order meant that, in effect, justifications were randomized. Justifications were coded as perceptual (i.e., reflecting shared perceptual properties), structural-functional (*i.e.,* reflecting non-perceptual properties, where the terms are viewed as the same kind of thing or do similar things), narrative-linked (*i.e.,* the participant described a series of actions or a scenario that reflected the response made), or unclear. Narrative-linked justifications of experimental tasks are associated with Williams syndrome, often describing what happened when performing the task in question (Bertrand, Mervis, Armstrong & Ayers, 1994). Perceptual and structural-functional justifications were specifically recorded because these reflected the kinds of similarity investigated in the study. Interrater reliability was 89% agreement, based on 6.25% of responses.

*Results*

In this section, there is first a short chronological age based analysis, followed by a more in-depth analysis based on verbal mental age. . The latter first deals with the categorization task alone, then categorization and comparison tasks considered together, and finally the comparison task in isolation. The categorization task is considered separately at first because understanding non-literal similarity rests on knowledge of literal categories (something is only non-literal with reference to the literal); considering the categorization and comparison tasks together may reduce the ability to find reliable interactions in the categorization task. There were no reliable differences between TD performance in *word* or *word+picture* conditions (all $F < 1$), so both were included in constructing the TD trajectories in the verbal mental age -based analysis. In addition, all the below analyses,

where they involved perceptually-similar items, were repeated excluding TD participants who had not received picture support – there were no changes in the pattern of reliable and non-reliable results.

The aim of the current study was to focus on developmental aspects of non-literal similarity, prompting the use of a developmental trajectories approach to study change with age. In order to perform this type of analysis, participant groups must have reasonably even distributions of ages. While this was true of mental age (as assessed by receptive vocabulary) in the WS group, with test ages spanning quite evenly from 6;1 to 17;6 years, it was not for chronological age, for which there was a more bimodal distribution of younger and older groups. Constructing trajectories in this case would simply connect the mean performance of the younger group with that of the older group. Therefore, in order to assess whether the abilities of the WS group were in line with chronological age, a group-matching approach was adopted. The chronological age analysis explored performance on the categorization task, because performance on this task can provide direct evidence for the development of semantic categories (while the comparison task would provide evidence of understanding similarity, which would not necessarily entail category knowledge). Owing to the bimodal age ranges of the participants, two analyses were conducted, for younger and older participants.

*CA analysis*

A sample of 15 TD participants was matched to the younger WS group on chronological age, $t(28) = .078$, $p = .938$. Data from this TD sample and the WS sample were analysed using a two-way mixed-design ANOVA, with Contrast (*literal-anomalous, literal-perceptual, literal-functional*) as the within-subjects factor and Group (*TD, WS*) as the between-subjects factor. There was a reliable main effect of Contrast, $F(2, 56) = 32.106$, $p < .001$, $\eta_p^2 = .534$ and a reliable interaction of Group and Contrast (see Figure 1), $F(2, 56) = 3.796$, $p < .05$, $\eta_p^2 = .119$. *Post-hoc* analysis of simple effects showed no reliable group differences on *literal-perceptual*, $F(1, 28) = 0.951$, $p = .338$, $\eta_p^2 = .033$, or *literal-*

*functional*, $F(1, 28) = 0.434$, $p = .516$, $\eta_p^2 = .015$, contrasts. However, the TD group made significantly more literal responses in *literal-anomalous* contrasts than the WS group, $F(1, 28) = 4.321$, $p < .05$, $\eta_p^2 = .134$, indicating that the WS group were not showing a CA-appropriate pattern of performance. If presented with a categorization such as "The moon is the same kind of thing as a star/shoe", participants with WS were less likely to respond with 'star' and more likely to respond with 'shoe', compared to the TD group, indicating that the young participants with WS were able to demonstrate less knowledge of literal categories than the young TD participants.

The older group of participants with WS was matched on chronological age to eight TD individuals, $t(14) = 0.003$, $p = .998$. A similar ANOVA to that described above was performed. Again, there was a significant main effect of Contrast, $F(2, 28) = 11.114$, $p < .001$, $\eta_p^2 = .443$ and a reliable interaction of Group and Contrast (see Figure 1), $F(2, 28) = 7.878$, $p < .01$, $\eta_p^2 = .360$. *Post-hoc* analysis of simple effects showed no significant group differences on *literal-perceptual*, $F(1, 14) = 0.336$, $p = .571$, $\eta_p^2 = .023$, or *literal-anomalous*, $F(1, 14) = 1.00$, $p = .334$, $\eta_p^2 = .067$, contrasts. However, the TD group made reliably fewer literal responses in *literal-functional* contrasts than the WS group, $F(1, 14) = 16.004$, $p < .01$, $\eta_p^2 = .533$, indicating that when presented with a categorization such as "The sun is the same kind of thing as the moon/an oven", participants with WS were less likely to respond with 'oven' and more likely to respond with 'moon' compared to the TD group. This, again, showed that the WS group had not responded in the pattern predicted by their chronological age. It should be noted, however, that the post-hoc ANOVAs used here were not Bonferroni-corrected, because we believe that such correction is too conservative, given the limited power of studies involving rare developmental disorders. However, we do not base any of our subsequent conclusions on these results.

*======== insert Figure 1 about here =========*

*Verbal mental age-based analysis*

Although there was a bimodal distribution of chronological age, verbal mental age was more continuously distributed, indicating that a trajectory design was appropriate to explore developmental relations of the experimental variables (see Thomas *et al*., in press). It should be noted that the verbal mental age ranges of the TD and WS groups do not overlap exactly. However, verbal mental age was rescaled to count in months from the youngest age measured according to the method given at http://www.psyc.bbk.ac.uk/research/DNL/stats/Rescaling_age2.html (i.e., subtracting the youngest age measured from every age). Initial analyses were undertaken, restricted to the categorization task, in order to investigate whether participants possessed category knowledge necessary for metaphorical understanding. These analyses were one-way ANCOVAs, with Group (*TD, WS*) as the between-subjects factor. Each subsequent analysis is a repeated-measures ANCOVA, with Task (*comparison, categorization*) as the within-subjects factor, and Group (*TD, WS*) as the between-subjects factor. The ANCOVA model included interaction terms between the VMA covariate and the between-subjects factors to explore whether the rate of development differed between groups. Simple effects of task were independent of the covariate of VMA, so these results are reported from an analysis that excludes the covariate. (Degrees of freedom may therefore differ between simple task effects and group effects or interactions.)

*Categorization task*

Summary: Developmental trajectories for the categorization task are depicted in Figure 2. The two groups showed a similar pattern of results for the literal perceptual contrast, indicating that the WS group showed typical emergence of conventional category knowledge. The WS group was also less likely than the TD group to class functionally similar items as the same kind of thing as the target word in the experimental sentences.

*Literal-perceptual contrast*: There was a greater number of literal responses with increasing mental age, in line with the emergence of semantic structure, $F(1,155) = 45.406$, $p < .001$, $\eta_p^2 = .227$, but there was no reliable interaction of VMA and group, $F(1,155) = 0.169$, $p = .682$, $\eta_p^2 = .001$.

*Literal-functional contrast*: VMA was a reliable predictor of literal responses, $F(1,155) = 17.847$, $p < .001$, $\eta_p^2 = .103$, and the WS group showed a steeper rise in literal responses with increasing VMA than the TD group, $F(1,155) = 5.641$, $p < .05$, $\eta_p^2 = .035$. This shows that, with increasing VMA, the WS group was less likely than the TD group to categorize functionally similar items as the same kind of thing as the target.

*Literal-anomalous contrast*: Again, the analysis showed that VMA was a reliable predictor of literal responses, $F(1,155) = 35.661$, $p < .001$, $\eta_p^2 = .187$, and there was a reliable interaction of VMA and group, $F(1,155) = 7.989$, $p < .01$, $\eta_p^2 = .049$, indicating a faster developmental increase in literal responses for the WS group than for TD group. A reliable group difference was revealed by the ANCOVA, $F(1,155) = 21.491$, $p < .001$, $\eta_p^2 = .122$, because the TD group made more literal responses than the WS group. In other words, the WS group was delayed given their level of receptive vocabulary and the TD group was at ceiling.

*======== insert Figure 2 about here =========*

*Categorization and comparison tasks*

Summary: Developmental trajectories for the categorization and comparison tasks are included in Figures 2-4. Both groups showed an emergence of understanding of non-literal similarity and the WS group showed a steeper relationship than the TD group between verbal mental age and literal responses in the literal-functional contrast.

*Literal-perceptual contrast*: There was an interaction of verbal mental age and task, $F(1,155) = 18.898$, $p < .001$, $\eta_p^2 = .109$, reflecting a faster developmental increase in literal responses in the categorization task than the comparison task. This is a key pattern that reflects the emergence of non-literal similarity: the comparison task was more likely to elicit non-literal choices. There was also a main effect of task, $F(1,157) = 13.269$, $p < .001$, $\eta_p^2 = .078$, due to a greater proportion of literal responses in the categorization task than in the comparison task, reflecting participants' understanding of non-literal similarity.

*Literal-functional contrast*: There was a reliable interaction of verbal mental age and group, $F(1,155) = 5.258$, $p < .05$, $\eta_p^2 = .033$, reflecting a faster developmental increase in literal responses for the WS group than for the TD group. The interaction of verbal mental age and task was also significant, $F(1,155) = 6.972$, $p < .01$, $\eta_p^2 = .043$, signifying a faster developmental increase in literal responses in the categorization task than the comparison task. The ANCOVA revealed no reliable group difference, $F(1,155) = 0.097$, $p = .756$, $\eta_p^2 = .001$, but there was a main effect of task, $F(1,157) = 19.458$, $p < .001$, $\eta_p^2 = .110$, due to a greater proportion of literal responses in the categorization task than the comparison task, again reflecting participants' understanding of non-literal similarity.

*Literal-anomalous contrast*: There was a reliable interaction of verbal mental age and group, $F(1,155) = 13.444$, $p < .001$, $\eta_p^2 = .080$, signifying a faster developmental increase in literal responses for the WS group than for TD group.

*Comparison task*

Summary: There were three contrasts that only occurred in the comparison task, which have particular relevance for the 'relational shift' debate. These reflect which comparisons participants preferred when neither option was literally similar. (For example, participants might be asked "Which is more like the sun, an oven or an orange?"). The relevant developmental trajectories are

shown in Figure 3. The TD group showed an emergent preference for functional, over perceptual, similarity with increasing verbal ability, but the WS group did not. Both groups showed improved recognition of perceptual and functional non-literal similarity with increasing verbal mental age.

*Perceptual-functional contrast*: This contrast and those that follow featured only in the comparison task, so there are no task effects to report. The interaction of group and verbal mental age was significant, $F(1,155) = 9.805$, $p < .01$, $\eta^2 = .059$. When the WS group was analysed separately, there was no reliable effect of verbal mental age on proportion of functional responses, $F(1,21) = 1.834$, $p = .190$, $\eta_p^2 = .080$. Considering the TD group alone, however, there was a significant *decrease* in the proportion of perceptual responses with increasing verbal mental age, $F(1,134) = 21.869$, $p < .001$, $\eta_p^2 = .140$, shown in Figure 5. Thus, the TD group increasingly preferred functional similarity with increasing verbal ability (e.g., for “sun”, preferring “oven” over “orange”), while the WS group did not.

*======== insert Figure 3 about here =========*

*Perceptual-anomalous contrast:* The interaction of group and verbal mental age was also significant, $F(1,155) = 6.797$, $p = .01$, $\eta_p^2 = .042$, reflecting a more rapid developmental increase in perceptual responses in the WS group. The TD group demonstrated a shallower increase because they were approaching ceiling performance (see Figure 6). There was a significant main effect of group, $F(1,155) = 21.307$, $p < .001$, $\eta_p^2 = .121$, indicating a delayed onset for the WS group relative to the TD group. In other words, the WS group’s ability to recognise perceptual non-literal similarity was delayed relative to the control group, given their verbal ability, but eventually reached the TD level because that group was at ceiling.

*Functional-anomalous contrast*: verbal mental age was a significant predictor of functional responses, $F(1,155) = 27.867$, $p < .001$, $\eta_p^2 = .152$, which increased with verbal mental age, but the interaction of group and verbal mental age was not reliable, $F(1,155) = 0.290$, $p = .519$, $\eta_p^2 = .002$. The ANCOVA revealed no reliable group difference, $F(1,155) = 2.586$, $p = .110$, $\eta_p^2 = .016$. As Figure 7 shows, both groups indicated similar developmental trajectories in recognizing functional non-literal similarity compared to anomaly.

*Justifications*

For the sake of brevity, justification data are not reported, but were broadly consistent with the rest of the data collected.

*Discussion*

There were two main aims of the experiment: the first was to investigate the relative emergence of perceptual and functional non-literal similarity in typically-developing children. The second was to explore the development of non-literal similarity in Williams syndrome, to gain insight on the interface of language and cognition. The results indicated that an understanding of both these types of non-literal similarity emerge similarly in typical development and WS. However, while TD children show an emergent functional preference over perceptual (object) similarity, such a preference does not emerge for individuals with WS, even though the requisite functional knowledge appears to be in place. The results also suggested that, at least in relation to vocabulary ability, individuals with WS appear to have a good understanding of both perceptual and functional non-literal similarity.

There are two key strands of evidence for the emergence of functional non-literal similarity. First, the trajectories for the *functional-anomalous* contrast of the comparison task showed that both groups tended to make a greater proportion of functional responses as verbal mental age increased. This relationship was very similar for both groups, suggesting that individuals with WS

develop an understanding of functional similarity in the normal way. Second, there is evidence from the *literal-functional* contrast. The WS group showed a similar pattern to that seen in the *literal-perceptual* contrast, often selecting functional responses in the comparison task, despite acquiring literal category knowledge with increasing verbal mental age. The TD data are different, however: although there is clear evidence of acquisition of literal category knowledge from the *literal-perceptual* contrast, with the *literal-functional* contrast the TD group showed no sign of favouring literal categorizations over functional ones with increasing mental age. The CA-analysis, too, showed that TD adults favoured literal categorizations over functional ones less than the adults with WS. This suggests that the conceptual structure of the TD children becomes organised on the basis of functional relations, in addition to literal ones, with advancing mental age. This point will be revisited in more detail later.

Critically, the TD group showed an increasing preference for functionally-related responses with verbal mental age in the *perceptual-functional* condition of the comparison task. It is important to bear in mind that the TD trajectory for the *functional-anomalous* condition of the comparison task indicates that even the younger TD children tended to demonstrate some degree of understanding of functional similarity. Therefore, it is an emergent functional preference that becomes apparent in typical development, rather than simply developing an understanding of functional relations *per se*.

This trend for an increasing preference for functionally-similar items is entirely absent for the WS group, however. If anything, the WS group showed a larger preference for perceptually-similar items in the comparison task with increasing verbal mental age, although this trend was not significant ($p$ = .190). A trivial explanation for the WS group's lack of emerging functional preference in the *perceptual-functional* contrast could be that the participants with WS received pictorial support in the experiment, whereas only half of the TD group did. In this way, the WS group could have been biased towards selecting perceptually-similar responses. However, this seems unlikely because pictorial support had no discernible effect on responding for the TD group in the *perceptual-functional* comparisons (effect size $\eta_p^2$ = .003). Moreover, omitting TD participants who had received

pictorial support from the analyses did not change the pattern of reliable and unreliable results. Furthermore, the WS group clearly showed an increase in preference for literal responses over perceptual ones in the categorization task (which was very similar to that of the TD group) suggesting that pictorial support did not bias responding for the participants with WS either.

Due to the finding that people with WS clearly do learn about functional similarity, this group difference in functional responding, even given receptive vocabulary, cannot be due to a lack of understanding of functional relations in the WS group. It could be that functional similarity becomes increasingly salient for TD children as they grow older, becoming similarly salient to literal similarity, while functional relations do not acquire such salience for people with WS. The trajectories for the *literal-functional* condition of the categorization task lend support to this notion, with the TD group giving similar proportions of literal and functional responses even as literal category knowledge is acquired with increasing mental age, while the WS group show only an increasing proportion of literal responses. An alternative explanation for this group difference is in terms of pragmatics. It could be that the TD group were attempting to make the exchange between participant and experimenter explanatorily useful, in a similar way to how teachers might explain novel ideas by making functional connections (see Harrison & Coll, 2008) while the WS group made no such attempt. Consistent with this explanation, children with WS have been reported to be impaired on all five of the pragmatic subtests of the Children's Communication Checklist (CCC; Bishop, 1998), including 'conversational context' and 'development of conversational rapport' (Laws & Bishop, 2004). It is these exactly these kinds of pragmatic skills that are key to the accurate deployment of figurative language. Moreover, children with WS are less likely than control participants to demonstrate inferences about the mental states and motivations of conversational partners (Reilly, Losh, Bellugi, & Wulfeck, 2004), in line with the suggestion that a lack of conversational engagement underlies the absence of emergent functional preference in WS.

A further possibility, raised earlier, is that the TD group's emergent functional preference reflects a restructuring of conceptual knowledge. This is not to say that concepts become organised

on the basis of functional relations instead of literal ones, just that knowledge may be restructured to *include* functional relations. Recently, Purser and colleagues (Purser, Thomas, Snoxall, & Mareschal, submitted) investigated the development of lexical-semantic knowledge in WS and typical development, using a categorization task. The task involved participants grouping toy animals according to questions asked, including functional relations, e.g., “Which can lay eggs?”, “Which can swim well?” The domain of animals was chosen to maximise the chances of success for the participants with WS, because it has been shown that individuals with WS as young as 10-years-old have comparable basic knowledge of this domain, relative to TD controls matched on verbal mental age (Johnson & Carey, 1998). TD participants made fewer categorization errors with increasing verbal mental age, presumably reflecting increasingly well-defined category boundaries. However no such trend was evident for the WS group, consistent with the suggestion that individuals with WS do not develop well-delineated conceptual structure.

Johnson and Carey (1998) investigated intuitive biological knowledge in WS and two TD groups, one matched on verbal mental age and the other a non-matched younger group. Biological knowledge was tested using two batteries based on a proposed distinction of (1) general knowledge in-line with the range of concepts available to typically developing preschool children and (2) folk-biological concepts thought to be acquired between the ages of six and twelve, requiring conceptual change for their construction. The WS group performed similarly to the verbal mental age -matched group on (1), indicating biological general knowledge appropriate for verbal mental age, but performed significantly worse on (2) and at a similar level to the younger children. The authors interpreted this pattern of results as evidence that the WS group had not acquired folk-biological concepts appropriate for verbal mental age, even though the requisite general knowledge was probably in place. It is possible, then, that the emergent functional preference seen in the TD (but not the WS) group in the current study reflects a similar restructuring of knowledge into functional categories. It should be noted that the types of knowledge addressed by Johnson and Carey are quite different from those addressed in the current study. However, the specific parallel with

Johnson and Carey's results is that the WS group appears to have the requisite knowledge on which to base functional categories, but is unable to demonstrate such functionally-organised concepts in making similarity judgments. Taken together, Johnson and Carey's (1998) results, along with those of Purser and colleagues (submitted) and the current study, suggest that individuals with WS might have relatively advanced figurative language without the normal underlying conceptual knowledge. Clearly this suggestion is based on indirect evidence and would be a fruitful topic of future research.

It is necessary to account for why people with WS struggle with the meanings of metaphors (Bertrand *et al.*, 1994), even though the WS group in the current study demonstrated a good understanding of both perceptual and functional non-literal similarity. As explained in the introduction, such understanding is but one necessary part of real-world metaphor comprehension. Another is to interpret the intention of the speaker, recognizing that a statement is not intended literally. In a possible parallel to the WS group's lack of emergent functional preference, Sullivan *et al*. (2003) found that even when individuals with WS demonstrated an understanding of second-order theory of mind and all the requisite knowledge for successful task performance, they were not able to use that information to judge whether statements were lies or jokes. When asked to account for their judgments, the participants with WS tended to focus on surface features of the task, e.g., appealing to the narrative of the task. Bertrand *et al*. (1994) found a similar surface-level focus when they asked people with WS to explain the meaning of proverbs and metaphors.

Future work could include investigation of whether pragmatic issues underlie the apparent lack of an emergent functional preference in WS. One way of examining this would be to use a similar paradigm to that employed in the current study, modified to involve puppet characters. In one condition, participants would be instructed that a puppet character was particularly interested in perceptual similarity. In another, the instruction would be that a different puppet character had a particular interest in functional similarity. The task would be to tell the experimenter how each character would respond. If, in the latter condition, participants with WS tended to select functional over perceptual responses in the comparison task, it would suggest that pragmatic difficulties do not

account for the absence of emergent functional preference. In addition, future research could determine whether the pattern of responding of the WS group seen in the current study is specific to that etiology, or to developmental delay more generally.

Returning to the main findings of the current study, the data reported here shed new light on the 'relational shift' debate: the WS group clearly showed a developing understanding of functional non-literal similarity, but showed no shift from preferring perceptual similarity to preferring functional (relational) similarity. The TD group, however, demonstrated just such an emergent functional preference. One other key group difference was that the TD group showed evidence of grouping concepts on the basis of functional similarity, in addition to literal similarity. It seems reasonable to suggest, on the basis of the data reported here, that such a reorganisation of conceptual information underlies the observed shift towards preferring functional similarity over perceptual similarity.

Since the WS group appeared not to develop a conceptual structure based on functional relations to the same extent, the evidence from the current study are consistent with the notion that where individuals with WS use figurative language, they may do so without fully understanding its meaning (cf. Bertrand *et al.*, 1994). In this way, superficial behaviour may not always reflect normally developing underlying processes. In this case, the hypersocial personality profile of people with WS, coupled with a good ability to learn vocabulary, overcomes weaker cognitive abilities to give the appearance of 'advanced' behaviour. Thus, although on the surface, language and cognition appear to diverge in WS, closer inspection favours a tight relationship between the two for the normal emergence of understanding figurative language.

**Acknowledgements**

We would like to thank the Williams Syndrome Foundation UK and the Williams Syndrome Association, and the children and adults with WS who participated in this study. We are grateful to the children and teachers of Manor Field Infant and Junior School, Basingstoke, UK, and to Claire Alford, Liz Daniel, and Klarika Vrublevskaja Floyd for their help with data collection. This research was supported by British Academy grant SG30602, MRC grant G0300188, and European Commission grant NEST-029088(ANALOGY).

**References**

Anderson, M. (1998). Mental retardation general intelligence and modularity. *Learning and individual differences, 10*, 159-178.

Annaz, D., Van Herwegen, J., Thomas, M. S. C., Fishman, R., Karmiloff-Smith, A., & Runbland, G. (in press). The comprehension of metaphor and metonymy in children with Williams syndrome. *International Journal of Language and Communication Disorders.*

Ansari, D., Donlan, C., Thomas, M. S. C., Ewing, S. A., Peen, T., & Karmiloff-Smith, A. (2003). What makes counting count? Verbal and visuo-spatial contributions to typical and atypical number development. *Journal of Experimental Child Psychology, 85*, 50-62.

Bellugi, U., Bihrle, A., Jernigan, T., Trauner, D., & Doherty, S. (1990). Neuropsychological, neurological, and neuroanatomical profile of Williams syndrome. *American Journal of Medical Genetics Supplement, 6*, 115–125.

Bellugi, U., Lichtenberger, L., Jones, W., Lai, Z., & St. George, M. (2000). The neurocognitive profile of Williams syndrome: A complex pattern of strengths and weaknesses. *Journal of Cognitive Neuroscience, 12:S*, 7-29.

Bellugi, U., Marks, S., Bihrle, A., & Sabo, H. (1988). Dissociation between language and cognitive function in Williams Syndrome. In D. Bishop & K. Mogford (Eds.), *Language development in exceptional circumstances*. Edinburgh, U.K.: Churchill Livingstone.

Bellugi, U., Wang, P. P., & Jernigan, T. L. (1994). Williams syndrome: An unusual neuropsychological profile. In S. Broman & J. Grafman (Eds.), *Atypical cognitive deficits in developmental disorders: Implications for brain function*. Hillsdale, NJ: Erlbaum.

Bertrand, J., Mervis, C. B., Armstrong, S. C., Ayers, J. (1994). *Figurative language and cognitive abilities of adults with Williams syndrome*. Paper presented at the 1994 National Williams Syndrome Professional Conference, La Jolla, CA.

Billow, R. M. (1981). Observing spontaneous metaphor in children. *Journal of Experimental Child Psychology, 31*, 430-445.

Bishop, D. V. M. (1998). Development of the Children's Communication Checklist (CCC): a method for assessing the qualitative aspects of communicative impairment in children. *Journal of Child Psychology and Psychiatry, 39*, 879-891.

Bowdle, B., & Gentner, D. (2005). The career of metaphor. *Psychological Review, 112*, 193-216.

Brock, J. (2007). Language abilities in Williams syndrome: a critical review. *Development and Psychopathology, 19*, 97-127.

Brock, J., Jarrold, C., Farran, E. K., Laws, G., & Riby, D. M. (2007). Do children with Williams syndrome really have good vocabulary knowledge? Methods for comparing cognitive and linguistic abilities in developmental disorders. *Clinical Linguistics & Phonetics, 21*, 673-688.

Donnai, D., & Karmiloff-Smith, A. (2000). Williams syndrome: From genotype through to the cognitive phenotype. *American Journal of Medical Genetics, 97*, 164-171.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test-Third Edition*. Bloomington, MN: Pearson Assessments

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *British Picture Vocabulary Scale II*. Windsor, U.K.: NFER-Nelson.

Elliot, C. D. (1990). *Differential Ability Scales*. New York: The Psychological Corporation.

Elliot, C.D., Smith, P., & McCulloch, K. (1996). *British Ability Scales*. NFER-Nelson, Windsor, UK.

Farran, E. K., & Jarrold, C. (2003). Visuo-spatial cognition in Williams syndrome: Reviewing and accounting for strengths and weaknesses in performance. *Developmental Neuropsychology, 23*, 173-200.

Gentner, D. (1988). *Metaphor as structure mapping: The relational shift. Child Development, 59*, 47-59.

Gentner, D., & Bowdle, B. F. (2001). Convention, form, and figurative language processing. *Metaphor and Symbol, 16*, 223-247.

Gentner, D., & Stuart. P. (1983) *Metaphor as Structure Mapping: What Develops*. BBN Technical Report No. 5479. Cambridge, MA: Bolt, Berankek and Newman

Gibbs, R. 1994. *The Poetics of Mind*. Cambridge: Cambridge University Press.

Glucksberg, S. & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review, 97*, 3-18.

Gosch, A. & Pankau, R. (1997). Personality characteristics and behaviour problems in individuals of different ages with Williams syndrome. *Developmental Medicine and Child Neurology, 39*, 527-533.

Goswami, U. (1996). Analogical reasoning and cognitive development. In H. Reese (Ed.), *Advances in child development and behaviour*, (pp. 91-138). New York: Academic Press.

Goswami, U. & Brown, A. L. (1989). Melting chocolate and melting snowmen: analogical reasoning and causal relations. *Cognition, 35*, 69-95.

Graesser, A. C., Long, D., & Mio, J. S. (1989). What are the cognitive and conceptual components of humorous texts? *Poetics, 18*, 143-164.

Harrison, A. G. & Coll, R. K. (2008). *Using analogies in middle and secondary science classrooms: The FAR guide--An interesting way to teach with analogies*. CA: Corwin Press.

Howlin, P., Davies, M., & Udwin, O. (1998). Cognitive functioning in adults with Williams syndrome. *Journal of Child Psychology and Psychiatry, 39*, 183-189.

Johnson, S. C. & Carey, S. (1998). Knowledge enrichment and conceptual change in folkbiology: evidence from Williams syndrome. *Cognitive Psychology*, 37, 156-200.

Jones, W., Bellugi, U., Lai, Z., Chiles, M., Reilly, J., Lincoln, A., & Adophs, R. (2000). Hypersociability in Williams syndrome. *Journal of Cognitive Neuroscience, 12*, 30-46.

Keil, F. C. (1986). Conceptual domains and the acquisition of metaphor. *Cognitive Development, 1*, 73-96.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press

Pollio, H.R., Barlow, J.M., Fine, H.J. & Pollio, M.R. (1977). *Psychology and the Poetics of Growth*. New York: Lawrence Erlbaum.

Purser, H. R. M. & Jarrold, C. (in press) Short- and long-term memory contributions to immediate serial recognition: Evidence from serial position effects. *Quarterly Journal of Experimental Psychology*.

Purser, H. R. M., Thomas, M. S. C., Snoxall, S., Mareschal, D., & Karmiloff-Smith, A. (submitted). Definitions versus categorisation: Assessing the development of lexico-semantic knowledge in Williams syndrome.

Ratterman, M. J., & Gentner, D. (1998). The effect of language on similarity: The use of relational labels improves young children's performance in a mapping task. In K. Holyoak, D. Gentner & B. Kokinov (Eds). *Advances in Analogy Research: Integration of theory and data from the cognitive, computational and neural sciences*. Bulgaria: New Bulgarian University.

Reilly, J., Losh, M., Bellugi, U., & Wulfeck, B. (2004). "Frog, where are you?" Narratives in children with specific language impairment, early focal brain injury, and Williams syndrome. *Brain and Language, 88*, 229-247.

Strømme, P., Bjørnstad, P.G., Ramstad, K. (2002). Prevalence estimation of Williams syndrome. *Journal of Child Neurology, 17*, 269-271.

Sullivan, K., Winner, E., & Tager-Flusberg, H. (2003). Can adolescents with Williams syndrome tell the difference between lies and jokes? *Developmental Neuropsychology, 23*, 85-103.

Thomas, M.S.C., Annaz, D., Ansari, D., Scerif, G., Jarrold, C., & Karmiloff-Smith., A. (in press). Using developmental trajectories to understand developmental disorders. *Journal of Speech, Language, and Hearing Research.*

Thomas, M. S. C., Dockrell, J. E., Messer, D., Parmigiani, C., Ansari, D., & Karmiloff-Smith, A. (2006). Speeded naming, frequency and the development of the lexicon in Williams syndrome. *Language and Cognitive Processes, 21*, 721-759.

Thomas, M. S. C. & Mareschal, D. (2001). Metaphor as categorisation: A connectionist implementation. *Metaphor and Symbol, 16*, 5-27.

Udwin, O. & Dennis, J. (1995). Psychological and behavioural phenotypes in genetically determined syndromes: A review of research findings. In G. O'Brien and W. Yule (Eds.) *Behavioural Phenotypes. Clinics in Developmental Medicine* No. 138. London: MacKeith Press.

Udwin, O., & Yule, W. (1990). Expressive language of children with Williams syndrome. *American Journal of Medical Genetics Supplement, 6*, 108-114.

Vosniadou, S. (1987). Contextual and linguistic factors in children's comprehension of nonliteral language. *Metaphor and Symbolic Activity, 2*, 1-11.

Vosniadou, S. (1987). Children and metaphors. *Child Development, 58*, 870-885.

Vosniadou, S. & Ortony, A. (1983). The emergence of the literal-metaphorical-anomalous distinction in young children. *Child Development, 54*, 154-161.

**Tables**

Table 1

*Participant characteristics (WS = Williams syndrome, TD = typically developing, CA = chronological age, VMA = verbal mental age, M = male, F = female). CA was used as VMA for the young TD participants; the maximum VMA for the BPVS (17;6) was used for the TD adults.*

| | Mean CA | Range | Mean VMA | Range |
|---|---|---|---|---|
| WS (young, N=15 (9 M, 6 F)) | 7;5 | 5;5-11;6 | 6;1 | 2;11-7;5 |
| WS (adult, N=8 (4 M, 4 F)) | 31;3 | 15;1-51;0 | 11;10 | 8;1-17;6 |
| TD (young, N=117 (59 M, 58 F)) | 8;4 | 4;8-11;3 | | |
| TD (adult, N=19 (9 M, 10 F)) | 26;4 | 18;3-39;2 | | |

Table 2

*List of stimuli*

| Target word | Literal match | Perceptual match | Functional match | Anomalous match |
|---|---|---|---|---|
| Moon | Star | Coin | Candle | Shoe |
| Elastic band | Parcel string | Donut | Trouser belt | Faucet |
| Sun | Moon | Orange | Oven | Chair |
| Eyes | Ears | Buttons | Camera | Wall |
| Nose | Mouth | Mountain | Trumpet | Bed |

**Figures**

Figure 1. Group differences in literal responses in the categorization task. WS = Williams syndrome, TD = typically developing. Error bars depict standard errors

Figure 2. Proportion of literal responses across verbal mental age in categorization and comparison tasks. WS = Williams syndrome, TD = typically developing.

Figure 2a. Literal/perceptual contrast

Figure 2b.: Literal/functional contrast

Figure 2c. Literal/anomalous contrast

Figure 3. Proportion of perceptual responses across verbal mental age in the comparison task: Perceptual/functional contrast. WS = Williams syndrome, TD = typically developing.

Figure 1

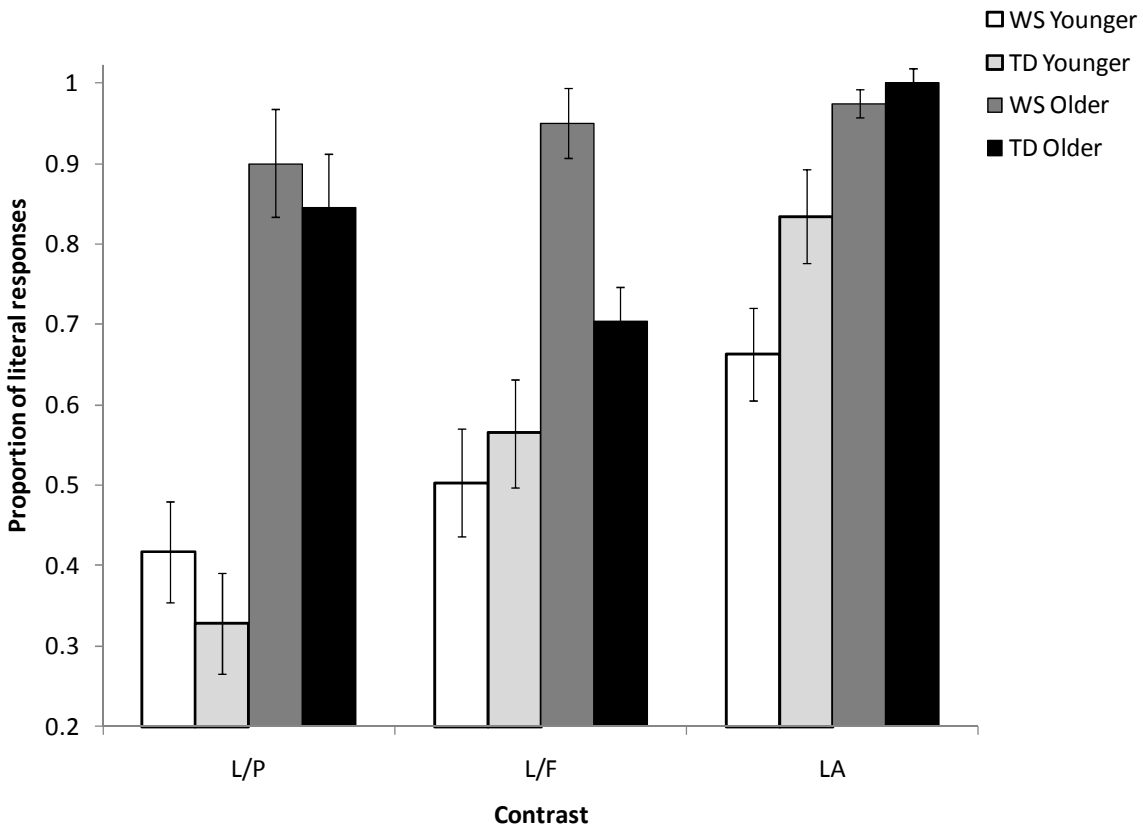

WS Younger
TD Younger
WS Older
TD Older
Proportion of literal responses
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
L/P
L/F
LA
Contrast

Figure 2a. Responses: Literal/perceptual contrast

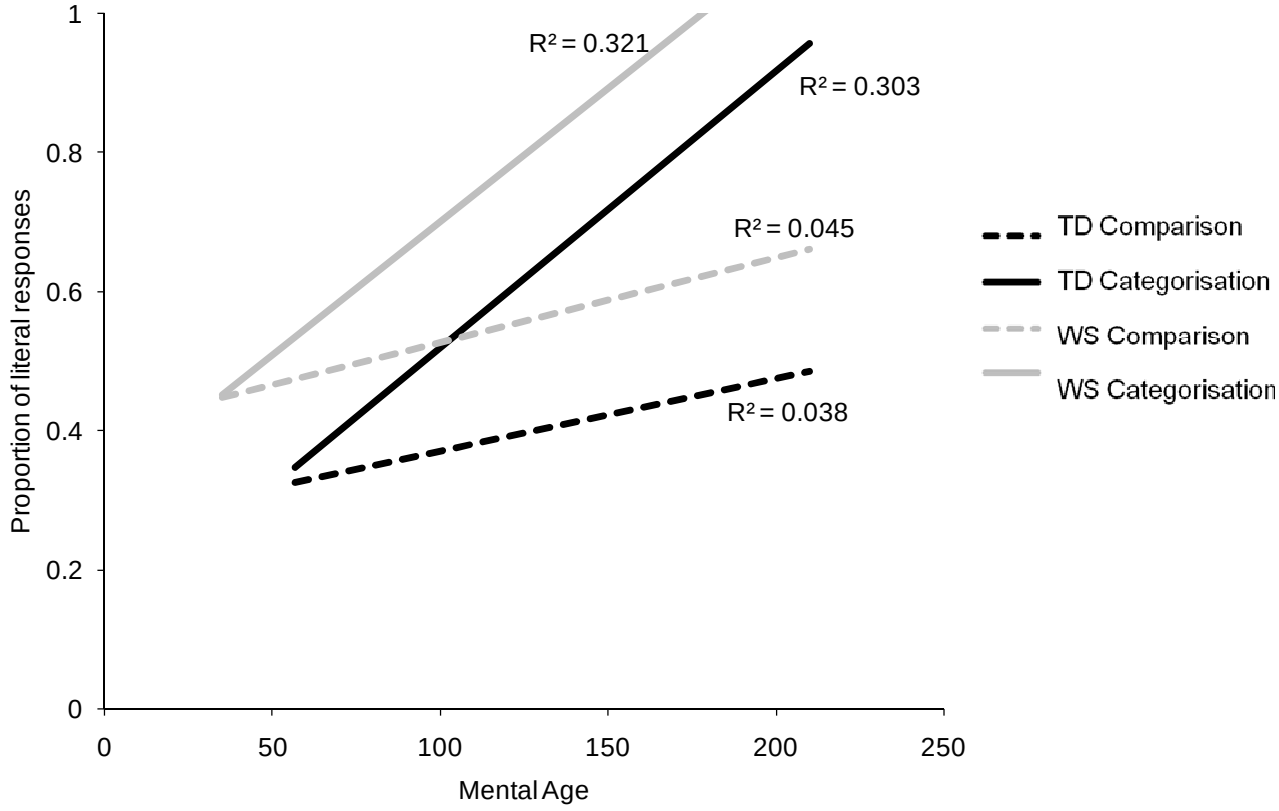


Figure 2b. Responses: Literal/functional contrast

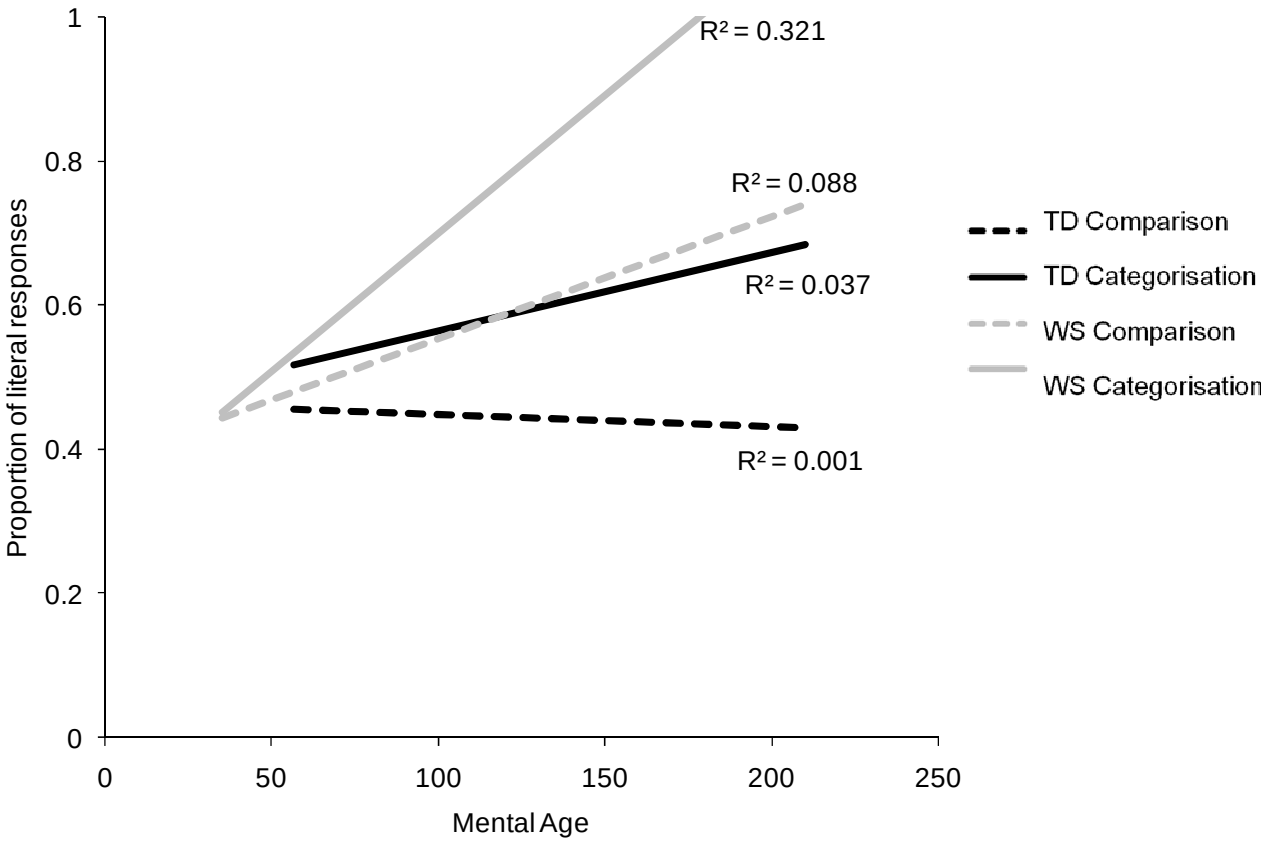


Figure 2c. Responses: Literal/anomalous contrast

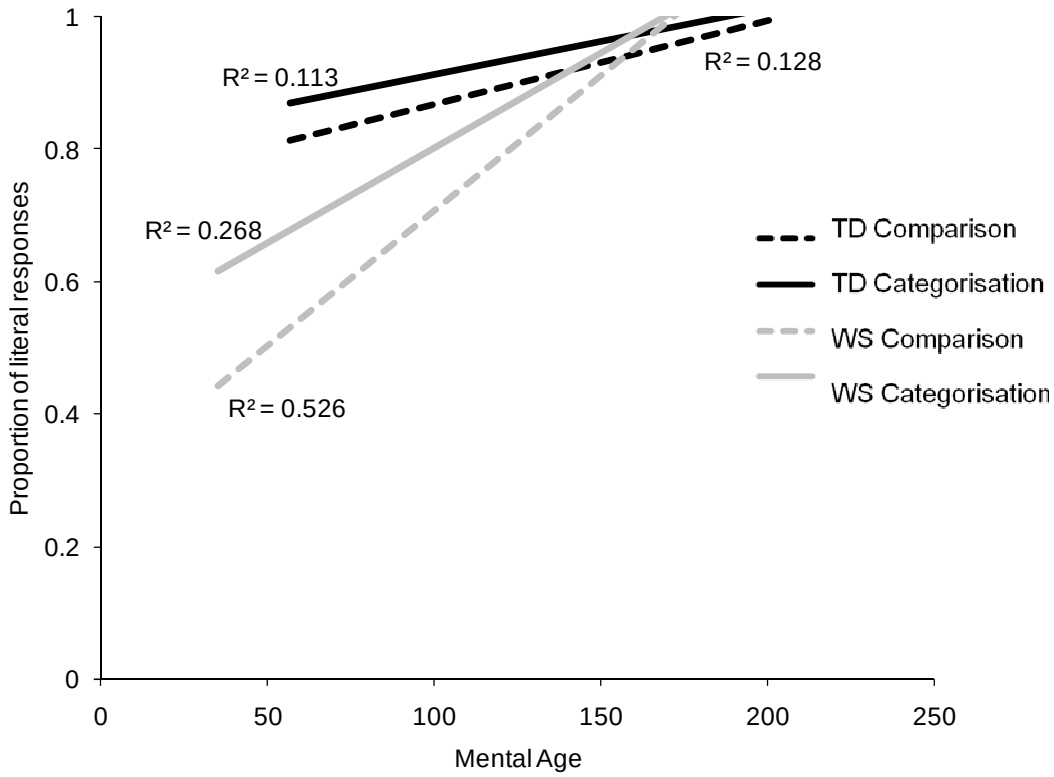
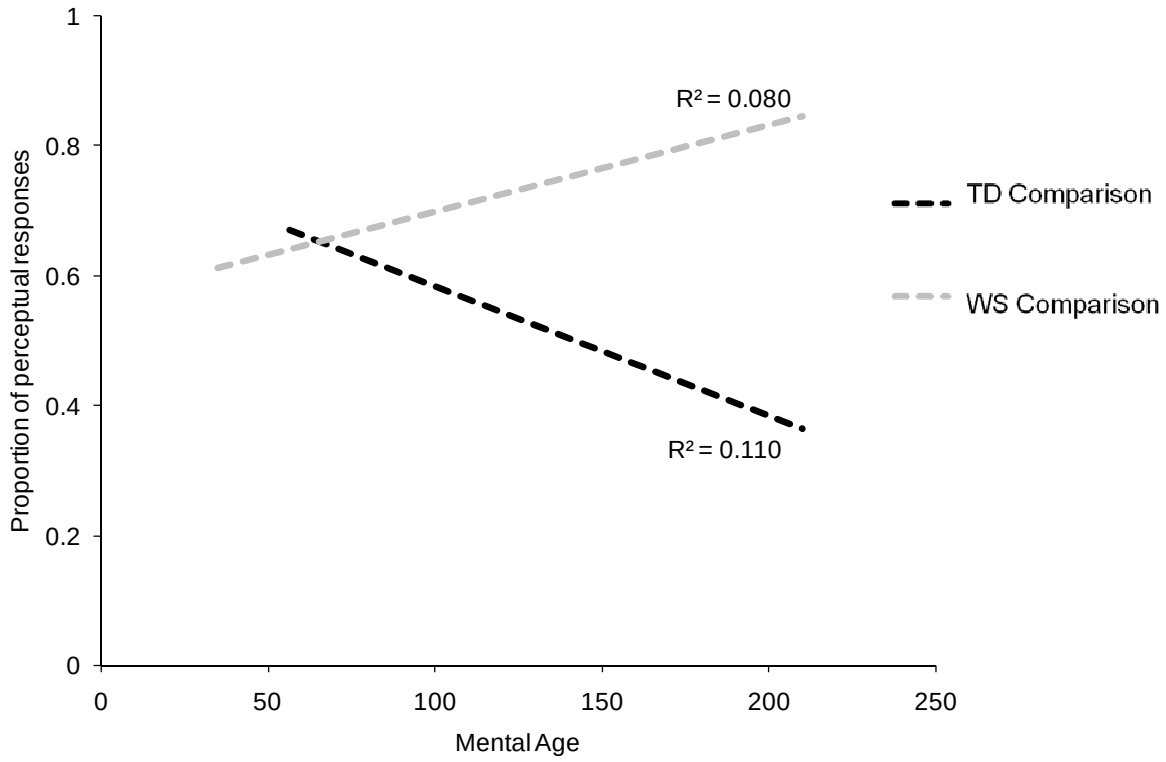
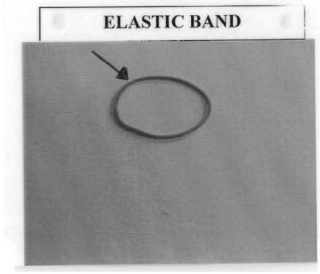


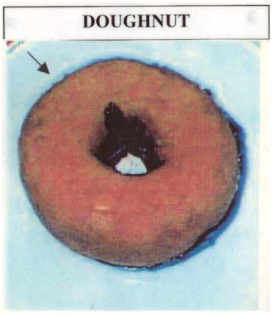
Figure 3. Responses: Perceptual/functional contrast

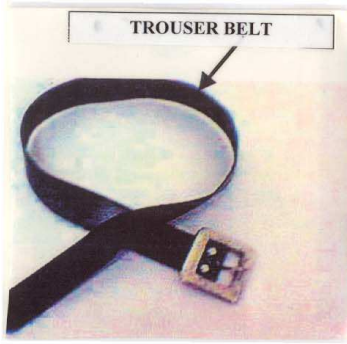


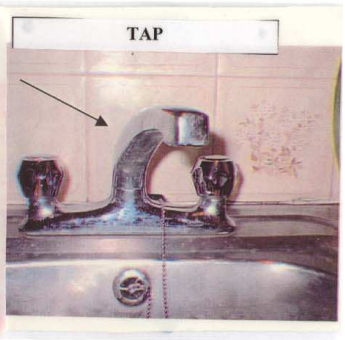
**Appendix 1. Pictures used in the experiment**

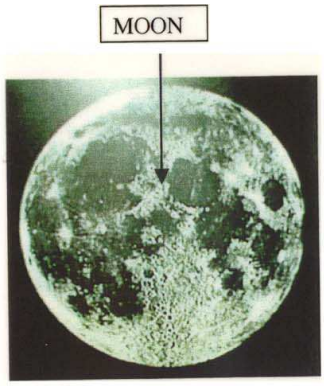


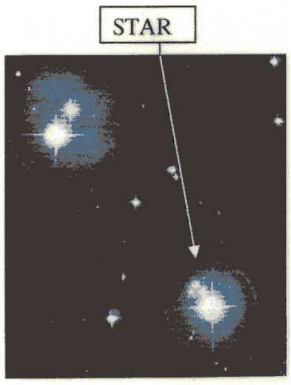


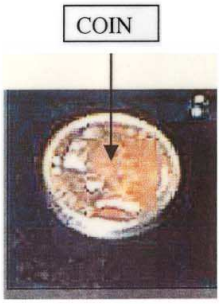


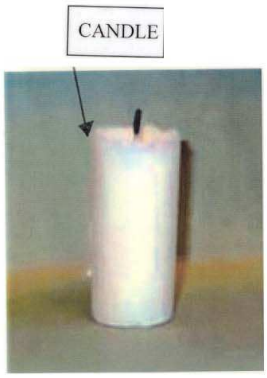



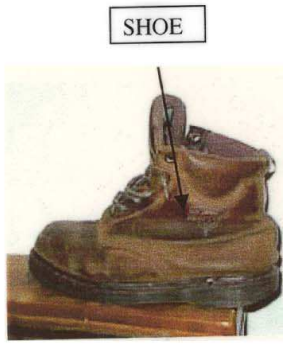



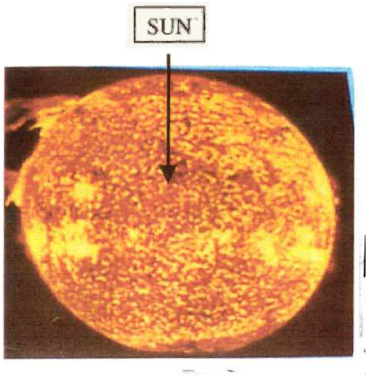



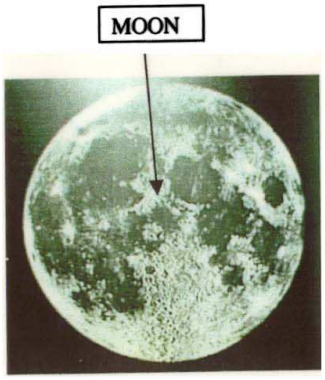



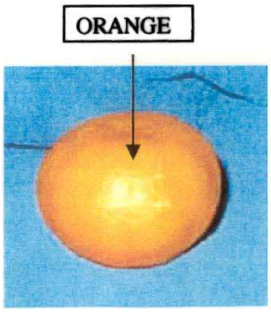



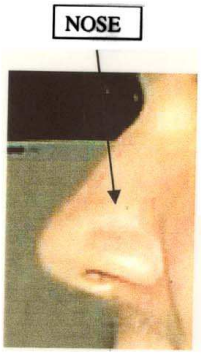

SUN

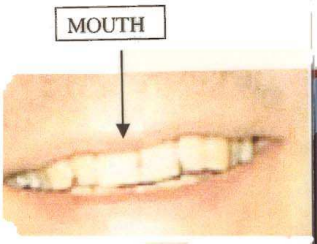

MOON

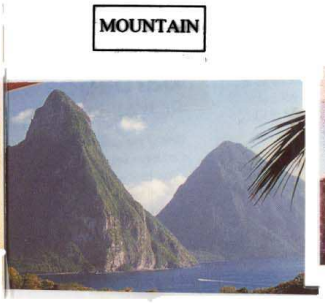

ORANGE

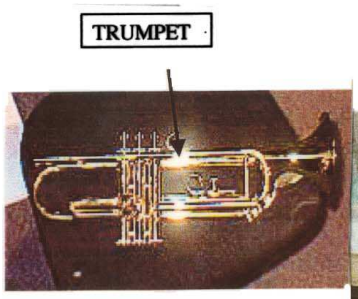

OVEN

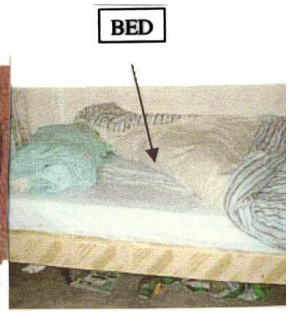

CHAIR

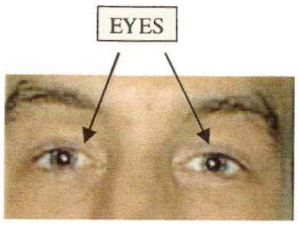

NOSE

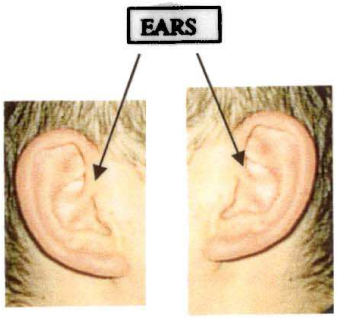

MOUTH

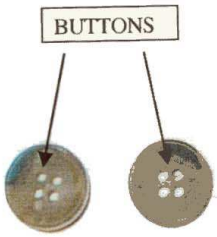

MOUNTAIN

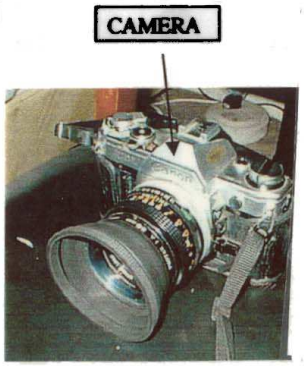

TRUMPET

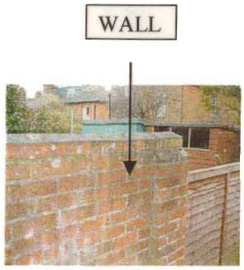

BED


EYES


EARS


BUTTONS


CAMERA


WALL