CHARACTERISTICS/RELATED CONDITIONS

DIAGNOSING AND PHENOTYPING VISUAL SYNAESTHESIA: A PRELIMINARY EVALUATION OF THE REVISED TEST OF GENUINENESS (TOG-R)

Julian E. Asher[1,2], Michael R.F. Aitken[2,3], Nasr Farooqi[2], Sameer Kurmani[2] and Simon Baron-Cohen[1,2]

([1]Department of Psychiatry, [2]Department of Experimental Psychology, [3]MRC Centre for Behavioural and Clinical Neuroscience, University of Cambridge, Cambridge, UK)

ABSTRACT

Synaesthesia, a neurological condition affecting approximately .05% of the population, is characterised by anomalous sensory perception: a stimulus in one sensory modality triggers an automatic, instantaneous, consistent response in another modality (e.g., sound evokes colour) or in a different aspect of the same modality (e.g., black text evokes colour). As evidence was limited to case studies based on self-report, the existence of synaesthesia was regarded with scepticism until the development of the Test of Genuineness (TOG) in 1987, which measures the consistency of stimulus-response linkage: synaesthetes typically score between 70-90% range, whereas controls typically score between 20-38%. However, the TOG had only limited ability to quantify the characteristics of visual synaesthesia. In this study, the revised Test of Genuineness (TOG-R), utilising the Pantone-based Cambridge Synaesthesia Charts, was given to 26 synaesthetes and 23 controls. Results confirmed that the TOG-R is equally accurate in the diagnosis of synaesthesia; synaesthetes scored significantly ($t_{47} = 16.01$, $p < .001$) higher (mean = 71.3%, SEM = 1.4%) than controls (mean = 33%, SEM = 2.0%). The TOG-R provides greater precision in quantifying the closeness of colour matches and enables a more detailed analysis of visual synaesthesia. Synaesthetes were phenotyped into broad- and narrowband based on their overall responsiveness to auditory stimuli, with bandwidth determined primarily by responsiveness to non-word stimuli. They were further sub-phenotyped based on responses to sub-groups of stimuli into word-colour (WC) and music-colour (MC). Development of this instrument has important implications for the diagnosis and phenotyping of visual synaesthesia.

Key words: synaesthesia, diagnosis, cognition, perception

INTRODUCTION

Synaesthesia – from the Greek *syn* (union) + *aisthesis* (sensation) – is a neurological condition characterised by anomalous sensory perception. For many synaesthetes, stimulation of one sensory modality triggers an automatic, instantaneous response in another modality: sound can evoke colours (e.g., Baron-Cohen et al., 1987, 1993) or taste (Ward and Simner, 2003). Some synaesthetes experience stimulus and response in different facets of the same modality: black letters or digits can evoke colours (e.g., Mattingley et al., 2001; Ramachandran and Hubbard, 2001). Synaesthetic percepts are almost always visual in nature; the most common phenotype is auditory-visual, or "coloured hearing", in which sounds trigger the perception of colours (Cytowic, 2002; Day, 2005).

Whilst many synaesthetes report no negative effects from their synaesthesia, there is growing evidence linking synaesthesia to perceptual and cognitive dysfunction. Rarely, synaesthetes find their percepts overwhelming, necessitating a restriction of their personal and occupational activities (Luria, 1968; Baron-Cohen et al., 1993). Studies of the "synaesthetic Stroop effect" have detected cognitive interference from synaesthetic percepts (e.g., Mattingley et al., 2001) and an ongoing study by Ward and Simner (2003) has found evidence of dyscalculia in auditory-visual synaesthetes. Conversely, synaesthesia is also associated with positive cognitive effects, such as enhanced recall (Smilek et al., 2002a). Whether the effects are positive or negative, it is clear that synaesthetic percepts lie outside the perceptual norm. Current evidence supports the theory that synaesthesia derives from altered neural connectivity (e.g., Baron-Cohen et al., 1993; Maurer, 1997; Grossenbacher and Lovelace, 2001).

Despite the publication of multiple case reports by investigators including Galton (1880), many researchers remained sceptical regarding the existence of synaesthesia; how could one verify that synaesthetes genuinely perceived colours when they heard sounds? Self-report is vulnerable to memory effects, suggestion and even outright fraud. Reluctance to rely upon self-report for diagnosis and the rise of behaviourism in psychology contributed to the sharp decline in active work in the field during much of the twentieth century (Marks, 1975; Harrison and Baron-Cohen, 1995).

It was only in 1987 that Baron-Cohen et al. (1987) developed an objective diagnostic test for synaesthesia, now known as the Test of Genuineness (TOG). E.P., a 76-year-old female auditory-visual synaesthete, and a control subject, a 27-year-old female lawyer, were asked to provide detailed descriptions of the colours evoked by 103 words and sounds; for example, E.P. described the word "Moscow" as "darkish grey, with spinach-green and

a pale blue in places.“ The control subject was instructed to memorize the colour-word associations for a re-test, whereas E.P. was unaware that she would be re-tested. Upon re-testing, the control subject showed 17% consistency after two weeks; E.P. showed 100% consistency after ten weeks. In follow-up testing with a random selection of the original stimuli eight months later, E.P. remained 100% consistent (Baron-Cohen et al., 1987).

Further investigation, including functional MRI studies of TOG-diagnosed auditory-visual synaesthetes which revealed activity in areas associated with colour processing (V4/V8) while hearing words (Gray et al., 1997; Nunn et al., 2002), confirmed the diagnostic validity of the TOG, and it has since become the “gold standard” diagnostic test for synaesthesia. Synaesthetes typically score between 70-90%, whereas control subjects score between 20-38% (e.g., Baron-Cohen et al., 1993, 1996). Though other tests have since been developed, they rely upon the same principles; both the pop-out test (Ramachandran and Hubbard, 2001) and the synaesthetic Stroop (e.g., Mills et al., 1999; Odgaard et al., 1999; Dixon et al., 2000; Mattingley et al., 2001) depend upon the consistency of synaesthetic percepts. Furthermore, the application of these tests is limited to small sub-groups of synaesthetes (those who ‘see’ their percepts ‘projected’ in front of them rather than in their mind’s eye for the pop-out test; grapheme-colour synaesthetes for the Stroop test). The TOG’s measurement of stimulus-response linkage consistency, however, is valid for all forms of synaesthesia, and since it does not depend on a particular stimulus set it can be used to create a diagnostic test for any type of synaesthesia [e.g., auditory-visual synaesthesia as above; lexical-gustatory synaesthesia (Ward and Simner, 2003)].

The TOG was not without its limitations. Reliance on verbal descriptions of synaesthetic percepts rendered the testing of children or other persons with limited verbal abilities difficult, hindering research into the developmental aspects of synaesthesia. Even extremely detailed verbal reports could not specify the hues of synaesthetically perceived colours, limiting depth of analysis; if a synaesthete reported a “red” percept, was it the same hue each time or did it vary, perhaps in relationship to certain critical aspects of the stimulus? Further exploration of the relationship between stimulus and response was not possible without a more precise way to characterise synaesthetic percepts.

Baron-Cohen et al. (1996) introduced the use of coloured swatches as an alternative to verbal reporting. Physical colour swatch sets were derived from sources such as paint swatches. Some investigators used computer monitors to display colour swatches, from which synaesthetes were asked to choose a colour matching that of the synaesthetic percept (e.g., Myles et al., 2003); alternatively, the synaesthete or researchers could adjust the colour of a patch until it matched (Smilek et al., 2001, 2002b). This modification enabled researchers to actually see the synaesthetically produced colours, facilitating experiments based on congruence/incongruence between presented stimuli and synaesthetic percepts (e.g., Smilek et al., 2001, 2002a; Mattingley et al., 2001; Myles et al., 2003) and colour contrast (Ramachandran and Hubbard, 2001).

These modifications brought with them their own set of limitations. Physical colour swatch sets were unique to each laboratory, complicating inter-study comparisons. In most cases only one set of swatches was available, so testing had to be conducted in the laboratory or required travel to the subject’s location.

Computerised colour swatch displays are inherently dependent on the equipment used; the need to control hardware and software factors such as monitor resolution, colour calibration and graphics software once again effectively limits testing to the laboratory or requires travel to the subject’s location and thus may necessitate immediate re-testing (Smilek et al., 2002), introducing potential memory effects. In both paper and computerised protocols, there is often no standardized system for grading the relationships between swatches or the closeness of a match.

One of the principal potential advantages of a computer-based remote testing protocol is the large range of colours available with modern software. While in theory the ability to select from millions of colours would increase precision, testing using a computerised colour set is very time-consuming and can result in subject fatigue (Merikle, 2003); investigators have also found that too large a selection of colours can prove overwhelming for some subjects (see discussion below). Moreover, not all potential subjects have access to computers capable of precisely displaying the maximum number of colours. Depending on their settings and associated hardware, computer monitors can be configured to a variety of colour depths, typically resulting in a palette somewhere between 256 distinct colours (VGA, video graphics array) and a virtually unlimited set of several million colours. The number of colours actually available to a subject is determined by the limitations of the subject’s computer, rather than by the investigators, and may vary widely. Even if it were practical to request that all participants select a similar colour palette prior to testing, variability in display hardware (e.g., between LCD – liquid crystal display – and traditional cathode-ray tube monitors) may lead to large differences in the colours displayed.

Adapting a computer-based remote protocol for non-visual stimuli would introduce additional technical problems, e.g. sound reproduction issues. These difficulties make computer-based methods suitable for initial screening or for diagnosing single synaesthetes but impractical for larger studies.

Researchers at the University of Waterloo have developed an Internet-based test that avoids potential confounding from equipment factors by administering the re-test immediately following the initial test (Merikle and Dixon, 2002). Subjects are shown single graphemes and use a mouse to scroll over a colour spectrum or an adjacent greyscale bar until they find a colour matching the percept evoked by that grapheme. While neutrals (such as khaki) are not included in the colour range, the researchers provide a way for subjects to note how closely the spectrum was able to approximate their percept for each grapheme. There are a total of 36 stimuli (26 letters and the digits 0-9).

The protocol works well for visual stimuli, though as noted above the process of finding a colour can be quite time-consuming which may explain the limited size of the stimulus set. The small number of stimuli makes the test an efficient initial screening assessment, but additional testing with a much larger stimulus set would be needed to establish a firm diagnosis. While the spectrum provided is reasonably broad, the actual number of colours seen by an individual subject is determined by the factors mentioned above.

More important, the immediate re-test introduces the possibility of memory effects, particularly in the context of a small stimulus set. A larger stimulus set would ameliorate this problem to some extent, though even a much larger set cannot completely eliminate memory effects (see discussion section below). Subject fatigue could also become an issue with a larger stimulus set. An Internet-based test also introduces potential sampling biases due to unequal online access among population groups, particularly older persons and those from lower socioeconomic groups, and may limit the ability to use a common protocol for all subjects – a particular issue for genetic studies, where all subjects should be assessed using the same protocol.

Overall, a computer-based remote testing protocol does not appear to offer a significant advantage and may indeed be disadvantageous due to the variability of hardware and access across subjects. The use of a printed chart delivered by post ensures that the effects of colour surrounds are consistent across subjects; that the whole range of colours is visible simultaneously; and that the investigators can control the number of colours visible to the subjects.

One must note that *any* remotely administered protocol has limited value in terms of generating data for direct inter-subject comparison. Lighting conditions can vary widely from one environment to another, a problem for both computer-based and paper tests. While a remote protocol could, for example, give preliminary indications of possible colour trends across subjects, rigorous inter-subject comparisons require testing in a laboratory environment.

In the study reported here, we conduct a preliminary evaluation of the diagnostic accuracy of a revised Test of Genuineness (TOG-R) for visual synaesthesia, and its ability to quantitatively characterise synaesthetic percepts and phenotypes.

Method

Subjects

Twenty-six synaesthetes [5 male, 21 female; mean age (years) = 48.4, SEM = 4.3] were recruited from our synaesthesia database. All synaesthetes who reported auditory-visual synaesthesia were contacted and invited to participate in this preliminary study, and a random sample was selected from those who responded. Nine synaesthetes had been tested using the original TOG and 17 were previously un-tested. Four female synaesthetes were tested in a laboratory environment and 22 synaesthetes (5 males, 17 females) were tested remotely with materials sent through the post. While this is a small sample group, it is the largest recruited thus far for the evaluation of a diagnostic instrument for synaesthesia; initial evaluation of the pop-out test involved two subjects (Ramachandran and Hubbard, 2001) and the largest evaluation of the synaesthetic Stroop involved 15 subjects (Mattingley et al., 2001). All synaesthetes received a professional A4 print of an fMRI image comparing synaesthetes and non-synaesthetes for their participation.

Twenty-three control subjects [4 males; 19 females; mean age (years) = 19.4, SEM = .15] were recruited. A medical history was taken and all control subjects were found to have normal colour vision; no history of ophthalmological (excepting spectacle or contact lens use), neurological, or psychiatric conditions or hallucinogenic drug use; and no history of visual or other symptoms consistent with any form of synaesthesia. All control subjects were tested under identical conditions in a laboratory environment and received £ 10 for their participation.

Written informed consent was given by all subjects and where appropriate by parents/guardians. The study protocol followed the standards laid out in the Declaration of Helsinki and was reviewed and approved by the University of Cambridge Psychology Research Ethics Committee and the Local Research Ethics Committee of the Cambridge University Hospital NHS Foundation Trust (Addenbrooke's).

Materials

Stimuli

Test stimuli consisted of a CD containing 99 unique sounds (51 word and 48 non-word sounds).

Each 8-second track contained a single sound (1-3 sec) followed by silence. The "word" sounds consisted of: words (days of the week, months of the year, Christian names, nouns, verbs, articles), numbers and letters. A range of concrete (giraffe, zebra) and less concrete (the, and) words was used. Homonyms (for/four, two/too/to) were excluded. A single male voice was used for all recordings.

The "non-word" sounds consisted of: musical instruments (Sample Cell Editor 3.1 and Soft Sample Cell 3.0) (Digidesign 2002), natural environmental sounds (animals, rain) (Wright 1990; *Phonak Sound CD 1,* Phonak Hearing Systems, 1999) and man-made environmental sounds (doorbell, car horn) (*Phonak Sound CD 1,* as above). Each instrumental sound consists of a note or chord on a single instrument; a range of notes on three major instruments (piano, violin and cello) are included as well as single notes or chords on a number of other instruments (xylophone, electric guitar). Sounds not available from the above sources were drawn from sound files in the public domain. In addition, three vocal exclamations ('whoa', 'aaah' and 'oooh') were drawn from sound files of male and female voices in the public domain.

The word and non-word sounds were randomly intermixed and divided into four blocks (three blocks of 25 and one block of 24) that were then sorted in a semi-random order, the only requirement being that the 24-sound block was last on both CD-A and CD-B (due to technical limitations). To further minimise potential memory effects, the sound order in each block was reversed for CD-B.

Cambridge Synaesthesia Charts (CSC)

The first edition of the Pantone© *European coated solid to process guide* for four-colour CMYK (Cyan, Magenta, Yellow, Black) printing was selected as the colour swatch source (Pantone, Inc. 2002). Each colour swatch measures 1.6 cm^2 with a total of 238 swatches, for a total of 241 potential colour choices (including white, transparent and opalescent/translucent, see below) and is accompanied by an identifying number based on its Pantone guide number. The swatches were arranged in an array of 34 rows of 7-swatch columns. The columns and rows were drawn directly from the Pantone guide, where their identifying numbers reflect relative CMYK similarity; columns reflect increasing saturation whereas rows reflect progressive alteration in hue (by alterations in the relative proportions of the four inks) (Figure 1).

Though often excluded from colour spectra, neutrals (browns, khakis and greys) were included due to their prevalence in previous reports from synaesthetes (e.g., Day, 2005). White was inadvertently omitted[1], as were the "transparent" or "opalescent" effects reported by some synaesthetes during pilot testing. Pilot subjects were instructed to write these in by the researchers, and amendments to that effect were incorporated into

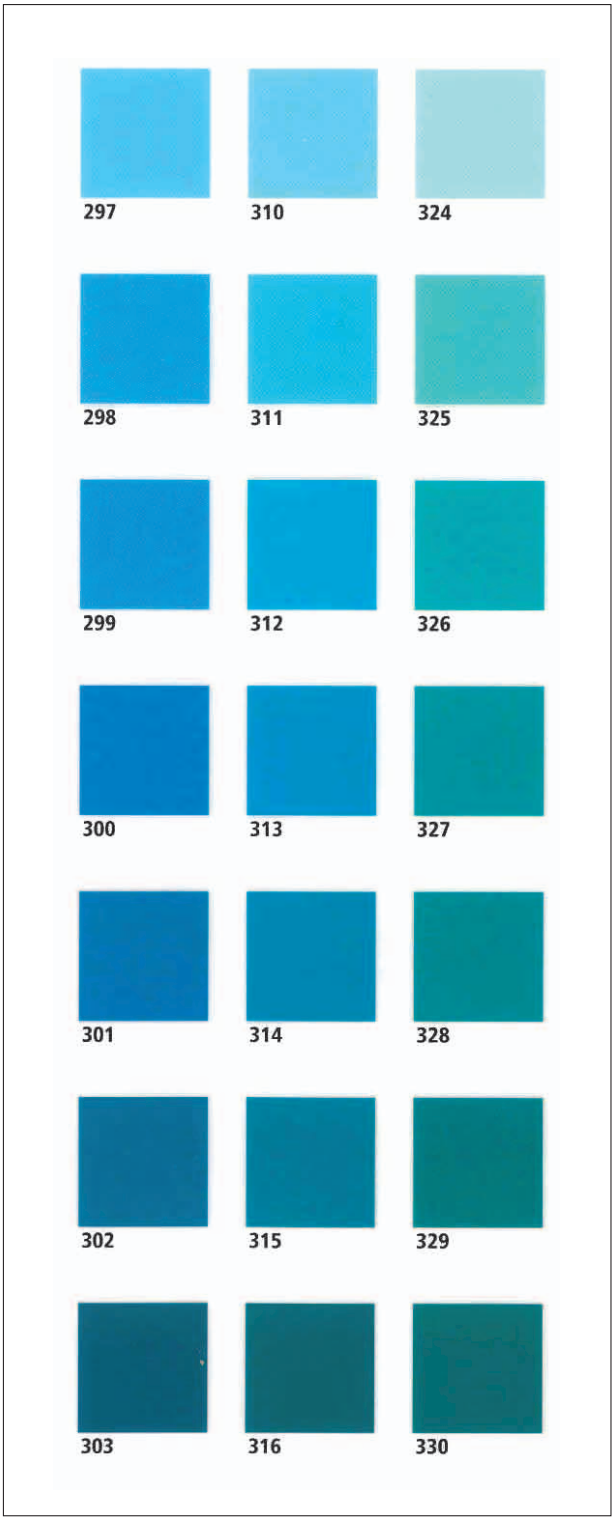

Fig. 1 – *Excerpt from Cambridge Synaesthesia Charts.* ©*2003 Julian Asher and Simon Baron-Cohen.*

[1]Ironically because Julian E. Asher, who designed the charts, is a visual synaesthete without white percepts.

the testing instructions for later subjects. The charts were printed on three A4 sheets by Cambridge University Press.

Procedure

Subjects were instructed to take the test in a room lit solely by artificial light and to lay the colour charts out on a flat surface according to a diagram provided by the investigators. Laboratory testing followed this protocol; identical apparatus and the same room were utilised for test and re-test. Remote subjects were instructed to utilise identical set-ups for test and re-test (e.g., the same room and lighting conditions).

The test CD can be played on a computer's CD-ROM drive or in a CD player. Laboratory testing used the same CD player for test and re-test; subjects did not wear headphones during testing. Remote subjects were permitted to take the test with or without headphones and were instructed to use an identical audio set-up for test and re-test.

Synaesthetes were instructed to choose the colour swatch closest to the dominant colour evoked by each sound (and major secondary colour if applicable for a maximum of two colours per sound) and to write its numbered code on the answer sheet. They were informed that they might not experience a colour for each sound (e.g., a music-colour synaesthete would not respond to the spoken words) and reminded that this was normal. Due to confusion on the part of some pilot subjects, the final protocol also included specific instructions not to guess in the event that they did not see a colour. If they did not see a colour, they were instructed to place a dash (—) in the answer space and to fast-forward the CD to the next track.

Control subjects were instructed to choose a single colour swatch after listening to each sound, and to write its numbered code on the answer sheet. They were explicitly instructed to try to remember the colour (rather than the number) they matched with each sound.

All subjects could pause the CD whilst choosing a colour and could replay a track as often as needed. If a subject became fatigued, a short break was permitted. All laboratory-tested subjects completed the test in one session; all remote subjects with the exception of one synaesthete reported completing the test in one session. The results for the subject whose testing was interrupted were indistinguishable from those of the other synaesthetes.

Synaesthetes experienced a minimum 1-month delay between test and re-test (mean interval = 167.4 days, SEM = 17.7) and were not explicitly informed that they would be re-tested (though those who had taken the original TOG were likely aware of the re-test). All control subjects experienced a 1-week delay between test and re-test and were explicitly informed that their colour recall would be evaluated on the re-test. Remote subjects were required to return their CD-A answer sheet before receiving CD-B[2].

Scoring

Performance was scored on the similarity of the choices recorded for a given sound on the two tests. The layout of the colour charts allows a simple proximity measure of similarity, as position in the array incorporates information concerning similarities of CMYK saturation and proportion; this was preferred to scoring on the basis of strict CMYK similarity due to the possible subtle influences of layout upon both choice and colour perception.

The colour charts were subdivided into seven colour groups (yellow, orange, red, violet, blue, green and neutral) determined by the point at which the swatches undergo a significant change in CMYK composition (i.e., from predominantly magenta to equal proportions of magenta and cyan). A single column between adjacent colour groups (2 columns between the blue and green groups) where the two major inks are present in equal proportions is regarded as part of both groups (e.g., a match in the yellow/orange region would be in-group for both yellow and orange groups). The yellow, orange, red and violet groups cover five vertical columns each (35 swatches). The use of CMYK proportions to determine group boundaries resulted in slight enlargement of the blue (42 swatches) and green colour groups (56 swatches). Neutrals are considered a separate group with 49 swatches; white, transparent and opalescent/translucent were also considered "neutral" for a total of 52. The potential inflationary effect of group size on score (via the slightly greater likelihood of an in-group match in a larger group) is negligible, due to the small proportion of swatches represented by differences in group size relative to the overall number of choices and the low score assigned to a simple in-group match.

An exact match (to the identical swatch or to a swatch ± 1 horizontal or vertical) scores three points (Figure 2). A near match (± 1 diagonally or ± 2 horizontally or vertically) scores two points. A match in the same colour group (that is not an exact or near match as defined above) scores one point. Two swatches that do not share a colour group score no points. Where two colours were reported for a sound, only the highest scoring match was counted.

[2]While as in all remote protocols it is theoretically possible for a remotely tested subject to "cheat", doing so would involve an extraordinary effort. A subject would need to copy their answers from CD-A and to decipher the track order. There is also the question of motivation; as there is no benefit from being diagnosed as a synaesthete, it is unclear why anyone would go to the effort required. Moreover, subjects are not informed that they will be re-tested.

1	1	1	1	1	1	1
1	1	1	2	1	1	1
1	1	2	3	2	1	1
1	2	3	EXACT 3 MATCH	3	2	1
1	1	2	3	2	1	1
1	1	1	2	1	1	1
1	1	1	1	1	1	1

Fig. 2 – *Example of Cambridge Synaesthesia Colour Charts scoring protocol for a potential match within a colour group.*

Reports of "white" and "transparent" percepts were scored as a near match (2 points) to each other, or to any of the five lightest members of the neutral group, and scored as an in-group match to all other neutrals. Reports of "no visual percepts" (marked with a dash, as described above) from a synaesthete were scored as a near match when they occurred for the same sound on both trials. Occasionally, a subject reported a number that did not correspond to any number on the colour charts. Such trials were excluded from all analyses. The score achieved by each subject was expressed as a percentage of the total possible points scored (i.e., as a proportion of the total number of sound pairings being compared * 3).

The printing process resulted in a few instances of unusually large shifts between adjacent swatches that are not accounted for by the CMYK proportions. The scoring of a match between two such swatches was slightly modified (in most cases scored as a colour group match rather than a near match). Exceptions were assessed independently by two persons with normal colour vision and included in the scoring protocol only where agreement was reached. These technical issues argue in favour of utilising the same or an identical model of printing press and paper to produce any additional charts to maximise comparability for cross-study comparisons.

The scoring criteria were implemented into an automated scoring algorithm written in VBA for Microsoft Excel. Further details are available from the authors.

Results

Two of the remotely tested synaesthetes received the revised instructions (instructing them not to 'guess' a colour when no percept was experienced) between test and re-test. Sound pairs for which they reported a colour on the first session, but no colour on the second session, were excluded from all scoring of matches.

Diagnosis

Overall Scores

Individual scores are shown in Figure 3. Synaesthetes scored higher (mean = 71.3%, SEM = 1.4%) than the control subjects (mean = 33%, SEM = 2.0%), a difference which is highly significant ($t_{40} = 16.01$, $p < .001$).

While this comparison is not a pure test of synaesthetic performance as the instructions given to the control subjects were slightly different (only synaesthetes have the opportunity to score matches for 'white' and 'transparent' reports or for reports of 'no colour') the scoring system ensures that synaesthetes can score no more than 2 (out of 3) for a sound pair resulting from a report of "white", "transparent" or "no colour". Given that the mean performance of the synaesthetes is above 67%, the exclusion of these trials from the scoring would only increase this group difference.

The control subjects were not strictly well-matched controls, but rather a group of individuals selected to estimate the "upper boundary" of performance that might be gained by a group of young, healthy subjects under optimal (identical) conditions attempting to use mnemonic strategy over a relatively short time interval. If anything, these circumstances bias scoring in favour of the control group. The failure of any of the control subjects to approach the scores of the synaesthete group confirms the test's diagnostic accuracy.

Figure 3 suggests that the score achieved discriminates clearly between the two populations. Taking the simple 95% confidence intervals for the population scores of ± two standard deviations

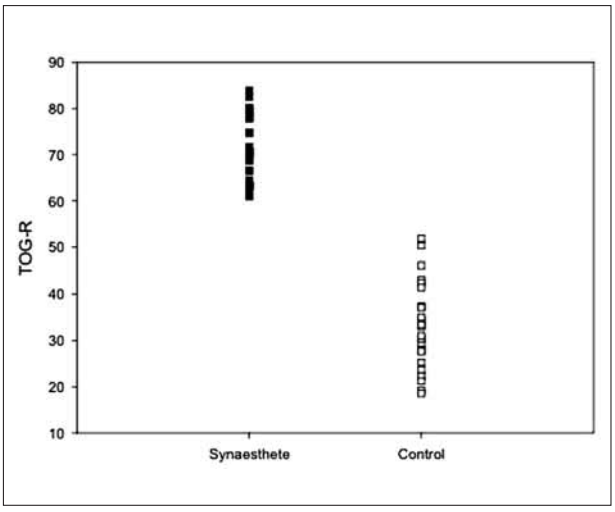

Fig. 3 – *Overall TOG-R scores of synaesthetes and controls.*

from the mean, we have a predicted range of 14.0% to 52.3% for the controls and 57.2% to 85.3% for the synaesthetes. A diagnostic threshold set between these two boundaries (e.g., 55%) would be very unlikely to result in misdiagnosis.

Block Effects

As described above, for technical reasons the block of 24 sounds was the last block presented in both sessions for all subjects; the remaining blocks of 25 sounds were shuffled between the two CDs. If this block attracted a higher score than the other blocks, this could be interpreted as indirect evidence for a memory process playing a role in the task. This pattern is indeed seen in the control subjects, with the performance on the final block of 24 trials scoring more highly than the remaining blocks (mean performance = 38.7%). A repeated measures Analysis of Variance (ANOVA) contrasting performance of the control subjects across blocks confirms that block influenced scoring in these subjects [$F(3, 66) = 5.41$, $p < .01$], with the final block attracting higher scores than all the preceding blocks [smallest $F(1, 22) = 7.80$, $p < .05$].

No such pattern was observed in the synaesthetes, with the final block attracting, if anything, a slightly lower than average score (mean performance = 70.3%). A mixed ANOVA confirmed that the block effect was significantly greater in the control group [$F(3, 141) = 4.91$, $p < .01$].

Although the impact on the overall score is small, the effect observed in the control group indicates the vulnerability of protocols involving a shortened test/re-test interval to memory effects. The fact that the effect was observable with a one-week interval raises important questions about the use of protocols involving immediate re-testing, and argues against using immediate re-testing to circumvent equipment-related confounders in online testing. These findings suggest that subjects initially assessed using an immediate re-testing protocol should be re-tested using a more rigorous protocol before their diagnosis is accepted as definitive, limiting the utility of such a test to initial screening. The argument for re-testing would be particularly strong for genetic studies where rigorous diagnostic standards are necessary to maximise the chances of finding a gene.

Phenotyping

Broad-band and Narrow-band Auditory-Visual Synaesthesia

While it is a well-known phenomenon that some synaesthetes respond to a broad range of stimuli whereas others respond only to a limited range of very specific stimuli (e.g., letters, numbers, days of the week and months of the year) it has not thus far been possible to quantify this distinction. The synaesthetes in the study group showed considerable variability in the range of stimuli to which they responded, with 37 to 100% of sounds (n = 198) producing a reported percept. The consistency of these percepts (measured via the average matching score across sound pairings where a percept was reported for both sounds) showed less variation, with subjects scoring 62 to 99% on these items.

Analysis of the proportion of sounds to which synaesthetes responded revealed two sub-groups within our sample of auditory-visual synaesthetes: eight "broad-band" synaesthetes reported percepts to a wide range of stimuli (80 to 100%) and sixteen "narrow-band" synaesthetes reported percepts to a narrower range of stimuli (37 to 59%). Three synaesthetes reported percepts to 59% and 71% of stimuli, and were classed as narrow-band for subsequent analyses. There is no simple relationship either overall, or within either group, between overall bandwidth and percept consistency (largest $r = -.211$, ns). While these findings are preliminary thus far and require confirmation with a larger sample, they indicate exciting possibilities for a more quantitative definition of synaesthetic phenotypes. In addition, our preliminary work with multiplex (> 3 synaesthetes) families has revealed evidence that familiality may extend to bandwidth.

For the purposes of comparing responses to words and non-words, the three vocal exclamations were excluded from the *n* leaving 51 word and 45 non-word stimuli. Although it is difficult to generalise from a small sample, it appears that variations in bandwidth are primarily a function of relative responsiveness to non-word stimuli. Only one of the synaesthetes responded to < 80% of word stimuli whereas response rate to non-word stimuli ranged from nil to 100%. The eight broad-band synaesthetes showed slightly greater consistency to word stimuli than to non-word stimuli (77.8% *vs.* 66.7%, $t_7 = 1.96$, ns). These differences may be due to the greater heterogeneity of the non-word stimuli, with different synaesthetes responding to certain sub-sets of non-word stimuli.

Sub-Phenotyping

Further analysis was performed on the responses of the narrow-band sub-group. As above, the three vocal exclamations were excluded from the analysis, leaving 45 non-word and 51 word stimuli. Eleven of the eighteen narrow-band synaesthetes were purely (0 non-word responses), and three were almost purely (≤ 4 non-word responses) responsive to words and were sub-phenotyped as word-colour (WC) (Table I). One synaesthete was almost purely (3 word responses) responsive to non-words and was sub-phenotyped as music (non-word) colour (MC). The remaining three synaesthetes are the most broad-band of this

TABLE I

Response rates of narrow-band synaesthetes to word (W) and non-word (NW) stimuli (n_W = 102; n_{NW} = 90 presentations) and resulting sub-phenotype designation: WC = word-colour; MC = music (non-word) colour; WC > MC = mixed, predominantly word-colour

Subject	W responses	NW responses	Sub-phenotype
N1	3	68	MC
N2	87	4	WC
N3	95	0	WC
N4	96	0	WC
N5	98	0	WC
N6	100	0	WC
N7	102	0	WC
N8	100	0	WC
N9	100	0	WC
N10	102	0	WC
N11	98	2	WC
N12	100	1	WC
N13	100	0	WC
N14	99	0	WC
N15	101	0	WC
N16	90	26	WC > MC
N17	84	38	WC > MC
N18	102	32	WC > MC

sub-group; they responded to slightly fewer words and many more non-words than the majority of the word-responsive group. All three responded to fewer non-words than the primarily non-word responsive synaesthete and were sub-phenotyped as mixed, predominantly word-colour (WC > MC).

The observed sub-phenotypes and the role of non-word sounds in determining overall bandwidth are not entirely unexpected. While they share some common initial auditory pathways, different brain regions are responsible for the linguistic and non-linguistic processing of sound. Though it is difficult to generalise from a small sample, as this is the first study to quantify synaesthetic phenotype it is noteworthy that the distribution of sub-phenotypes accords with previously observations based on self-report from synaesthetes, which indicate that WC synaesthesia is the most common auditory-visual sub-phenotype (Cytowic, 2002; Day, 2005). While such an analysis is beyond the scope of the present paper, a similar protocol could be used to partition the group into more specific sub-phenotypes, i.e. phonemic and graphemic WC synaesthetes. Further clarity regarding synaesthetic phenotype will play an important role in advancing our understanding of synaesthesia, particularly in terms of the genetics of synaesthesia.

Response Variance to Word and Non-Word Stimuli

Individual TOG-R scores for the 45 word and 51 non-word stimuli were calculated to examine the effect of stimulus type on overall score. A linear relationship between the two sub-scores is observed in the control subjects ($r = .701$, $p < .01$), but no such relationship is seen in either of the synaesthetic sub-groups ($|r| = .1$, ns) (Figure 4). However, a larger sample is needed to further evaluate this issue.

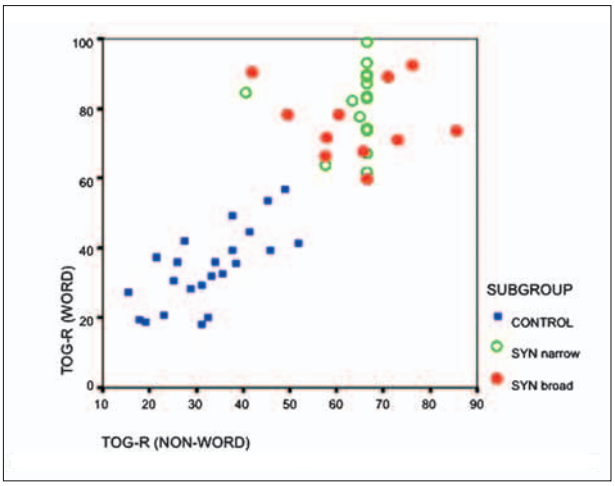

Fig. 4 – *Relationship between non-word and word TOG-R scores for controls and synaesthetes (broad-band and narrow-band sub-phenotypes).*

Consistency Within Sub-Phenotypes

A modified TOG-R score (data not shown) was calculated for the fifteen subjects categorised as WC in Table I to evaluate consistency within the sub-phenotype. As above, the score was expressed as a percentage of total possible points scored on sounds that evoked a response from a given subject on both presentations, but in this case non-word sounds were excluded from the calculation. The score range (61 to 99%) paralleled that observed for the overall consistency scores for the synaesthete group as a whole.

Additional sub-phenotypic analyses incorporating a larger subject group and extending to additional sub-phenotypes would be of interest in further defining the differences between synaesthetic phenotypes.

Qualitative Data

The majority of synaesthetes (23 out of 26) reported that the colour range presented in the colour charts are adequate. One subject reported that there are too few colours, and two reported that the colour selection was large enough to be overwhelming. One flaw of the Pantone system is its difficulty reproducing some of the warmer shades of red, which was remarked upon by both synaesthetes and controls during testing. Some synaesthetic subjects (7 out of 26) also reported fatigue from the effort of focussing on their percepts and trying to choose an appropriate colour.

Discussion

In this study, we have reported the results of our preliminary evaluation of the TOG-R and found that the TOG-R both replicates the diagnostic accuracy of the original TOG and offers

increased quantitative specification of synaesthetic phenotypes and sub-phenotypes.

By combining the strengths of the original TOG (simplicity, flexibility and ease of conducting remote testing) with the advantages of the modifications pioneered since its development (more detailed characterisation of visual synaesthetic percepts), the TOG-R represents a significant step forward in the diagnosis and phenotyping of synaesthesia, with particularly important implications for developmental and genetic studies. There has been an increasing interest in the developmental aspects of synaesthesia, including how and when synaesthesia first manifests in children, but research has previously been thwarted by the difficulty in diagnosing and monitoring the progress of young children who may be uncomfortable in a laboratory situation. The TOG-R will enable testing to be conducted in a familiar environment without the need for a researcher to travel to the subject's location, while maintaining increased precision. The ability to study children over time will enable researchers to evaluate the development of synaesthetes, including the possible existence of a critical period for the establishment of synaesthetic connections. It may also open up the possibility of definitively answering some of the questions surrounding the co-incidence of synaesthesia with other neurodevelopmental conditions, notably autism, Asperger syndrome and dyslexia, and will assist with genetic studies.

Of particular import to genetic studies of synaesthesia is the remote testing capability of the TOG-R, which significantly increases the number of potential subjects without sacrificing precision. As the success of a genetic study depends upon finding a common genetic factor among affected persons, it is vital for all potential subjects to be accurately phenotyped according to a set of common diagnostic criteria prior to inclusion in (or exclusion from) the study sample. Because synaesthesia is a condition with a low (.05%) estimated population prevalence (Baron-Cohen et al., 1996) an easily used, high-throughput remote testing protocol is vital for screening potential subjects; obtaining sufficient statistical power to detect a gene may well necessitate the recruitment of subjects worldwide. Remote testing capability will also play an important role in the assessment and investigation of rare phenotypes.

Trends in the colours of synaesthetic percepts (both within and between subjects) are much more easily perceived and analysed using the colour charts, and when combined with testing in a laboratory setting will enable a much more detailed analysis and fuller characterisation of visual synaesthesia. This has particularly important implications for genetic studies; our preliminary studies have revealed some hitherto unknown patterns in synaesthetic percepts. While it is clear that synaesthetic percepts are not identical among family members (Baron-Cohen et al., 1996) there may be subtler underlying trends within a family. Though further investigation is required, our preliminary investigations have revealed a family in which none of the four synaesthetes has any percepts in the violet region of the colour charts, indicating that familiality may extend to perceptual colour spectra. Since the TOG-R can be adapted for use with any form of visual synaesthesia, it will also be possible to compare colour trends across phenotypes, an area of particular interest when members of a single family display different visual synaesthetic phenotypes (e.g., auditory-visual and gustatory-visual).

The TOG-R also makes possible a deeper understanding of the differences between synaesthetic phenotypes, which have thus far been only roughly defined. In particular, though such an analysis is beyond the scope of the current paper, it offers the ability to compare colour spectra between "broad-band" and "narrow-band" variants of a single phenotype (e.g., auditory-visual synaesthetes who respond to all words *vs*. those who respond only to letters, numbers, days of the week and months of the year) as well as between phenotypes. Are narrow-band synaesthetes more likely to focus on primary colours? Does greater bandwidth necessarily mean a wider colour spectrum, or is it more likely to manifest as more combinations or patterns of relatively few colours?

The colour charts enable researchers to choose a particular set of stimuli, such as words beginning with a particular phoneme or grapheme, and to directly compare the quantitative differences and trends in the colours evoked. Of particular interest is the evidence from case reports that indicates that certain colours for particular stimuli are sometimes reported from multiple subjects, particularly for vowels (Marks, 1975). Is there something fundamental about these colours? Evaluating these trends with the colour charts may offer additional insight into the psycholinguistics of synaesthesia, and facilitate further investigation of the questions some researchers have posed about the role synaesthesia may play in metaphor (Ramachandran and Hubbard, 2001).

The TOG-R will also facilitate a more detailed analysis of stimulus-response relationships. Using the colour charts, the effect of varying particular aspects of a stimulus (such as volume with auditory stimuli or typeface with graphemic stimuli) on synaesthetic percepts can be directly observed.

While enabling a much more detailed analysis of the information elicited, utilisation of the TOG-R does result in some information loss. Synaesthetes are instructed to choose a maximum of two colours, while as indicated by EP's description of "Moscow" three or more colours may be present in a given percept (Baron-Cohen et al., 1987). Visually salient texture plays an

important role in many synaesthetes' percepts; E.P. describes "Daniel" as "shiny" as well as "deep purple, red and blue" (Baron-Cohen et al., 1987). In addition, some synaesthetes perceive shaped rather than abstract percepts, or perceive their percepts arranged across a particular mental "geography". While texture, shapes and geography are inadequately represented by the colour charts, they have evaded successful characterisation on any test thus far.

We have not thus far been able to adapt the TOG-R for use over the Internet due to the same limitations encountered by online versions of the original TOG: it is not currently possible to ensure that subjects take the test under the same technical conditions whilst avoiding the memory effects introduced by immediate re-testing. However, we are investigating ways of addressing these issues.

The creation of a common diagnostic instrument will greatly facilitate hitherto elusive cross-study comparisons, ensuring that the synaesthetes evaluated by one group are equivalent to those evaluated by another group, and allowing the characterisation of synaesthetes into phenotypes with a common meaning across research groups. Some researchers believe that there are significant differences in the distribution of synaesthetic phenotypes in different populations; we may now be able to address this question more definitively (Cytowic, 2002). The colour charts also make remote testing less subject to linguistic and cultural differences that may impact verbal descriptions, and can assist researchers in surmounting linguistic barriers. To this end, the colour charts have been distributed to research groups in the United States, the United Kingdom, Canada and Ireland[3].

Methods of diagnosing synaesthesia have undergone tremendous change since the time of the original case reports. Emphasis is shifting from verifying the existence of synaesthesia to a deeper analysis of synaesthetic percepts, the genetics of synaesthesia and the neural mechanisms underlying the condition. The TOG-R, a flexible diagnostic instrument combining the diagnostic accuracy and remote testing capability of the original TOG with the increased precision of previous colour chart methods, offers researchers a powerful tool for expanding the depth and breadth of synaesthesia research.

Acknowledgements. We dedicate this paper to the memory of Professor Jeffrey Gray, whose contributions to the field of synaesthesia research were greatly appreciated.

We are grateful to J.D. Mollon for helpful discussions, to J. Lawson, B.C.J. Moore and I. Cross for their advice and assistance in the preparation of the stimulus CD, to A.P. Monaco for advice and support, and to J. Hannah and P. Naimi for administrative support.

Julian E. Asher is a Cambridge Overseas Trust Scholar and the recipient of an Overseas Research Studentship, and has received support from the Wellcome Trust Centre for Human Genetics at Oxford. Simon Baron-Cohen was the holder of a research grant from the James S. McDonnell Foundation. Michael R.F. Aitken is a researcher in the MRC Centre for Behavioural and Clinical Neuroscience.

[3]Interested research groups should contact the corresponding author.

References

Baron-Cohen S, Burt L, Smith-Laittan F, Harrison J and Bolton P. Synaesthesia: Prevalence and familiality. *Perception, 25:* 1073-1079, 1996.

Baron-Cohen S, Harrison J, Goldstein LH and Wyke M. Coloured speech perception: Is synaesthesia what happens when modularity breaks down? *Perception, 22:* 419-426, 1993.

Baron-Cohen S, Wyke MA and Binnie C. Hearing words and seeing colours: An experimental investigation of a case of synaesthesia. *Perception, 16:* 761-767, 1987.

Cytowic RE. *Synesthesia: A Union of the Senses.* Cambridge, MA: MIT Press, 2002.

Day S. Some demographic and socio-cultural perspectives of synaesthesia. In Robertson LC and Sagiv N (Eds), *Synesthesia: Perspectives from Cognitive Neuroscience.* Oxford: Oxford University Press, 2005.

Dixon MJ, Smilek D, Cudahy C and Merikle PM. Five plus two equals yellow. *Nature, 406:* 365, 2000.

Galton F. Visualised numerals. *Nature, 21:* 494-495, 1880.

Gray JA, Williams SCR, Nunn J and Baron-Cohen S. Possible implications of synaesthesia for the hard question of consciousness. In Baron-Cohen S and Harrison J (Eds), *Synaesthesia: Classic and Contemporary Readings.* Cambridge: Blackwell Publishers, 1997.

Grossenbacher PG and Lovelace CT. Mechanisms of synesthesia: Cognitive and physiological constraints. *Trends in Cognitive Sciences, 5:* 36-41, 2001.

Harrison J and Baron-Cohen S. Synaesthesia: Reconciling the subjective with the objective. *Endeavour, 19:* 157-60, 1995.

Luria A. The mind of a mnemonist. Cambridge, MA: Harvard University Press, 1968.

Marks LE. On colored-hearing synaesthesia: Cross-modal translations of sensory dimensions. *Psychological Bulletin, 82:* 303-331, 1975.

Mattingley JB, Rich AN, Yelland G and Bradshaw JL. Unconscious priming eliminates automatic binding of colour and form in synaesthesia. *Nature, 410:* 580-582, 2001.

Maurer D. Neonatal synaesthesia: Implications for the processing of speech and faces. In Baron-Cohen S and Harrison J (Eds), *Synaesthesia: Classic and Contemporaryn Readings.* Cambridge, MA: Blackwell Publishers, 1997.

Merikle PM. Personal communication, May 2003.

Merikle PM and Dixon MJ. *http://watarts.uwaterloo.ca/~src/home.htm,* 2002.

Mills CB, Boteler EH and Oliver GK. Digit synaesthesia: A case study using a Stroop-type test. *Cognitive Neuropsychology, 16:* 181-191, 1999.

Myles KM, Dixon MJ, Smilek D and Merikle PM. Seeing double: The role of meaning in alphanumeric-colour synaesthesia. *Brain and Cognition, 53:* 342-345, 2003.

Nunn JA, Gregory LJ, Brammer M, Williams SC, Parslow DM, Morgan MJ, Morris RG, Bullmore ET, Baron-Cohen S and Gray JA. Functional magnetic resonance imaging of synesthesia: Activation of V4/V8 by spoken words. *Nature Neuroscience, 5:* 371-375, 2002.

Odgaard EC, Flowers JH and Bradman HL. An investigation of the cognitive and perceptual dynamics of a colour-digit synaesthete. *Perception, 28:* 651-664, 1999.

Ramachandran VS and Hubbard E. Psychophysical investigations into the neural basis of synaesthesia. *Proceedings of the Royal Society of London B, 268:* 979-983, 2001.

Smilek D, Dixon MJ, Cudahy C and Merikle PM. Synesthetic photisms influence visual perception. *Journal of Cognitive Neuroscience, 13:* 930-936, 2001.

Smilek D, Dixon MJ, Cudahy C and Merikle PM. Synesthetic color experiences influence memory. *Psychological Science, 13:* 548-552, 2002a.

Smilek, D, Moffatt BA, Pasternak J, White BN, Dixon MJ and Merikle PM. Synaesthesia: A case study of discordant monozygotic twins. *Neurocase, 8:* 338-342, 2002b.

Ward J and Simner J. Lexical-gustatory synaesthesia: Linguistic and conceptual factors. *Cognition, 89:* 237-261, 2003.

Wright B. *Timberwolf in the Tall Pines.* Salem: Rykodisc, 1990.

Julian Asher, Department of Psychiatry, University of Cambridge, Douglas House, 18b Trumpington Road, Cambridge CB2 2AH, UK. e-mail: jea41@cam.ac.uk