# Do Designers Show Categorical Perception of Typefaces?

*Mary C. Dyson*

UNIVERSITY OF READING — DYSON, 193-220 — *VISIBLE LANGUAGE* 45.3

ABSTRACT

Readers need to easily discriminate between different letters, so typefaces are designed to make these differences distinctive. But there is also a uniformity of style within a typeface. These styles are recognized by typographic designers and may be categorized to enable more efficient discrimination among typefaces. The manner in which designers perceive typefaces is explored using the paradigm of Categorical Perception (CP). A continuum of fonts is created by interpolating between two typefaces, and two tasks (identification and discrimination) are used to test for CP. As the application of CP to typefaces is a new approach, various methodological issues are pursued. The experiments reveal that the conditions required to demonstrate CP are quite specific and CP was only evident in Times and Helvetica and not Garamond and Bodoni. Possible reasons for this difference are the characteristics of the two typefaces and their context of use. Speculation as to the purpose of CP in non-designers raises the under-researched question of how we identify letters in different typefaces when reading.

## INTRODUCTION

Text typefaces are usually designed to be legible and unobtrusive. Readers are skilled at coping with variability in letterforms, recognizing letters regardless of the typeface. But experts learn to distinguish among concepts in their domain in a way that novices cannot do (Goldstone, 1994), e.g., ornithologists can identify many species of birds. Typographic designers are taught to categorize typefaces, through historical classifications or analyses of design features. How does this affect their perceptual abilities?

### CATEGORICAL PERCEPTION

Categorical Perception (CP) is a psychophysical phenomenon whereby we perceive categories where none exist in the world. A clear example is color perception as the color spectrum is a natural continuum of light frequencies but we perceive discrete categories, rather than gradually changing hues (Bornstein, 1987). The origins of CP are in speech perception and identification of phoneme boundaries from an acoustic continuum (Liberman, Harris, Hoffman and Griffith, 1957). However, the prevalence of CP is evident in the examples covered by Harnard (1987) which include CP in humans and animals, infants and adults, and different senses. Since Harnard's compilation, CP has been demonstrated in many other visual domains, e.g., familiar objects (Newell and Bülthoff, 2002), orientation of lines (Quinn, 2004), American Sign Language (Emmorey, McCullough, and Brentari, 2003) and various aspects of face perception. These include facial identity (Beale and Keil, 1995), facial expressions (Calder, Young, Perrett, Etcoff and Rowland, 1996), gender (Bülthoff and Newell, 2004), race (Levin and Angelone, 2002), familiar versus unfamiliar faces (e.g., Campanella, Hanoteau, Seron, Joassin and Bruyer, 2003; Bülthoff and Newell, 2004; Angeli, Davidoff and Valentine, 2008) or newly-learned faces (Viviani, Binda and Borsato, 2007).

Faces are a useful source of comparison for typefaces as unlike colors, which correspond to changes in a single dimension (the wavelength of light), they are multidimensional, like typefaces. Faces vary in the size and shape of various features, the relative distance between them, etc. Typefaces vary in weight, contrast, proportions, basic shapes, terminals and serifs (Baines and Haslam, 2005). Also, faces and typefaces do not exist as continuous shapes within the real world; there are no naturally occurring continua. This presents challenges in creating appropriate visual material for testing. Although software tools have enabled the creation of continua between multidimensional shapes (Campanella et al, 2003), applying these tools to generate typeface continua requires careful consideration. It is important that the interpolation procedure produces a linear continuum (equal physical steps). This is straightforward with color or orientation, but not with multidimensional stimuli (Newell and Bülthoff, 2002). With typefaces, the results of the interpolation must be a set of plausible fonts.

### TRAINING

Although the shapes of letters may have been selected to be similar to the shapes in our environment to exploit our general object recognition mechanisms (Changizi, Zhang, Ye and Shimojo, 2006), typefaces are not universal and styles change over time. The ability to distinguish among them must therefore be learned. Evidence for the effect of design training on perceptual abilities comes from a study using Art students at Rhode Island School of Design who took a color theory course. These students were superior to non-experts in some aspects of color perception (Burns and Shepp, 1988).

The influence of training on the learning of categories has also been demonstrated in the field of music. A study found that trained musicians show categorical perception of musical intervals and those without musical training do not (Burns and Ward, 1978). A later study included listeners with a wider range of 'musicality,' measured using three tests, and found that the degree of CP appears to relate to the level of musicality (Howard, Rosen and Broad, 1992). They conclude that CP may develop through a process of learning, so the extent to which CP is evident may be influenced by the amount of training and level of exposure to relevant stimuli.

Goldstone (2004) investigated the nature of the development of categories through training participants to make perceptual discriminations of the saturation, brightness and size of squares. The experiments explored whether the training resulted in the acquisition of distinctiveness or similarity. Acquired distinctiveness is an increase in perceptual sensitivity to differences that are relevant to a categorization (Gibson, 1969). With typefaces, this would mean learning the specific dimensions along which they vary (as described above) to improve discrimination. This contrasts with acquired similarity where discrimination is poorer due to learning to be less sensitive to differences that do not affect categorization. This difference is important as categories are considered a means of enabling more efficient processing (Harnard, 1987).

If design training facilitates categorization at the expense of less sensitivity to differences in visual details, this would be counter-productive. Differences in the perception of semantic quality of type between designers and non-designers have been found (Bartram, 1982), which demonstrate that there are qualitative differences in perception due to training.

### TESTING FOR CP

The CP paradigm is a way of exploring qualitative differences in a quantitative way (Harnard,1987). Different categories reflect qualitative differences. Two tasks are used to test for CP: identification and discrimination. Participants are asked to identify examples from along the continuum of equal steps and a category boundary is determined where half the responses are at one end of the continuum and half the other. Discrimination is measured by comparing two examples which may come from the same point on the continuum or from different points. If participants find it harder to tell the difference between two examples that both fall on one side of the category boundary, as compared with two examples that straddle the boundary, then there is evidence of CP. Equal physical differences are perceived as larger or smaller depending on whether or not they are in the same category.

Three criteria have been proposed as tests for CP (Studddert-Kennedy, Liberman, Harris and Cooper, 1970), although not all the studies mentioned above use all three. They are (i) consistent identification within categories and a steep gradient between categories; (ii) peak of discrimination at category boundary with poorer discrimination within categories; (iii) correspondence between discrimination performance and predicted discrimination based on identification data.

There is some latitude in the extent to which discrimination can be predicted from identification performance. Absolute CP is not usually found, typically discrimination is reasonable within categories (Burns and Campbell, 1994). The alternative to CP is more continuous perception where there are similar levels of discrimination performance within and between categories, i.e., across the whole continuum. In such cases, discrimination is not predictable from the boundary established in the identification task. In looking at orientation perception, Quinn (2004) found CP and more continuous perception: near vertical lines are perceived categorically, and near horizontal lines are perceived more continuously. The reason for this asymmetry is not resolved.

The nature of the discrimination task can affect the extent to which CP is demonstrated. Where there is greater consistency from trial to trial (lower stimulus uncertainty), and the contribution of memory is minimized (Macmillan, 1987), then people may be able to make finer discriminations (Burns and Ward, 1978). Hence, using fixed discrimination (where the same two examples are repeated within a block of trials) may be less likely to demonstrate CP than roving discrimination (where pairs of examples come from different points over the whole continuum and therefore examples change from trial to trial).

### PILOT TESTS

As CP may only occur in an experimental setting under certain conditions (as described above), a series of pilot tests was carried out to look for evidence of CP in typefaces. These explored the discrimination task, varying the nature of the visual material and how it was presented on screen. The pilots started with the same single letter string (hamburefonsiv) for each example, then introduced jumbled strings (e.g., mvoesfaurhnib , fasvumebhrion), and finally adopted paragraphs of letter strings. The size of type was reduced from 48pt to 36pt to cater for the increase in number of letters. Other variations that were explored included whether examples were displayed simultaneously (one above the other) or sequentially, and the timings (i.e., length of time examples were displayed on the screen and interval between displaying two examples).

The two aims of the modifications to the task were to achieve a level of performance that was somewhere between chance and 100 per cent correct and to reduce the variability among participants. The pilot data showed little consistency across participants as to which discriminations were easier. When participants were asked at the end of the experiment what strategy they were adopting to do the task, a possible reason for this inconsistency emerged. Participants were commonly focusing on specific elements of letters or specific letters, but the target of the focus varied (e.g., left curve of the 'e' and terminal of the 'f'; first used 'a' then moved on to shape of the 's').

Although this attention to detail may be a natural response (especially from typographers) when small details are changing, it was not considered to be an optimum strategy. Features of other letters would have provided a greater number of possible cues for discrimination and sometimes better cues. It was hypothesized that the combination of cues that differentiate typefaces (e.g., stroke width, shape of serif, contrast between thick and thin strokes) would provide a stronger basis for categorization. In exploring the categorization of speech, Gerrits and Schouten (2004) used an experimental design that aimed to motivate listeners to focus on the speech signal as a whole. Another reason for changes to the current task was therefore to encourage participants to look at the material as a whole (i.e., globally), rather than focus on particular details. Presenting a larger volume of text for a shorter time was considered a possible means of eliciting this global way of looking.

#### RESEARCH QUESTIONS

The research initially planned to compare designers and non-designers to discover whether CP is limited to designers. However, the difficulty of establishing the precise conditions in which to test made this comparison untenable until reliable evidence for CP in designers was established. This therefore became the aim of the research. Exploring whether CP could be demonstrated involved various methodological questions:

- → Is a fixed discrimination task less sensitive a measure of CP than a roving discrimination task because of the greater consistency from trial to trial (less stimulus uncertainty)?
- → Can identification performance predict discrimination?
- → Are similar results found for a continuum with two more similar typefaces (i.e., two serif faces) compared with a continuum from serif to sans serif?
- → Do individual participants show greater correspondence between their identification and discrimination performance than averaging across participants, as the location of category boundaries may vary from person to person?

### GENERAL METHOD

#### PARTICIPANTS

Participants were volunteer students within the Department of Typography & Graphic Communication at the University of Reading, UK. They ranged from students in their second or third year of the undergraduate BA course, to Masters students in Typeface Design, Book Design and Information Design, to PhD students. They had all received some education in typographic or graphic design, but the nature and extent of their teaching and learning varied. Due to the rather limited pool of volunteers, no comparisons of different levels of design training were made. This might also have conflicted with the ethical demands of the study. It was made clear to participants that the project, their participation and the results were not part of any assessment on their programs. The research project received ethical review according to the procedures specified by the University of Reading Research Ethics Committee and was allowed to proceed.

Figure 1: Twelve points on the Timevetica continuum from Times (1) to Helvetica (12)

1 hamburefonsivHBUREFOSI.,
2 hamburefonsivHBUREFOSI.,
3 hamburefonsivHBUREFOSI.,
4 hamburefonsivHBUREFOSI.,
5 hamburefonsivHBUREFOSI.,
6 hamburefonsivHBUREFOSI.,
7 hamburefonsivHBUREFOSI.,
8 hamburefonsivHBUREFOSI.,
9 hamburefonsivHBUREFOSI.,
10 hamburefonsivHBUREFOSI.,
11 hamburefonsivHBUREFOSI.,
12 hamburefonsivHBUREFOSI.,

1 hamdurefonsivHAMDUREFONSIV.,
2 hamdurefonsivHAMDUREFONSIV.,
3 hamdurefonsivHAMDUREFONSIV.,
4 hamdurefonsivHAMDUREFONSIV.,
5 hamdurefonsivHAMDUREFONSIV.,
6 hamdurefonsivHAMDUREFONSIV.,
7 hamdurefonsivHAMDUREFONSIV.,
8 hamdurefonsivHAMDUREFONSIV.,
9 hamdurefonsivHAMDUREFONSIV.,
10 hamdurefonsivHAMDUREFONSIV.,
11 hamdurefonsivHAMDUREFONSIV.,
12 hamdurefonsivHAMDUREFONSIV.,

**Figure 2: Twelve points on the Garadoni continuum from Garamond (1) to Bodoni (12)**

## PREPARATION OF MATERIALS

To apply a test for CP to typefaces requires the creation of a continuum of equal steps between two existing typefaces. The experiments reported below used two continua: one from Times to Helvetica (Experiments 1, 2a and 2b) and one from Garamond to Bodoni (Experiments 3, 4a and 4b). These continua were developed by students and graduates of the MA Typeface Design program who have the technical and design skills to produce suitable letter forms. These students were not participants in the experiments.

Times New Roman and Helvetica were chosen for the first continuum (labeled Timevetica) because they are common examples of a serif and sans serif typeface. The second continuum, between Stempel Garamond and Berthold Bodoni (Garadoni), was introduced to extend the research beyond a single pair of typefaces and explore two serif typefaces. It also avoided some of the difficulties in creating a transition from a serif to no serif. The 12-step continua were produced in Fontlab using linear interpolation; this number of steps provides the opportunity to test quite subtle differences.[1]

When mixing typefaces, some letters at the two end points of the continuum do not share the same structure and have a different number of points to represent the curves. Therefore, in order to be able to interpolate, points were manually added to the Bézier curves, with attention paid to their position as this is critical in determining the resulting intermediate shapes. These were placed so as to result in shapes that would appear natural, i.e., might be found in other typefaces, judgments that are typically part of the process of type design. Letters were either manually or automatically hinted to improve the appearance of individual forms. The complexity of this procedure and the incompatibility of some letters constrained the number of letters that were produced for the test material. However, these were considered sufficient given the exploratory nature of the experiments.

Timevetica included the letters that spell 'hamburefonsiv.' These letters cover some of the different shapes within letters (e.g., round, vertical and diagonal forms). Upper case forms were created for nine of these letters (four were excluded due to problems with interpolation) and punctuation (full point and comma) was added to simulate sentences. The continuum from Times (font 1) to Helvetica (font 12) is illustrated in Figure 1. Due to the idiosyncrasies of the typefaces, Garadoni swapped the letter 'b' for 'd,' using 'hamdurefonsiv,' and included all letters in upper case (*figure 2*).

1. In order to interpolate 10 steps between the two end points, a slight adjustment was made to the size of the last two steps.

A freely available dynamic text tool[2] was used to generate dummy text paragraphs of 60 'words' in 36pt using the limited character sets. Capitalization was introduced at the end of 'sentences' and the paragraph ended in a full point. Two sets of thirty six paragraphs were generated for the experiments: one set was used for Experiments 1, 2 and 2a (Timevetica) and the second for Experiments 3, 4a and 4b (Garadoni). Two screen shots of sample paragraphs from points 2 and 5 on the Garadoni continuum are shown in Figure 3, illustrating examples of the visual material used in the discrimination task. All experiments were run using Inquisit Millisecond software which controlled the timing and presentation of material, recorded responses and gave feedback in the discrimination task.

### EQUIPMENT

Experiments 1, 2 and 2a used a Research Machines PC to present the experimental material on a CTX (CRT) monitor with an 18″ viewing image, a screen resolution set to 1280 × 1024, and ClearType was switched on. Experiments 3, 4a and 4b moved onto later technology, using a Dell Latitude D820 laptop. This has a TFT-LCD of 8.7″ by 13.56″ and a diagonal of 15.4″, a resolution of 1280 × 80 and ClearType was again switched on. In all experiments, participants were able to choose where they sat in relation to the monitor, while able to reach the keyboard.

2. Adhersiontext™ was developed by Miguel Sousa as a tool for use in designing type. It is possible to typeset a paragraph of dummy text based on a specified set of letters to make judgments on letter shapes in context, rather than as single isolated letters. Available at http://www.adhesiontext.com/ [accessed May 24, 2011]

Figure 3: Two sample paragraphs using the fonts at point 2 (above) and point 5 (opposite) on the Garadoni continuum

Unverifiedness, do, mauveine dudism inns, middens marmarosis hi hards anisado, hard. Midi in mans, hushed, en, aah overharshness, hided divi hind, sun, divined rinds duodedenums is. Sea mans, ram afro hush if, neossine, muras, nomos murrine a uniramous sassed sudsier, is, find add. Am sis, onerousness, unerased. Dravidian, roominess. Hums, a ha a dean dromomania, dare, a roe, unremarried forfars.

A adamas evenmindedness, maven hosed. Randomish reified, is, faroeish. Of, def vies. Deveined, shun, had he, unrehearsed, he, shoe, firs. Individuous, hods, is, dandier, mands, hinderers, humifuse morned. Div dues, suer, muse, undefrauded oh mod, unhood memoire underseam, hoarier, minivers. Arms arshins, ninons a mini, um, adrenin, urn, nonoffensiveness or normed hi marasmus foreseason a a ford sourishness dominium dimmish.

## EXPERIMENT 1: IDENTIFICATION OF TIMEVETICA FONTS

### PROCEDURE

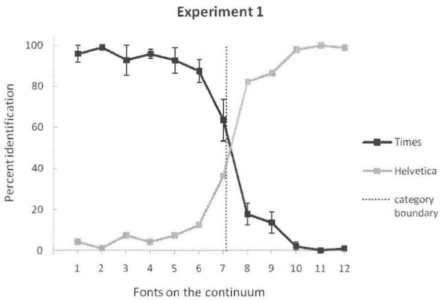
Participants were shown a paragraph and asked to say whether it was most like Times or most like Helvetica and to indicate their response by pressing one of two keys on the keyboard. Half the participants pressed the 'A' key to indicate more like Times and the 'L' key to indicate more like Helvetica; the other half used the reverse pairing. These particular letters were chosen to be separate from each other on the keyboard. The paragraph remained on the screen until the participant responded and the next paragraph appeared automatically after 600 msec. Two sheets of paper provided printed examples of 'hamburefonsiv' in Times and Helvetica, for participants to refer to if necessary. These remained in view throughout the experiment and were located to the left and right of the keyboard as reminders of the appropriate keys.

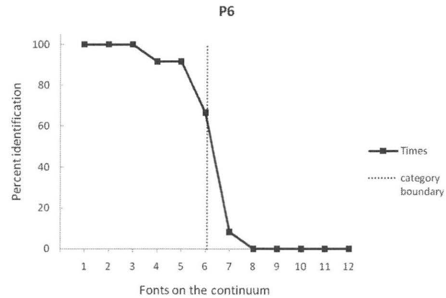
Six practice trials, which were not included in the data analysis, familiarized participants with the task and the pairing of responses and keys. Following this practice, trials were grouped into blocks of 36, with 3 examples from each of the 12 points on the continuum in each block. Eight participants each completed a total of 4 blocks resulting in 12 trials for each point. The order of trials within the block was randomized for each participant, as was the selection of paragraphs.

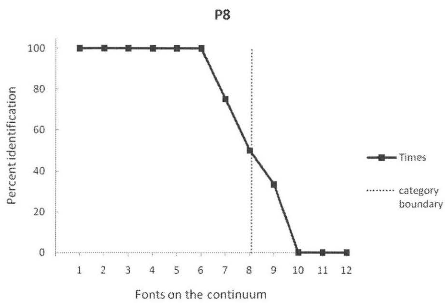
### RESULTS

Figure 4 combines the results of the eight participants to show the percent identification of the examples as Times, across each of the 12 points on the continuum. The bars indicate the standard error of the means, i.e., the variation among participants. The corresponding identification as Helvetica is included to provide a complete picture, but is redundant as it mirrors the Times data. The category boundary is not precisely where the two graphs cross (at 50% correct identification), but is calculated to be at 7.11, slightly towards Times. This calculation takes into account the individual points, i.e., the deviation from a smooth curve.[3] The results from two individual participants are shown in Figures 5 and 6. These illustrate the variation in the location of the boundary between participants. The participant in Figure 5 has a category boundary at 6.08, closer to

3. The formula for calculating the category boundary is $(\Sigma Y/100)+0.5$ where Y=percent identification, i.e., the sum of the percent identification scores for each point on the continuum, divided by 100, plus 0.5.



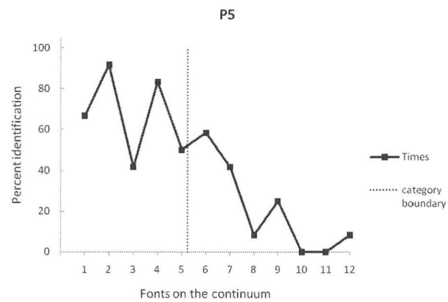




**Figure 4 (top left): Average identification of fonts across the Timevetica continuum**

**Figure 5 (top right): Identification of fonts across the Timevetica continuum by a single participant**

**Figure 6 (bottom left): Identification of fonts across the Timevetica continuum by a single participant**

**Figure 7 (bottom right): Participant with unsystematic identification of fonts across the Timevetica continuum**

the Times end of the continuum than the average, whereas the category boundary of participant 8 (*figure* 6) is at 8.08, closer to the Helvetica end of the continuum.

The results overall show clear evidence of a switch from Times to Helvetica, i.e., participants are able to correctly identify (categorize) near the two ends of the continuum but vary slightly as to where they switch from Times to Helvetica. One participant did not show this systematic identification (*figure* 7).

## EXPERIMENT 2: DISCRIMINATION OF TIMEVETICA

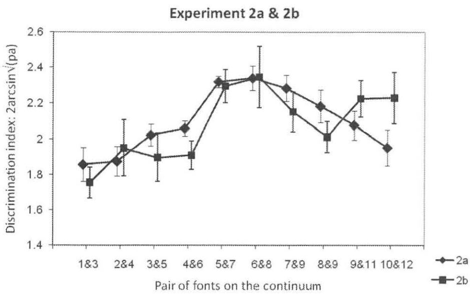
In order to explore whether the predictability of what would be presented from trial to trial would affect the nature of discriminations, two versions of the experiment were conducted with independent groups of 10 participants. These participants had not taken part in Experiment 1. Experiment 2a used a roving discrimination task where each trial within a block could come from any point on the continuum. In experiment 2b, each block of trials was fixed around one point on the continuum.

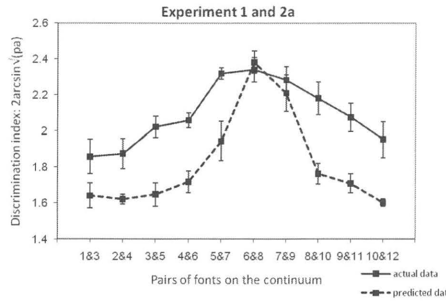
### PROCEDURE

Participants were shown one paragraph for 1000msec followed by a second paragraph (of different words) also for 1000msec; these were separated by a gap of 750msec. They were required to indicate whether the two paragraphs were in the same font or different fonts. A two-alternative forced-choice task was used with a scale from 'sure same' to 'same' to 'different' to 'sure different' using 4 keys on the keyboard ('A', 'S', 'K', 'L'). This scale allows the participant to adopt different criteria reflecting their confidence in the judgment. They were given feedback on their response with either a green tick or a red cross.

In different trials, fonts were two steps apart on the continuum (e.g., font 3 and font 5) and each of the 10 pairings (of the 12 points) were presented in both orders (i.e., 3 followed by 5; 5 followed by 3). A practice set of 8 trials with feedback ensured that participants were responding to the font (and not the content of the paragraphs) and familiarized them with the scale and keyboard. The next trial followed automatically 1300msec after the feedback disappeared from the screen.

In Experiment 2a, each of 10 blocks contained all possible 32 trials, i.e., 12 same trials from each point on the continuum and 20 different trials (10 pairings in each order). The order of trials was randomized within each block for each participant. Experiment 2b had the same number of blocks and trials, but each block used the same two points on the continuum. For example, within a block, same trials of font 3 and same trials of font 5 were combined with different trials where font 3 was followed by font 5 and font 5 was followed by font 3. This resulted in Experiment 2b having a different ratio of same to different trials, but calculation of a discrimination index, p(A), takes account of any bias that could be introduced by the weighting of trials. In Experiment 2b, the order of blocks was randomized for each participant.



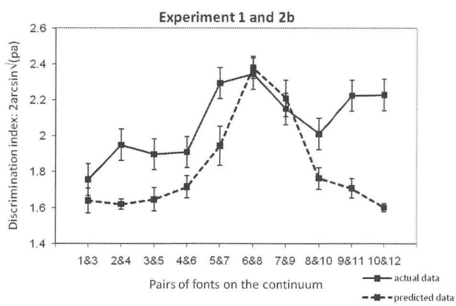




**Figure 8 (top left): Discrimination between fonts on the Timevetica continuum with roving (2a) and fixed (2b) task**

**Figure 9 (top right): Comparison of predicted discrimination, based on Experiment 1, and actual discrimination from Experiment 2a (roving)**

**Figure 10 (bottom): Comparison of predicted discrimination, based on Experiment 1 and actual discrimination from Experiment 2b (fixed)**

## RESULTS

The discrimination index was calculated for each participant and this was transformed by 2 arcsin√p(A) into a form suitable for statistical analysis (McNicol, 1972). A higher value indicates better discrimination and chance is at $\pi/2$ (1.571) and perfect discrimination is equal to $\pi$ (3.141). The results for 2a and 2b are plotted together in Figure 8, showing the average discrimination index for the various pairings on the continuum. The bars indicate the standard errors of the means. Analysis of Variance shows that there is a statistically significant difference between the font pairings at different points on the continuum ($F(9,162)=7.67$, $p<0.0001$). There is no overall effect of roving versus fixed discrimination (comparing Experiment 2a and 2b) so discrimination is not better when blocks of trials are fixed at one point on the continuum, compared to mixing trials across the whole continuum. More importantly, there is no significant interaction between points on the continuum and the experimental condition. The pattern of results indicates that discrimination is better around the middle of the continuum than at the ends (in particular the Times end of the continuum). This is the form typically associated with CP, i.e., poorer discrimination when pairs fall well within a category and better discrimination when pairs straddle category boundaries (see, for example, Burns and Ward, 1978). Both forms of the discrimination task appear to show CP.

Using the formula provided by Burns and Ward (1978), the identification data from Experiment 1 is used to predict discrimination performance. Figure 9 compares the predictions with discrimination data from Experiment 2a and Figure 10 with data from Experiment 2b. As the predicted and actual data come from two different groups of participants, it is not possible to calculate correlations. However, there appears to be a reasonable correspondence in both figures. The lack of an interaction between the discrimination task and the points on the continuum would suggest a similar relationship between each task and the predicted discrimination.

## EXPERIMENT 3: IDENTIFICATION OF GARADONI FONTS

### PROCEDURE

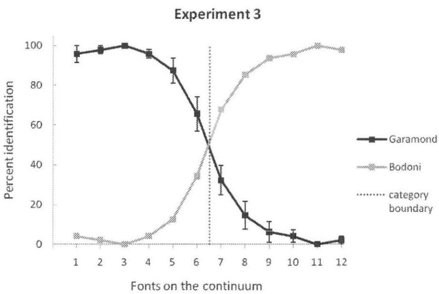
The procedure was identical to Experiment 1, except the continuum was changed to Garadoni; participants were asked to say whether the paragraph was most like Garamond or most like Bodoni. Printed examples were provided of the two typefaces. The same eight participants who completed Experiment 1 with Timevetica also did this identification task with Garadoni. The order of the two continua was alternated to counteract any order effects.

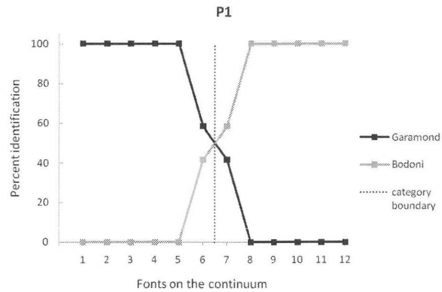
### RESULTS

The combined results are shown in Figure 11 and the category boundary is indicated at 6.52, which is extremely close to the physical center of the continuum. As with Experiment 1, virtually all participants were systematic in their identification (*see for example figure 12*). There was just one exception with more erratic identification towards the extremes (*figure 13*).

## EXPERIMENT 4: DISCRIMINATION OF GARADONI

A discrimination task similar to Experiment 2a was carried out using the Garadoni continuum. As there was no difference between roving and fixed discrimination in Experiment 3, a single condition (roving) was used for further experiments. Initially the conditions of Experiment 2a were repeated, using fonts that were two steps apart on the continuum, and the same timings. Pilot work indicated that this task was too



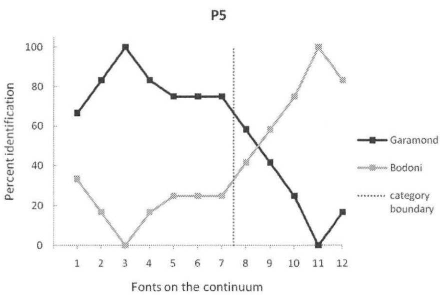




**Figure 11 (top left): Average identification of fonts across the Garadoni continuum**

**Figure 12 (top right): Identification of fonts across the Garadoni continuum by a single participant**

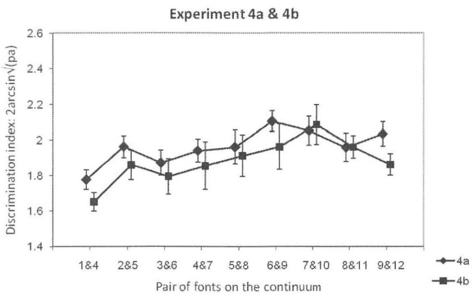
**Figure 13 (bottom): Participant with unsystematic identification of fonts across the Garadoni continuum**

difficult as performance was close to chance. In Experiment 4a the number of steps between examples was therefore increased to three to make the task easier. Although performance did improve, there was little indication of CP, i.e., no obvious peak in discrimination performance close to the category boundary. As participants still reported focusing on specific details, as found in the initial pilot work, experiment 4b explored another variation on the task. The duration of examples and the interval between them was reduced in an attempt to discourage an analytical strategy.

### PROCEDURE

As experiments 4a and 4b used fonts that were 3 steps apart on the continuum (e.g. font 3 and font 6), this resulted in 9 pairings. There were 9 blocks of 30 trials, each with 12 same trials and 18 different trials (9 pairings in each order). The order of trials within blocks was randomized for each participant. Eight practice trials preceded the first block. The rating scale for responses and feedback on these responses were identical to those used in Experiment 2.

In Experiment 4a, the duration of each paragraph was 1000msec, separated by a gap of 750msec. The time between responding and seeing the first paragraph of the next trial was 1300msec. In Experiment 4b, the duration was reduced to 500msec, with an interval of 300msec. The pause before the next trial was kept at 1300msec.



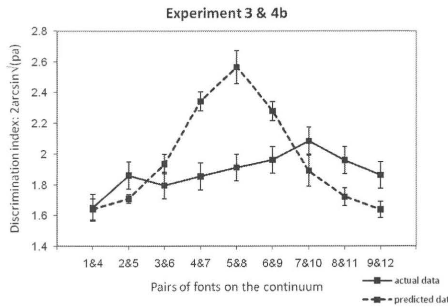




**Figure 14 (left): Discrimination between fonts on the Garadoni continuum with paragraphs presented for 1000msec (4a) and 500msec (4b)**

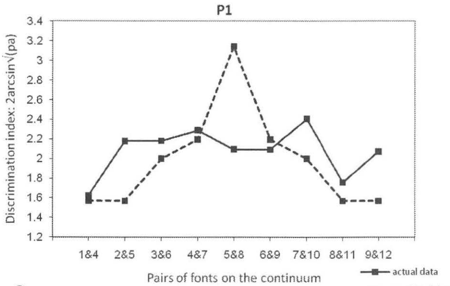
**Figure 16 (right): Comparison of predicted discrimination, based on Experiment 3, and actual discrimination from Experiment 4b (500msec presentation)**

Experiment 4a used 12 participants. Experiment 4b was able to recruit 7 of the 8 participants who had completed Experiment 3, identification of Garadoni. The missing participant was the person who appeared to have difficulty with the identification task as responses were less systematic (*figure 13*). This experiment was conducted more than two months after they had done the identification task so there would be no short-term carry-over effects on their discrimination task.

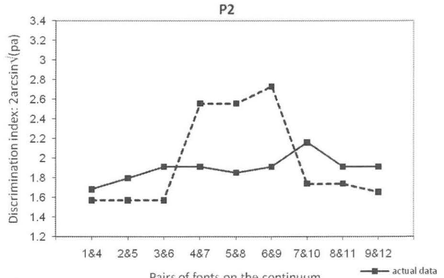
## RESULTS

The results for 4a and 4b are plotted together in Figure 14. Analysis of Variance of the two experiments finds a significant effect of the pairing of fonts on the continuum (F(8,136=5.09, p<0.0001). Although discrimination is generally better with the longer duration, this difference is not significant and there is no interaction between duration and font pairings. Reducing the time available to look at the paragraph does not lead to a pattern that would suggest CP, i.e., there is no single peak around the category boundary (6.52). Instead there are suggestions of additional smaller peaks closer to each end of the continuum.

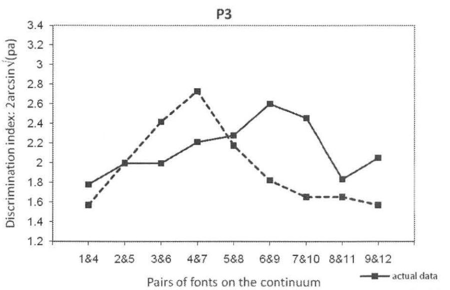
As 7 participants completed both the identification and discrimination tasks, predicted discrimination can be calculated for each individual and compared with their actual data. Figures 15a-g illustrates the large degree of variation among participants in both the predicted discrimination (different locations for the category boundary based on their identification data) and actual discrimination. The only participant data that comes close to indicating a single, defined peak in their actual discrimination data is P3 (*figure 15c*). However, the location of this peak



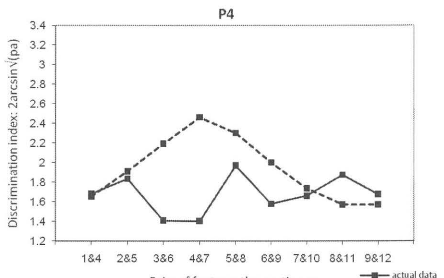

a.



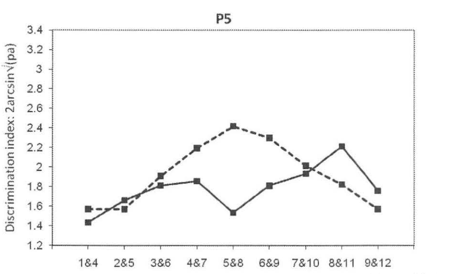

b.



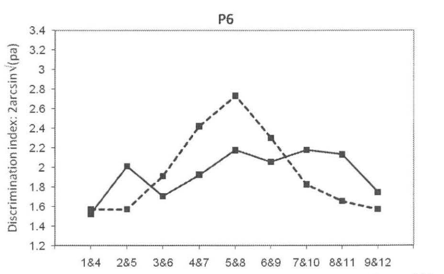

c.



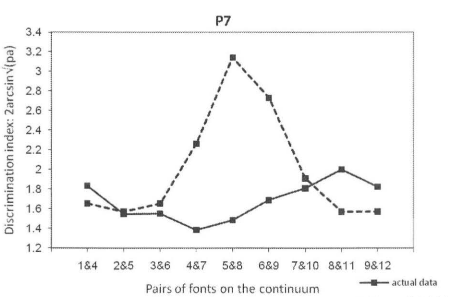

d.




e.




f.




g.

**Figure 15a-g: Individual results of 7 participants comparing predicted discrimination, based on Experiment 3, and actual discrimination from Experiment 4b (500msec presentation)**

is shifted from the location that is predicted from their identification data. Given the intervening period between the two experiments, it is conceivable that the participant's categorization criteria changed.

Looking at the data averaged across all 7 participants (*figure 16*) confirms that the identification data does not predict discrimination. The Pearson Product-Moment Correlation is virtually 0 (r=0.04).

## DISCUSSION

The results indicate that designers do demonstrate CP, but the circumstances that produce this mode of responding may be quite specific. The fact that CP was not found with a second pair of typefaces raises questions as to which factors might be responsible for this difference. However, before looking at characteristics of the typefaces, the methodological findings are discussed.

The consistent pattern of discrimination across roving and fixed tasks, both indicating CP, suggest that the fixed task may also have included stimulus uncertainty. This may have resulted from the relatively short time available to look at each paragraph (1000msec) combined with the difference in the content (letters) within the two sequential paragraphs (*see figure 3*). Most visual CP paradigms (e.g., faces or lines) compare the same configurations at different points on the continuum, rather than different configurations at different points on the continuum. Faces always have two eyes, one on each side of the nose, and a mouth below the nose, whereas letters naturally occur in different orders. The reason to use different content was to prevent participants focusing on only one letter (in a known location) but this may have introduced uncertainty that induced CP, even in a fixed discrimination task.

An alternative explanation is that although the interpolation produced physically equal steps, these were not perceptually equal (even without the intervention of CP). Goldstone (1994) equated the perceptual differences between adjacent stimuli before testing for perceptual learning. In Experiment 2 of the current paper, the fixed discrimination was intended to provide an index of perceptual differences with which to compare the roving discrimination results, i.e., Experiment 2b would serve as a control. However, there were no differences between these two conditions.

Although there may be some perceptual inequalities across the pairings on the continuum, it is unlikely that these would coincide so closely with the category boundary and yet not be influenced by the boundary, i.e., that there would be a larger perceptual difference between 6 and 8 that is unrelated to the category boundary. Looking at the data (*figure 8*) suggests that if there is inequality in the size of steps, this may lie towards the Helvetica end of the continuum. In fixed discrimination, fonts 10 and 12 are better discriminated than in roving discrimination. This may result from 10 having the vestiges of serifs and 12 having no serifs (*see figure 1*).

Further evidence to support CP of Timevetica is the correspondence between actual and predicted discrimination performance. This degree of correspondence is in line with other CP data (e.g., Burns and Campbell, 1994). As Garadoni was not perceived categorically, no such correspondence was found, despite looking at individual participant data (*figures 15a-g*). With both continua, the location of category boundaries varied from individual to individual (Timevetica ranging from 5.25 to 8.08; Garadoni from 5.5 to 7.5). Therefore the variability among individuals cannot account for the lack of CP.

To consider the possible explanations for why Garadoni does not demonstrate CP requires examination of the differences between the two continua and the 4 typefaces. The most obvious are:

- → pairing of a sans serif with a serif typeface as opposed to two serif faces
- → the degree of difference between the two end points (in part due to the serif/sans serif distinction)
- → familiarity with the typefaces

The first two characteristics are likely to have influenced the overall level of performance. This is confirmed by the slightly higher average discrimination performance for Experiment 2a (2.096) compared with Experiment 4a (1.963), despite the increase in step size for Experiment 4a. Following this adjustment, it is unlikely that such a small difference in difficulty would hinder CP. However, there is a qualitative difference between a serif and sans serif face, and this may be what is required to provide a sufficiently robust basis for categorization. Although a linear continuum was created, the presence or absence of serifs is essentially a binary division.

But in addition, exposure to the two pairs of typefaces is likely to be very different. Even if designers tend not to choose Times or Helvetica for their own work, because they have many more typefaces at their disposal, their exposure to Times and Helvetica is likely to be greater than to Garamond and Bodoni. This might mean that they have a well-established, over-learned perceptual category for Times and Helvetica but not for Garamond and Bodoni. This explanation has been used to account for differences between speech and non-speech sounds, which were removed after training in categorizing the non-speech sounds (Aravamudhan, Lotto and Hawks, 2008).

A number of ways of deciding among the alternative explanations for the difference between continua are possible: introduce some form of perceptual training for Garadoni and re-test; create a new continuum between less popular serif and sans serif typefaces; create a new continuum between two popular serif faces.

Although the Garadoni continuum did not peak at the category boundary established in the identification task, there were differences in the pairings across the continuum, and small peaks at fonts 2 & 5 and possibly 6 & 9 and 7 & 10. These steps would therefore appear to be perceptually larger. As discussed above, this could be an artifact of the continuum, or it is possible that one or more additional categories emerge, i.e., distinct typefaces are perceived, different from the two ends. Newell and Bülthoff (2002) introduce this possibility to explain why some object pairs were not perceived as categorical. In their case, a new object may emerge on a continuum from a spindle to a barrel that may be perceived as a cylinder. With Garadoni, one possible candidate for an additional category is Baskerville, with formal attributes that are transitional between Garamond ('Old Face') and Bodoni ('Modern' face) (Luna, 1992). Monotype Baskerville is illustrated in the second line of Figure 17 with Garamond above and Bodoni beneath. The result of perceiving another category would be additional peaks in discrimination performance and no observable effects of CP (Newell and Bülthoff, 2002).

## CONCLUDING REMARKS

Although evident with only one typeface continuum, this work has demonstrated that typefaces can be perceived categorically. As there are no natural categories for typefaces, even typographers are unlikely to be born with the ability to distinguish among typefaces. It is therefore probable that typographic training, attending to differences among typefaces, can establish categories that lead to efficient discrimination.

The question remains as to whether non-designers develop similar categories through exposure to common typefaces, without the training. One reason to question the usefulness to non-designers of developing CP of typefaces is found at the start of this paper. In normal reading, we process letters to create abstract letter hamdurefonsiv

hamdurefonsiv

hamdurefonsiv

**Figure 17: Garamond (top), Monotype Baskerville (middle), Bodoni (bottom)**

identities (Besner, Coltheart, and Davelaar, 1984). Therefore for those people who do not work closely with type, there appears to be no benefit from acquiring distinctive categories for typefaces. Readers need to decrease their sensitivity to differences that are not relevant to categorizing the letter identity, i.e., are not diagnostic of the letter structure (Walker, 2008).

However, research on font tuning (Sanocki, 1987, 1988; Gauthier, Wong, Hayward and Cheung, 2006; Wong and Gauthier, 2007; Walker, 2008) seeks to explain how the particular characteristics of a font can be used to facilitate letter recognition. Sanocki suggested that the perceptual system can become tuned to a particular font over time and a set of font parameters are developed. It may be the case that these parameters are encapsulated into a typeface category. Early results from the current author's pilot work, still in progress, suggest that typefaces may need to be quite different from each other to show the effects of font tuning.[4] This is consistent with one of the explanations for finding CP in Timevetica and not Garadoni.

The connection between the early stages of the reading process (letter identification) and CP of typefaces is highly speculative. But such tentative suggestions may be the starting point for new research directions and also aim to bridge a gap between design practice and psychological theories of reading.

4. Font tuning effects have not been found for Garamond and Bodoni, but this may be because the experimental method is not sufficiently sensitive to demonstrate the effects. Other methods are currently being explored.

## ACKNOWLEDGMENTS

I wish to thank: Malou Verlomme, Daniel Rhatigan and Jo De Baerdemaeker who developed and modified the typeface continua; José Marconi Bezerra de Souza for help in carrying out some of the experiments; Anthony J Watkins for directing me to research in the auditory domain; Androniki Sioki for helpful suggestions; anonymous students who gave their time.

## REFERENCES

**Angeli, A., J. Davidoff and T. Valentine. 2008.** Face Familiarity, Distinctiveness, and Categorical Perception. *Quarterly Journal of Experimental Psychology*, 61.5, 690–707.

**Aravamudhan, Radhika, Andrew J. Lotto and John W. Hawks. 2008.** Perceptual Context Effects of Speech and Nonspeech Sounds: The Role of Auditory Categories. *The Journal of the Acoustical Society of America*, 124.3, 1695–1703.

**Baines, Phil and Andrew Haslam. 2005.** *Type & Typography*. Second ed. London, UK: Laurence King.

**Bartram, David. 1982.** The Perception of Semantic Quality in Type: Differences between Designers and Non-Designers. *Information Design Journal*, 3.1, 38–50.

**Beale, J.M. and C.F. Keil. 1995.** Categorical Effects in the Perception of Faces. *Cognition*, 57, 217–239.

**Besner, Derek, Max Coltheart and Eileen Davelaar. 1984.** Basic Processes in Reading: Computation of Abstract Letter Identities. *Canadian Journal of Psychology*, 38, 126–134.

**Bornstein, Marc H. 1987.** Perceptual Categories in Vision and Audition. In Harnard, Stevan, editor. *Categorical Perception: The Groundwork of Cognition*. Cambridge, UK: Cambridge University Press, 287–331.

**Bülthoff, I. and F.N. Newell. 2004.** Categorical Perception of Sex Occurs in Familiar but Not Unfamiliar Faces. *Visual Cognition*, 11.7, 823–855.

**Burns, Barbara and Bryan E. Shepp. 1988.** Dimensional Interactions and the Structure of Psychological Space: The Representation of Hue, Saturation, and Brightness. *Perception & Psychophysics*, 43.5, 494–507.

**Burns, E.M. and S.L. Campbell. 1994.** Frequency and Frequency-Ratio Resolution by Possessors of Absolute and Relative Pitch—Examples of Categorical Perception. *Journal of the Acoustical Society of America*, 96.5, 2704–2719.

**Burns, E.M. and W.D. Ward. 1978.** Categorical Perception—Phenomenon or Epiphenomen: Evidence from Experiments in the Perception of Melodic Intervals. *Journal of the Acoustical Society of America*, 63, 456–468.

**Calder, A.J., A.W. Young, D.I. Perrett, N.L. Etcoff and D. Rowland. 1996.** Categorical Perception of Morphed Facial Expressions. *Visual Cognition*, 3.2, 81–117.

**Campanella, S., C. Hanoteau, X. Seron, F. Joassin and R. Bruyer. 2003.** Categorical Perception of Unfamiliar Facial Identities, the Face-Space Metaphor, and the Morphing Technique. *Visual Cognition*, 10.2, 129–156.

**Changizi, Mark, Qiong Zhang, Hao Ye and Shinsuke Shimojo. 2006.** The Structures of Letters and Symbols Throughout Human History Are Selected to Match Those Found in Objects in Natural Scenes. *The American Naturalist*, 167.5, E117–E139.

**Emmorey, K., S. McCullough and D. Brentari. 2003.** Categorical Perception in American Sign Language. *Language and Cognitive Processes*, 18.1, 21–45.

**Gauthier, Isabel, Alan C.-N. Wong, William G. Hayward and Olivia S. Cheung. 2006.** Font Tuning Associated with Expertise in Letter Perception. *Perception*, 35, 541–559.

**Gerrits, E. and M.E.H. Schouten. 2004.** Categorical Perception Depends on the Discrimination Task. *Perception & Psychophysics*, 66.3, 363–376.

**Gibson, Eleanor Jack. 1969.** *Principles of Perceptual Learning and Development.* Englewood Cliffs, N.J: Prentice-Hall.

**Goldstone, Robert. 1994.** Influences of Categorization on Perceptual Discrimination. *Journal of Experimental Psychology: General*, 123.2, 178–200.

**Harnard, Stevan. 1987.** *Categorical Perception: The Groundwork of Cognition.* Cambridge, UK: Cambridge University Press.

**Howard, D., S. Rosen and V. Broad. 1992.** Major Minor Triad Identification and Discrimination by Musically Trained and Untrained Listeners. *Music Perception*, 10.2, 205–220.

**Levin, D.T. and B.L. Angelone. 2002.** Categorical Perception of Race. *Perception*, 31.5, 56–578.

**Liberman, A.M., K.S. Harris, H.S. Hoffman and B.C. Griffith. 1957.** The Discrimination of Speech Sounds within and across Phoneme Boundaries. *Journal of Experimental Psychology*, 54.5, 358–368.

**Luna, Paul. 1992.** *Understanding Type for Desktop Publishing.* London, UK: Blueprint.

**Macmillan, Neil A. 1987.** Beyond the Categorical/Continuous Distinction: A Psychophysical Approach to Processing Modes. In Harnard, Stevan, editor. *Categorical Perception: The Groundwork of Cognition.* Cambridge, UK: Cambridge University Press, 53–85.

**McNicol, D. 1972.** *A Primer of Signal Detection Theory.* London, UK: George Allen & Unwin Ltd.

**Newell, Fiona N. and Heinrich H. Bülthoff. 2002.** Categorical Perception of Familiar Objects. *Cognition*, 85, 113–143.

**Quinn, Paul C. 2004.** Visual Perception of Orientation Is Categorical near Vertical and Continuous near Horizontal. *Perception*, 33.8, 897–906.

**Sanocki, Thomas. 1987.** Visual Knowledge Underlying Letter Perception: Font-Specific Schematic Tuning. *Journal of Experimental Psychology: Human Perception and Performance*, 13.2, 267–278.

**Sanocki, Thomas. 1988.** Font Regularity Constraints on the Process of Letter Recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 14.3, 472–480.

**Studdert-Kennedy, M., A.M. Liberman, K.S. Harris and F.S. Cooper. 1970.** Theoretical Notes Motor Theory of Speech Perception—a Reply to Lanes Critical Review. *Psychological Review*, 77.3, 234–249.

**Viviani, P., P. Binda and T. Borsato. 2007.** Categorical Perception of Newly Learned Faces. *Visual Cognition*, 15.4, 420–467.

**Walker, Peter. 2008.** Font Tuning: A Review and New Experimental Evidence. *Visual Cognition*, 16.8, 1022–1058.

**Wong, Alan C.-N. and Isabel Gauthier. 2007.** An Analysis of Letter Expertise in a Levels-of-Categorization Framework. *Visual Cognition*, 15.7, 854–879.

## AUTHOR NOTE

**MARY DYSON** is a Senior Lecturer in the Department of Typography & Graphic Communication at the University of Reading, in the UK. She has a background in experimental psychology and uses this training to investigate how readers interact with various aspects of document design. More recent work draws on her PhD in perception and looks for parallels between our perception of letters and typefaces (by readers and designers) and other perceptual activities.