Dyslexic participants show intact spontaneous categorization processes

Dimitris S. Nikolopoulos
Department of Psychology
University of Crete

Emmanuel M. Pothos
Department of Psychology
Swansea University

in press, ***Dyslexia***.

Nikolopoulos, D. S. & Pothos, E. M. (2009). Dyslexic participants show intact spontaneous categorization processes. ***Dyslexia***, 15, 167-186.

Word count: 7362; Running head: Spontaneous categorization in dyslexia

Please address correspondence regarding this article to Dimitris Nikolopoulos, Department of Psychology, University of Crete, Rethymnon 74100, Greece, or to Emmanuel M. Pothos, Department of Psychology, Swansea University, Swansea SA2 8PP, UK. Electronic mail may be sent at: nikolopoulos@psy.soc.uoc.gr or to e.m.pothos@swansea.ac.uk

We examine the performance of dyslexic participants on an unsupervised categorization task, against that of matched non-dyslexic control participants. Unsupervised categorization is a cognitive process critical for conceptual development. Existing research in dyslexia has emphasized perceptual tasks and supervised categorization tasks (for which intact attentional processes are paramount), but there have been no studies on unsupervised categorization. Our investigation was based on Pothos and Chater's (2002) model of unsupervised categorization and the corresponding methodology for analyzing results. Across all performance indices and various data processing options we could identify no difference between dyslexic and non-dyslexic participants.

1. Introduction

Dyslexia is a complicated condition, not least because of disagreement as to whether it is better understood as a unitary condition, as opposed to a collection of related (but distinct) conditions (e.g., Bishop & Snowling, 2004; Bradley & Bryant, 1983; Miles, 1999). The attempts to characterize the cognitive deficits associated with dyslexia have, unsurprisingly, emphasized language-related deficits. For example, an influential research tradition has examined phonological deficits as a possible explanation for reading difficulties and dyslexia (e.g., Galaburda et al., 2006; Vellutino et al., 2004). Other researchers have suggested that the underlying causes of dyslexia are not phonological, but rather cognitive. For example, Nicolson, Fawcett and colleagues (e.g., Nicolson & Fawcett, 1990, 2007) argued that the origins of dyslexia have to do with a difficulty to automatise behavior. With respect to reading, difficulty in automatization translates to difficulty in (linguistic) fluency and hence dyslexia. In support of their hypothesis, Nicolson and Fawcett reported results showing dyslexics to have problems automatising skills in competence domains completely irrelevant to dyslexia (e.g., Nicolson & Fawcett, 2000). Another notable theory of dyslexia whose emphasis is not phonology is Stein's magnocellular deficit one, according to which difficulties with integrating information between the two visual pathways cause the problems in reading which are the basis of dyslexia (e.g., Stein, 2001).

In this vein, some researchers have examined non-linguistic deficits associated with dyslexia and, at the same time, tried to understand more carefully the cognitive processes in dyslexic participants which are actually intact: after all, dyslexics are able to function (mostly) without impairment in our complex modern world, their conceptual development appears identical to that of non-dyslexic people (e.g., Sylva-Pereyra et al., 2003), and they are routinely able to display the same level of intellectual achievement as non-dyslexic people (Miles, 1999).

Dyslexics do appear to have some impairment in their perceptual system. For example, Facoetti and Molteni (2001; Facoetti et al., 2000) found an asymmetric distribution of attention in a target identification task for children with a specific reading disorder, compared to normally reading children. More recently, Ahissar et al. (2006) reported that dyslexics were less able (compared to controls) to modulate their attention away and towards specific stimuli in various perceptual tasks. By contrast, when it comes to learning processes, the evidence indicates that dyslexic participants perform comparably to non-dyslexic ones. Kelly, Griffiths, and Frith (2002) examined the performance of dyslexic participants in a serial reaction time task. In a serial reaction time task, a target (typically a dot) appears on a computer screen, and participants have to identify the corresponding screen region (e.g., top left quadrant etc.). Unbeknownst to participants, the sequence of target locations is deterministic and typically identification of the target speeds up with practice. This result is taken to indicate that participants gradually learn the sequence of locations. In Kelly et al's results, while the dyslexic group responded on average slower than the non-dyslexic one (cf. Nicolson & Fawcett, 1994), in both groups there was evidence of awareness of the sequence of locations (but see Vicari et al., 2003, for a different result). More recently, Pothos and Kirk (2004; cf. Pothos, 2007) employed a learning task which could be instantiated with stimuli in different formats. Where the stimuli corresponded to sequences of shapes, dyslexic participants were impaired compared to controls. Where the stimuli appeared as embedded arrangements of shapes, dyslexic participants performed comparably to controls. This result was interpreted as showing that dyslexic people have generally intact learning processes, however, learning is often inhibited by problems with adequately perceiving the stimulus domain (as could be the case, for example, when it comes to linguistic development).

Our brief discussion above is hardly meant to correspond to an exhaustive review, rather our aim is simply to illustrate the range of cognitive processes which have been examined in association with dyslexia—and so motivate the emphasis of the present paper, which is categorization processes. Categorization processes are at the heart of our conceptual understanding of the world, and so of obvious importance for characterizing dyslexia. Research linking categorization processes and dyslexia is a lot less extensive, compared to the research traditions relating to perceptual and learning deficits. Petkov et al. (2005) reported that dyslexics performed worse (compared to non-dyslexic controls) on a grouping task of auditory stimuli. Their task required participants to listen to a sequence of tones and attempt to group a target tone with a reference one. Pernet et al. (2006; see also Pernet, Celsis, & Demonet, 2005) compared dyslexics and non-dyslexics on a simple categorization task. Participants saw two items at a time and they had to decide whether they belonged to the same category or not (three different categories were employed, Latin letters, geometrical figures, and Korean letters). Pernet et al. reported that dyslexics showed lower performance compared to controls.

Should results like those reported by Petkov et al. (2005) and Pernet et al. (2006) be taken to indicate that dyslexics indeed have a categorization deficit? Such a conclusion, if shown to be general, would have far-reaching implications, insofar that categorization is a process fundamental for the development of normal conceptual understanding of the world and the acquisition of knowledge. Categorization, however, is an extremely complex process, so the impairments briefly summarized above should not necessarily be attributed to a categorization impairment per se. This is the focus of the present article: to discuss various aspects of the categorization process, and then examine the performance of dyslexic participants and non-dyslexic controls in categorization, in a way that is not confounded by possible problems (of the dyslexics) with perceptual/ linguistic tasks.

Studies such as those of Petkov et al. (2005) and Pernet et al. (2006) can be thought of as concerning supervised categorization, the process of learning a particular set of categories. In supervised categorization, participants typically see a set of novel stimuli and they are told that each stimulus belongs to an imaginary category; category membership is indicated with linguistic labels. Their task is to discover the correct category assignment, with the help of corrective feedback. For example, when a participant sees a stimulus for the first time she will have to guess its category assignment and the experimenter will provide some information on whether the guess was correct or not. With subsequent presentations, the participant will be a little wiser as to which labels correspond to which items, and eventually will learn the required categorization. In practical terms, supervised categorization is the process that allows, for example, children to learn categories from adults or adults to learn novel concepts and categories. The studies of Petkov et al. (2005) and Pernet et al. (2006) are broadly analogous to studies of supervised categorization in that participants have to classify novel stimuli relative to some experimenter-defined classification. For example, even though Pernet et al. employed familiar stimuli, there was a particular normative assignment of stimuli to categories, against which participant performance was assessed (cf. Smits et al., 2002, for another example of supervised categorization research with familiar stimuli).

Intact supervised categorization is clearly important for normal conceptual development. However, in general, it will depend on intact attentional processes, since when learning a categorization for a novel set of objects not all dimensions of physical variation may be equally important. Even in the case of categorizing familiar stimuli (as with Pernet et al.), the assignment of stimuli to categories requires participants to perceive the stimuli in a particular way (Pernet et al. made a distinction between Latin and Korean letters, as opposed to, for example, letters of any kind and geometric shapes). For example, consider the items in Figure 1. Here and elsewhere, each dot corresponds to an object in psychological space, such that greater proximity indicates greater similarity (e.g., Shepard, 1987). In the categorization shown in Figure 1, dimension *y* is irrelevant and should be ignored by an efficient cognitive system; indeed, this is what is typically reported (e.g., Ashby, Queller, and Berretty, 1999). The fact that the cognitive system ignores dimensions that do not contribute to a required categorization for a set of items can be motivated theoretically (cf. Goodman, 1972; Pothos, 2005a). Moreover, attentional weighting is an integral part of influential computational models of supervised categorization (Minda & Smith, 2002; Nosofsky, 1988, 1989) and a mechanism that has received extensive empirical support as a cognitive process. Therefore, supervised categorization requires attentional weighting; if attentional processes are impaired in dyslexics, we would reasonably expect supervised categorization to be impaired as well. (Note that the studies of Pernet et al., 2006, and Petkov et al., 2005, somewhat deviate from the paradigmatic case of a supervised learning study described above and it would be important to replicate in future research the tentative conclusion here: that dyslexics have difficulty with supervised categorization tasks).

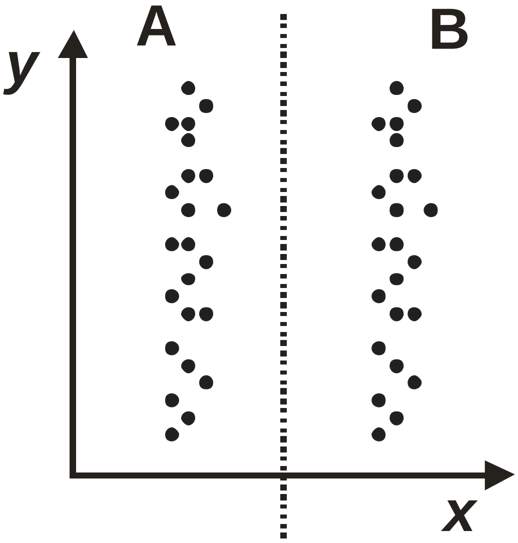


Figure 1. An illustration of the fact that supervised categorization requires selective attention: to learn to divide the items above into the A and B categories, dimension *y* needs to be ignored and the items need be processed along dimension *x*.

Supervised categorization requires some linguistic processes as well, however rudimentary. Learners have to associate labels with objects and the labels are always in linguistic form. If dyslexics have problems recognizing or differentiating between the available linguistic labels, they might likewise have difficulty making progress in a supervised categorization problem at the same pace as non-dyslexic participants.

Overall, in terms of understanding possible categorization deficits in dyslexia, we clearly have a problem: supervised categorization is closely confounded with cognitive processes that are known to be impaired in dyslexia: selective attention and recognition of linguistic labels. Therefore, arguably, the results of Petkov et al. (2006) and Pernet et al. (2006) do not show a categorization problem per se, but rather are simple manifestations of attentional / linguistic problems of dyslexics (note, again, that these investigators employed very simplified supervised categorization paradigms). Ideally, we would like to study dyslexic participants with a categorization task that is not confounded with either attention or linguistic competence. With respect to the latter, it is also worth noting that Haslam et al. (2007) have provided some evidence that spontaneous categorization ability may be unrelated to linguistic ability—our approach is analogous to their, but in the context of dyslexia and using a normative measure of categorization performance (Haslam et al. employed participants whose linguistic ability was deteriorating due to semantic dementia).

Unsupervised categorization may provide a solution and is the focus of the present investigation. In unsupervised categorization there are no set categories to be learned: participants are presented with a set of usually novel objects and are asked to divide them in any way that seems natural and intuitive. Unsupervised categorization is a cognitive process most linked with category coherence, our ability to recognize certain groupings of stimuli as more intuitive than others (Love, Medin, & Gureckis, 2004; Milton & Wills, 2004; Milton, Longmore, & Wills, in press; Murphy & Medin, 1985; Pothos & Chater, 2002; 2005). Computationally, it is a difficult process to study since for as few as 10 objects there are about 100,000 different alternative classifications (Medin & Ross, 1997).

Importantly, in unsupervised categorization there are no correct or wrong answers and therefore little role for selective attention. Also, participants do not need to identify different groupings with a linguistic label. Therefore, unsupervised categorization allows us to examine the intactness of categorization processes in dyslexic participants, independently (to a large extent) of attentional and linguistic deficits. The unsupervised categorization framework we employed in the present investigation is that of Pothos and Chater (2002, 2005). These researches developed a model to predict how naïve observers should classify a set of stimuli. This model is based on Rosch and Mervis's (1975) intuition that we should prefer groupings that minimize between category similarity while maximizing within category similarity, and the simplicity principle of perceptual organization (e.g., Chater, 1999; Hochberg & McAlister, 1953). The use of the simplicity principle is motivated from the intuitive resemblance between unsupervised categorization and perceptual organization and allows translating Rosch and Mervis's idea into a specific computational framework.

The 'simplicity' model of Pothos and Chater (2002) effectively examines how much the similarity information in a set of objects can be simplified by using categories. The similarity information can be simplified by using an operational definition for categories: that the similarities between objects in the same category should be greater than the similarities between objects in different categories. Therefore, for a set of objects, if categories can be found for which there are many such constraints, then the similarity information for these objects can be simplified considerably; for example, this would be the case for data sets A and B in Figure 2. By contrast, it is possible that we will not be able to identify such categories, as, for example, in the case of data set D in Figure 2. Pothos and Chater's (2002) model computes the codelength to describe the similarity structure of a set of items with categories, relative to the similarity structure of a set of items without categories, as a percentage: the lower this percentage, the more it is possible to simplify the description of the similarity structure of a set of items using categories. For example, the codelength of data set A is about 50%, while the codelength of data set D 80%. Pothos and Chater's (2002) model is specified within the minimum description length framework of algorithmic complexity (Risannen, 1978, 1987) and involves other considerations that are not presently relevant. Naïve observers typically do prefer to produce classifications of lower codelengths in unsupervised classification experiments (Pothos & Chater, 2002, 2005).

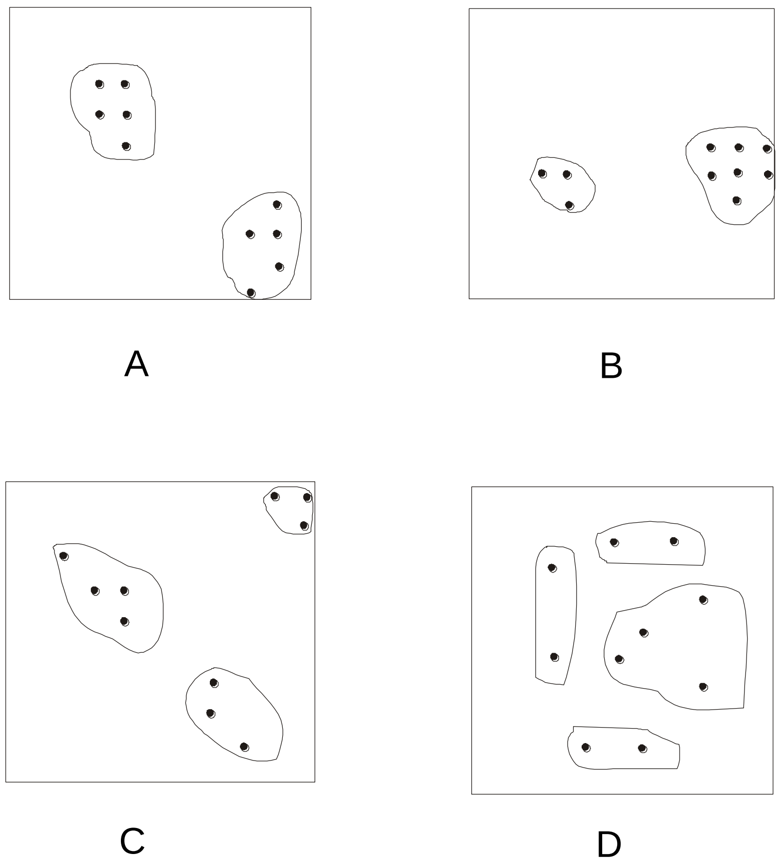


Figure 2. Arrangements of objects in psychological space, so that each arrangement varies in intuitiveness (adapted from Pothos & Chater, 2002). The groupings of objects are the ones predicted by the simplicity model to be most intuitive.

In a typical unsupervised categorization experiment, participants receive a set of objects and are asked to divide them into categories that appear natural and intuitive (see Milton & Wills, 2004, for alternative paradigms). Subsequently, the simplicity model can be applied to compute the codelengths associated with the classifications produced by the participants—these codelength values will be a measure of how optimal participants' performance on the unsupervised categorization task is.

2. Experimental investigation

2.1 Participants

239 native speakers of the Greek language with regular school attendance participated in the study. Participants were recruited from 6 primary schools (in different areas of Greece), as well as a high school—details are shown in Table. 1.

Table 1. Descriptive statistics of all children participating in the study. Due to a transcription error the origins of one of our participants were not available.

	N^{o} of recruited pupils	**Age in months**		
Grade		Mean	SD	Range
1st Primary	5	81.4	(4.5)	76 - 87
2nd Primary	46	92.3	(3.8)	86 - 103
3rd Primary	50	103.5	(3.9)	97 -115
4th Primary	29	114.8	(3.4)	109 - 123
5th Primary	40	127.5	(5.0)	120 - 149
6th Primary	41	141.0	(8.1)	114 -168
1st High Sch.	28	151.1	(5.2)	143 - 170
TOTAL	239			

2.2 Materials and Procedure

Participants were assessed with a range of academic, cognitive and (meta)linguistic tasks, as summarized in Table 2 and described in more detail below. The majority of these tasks were administered for the purposes of establishing dyslexia. Note that examinations of dyslexia in Greek have been less developed, so that a rather extensive assessment is required before establishing dyslexia with Greek speakers. Also, participants carried out a spontaneous categorization task, which corresponded to the dependent variable of interest. Below, we briefly summarize the tasks employed; for more details on the dyslexia assessment procedure the reader should refer to Nikolopouloς and Goulandris (2000), Nikolopoulos, Goulandris and Snowling (2003) and Nikolopoulos, Goulandris, Hulme and Snowling (2006); more details on the spontaneous categorization task can be found in Pothos and Chater (2002) and Pothos and Chater (2005).

Parental consent was sought prior to testing. When participants were assessed individually, this was done in a quiet room near their classroom. The individualized assessments were carried out during two testing sessions, each one lasting for about an hour and a half. Group testing always took place in the participants' classroom. Details on which tasks were administered individually/ to groups are shown in Table 2. Finally, where this was possible, the different assessments were presented in a randomized order.

Table 2. Overview of Testing – Testing Domains and Individual Tests

ACADEMIC		**COGNITIVE / (META)LINGUISTIC**	
READING SKILLS		PHONOLOGICAL AWARENESS	
Reading Real Words Task	(Ind.)	Phoneme Substitution Task	(Ind.)
Reading Pseudowords Task	(Ind.)	Spoonerisms Task	(Ind.)
SPELLING SKILLS		PHONOLOGICAL PROCESSING	
Spelling Real Words Task	(Gr.)	Color Naming Task	(Ind.)

Spelling -Months of the Year Task	(Gr.)	Object Naming Task	(Ind.)
ARITHMETIC		Digit Naming Task	(Ind.)
Basic Number Skills Task	(Gr.)	Letter Naming Task	(Ind.)
		MEMORY SKILLS	
		Pseudoword Repetition Task	(Ind.)
		Digit Recall Task	(Ind.)
		Recalling the Order of the Months	(Gr.)
		SYNTACTIC SKILLS	
		Syntactic Awareness Task	(Ind.)
		SPONTANEOUS CATEGORIZATION	
		Spontaneous Categorization Task	(Ind.)

(Ind.) = Individual Testing / (Gr.) = Group Testing

2.2.1 Assessment of Reading ability. Reading ability was assessed using two timed reading tests for real words and pseudowords. The first test involved 131 real words and the second 96 pseudowords. Word length, word frequency, and phonological complexity (presence of consonant clusters) were the three word selection criteria, so that different words (and pseudowords) were easier or more difficult to read. Participants were asked to read aloud the 131 words and 96 pseudowords. A Speed Criterion was derived by measuring the time taken (with a stopwatch) to read all words and an Accuracy Criterion was derived by observing the number and nature of children’s errors. Both tests have been previously used in studies assessing reading performance in Greek (Nikolopouloς & Goulandris, 2000; Nikolopoulos, Goulandris & Snowling, 2003; Nikolopoulos, Goulandris, Hulme & Snowling, 2006).

2.2.2 Assessment of Spelling Ability. Spelling ability was assessed using two spelling measurers: a) a spelling test of real words, as used in Nikolopoulos et al. (2006) and b) spelling the 12 months of the year. The spelling test consists of six sets of 12 words of graded spelling difficulty, so that the words in each set were chosen to be approximately suitable for each grade of the Greek elementary school (six grades in total). Participants were asked to spell all 72 words in the spelling test. Each trial in the task consisted of dictating a single word and providing participants with a short sentence containing the word (e.g., car → this car is very fast). Participants were allocated one point for each word spelled correctly.

2.2.3 Assessment of Basic Number Skills. Basic number skills were assessed using a version of the British Abilities Scales (BAS) Arithmetic sub-test (Elliott , Murray, & Pearson, 1983), adapted for Greek children. The test starts with a set of very easy, single-digit mathematical operations (e.g., additions: 2+7; subtractions: 6-3; multiplications: 2x7; and divisions: 6/2) and progresses to more difficult two-digit operations (e.g., additions with carrying, divisions where the divisor is bigger than the dividend, or operations involving fractions or decimals).

2.2.4 Assessment of Nonverbal Ability. Raven's Standard Progressive Matrices Test (Raven, 1987) was used to assess nonverbal ability. This was our main measure of cognitive ability. Although Raven's scores were not used in assessing dyslexia, they served as an important covariate in examining possible relations between spontaneous categorization and dyslexia. Scores on the Raven's test are often referred to below simply as IQ.

2.2.5 Assessment of Phonological Awareness Skills. Phonological awareness was assessed on the basis of a phoneme substitution task and a spoonerisms one. These tasks have been found to predict reading ability in Greek in both normal (Nikolopoulos et al. 2006) and dyslexic children (Nikolopoulos, Goulandris & Snowling, 2003). In the phoneme substitution task, children were asked either to exchange the initial phoneme of a given word with another phoneme provided by the examiner (e.g., νερό = /nεıræ/ – /g/ → /gεıræ/), or to substitute a prespecified phoneme in words that contained this phoneme twice in different positions (e.g., πα**τ** **τ**α = /paıtata/) for a new phoneme (e.g., change the phoneme /t/ with the phoneme /χ/ → /paıχaχa/). A total of 15 words were used in this test: 10 words in the first part and 5 words in the second part. For the spoonerism task, children were asked to exchange the first phoneme in each of 10 word pairs (e.g., **μ**αχαίρι–**π**ηρούνι = /maxεri–piruni/ → /**p**aχεri-**m**iruni/). In five word pairs the phoneme exchange had to be made from open CV words (e.g., as above), whereas in the other five word pairs it was made from words having a consonant cluster in the initial position (e.g., χρόνια–πολλά= /[1]χræni[1] pæıla/ → /ı**p**ræni[l] χæıla/). Children were allocated two points for each correct word pair, that is one point for each phoneme exchanged correctly.

2.2.6 Assessment of Phonological Processing Skills. Phonological processing skills were assessed with a rapid naming (RAN) task. It consisted of four components, each one containing five items repeated 10 times; the order of items was randomized in each component. The first component examined naming speed for simple words (umbrella, ball, scissors, tab, key), the second for colors (red, blue, yellow, brown, black), the third for digits (9, 2, 7, 4, and 5), and the last for letters (ε, σ, ο, λ, and β). Participants were asked to name all 50 items as quickly as possible without making errors. The time taken to name all 50 items was recorded for each category, as was the number of uncorrected errors.

2.2.7 Assessment of Syntactic Skills. Syntactic skills were assessed using a Greek translation of the Sentence Assembly Subtest, which is part of the Clinical Evaluation of Language Fundamentals–Revised assessment (Semel et al., 1987). The subtest evaluates children's awareness of syntactical and grammatical constraints. Participants were presented with random sequences of words and short phrases and they were asked to re-arrange these so as to produce meaningful sentences, in two ways (e.g., kicked, the girl, the boy → *The boy kicked the girl*). Note that in Greek there are alternative acceptable word orders for the same sentence. The subtest was composed of a total of 21 strings of words/phrases. Participants were allocated one point for each correct sentence, so that the maximum score was 42.

2.2.8 Assessment of Memory Skills. Three tasks were employed. First, we used a pseudoword repetition task, in which participants were asked to recall in the correct order a sequence of pseudowords spoken by the examiner. Memory was assessed separately with pseudowords having two, three, and four syllables. In each case, the two first memory trials involved repeating a list of only two words, the second two memory trials lists of three words etc. Testing was discontinued after two consecutive unsuccessful trials at a given sequence length. Participants were allocated one point for each list which was recalled in the correct order. Second, we used a digit recall task, in which participants were asked to recall in the correct serial order a sequence of digits spoken by the examiner. In the first section of the test, participants were asked to recall the digits in the same order as that spoken by the examiner, while in the second to repeat all the digits backwards. In the first two trials, participants had to recall a list of two digits, in the second two trials a list of three digits, etc.; the last two trials involved lists of eight digits. As before, testing was discontinued after two consecutive unsuccessful trials at a given sequence length. Participants were allocated one point for each list recalled in the correct order. Finally, participants were asked to write the 12 months of the year in the correct order, as a test of the ability to retrieve verbal codes from long-term memory (cf. Miles, 1983). One point was allocated for each month written in its correct position in the sequence, so that the maximum possible score in this task was 12.

2.2.9 Spontaneous Categorization. We selected a task which would correspond to a naturalistic grouping process as closely as possible.

We employed stimuli created on the basis of real starfish, so as to make them less abstract/ unreal (Figure 3). They varied along two dimensions, overall size and the size of a central distinct blob. We wished to avoid unidimensional, schematic stimuli, since such materials may lead participants to approach the categorization task in a contrived manner. Participants were given no information as to what the stimuli corresponded to, since spontaneous categorization for biological kinds may be different compared to spontaneous categorization for artifacts: accordingly, stimuli were presented in a neutral way, as ‘objects’. The coloring of the central blob was distinct from that of the rest of the stimuli, so as to enhance the perception that the central blob and overall size were two independent dimensions of variation for the stimuli. The overall stimulus size varied from 110mm to 200mm when printed on A4 sheets of paper in steps of on average 10mm: the same difference in overall size would be more conspicuous for smaller stimuli than for bigger ones (in accordance with Weber’s law in psychophysics), so that successive bigger stimuli differed by as much as 15mm whereas the smaller ones differed by as little as 7mm. In a similar way, the internal blob varied from 2mm to 40mm in steps of as little as 2mm and as great as 6mm (average: 4.2mm). The stimuli were designed with Corel Draw 8.0. Both dimensions were parameterized on a 1 to 10 scale.

Three different category structures were specified, each one consisting of 16 stimuli. A category structure is a collection of stimuli that participants were asked to categorize. In different category structures the similarity relations between the stimuli were different and likewise the most appropriate classification was different too. The three category structures are presented below (Figure 4), in terms of the 1-10 parameterizations of the two dimensions of physical variation of the stimuli.

Each stimulus was individually printed on a sheet of A4 in color.

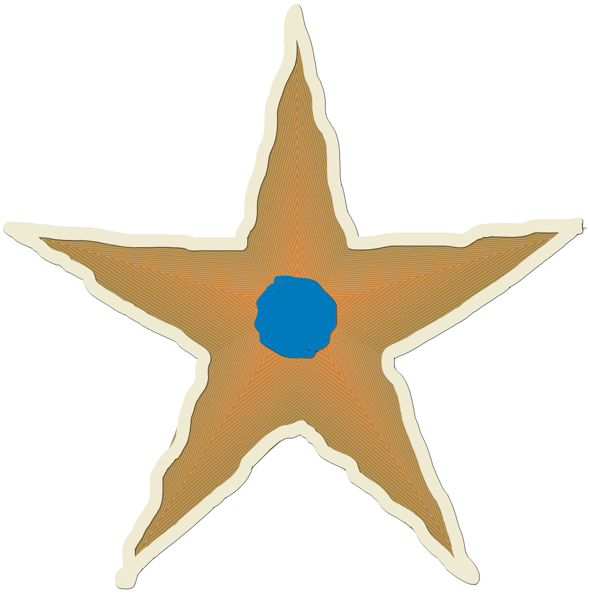
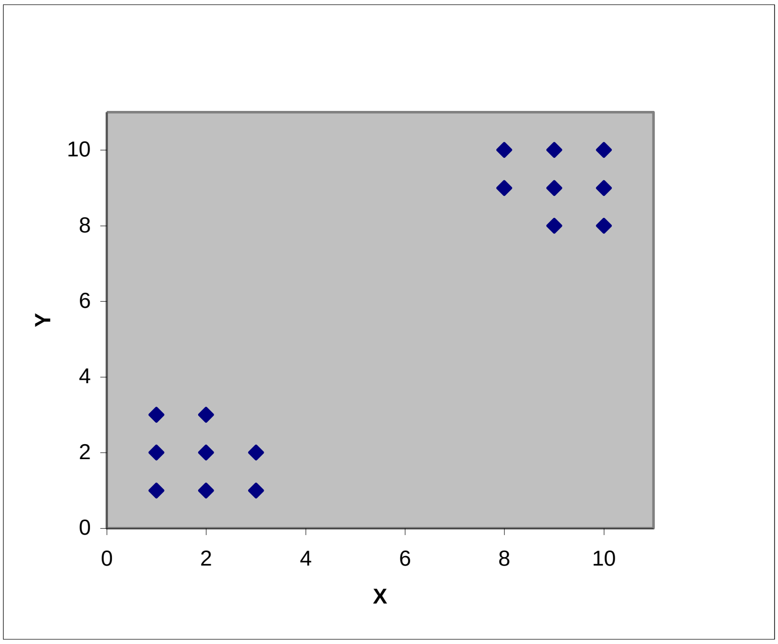
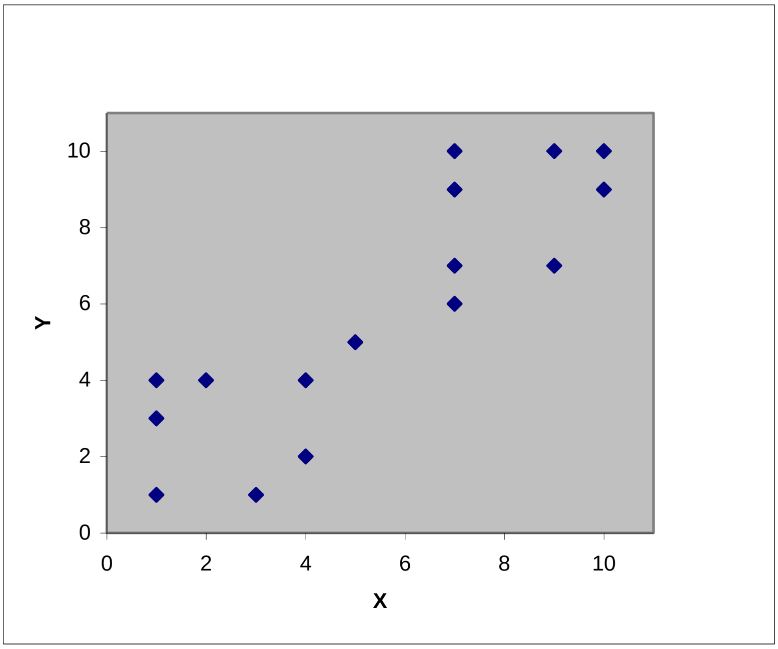


Figure 3. An example of the stimuli employed.


Y
X
0
2
4
6
8
10


Y
X
0
2
4
6
8
10

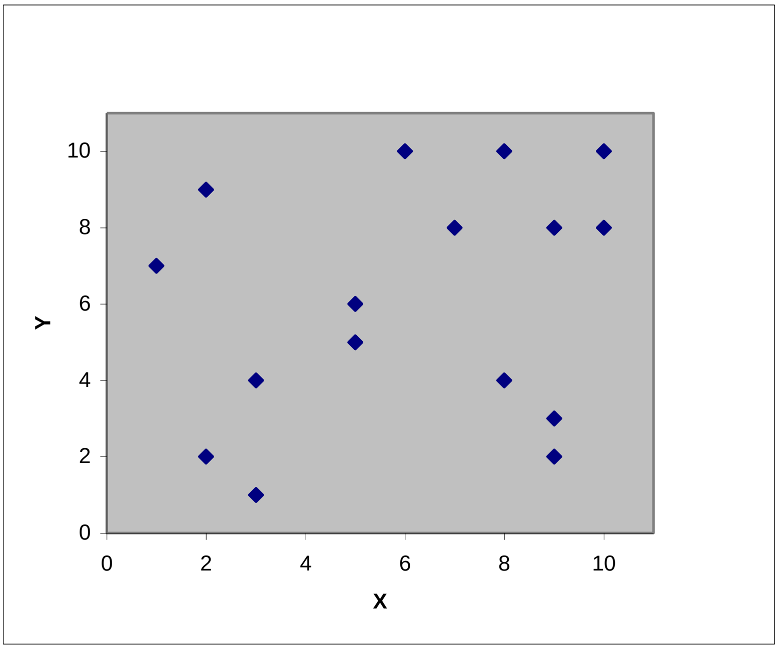


Figure 4. The three category structures employed in the experimental investigation. The first data set is referred to as the 'two clusters' one, the second as 'two clusters with noise', and the third as 'noise'.

An instructions sheet was given to participants, informing them that they were about to receive three sets of stimuli. Participants read that they should lay out the stimuli in each set in front of them and inspect them, before arranging them into groups that seemed 'natural and intuitive'. They also read that more similar objects should end up in the same group, and that they could use as many groups as they thought were necessary but no more. The stimuli in each set were stored in a folder; participants indicated their groupings by putting the corresponding stimuli into piles. Occasionally participants would ask further guidance as to how they should go about the task of grouping. They were simply reminded of the instructions. After a participant had finished grouping the stimuli in the first set, the experimenter put away the stimuli (in a way that the groupings were preserved) and presented the participant with the stimuli in the second set etc. The order in which each participant received the three data sets was randomized. The experiment lasted for about five minutes.

3. Results

3.1 Measures

We computed an academic-cognitive severity index on the basis of 14 criteria, based on the tasks outlined above. (We refer to our index as academic-cognitive severity, rather than dyslexia severity, since to establish dyslexia we require high scores on the severity index *and* normal/ high cognitive ability). The 14 criteria were: 1. Reading Real Words (Speed), 2. Reading Real words (Accuracy), 3. Reading Pseudowords (Speed), 4. Reading Pseudowords (Accuracy), 5. Spelling - Real Words, 6. Spelling – Months of Year, 7. Basic Number Skills, 8. Phoneme Substitution, 9. Spoonerisms, 10. Rapid Automatized Naming (Total of all 4 sub-tests), 11. Recalling Digits, 12. Recalling Pseudowords, 13. Syntactic Awareness, 14. Recalling Correct Order of the Months Tasks. The severity index score for each participant was based on allocating 'severity points' to the participant, depending on the number of standard deviations the participants' score differed from the mean performance of participants in the same age group (for each assessment task). Table 3 shows the allocation of severity points as a function of deviation from the mean score (for each assessment task). As can be seen in Table 3, a participant with average performance is expected to accumulate 14 severity points. Children with reading difficulties/ dyslexia risk would have an overall severity score greater than 14, while the score of more able children would be less than 14 points.

Table 3. Allocating severity points on the basis of performance on each of the tasks employed to assess academic-cognitive severity.

Good Performance	*+ 2 s.d.*	*0 severity points*
	+ 1 s.d.	*.5 severity points*
Average Performance		*1 severity point*
Low Performance	*- 1 s.d.*	*1.5 severity points*
	- 2 s.d.	*2 severity points*

On the basis of the academic-cognitive severity index, together with information about nonverbal IQ (assessed with the Raven's Matrices test), we classified our participants into three categories (i.e., we adopted the discrepancy definition of dyslexia): *dyslexics*, characterized by high academic-cognitive severity (severity score above 14 points) and average or high cognitive performance (Raven's Classifications: Average III-, Average III+, Above average, Superior); *low ability*, characterized by high academic-cognitive severity (severity score below 14 points), but also low cognitive performance (Raven's Classifications: Below, Below IV-, Impaired); and *high ability*, characterized by low academic-cognitive severity (severity score below 14 points) and average or good cognitive performance (Raven's Classifications: Average III-, Average III+, Above average, Superior). In this way we sought to avoid the confound between dyslexia and generally low academic performance (note that the potentially confounding effect of IQ is taken directly into account in the analyses as well).

Classification performance was measured on the basis of Pothos and Chater's (2002) model of unsupervised categorization. For each participant we considered his or her classification for the three stimulus sets, identified above as two clusters, two clusters with noise, and noise (Figure 4). Applying the simplicity model, we computed three percentage values that indicate how well the participants' classifications capture the similarity structure of the stimuli, for each of the three data sets. We will refer to these percentage values as codelengths. Lower codelengths imply 'better' classification performance. More specifically, lower codelengths imply that the participant's classification is closer to the best possible classification for a data set.

How good can the classification performance of a participant be? This depends on the actual data set. Depending on how well-separated the stimuli are, the least possible codelength might be lower (=better). For example, in the two clusters data set, the best possible classification is associated with a codelength of 50.2% (this value is near the lowest possible codelength value for 16 items; the two clusters data set corresponds to an extremely intuitive category structure). The two clusters with noise stimulus set is meant to correspond to stimuli for which there are some intuitions about a well-formed classification, but this classification is amidst noise (that is, there are stimuli that have no clear-cut classification). In this case, the best possible classification is associated with a codelength of 59.4% . Finally, the stimuli in the noise data set are semi-randomly arranged and the best possible classification is associated with a much higher codelength, 72.1%.

3.2 Analyses

We considered data from all 240 participants, whose average age was 118 months (9.8 years); age had a standard deviation of 21.4 months and ranged from 76 months to 170. Their average IQ score was 4.66, with a standard deviation of 1.75, a minimum of 1 and a maximum of 8. For the two groups of interest, (high ability, N=71; dyslexics, N = 119), average age and IQ were well-matched, as seen in Table 4. Below, we consider explicitly the possible confounding role of IQ and age in our comparisons between dyslexics and non-dyslexics, by partialling out variance due to IQ and age. Note that we identified a large number of dyslexic participants simply because we had carried out a preliminary screening of the student population we had access to for academic problems (mostly arithmetic and spelling). Subsequently, detailed participant assessment was performed only for those participants who were judged likely to be dyslexic, and their matched controls.

Table 4. Average age and IQ for the dyslexic (N=119), high ability (N=71), and low ability (N=50) participants in the sample. Age is measured in months. Raven's scores correspond to mapping a simple ordinal scale (1, 2, 3 etc.) to the classifications from the Raven's matrices test, so that lower scores correspond to lower IQ. Next to each mean the standard deviation is shown.

	Age	IQ
High ability	109.4/ 18.7	5.59/ 1.25
Dyslexics	119.1/21.9	5.16/ 1.26
Low ability	128.9/18.6	2.16/0.82

The main objective was to assess the extent to which classification performance of the non-dyslexic controls was different from that of dyslexics. In terms of the assignment of participants into different categories, on the basis of our dyslexia assessment, this concerned a comparison of classification performance between participants classified as dyslexics (N=119) and ones classified as high ability (N=71), thereby eliminating participants who were classified as low ability but not dyslexic (N=50). Classification performance was examined in terms of codelength values. Note that an alternative possibility would be to simply examine whether the classifications of dyslexic and non-dyslexic participants are similar, however, we don't believe such an approach to be appropriate. The key question here is not whether dyslexics and matched controls produce the same (or similar) classifications, but rather whether their classifications are equally optimal, relative to the measure of category intuitiveness postulated in the simplicity model. Accordingly, if dyslexic participants managed to classify the given stimuli in a way different to that of non-dyslexic controls, but equally optimal, we would still conclude that dyslexics have intact spontaneous classification processes. However, such a possibility is unlikely. Our experience with the simplicity model is that for structured datasets (that is, datasets for which there are some obvious clusters), classifications with similar and good codelengths will be likewise similar as well. Unfortunately, we have not been able to prove this result rigorously so far. Nonetheless, we have never observed two classifications with similar and good codelengths that are very different. Of course, the extent to which codelength is a valid normative measure of classification performance is an assumption, which may be refuted in future work.

We ran three between participant t-tests, whereby the dependent variables were classification performance on the two clusters, two clusters with noise, and noise stimulus sets. In all cases, the t-tests were not significant: $t(188)=1.29$, $p=.2$; $t(188)=0.02$, $p=.99$; $t(188)=1.22$, $p=.22$. Note that the average classification scores follow the ordering predicted by the model (the highest average codelength was achieved in the two clusters data set, the next highest in the two clusters with noise data set, and the worst in the noise data set).

As with all null results, care is needed to establish confidence in the result. First, and most importantly, note that, as can seen in Table 5, the means between dyslexic participants and high ability participants are *nearly identical*. To appreciate how small these differences are, consider that the codelength for the two clusters data set is about 50%. A random classification comprised of two clusters for this dataset would be associated with a codelength as high as 100.8%. Of course, we would not expect dyslexic participants to generate completely random classifications. However, in the context of a difference of 50 percentile units between the best and the worst possible classifications, a difference of 2.3 percentile units appears extremely small. Similar points apply to the other datasets (although slightly less so for the noise dataset, since in that case the difference between the best possible classification and a random classification is smaller). Second, we can compute the statistical power (for detecting a difference which exists) for the three comparisons. In order to do this, we have to specify the least difference in codelength for each dataset which we would consider meaningful. For example, consider the two clusters dataset. The *observed* difference was only 2.3 units. Even if such a difference were to be found significant, it is so small that we should conclude that the categorization performance of dyslexic participants is effectively equivalent to that of the matched controls. On the basis of previous studies where the ‘codelength’ measure has been used, we suggest that a meaningful difference in codelengths for a dataset would be no less than a fourth of the difference between the best possible codelength for a dataset and the worst possible codelength (which we here assume to be 100%). This is a fairly conservative estimate of what should be considered a meaningful difference in such studies. We can then use the pooled standard deviation for each dataset to compute a corresponding effect size and so the power for detecting such a difference value to be significant (at the 0.05 level). Such a power computation would inform of the likelihood of identifying a *meaningful* difference as significant, given the parameters of our study (sample sizes, standard deviations; the same approach for supporting a null hypothesis was adopted by Pothos, 2005b). In all cases power was .99. This may seem high, but it should be fully expected given the large sample size and the relatively small standard deviations. Note that an alternative possibility would be to compute effect sizes on the basis of the observed mean differences, rather than the meaningful mean differences. We think such an approach is misleading: as the actual difference between two means approaches 0, the corresponding effect size and power would both be zero. However, in such a case we would not wish to conclude that the experiment has low power, but rather that we have very high confidence in the observed null result (see also Pothos, 2005b).

We next consider three ways in which this null result might be misleading. We consider, and reject, each possibility in turn. First, it is possible that age and/or cognitive ability are confounding variables, that obscure an underlying difference in classification performance between dyslexics and high ability participants. Note that the academic-cognitive severity index we computed (as opposed to the indices simply assigning participants into different groups), correlated negatively with IQ (r=-.59,

$p<.01$) and positively with Age ($r=.26$, $p<.01$), highlighting the importance of these factors in dyslexia studies. However, there were no correlations between either age or IQ and classification performance (for each of the three stimulus sets). Moreover, for each of the three t-tests above, we ran a corresponding ANCOVA, where age and IQ were included as covariates, dyslexia (dyslexics vs. high ability participants) was the independent variable, and classification performance was the dependent variable (a separate ANCOVA was run for each of the three stimulus sets). None of the ANCOVA's was significant. For the two clusters stimulus set, $F(3, 186) = 0.46$, $p=.50$; for two clusters with noise, $F(3, 186) = 0.007$, $p=.93$; for the noise data set, $F(3, 186) = .34$, $p=.56$.

Second, it is possible that the way we computed dyslexia (which, recall, distinguishes between dyslexic participants and participants with low ability), was particularly stringent. In other words, if we were to relax a little bit the criterion between dyslexia and simply poor ability, we might obtain a difference in classification performance. We therefore computed a less stringent 'dyslexia' variable, whereby all our participants were classified as either poor or normal readers. The new N was 240, with 71 participants classified as normal readers and 169 as poor readers. Consistently with the approach adopted above, we ran three independent-samples t-tests to compare classification performance on each of the three stimulus sets, between poor and normal readers. None of the t-tests was significant, $t(238) = 1.62$, $p=.10$; $t(238) = 0.13$, $p=.89$; $t(238) = 1.51$, $p=.13$. ANCOVA's with IQ and age as covariates, as outlined above, were also not significant.

Third, the extent of the data collection necessitated the use of a large number of experimenters, nine in total. These experimenters were final year undergraduate students at the Department of Psychology of the University of Crete, who were collecting data for their final year dissertation. Although they had considerable training in valid and appropriate data collection methods (cf. Orne, 1962), we cannot preclude the possibility that some of them might have tried to encourage participants to produce well-formed clusters. Indeed, ANOVA's with classification performance as the dependent variable and examiner as the independent variable, were significant for all three stimulus sets (we considered data from all participants, not just dyslexics vs. high ability participants): for the two cluster data set, $F(8,231) = 2.60$, $p = .01$; for the two clusters with noise data set, $F(8,231) = 3.33$, $p =.001$; for the noise data set, $F(8,231) = 3.30$, $p = .001$. We therefore, eliminated data from the three examiners for whom average classification scores were highest (these were the same for the three data sets, underscoring the possibility that some examiners may have been particularly encouraging towards their participants). Doing this, ensured that the 'examiner' effect in the three ANOVA's above became non-significant. Subsequently, we repeated the procedure of running t-tests to compare classification performance on the three stimulus sets between dyslexics (new N = 96) and high ability participants (new N = 54). None of the t-tests were significant; concerning performance with the two clusters data set: $t(148) = 1.64$, $p=.10$; for the two clusters with noise data set, $t(148) = 0.69$, $p=.49$; for the noise data set, $t(148) = 1.36$, $p=.18$. ANCOVA's with age and IQ as covariates were also non-significant.

Table 5. Mean classification performance between dyslexic participants and high ability participants. Classification performance is assessed using Pothos and Chater's (2002) model of unsupervised categorization. The lower the percentage, the 'better'

the produced classification. Percentage values vary from around 50% (a value which would correspond to an extremely intuitive classification) to 100% (which would correspond to either a classification that entirely fails to capture any cluster structure in a data set, or a classification on a data set that does not have any cluster structure in itself). The value of one standard deviation is given next to each mean.

	Two clusters	Two clusters with noise	Noise
Dyslexics	84.6/12.7	88.2/7.6	93.9/5.1
High ability participants	86.9/11.4	88.3/7.9	94.8/4.9

4. Discussion and conclusions

There has been extensive research examining possible perceptual and learning deficits of dyslexic participants relative to controls (e.g., Facoetti & Molteni, 2001; Pothos & Kirk, 2004). Categorization processes have been relatively under-researched, a problematic situation considering the importance of categorization in a person's normal conceptual and intellectual development. The few studies examining categorization in dyslexic participants did not report encouraging results: using a simplified supervised categorization paradigm, both Pernet et al. (2006) and Petkov et al. (2005) reported categorization deficits for dyslexic participants.

The purpose of this article was first to explain that supervised categorization is a poor test of categorization performance with dyslexic participants, since it is confounded both with selective attention and linguistic processes. Accordingly, an examination of categorization processes in dyslexic participants is most appropriately carried out within an unsupervised categorization paradigm. The second purpose of the article was to present the experimental and computational methodology for such an investigation (based on Pothos & Chater's, 2002, model) and so carry out a comparison of dyslexic and non-dyslexic participants on an unsupervised categorization task.

Our results can be straightforwardly summarized: we found no evidence that dyslexic participants performed any differently from non-dyslexic participants. As the methodology in unsupervised categorization becomes more sophisticated (e.g., Milton et al., in press; Pothos & Chater, 2005), it will be possible to carry out comparisons between dyslexic and non-dyslexic participants in alternative unsupervised categorization tasks. For the time being, we can conclude that there is no evidence for a deficit in unsupervised categorization in dyslexics. Our results are in correspondence with the conclusion of Haslam et al. (2007), who also reported that spontaneous categorization performance and language ability seem unrelated. An interesting general question arising from such research is exactly what is the relation between language and our conceptual understanding of the world.

Acknowledgements

This research was partly supported by ESRC grant R000222655 and EC Framework 6 grant contract 516542 (NEST) to the second author. We would like to thank Alan Beaton, Angela Fawcett, Glyn Humphreys, Padraic Monaghan, Akis Simos, and Andy Wills for their helpful advice.

References

Ahissar, M., & Lubin, Y., Putter-Katz, H., & Banai, K. (2006). Dyslexia and the failure to form a perceptual anchor. Nature Neuroscience, 9, 1558-1564.

Ashby, F. G., Queller, S., & Berretty, P. M. (1999). On the dominance of unidimensional rules in unsupervised categorization. Perception & Psychophysics, 61, 1178-1199.

Bishop, D. V. M. & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: same or different? Psychological Bulletin, 130, 858-886.

Bradley, L. & Bryant, P. E. (1983). Categorizing sounds and learning to read: A causal connection. Nature, 301, 419-421.

Elliot, C. D., Murray, D. J. and Pearson, L. S. (1983). British Abilities Scales. Windsor: NFER-Nelson.

Facoetti, A. & Molteni, M. (2001). The gradient of visual attention in developmental dyslexia. Neuropsychologia, 39, 352-357.

Facoetti, A., Paganoni, P., Turatto, M., Marzola, V., & Mascetti, G. (2000). Visual-spatial attention in developmental dyslexia. Cortex, 36, 109-123.

Galaburda, A. M., LoTurco, J., Ramus, F., Holly Fitch, R., & Rosen, G. D. (2006). From genes to behavior in developmental dyslexia. Nature Neuroscience, 9, 1213-1217.

Goodman, N. (1972). Seven strictures on similarity. In N. Goodman, Problems and projects (pp. 437-447). Indianapolis: Bobbs-Merrill.

Haslam, C., Wills, A. J., Haslam, A., Kay, J., Baron, R., McNab, F. (2007). Does maintenance of colour categories rely on language? Evidence to the contrary from a case of semantic dementia. Brain and Language, 103, 251-263.

Hochberg, J. E., & McAlister, E. (1953). A quantitative approach to figural goodness. Journal of Experimental Psychology, 46, 361-364.

Kelly, S. W., Griffiths, S., & Frith, U. (2002). Evidence for implicit sequence learning in dyslexia. Dyslexia, 8, 43-52.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. Psychological Review, 111, 309-332.

Medin, D. L. & Ross, B. H. (1997). Cognitive psychology. (2nd Ed.). Fort Worth: Harcourt Brace.

Miles, T. R. (1983). Bangor Dyslexia Test. LDA Camps.

Miles, T. R. (1999). Dyslexia: A hundred years. Second edition. Buckingham: Open University Press.

Milton, F. & Wills, A. J. (2004). The influence of stimulus properties on category construction. Journal of Experimental Psychology: Learning, Memory, and Cognition, 30, 407-415

Milton, F., Longmore, C. A., & Wills, A. J. (in press). Processes of overall similarity sorting in free classification. Journal of Experimental Psychology: Human Perception and Performance.

Minda, J. P. & Smith, J. D. (2002). Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 28, 275-292.

Murphy, G. L. & Medin, D. L. (1985). The Role of Theories in Conceptual Coherence. Psychological Review, 92, 289-316.

Nicolson, R. I. & Fawcett, A. J. (2007). Procedural learning difficulties: reuniting the developmental disorders? Trends in Neurosciences, 30.

Nicolson, R. I. & Fawcett, A. J. (2000). Long-term learning in dyslexic children. European Journal of Cognitive Psychology, 12, 357-393.

Nicolson, R. I. & Fawcett, A. J. (1990). Automaticity: A new framework for dyslexia research? Cognition, 35, 159-182.

Nikolopoulos, D. & Goulandris, N. (2000). The cognitive determinants of reading and spelling abilities in a regular orthography". In M. Perkins and S. Howard (Eds.*)* New Directions in Language Development and Disorders. Plenum Publishing Corporation, England.

Nikolopoulos, D., Goulandris, N., & Snowling, M. (2003). Developmental dyslexia in Greek. In N. Goulandris (Ed.). Dyslexia in Different Languages: Cross-linguistic Comparisons. Whurr Publishers, London, UK.

Nikolopoulos, D., Goulandris, N., Hulme, C., & Snowling, M. (2006). The cognitive bases of learning to read and spell in Greek: Evidence from a longitudinal study. Journal of Experimental Child Psychology, 94, 1-17.

Nosofsky, R. M. (1988). Similarity, frequency, and category representation. Journal of Experimental Psychology: Learning, Memory and Cognition, 14, 54-65.

Nosofsky, R. M. (1989). Further tests of an exemplar-similarity approach to relating identification and categorization. Journal of Experimental Psychology: Perception and Psychophysics, 45, 279-290.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. American Psychologist, 17, 774-783.

Pernet, C., Valdois, S., Celsis, P., Demonet, J. F. (2006). Lateral masking, levels of processing and stimulus category: A comparative study between normal and dyslexic readers. Neuropsychologia, 44, 2374-2385.

Pernet, C., Celsis, P., & Demonet, J. F. (2005). Selective response to letter categorization within the left fusiform gyrus. NeuroImage, 28, 738-744.

Petkov, C. I., O'Connor, K. N., Benmoshe, G., Baynes, K., & Sutter, M. L. (2005). Auditory perceptual grouping and attention in dyslexia. Cognitive Brain Research, 24, 343-354.

Pothos, E. M. (2005a). The rules versus similarity distinction. Behavioral & Brain Sciences, 28, 1-49.

Pothos, E. M. (2005b). Expectations about stimulus structure in implicit learning. Memory & Cognition, 33, 171-181.

Pothos, E. M. (2007). Theories of Artificial Grammar Learning. Psychological Bulletin, 133, 227-244.

Pothos, E. M. & Kirk, J. (2004). Investigating learning deficits associated with dyslexia. Dyslexia, 10, 61-76.

Pothos, E. M. & Chater, N. (2002). A Simplicity Principle in Unsupervised Human Categorization. Cognitive Science, 26, 303-343.

Pothos, E. M. & Chater, N. (2005). Unsupervised categorization and category learning. Quarterly Journal of Experimental Psychology, 58A, 733-752.

Ravens (1987). Manual for Raven's Progressive Matrices and Vocabulary Scales, Section 3. Standard Progressive Matrices. London: H. K. Lewis and Co.

Rissanen, J. (1978). Modeling by shortest data description. Automatica, 14, 465-471.

Rissanen, J. (1987). Stochastic complexity. Journal of the Royal Statistical Society, Series B, 49, 223-239.

Rosch, E. & Mervis, C. B. (1975). Family Resemblances: Studies in the Internal Structure of Categories. Cognitive Psychology, 7, 573-605.

Semel, E., Wiig, E. H., Secord, W. (1987). Clinical Evaluation of Language Fundamentals-Revised. Examiners Manual. The Psychological Corporation, Harcourt Brace & Company, Publishers.

Shepard, R. N. (1987). Toward a Universal Law of Generalization for Psychological Science. Science, 237, 1317-1323.

Smits, T., Storms, G., Rosseel, Y., & De Boeck, P. (2002). Fruits and vegetables categorized: An application of the generalized context model. Psychonomic Bulletin & Review, 9, 836-844.

Stein, J. (2001). The sensory basis of reading problems. Developmental Neuropsychology, 20, 509-534.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? Journal of Child Language and Psychiatry, 45, 2-40.

Vicari, S., Marotta, L., Menghini, D., Molinari, M., & Petrosini, L. (2003). Implicit learning deficit in children with developmental dyslexia. Neuropsychologia, 41, 108-114.