Pareto-Optimality of Cognitively Preferred Polygonal Hulls for Dot Patterns

Antony Galton

School of Engineering, Computing and Mathematics, University of Exeter, UK

Abstract. In several areas of research one encounters the problem of generating an outline that is in some way representative of the spatial distribution of a pattern of dots. Several different algorithms have been published which can generate such outlines, but the detailed evaluation of such algorithms has mostly concentrated on their computational and mathematical properties, while the adequacy of the resulting outlines themselves has been left as a matter of informal human judgment. In this paper it is proposed to investigate the perceptual acceptability of outlines independently of any particular algorithm for generating them, in order to determine objective criteria for evaluating outlines from the full range of possibilities in a way that is conformable to human intuitive assessments. For the sake of definiteness it is assumed that the outline to be produced is a simple closed polygon whose vertices are elements of the given dot pattern, all remaining elements of the dot pattern being in the interior of the polygon. It is hypothesised that to produce a cognitively acceptable outline one should seek simultaneously to minimise both the area and perimeter of the polygon, and that therefore the points in area-perimeter space corresponding to cognitively optimal outlines will lie on or close to the Pareto front. A small pilot study was conducted, the results of which lend strong support to the hypothesis. The paper concludes with some suggestions for further more detailed investigations.

1 Introduction

When presented with a two-dimensional pattern of dots such as the one shown in Figure 1, and asked to draw a polygonal outline which best captures the shape formed by the pattern, people readily respond by drawing outlines such as those shown in Figure 2. Interestingly, on first encountering this task, people often tend to imagine that there is a unique solution, 'the' outline of the dots; but they will very quickly be persuaded that there is typically no unique best answer. Only the convex hull has any claim to uniqueness, but in very many cases (such as the example shown), the convex hull is a bad solution to the task, since it does not capture the shape that we humans perceive the dots as forming. This is illustrated in Figure 3, where two distinct point-sets, having the shape of the letters 'C' and 'S', have the same convex hull.
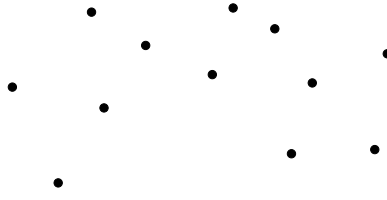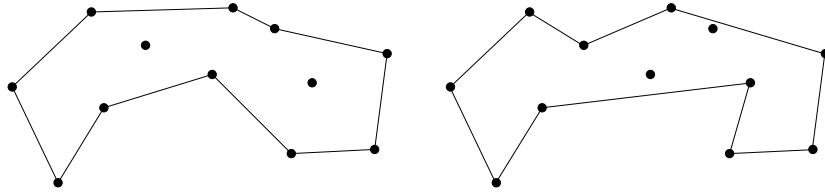
Fig. 1. A simple dot pattern

Fig. 2. Example outlines for the same dot pattern

The problem of representing a dot pattern by means of an outline which in some way captures the shape defined by the pattern, or the region of space occupied by the dots, has been investigated over a number of years by researchers from a number of different disciplines [1–6]. There may be several distinct motivations underlying such investigations; for example:

– *Map generalisation.* What appears at one level of detail as a set of discrete points may be better represented, at a coarser level of detail, as a region. The region indicates the general location and configuration of the points, but does not indicate how many points there are or their individual positions.
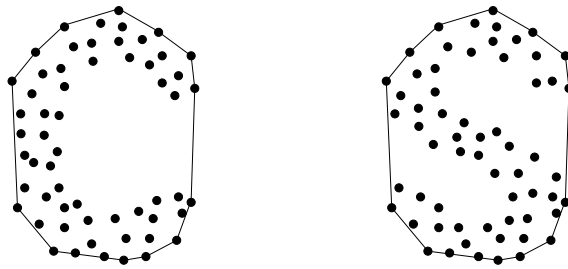

Fig. 3. Point-sets with the same convex hull

– *Region approximation* for storage and retrieval efficiency. Geographical regions typically have complex sinuous outlines which place a high load on

storage and retrieval capacity when represented digitally. Web-based digital gazetteers require efficient ways of recording the locations associated with region names; but traditional approximations such as bounding boxes, centroids, or convex hulls are far too crude for most purposes, and detailed information concerning the region's boundary may be unnecessarily complex, even if available. What is needed is an approximation to a region which may be efficiently generated from available information about points known to lie inside (or outside) the region [7, 8].

– *Gestalt perception.* As humans, we typically perceive a cluster of points as occupying some two-dimensional region in the visual field, and can describe, at least roughly, the outline of the region they occupy. If we want to emulate this capacity of the human visual system in a computer vision system, we need to be able to compute the region from the points.

– *Representation and reasoning about collective phenomena* such as flocking animals, traffic, or crowds. In such cases the 'ground truth' consists of a set of individuals each with its own typically point-like location at any given time, but for many purposes it is desirable to think of the phenomenon as a single entity spread out across a region which may be thought of as the 'spatial footprint' of the collective [9].

Different motivations may result in different criteria for evaluating the quality of the outlines produced by any proposed method, but it is striking that, for the most part, existing literature on the problem has had little to say about these criteria, focusing rather on the technical details and computational characteristics of different algorithms. The main aim of this paper is to redress this imbalance by focussing on the question of evaluation criteria rather than particular algorithms.

2 Previous Work

As mentioned above, there is already a considerable body of work, much of it in the pattern analysis, computer vision, and geographical information science communities, on defining the shape of dot patterns. A typical paper in this area will propose an algorithm for generating a shape from a pattern of dots, explore its mathematical and/or computational characteristics (e.g., computational complexity), and examine its behaviour when applied to various dot patterns. The evaluation of this behaviour is typically very informal, often amounting to little more than observing that the shape produced by the algorithm is a 'good approximation' to the perceived shape of the dots. While lip-service is generally paid to the fact that there is no objective definition of such a 'perceived shape', little is said about how to verify this, or indeed, about exactly what it means.

The much-cited work of Edelsbrunner *et al.* [1], introduces the notion of α-*shape*: whereas the convex hull of a point-set S is the intersection of all closed half-planes containing all the points of S, their 'α-hull' is the intersection of all closed discs of radius $1/\alpha$ containing all points of S (for $\alpha < 0$ the closed disc of radius $1/\alpha$ is interpreted as the complement of an open disk of radius $-1/\alpha$, and

for $\alpha = 0$ it is a half-plane). The α-shape is a piecewise linear curve derived in a straightforward manner from the α-hull. For certain (typically small negative) values of α, the α-shape can come close to capturing the cognitively salient aspects of the overall distribution of points. The authors go into considerable details concerning the mathematical properties of these shapes, but almost the only thing stated by way of evaluating the adequacy of the shapes produced by the algorithm is that 'α-shapes . . . seem to capture the intuitive notion of "finer" or "cruder shape" of a planar pointset'.

Similar reticence is shown by others who have followed. Garai and Chaudhuri [2] propose a 'split-and-merge' procedure, which starts by constructing the convex hull of the points, and then successively inserts extra edges or smooths over zigzags. The splitting procedure results in a highly jagged outline, which is then made smoother by the merging procedure. But again, the authors say almost nothing on the evaluation of the results of the algorithm, although it is clear that one purpose of reducing the jaggedness of the outline is to improve its cognitive acceptability.

Melkemi and Djebali [3] propose the $\mathcal{A}$-*shape*: Given a finite set of points P and a set $\mathcal{A}$ disjoint from P, the $\mathcal{A}$-shape of P is obtained from the Voronoi diagram for $\mathcal{A} \cup P$ by joining any pair of points $p, q \in P$ whose Voronoi cells border each other and also the Voronoi cell of a point in $\mathcal{A}$. The edges pq are the '$\mathcal{A}$-exposed' edges of the Delaunay triangulation of $\mathcal{A} \cup P$. The $\mathcal{A}$-shape was introduced 'with the aim of curing the limits of α-shape', and the authors have only a little more to say on its evaluation: their explicit aim is to look for a 'polygonal representation' that 'must reflect the forms perceived by a human observer of the dot patterns set'.

Chaudhuri *et al* [4] also make explicit reference to human visual perception. Their r-shape is obtained by constructing the union U_r of all disks of radius r centred on points of P, and then, for $p, q \in P$, selecting edge pq if the boundaries of the discs centred on p and q intersect on the boundary of U_r; the r-shape of P is the union of the selected edges. In the same paper they discuss the s-shape, obtained by partitioning the space into a lattice of $s \times s$ squares and then taking the union of those squares which contain points of P. They confine their attention to *regular dot patterns*, in which 'the points are clearly visible as well as fairly densely and more or less evenly distributed' (unlike, for example, our Figure 1). For such patterns they say that 'one can perceive the border of the point set', and see their problem as 'extracting the border that is compatible with the perceived shape of the input pattern'; they also speak of 'the intuitive shape of the dot pattern'. This way of speaking seems to imply that there is a *unique* perceived shape, but they acknowledge that '"perceptual structure" of [a dot pattern] S cannot be defined uniquely', adding that it 'will vary from one person to another to a small extent'. But no attempt is made to determine the extent of such variation, and in evaluating the results little is said beyond the statement that 'if ε [a real-valued scaling factor used in their algorithms] lies in the range 0.3–0.5, the extracted border is compatible with the perceptual border of the dot pattern' — and again, no quantitative measure of degree

of compatibility is given. The remainder of their evaluation concerns intrinsic features of the algorithm such as its computational complexity.

Galton and Duckham [5] proposed three different algorithms for generating a region (called a 'footprint') from a set of points. One, the 'swinging arm' method, generalises the 'gift-wrap' algorithm for constructing convex hulls; a line segment of length r is swung about an extremal point of the set until it encounters another point in the set; the two points are joined, and the procedure repeated from the second point, until a closed shape is produced. Additional components of the footprint will be obtained if points in the set lie outside the first component. Similar results can be obtained by joining all pairs of points separated by at most r and then selecting the peripheral joins, resulting in the 'close pairs' method. In the third algorithm, a region is produced by successively removing the longest exterior edges from the Delaunay triangulation of the points, subject to the condition that the region remains connected and its boundary forms a Jordan curve. In this work, more attention was paid to the question of evaluation criteria, and nine questions were listed that could be used to help classify different types of solution to the general problem of associating a region with a set of points. But like the work previously reviewed, this paper shied away from any detailed examination of the concept of 'perceived shape' other than noting that any such examination must 'go beyond computational geometry to engage with more human-oriented disciplines such as cognitive science'.

Moreira and Santos [6] proposed a 'concave hull' algorithm which is an alternative generalisation of the gift-wrap algorithm, in which at any stage only the k nearest neighbours of the latest point added to the outline are considered as candidates for the next addition. They state the problem as that of finding 'the polygon that best describes the region occupied by the given points', and acknowledge that the word 'best' here is ambiguous, what counts as a best solution being application dependent; but evaluation of the algorithm is largely confined to its computational characteristics and not the adequacy of the results, for which they do little more than refer to the criteria listed in [5]. Outputs from this algorithm (for Pattern 5 in Appendix A) are shown in Figure 4.

In work currently in press, Duckham *et al.* [10] present more detailed evaluation for the Delaunay-based method first presented in [5], leading to a conclusion that 'normalized parameter values of between 0.05–0.2 typically produce optimal or near-optimal shape characterization across a wide range of point distributions', but it is acknowledged that what 'optimal' means here is both underspecified and somehow connected with 'a shape's "visual salience" to a human'. The actual evaluation presented in [10] takes the approach of starting with a well-defined shape, generating a dot pattern from it, and then testing the algorithm's efficacy at reconstructing the original shape.

The purpose of the present paper is to take some first steps towards establishing some principles for evaluating any proposed solution to the problem of determining an outline for a set of points. Whereas previous work has mostly been concerned with proposing particular algorithms for generating outlines, here I propose that, independently of any particular algorithm, we consider a
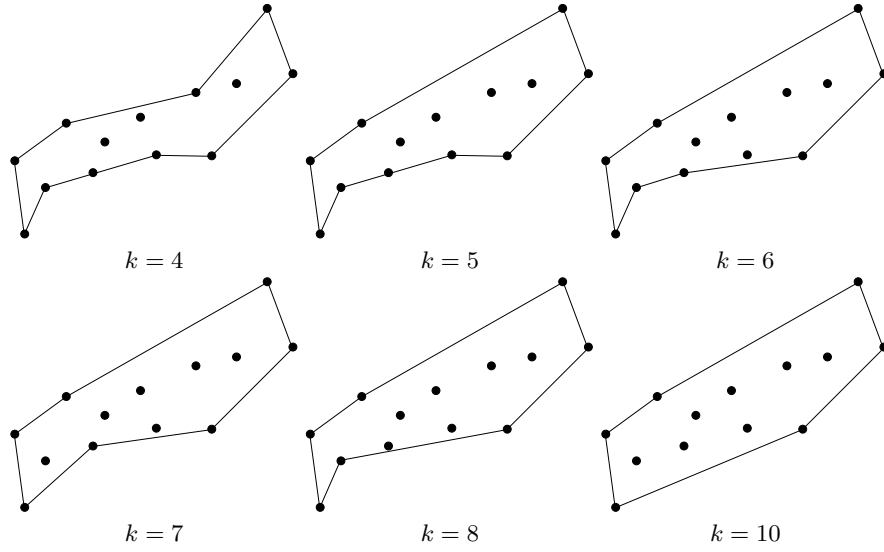
Fig. 4. Polygonal hulls generated by the Concave Hull algorithm [6].

full range of possible outlines, and try to determine what features, describable in objective (e.g., geometrical) terms, influence cognitive judgments as to the suitability of an outline as a depiction of 'the' shape defined by the set of points.

3   The Scope of the Inquiry

In order to bring the treatment to manageable proportions, we first make some assumptions about the kind of solution that is being sought. Many, though by no means all, of the published algorithms produce outlines satisfying the following criteria:

1. The outline is a polygon whose vertices are members of the dot pattern.
2. Any member of the dot pattern which is not a vertex of the polygon lies in the interior of the polygon.
3. The boundary of the polygon forms a Jordan curve (so in particular no point is encountered more than once in a full traversal of the boundary).

We shall call such outlines *polygonal hulls* of the underlying dot pattern; for brevity, we shall usually just refer to them as 'hulls'. The outlines shown in Figures 2 and 4 are of this kind. We exclude from consideration curvilinear outlines, outlines which exclude one or more points of the dot pattern, outlines which include all points of the dot pattern in their interior, outlines which are topologically non-regular, self-intersecting outlines, etc. Examples of two such excluded outlines are shown Figure 5.
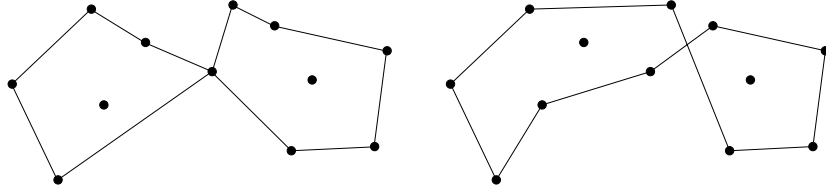
Fig. 5. Two non-examples: these do not count as polygonal hulls

It is obvious that the vertices of the convex hull for any dot pattern will appear as vertices of all of the polygonal hulls for that pattern; and moreover, in any of the polygonal hulls, the convex-hull vertices will appear in the the same sequential order around the perimeter. In general we may represent a dot pattern as $K \cup I$, where K is the set of vertices of the convex hull and I is the set of dots in the interior of the convex hull. Let the clockwise ordering of convex hull vertices be $p_1, p_2, \ldots, p_k$, and let the interior dots be $q_1, q_2, \ldots, q_{n-k}$. Then the sequence of vertices of any polygonal hull for the dot pattern will consist of $p_1, p_2, \ldots, p_k$ in that order, interspersed with some selection from $q_1, q_2, \ldots, q_{n-k}$ in some order.

How many polygonal hulls are there for a pattern of n dots? We can easily calculate an upper bound. From the above observations, for an n-point dot pattern whose convex hull has k vertices, we can select a polygonal hull by a sequence of four choices: (1) choosing how many interior dots q_i will be vertices of the hull (say r dots, where $0 \leq r \leq n - k$); (2) choosing which r of the $n - k$ available interior dots will be vertices of the hull ($^{n-k}C_r$ choices); (3) in the clockwise traversal of the hull starting from p_1, choosing which r of the $k + r - 1$ remaining vertices will be assigned to interior dots ($^{k+r-1}C_r$ choices); (4) choosing in which order the r interior dots will be assigned to the r vertex positions chosen at the previous step ($r!$ choices). Not every combination of such choices will lead to a polygonal hull (the perimeter of the resulting polygon may be self-intersecting, or some of the dots may lie outside the polygon), but each polygonal hull will arise from exactly one combination of choices. Thus the number of polygonal hulls for an n-point dot pattern with a k-vertex convex hull is at most

$$\sum_{r=0}^{n-k} {}^{n-k}C_r \, {}^{k+r-1}C_r \, r!.$$

For the case $n = 12$ and $k = 7$, this comes to 86,276; but the 12-point dot pattern shown in Figure 6, with seven vertices in its convex hull, actually has only 5674 polygonal hulls, approximately 6.6% of the upper bound. Even so, the number of polygonal hulls does grow rapidly as the number of dots increases, and for large values of n it becomes impracticable to compute all of them (with $n = 16$ we are already talking days rather than hours or minutes in the worst case). In reality, however, only a tiny fraction of the polygonal hulls are worth considering as good candidates for the 'perceived shape' of the dot pattern.
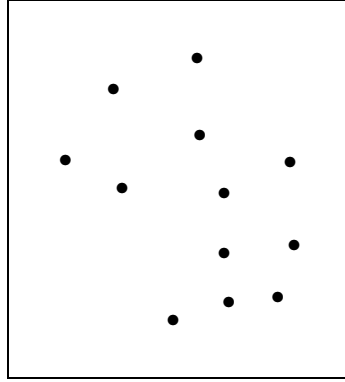
Fig. 6. Dot Pattern 1

Figure 7 illustrates three of the 5674 polygonal hulls for the dot pattern in Figure 6. The leftmost one is the convex hull. This is easily defined, has well-known mathematical and computational properties, and might be considered as a useful representation of the dot pattern for some purposes; but as already noted, it does not usually capture the perceived shape of the pattern. The rightmost one provides a very jagged outline which does not correspond to anything that we readily perceive when observing the dots on their own. The middle hull, on the other hand, does seem to capture pretty well a shape that we can readily perceive in the dots. It is certainly not unique in doing so, however, and in the pilot study reported below, only 2 out of 13 subjects drew this as their preferred hull for this pattern of dots.
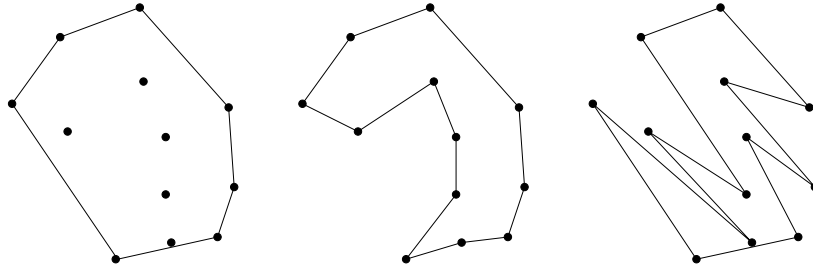

Fig. 7. Example polygonal hulls for Dot Pattern 1

What factors make a polygonal hull acceptable as a representation of the 'perceived shape' of a dot pattern? The problem with the convex hull is that it will often include large areas devoid of dots; these are the perceived concavities in the shape, and the convex hull completely fails to account for them. Of all

possible hulls, the convex hull simultaneously maximises the area while minimising the perimeter. It is the maximality of the area which causes the problem, since this correlates with the inclusion of the empty spaces represented by the concavities in the perceived outline. At the other extreme, the jagged figure on the right does very well at reducing the area, but at the cost of a greatly extended perimeter. The middle figure seems to strike a better balance, with both area and perimeter taking intermediate values, as shown in Table 1.

Table 1. Area and perimeter measurements for the hulls in Figure 7 (units of measurement arbitrary)

	Area	Perimeter
Hull 1	42761.0	783.5
Hull 2	27163.0	962.5
Hull 3	21032.0	1599.3

A cognitively acceptable outline should (a) not contain too much empty space, and (b) should not be too long and sinuous. This suggests that to produce the optimal outline we should seek to simultaneously minimise both the area and the perimeter. These are, of course, conflicting objectives, since the minimum perimeter (that of the convex hull) corresponds to the maximum area. In the language of multi-objective optimisation theory [11], we seek *non-dominated* solutions. A polygonal hull with area A_1 and perimeter P_1 is said to *dominate* one with area A_2 and perimeter P_2 (with respect to our chosen objectives of minimising both area and perimeter) so long as

$$(A_1 \leq A_2 \wedge P_1 < P_2) \vee (A_1 < A_2 \wedge P_1 \leq P_2).$$

The hulls which are not dominated by any other hulls form what is known as the *Pareto set*. When plotted in area-perimeter space ('objective space') they lie along the *Pareto front*. This shows up in the graphs as the 'south-western' frontier of the set of points corresponding to all the hulls for a given dot pattern. Area-perimeter plots for all eight dot patterns used in the pilot study described below can be found in Appendix B. In these figures, area is plotted along the horizontal axis, perimeter along the vertical; the convex hull, with maximal area and minimal perimeter, corresponds to the point at the extreme lower right.

In light of the above considerations, we propose the following

Hypothesis: *The points in area-perimeter space corresponding to polygonal hulls which best capture a perceived shape of a dot pattern lie on or close to the Pareto front.*

The next section describes a pilot study which was carried out as a first step in the investigation of this hypothesis.

4 Pilot Study

A small pilot study was carried out to gain an initial estimation of the plausibility of the hypothesis. Eight dot patterns were presented to 13 adult subjects, who were asked to draw a polygonal outline which best captures the shape formed by each pattern of dots. An example dot pattern with two possible polygons was shown (these are our Figures 1 and 2), and more precise rules given as follows:

1. The outline must be a simple closed polygon whose vertices are members of the dot pattern; that is, it must consist of a series of straight edges joining up some or all of the dots, forming a closed circuit.
2. You do not have to include all the given dots as vertices of your outline; but any dots that are not used must be in the interior of the polygon formed, not outside it.
3. The outline must not intersect or touch itself; so outlines such as the two below are not allowed: [here the two non-examples of Figure 5 were given].

The eight dot patterns used in the pilot study are shown in Appendix A.

The results of the pilot study are tabulated in Table 2. The rows of the table correspond to the eight dot patterns. For each dot pattern the following data are given:

- The number of dots in the pattern.
- The total number of polygonal hulls for the pattern.
- The number of Pareto-optimal polygonal hulls for the pattern.
- The maximum number of dominators for any individual polygonal hull.
- The number of *distinct* hulls generated by the subjects: the relevance of this figure is that it shows that the subjects provided a variety of different responses — for none of the dot patterns were there just one or two 'obvious' outlines to draw.
- The number of subjects who responded with a Pareto-optimal hull.
- The mean *relative domination* of the responses — this quantity is explained below.

Our hypothesis was that hulls corresponding to some 'perceived shape' of the dot pattern should lie on or close to the Pareto front in the area-perimeter plot. Totalling the figures in the penultimate column of the table, we see that 57 out of the total 104 responses were Pareto-optimal. The figures in the fourth column give the number of Pareto-optimal hulls available for that dot pattern, an indication of the size of the 'target' if our hypothesis is correct. The fifth column in the table shows the maximum number of hulls by which any given hull for that dot pattern is dominated: it will be seen that this always falls short of the total number of hulls, but not usually by much.

A measure of the extent to which a hull falls short of being Pareto-optimal is given by the 'relative domination', that is, the ratio of the number of hulls which dominate it to the the maximum number of hulls that dominate any one hull for that dot pattern. The relative domination for any individual hull is thus obtained

Table 2. Results of pilot study involving 13 subjects and 8 dot patterns

Pattern	No. of dots	No. of hulls	Pareto-opt. hulls	Max. no. of dominators	Distinct responses	Pareto-opt. responses	Mean rel. dom.
1	12	5674	43	5186	9	8	0.000252
2	12	14095	81	13023	12	5	0.002640
3	11	1246	38	996	8	10	0.004943
4	12	1826	23	1632	10	7	0.002168
5	13	74710	61	73205	12	6	0.000139
6	11	3303	29	3024	12	4	0.000738
7	11	3637	36	3322	11	6	0.003473
8	11	8308	72	7630	5	11	0.000323

by dividing the number of dominators of that hull by the number of dominators of a maximally dominated hull. The relative domination ranges from 0 for a Pareto-optimal hull to 1 for a maximally-dominated hull. For the hypothesis to be corroborated, we should expect the relative domination of subjects' responses to be consistently close to 0, and this is indeed what we find. The rightmost column of the table shows the mean relative domination across all thirteen subjects, for each dot-pattern. The highest *individual* value for the relative domination was 0.008578, for a response to dot pattern 2 which was dominated by 118 out of the 14,095 hulls for that pattern. Compare this with the jagged rightmost hull in Figure 7, which has a relative domination of 0.2347.

If Pareto-optimality had no influence on the subjects' selection of polygonal hulls, we should expect the relative frequency of Pareto-optimal hulls selected for any of the dot patterns in the pilot study to approximate the relative frequency of Pareto-optimal hulls in the full set of hulls for that pattern. For example, for pattern 3, only 3% of the hulls are Pareto-optimal, which means that we should expect a Pareto-optimal hull to be chosen by $0.03 \times 13 \approx 0.4$ subjects. Summing the corresponding values for all the dot patterns, we would expect about 1.1 out of the 104 responses to lie on the Pareto front, on the hypothesis that Pareto-optimality is not a relevant factor. This should be compared with the 57 Pareto-optimal responses actually observed. A chi-squared test gives $\chi^2 = 2872$, considerably larger than the value of 10.827 required for statistical significance at the 0.1% level. From our observations, the chance that Pareto-optimality has no influence on subjects' choices is effectively zero.

In Appendix C are shown, for each dot pattern, the points on the Pareto front (small dots), and the points corresponding to the hulls chosen by the subjects in the pilot study (circles). Comparing these with the full set of hulls illustrated in Appendix B, one obtains a good idea of how closely the hulls drawn by human subjects to represent the perceived shape of the dot pattern adhere to the Pareto front.

In conclusion, the results of the pilot study lend considerable support to the hypothesis that the perceived shape of a dot pattern will tend to be Pareto-optimal with respect to minimising both area and perimeter.

5 Next Steps

The pilot study reported here is limited in both scale and scope. There are many possibilities for further work to examine a range of additional factors with larger-scale experiments. Here we list a number of such possibilities.

1. *Choice of dot patterns.* The dot patterns used in the pilot study were chosen on the basis of an informal idea that they were in some way 'interesting'. As such, they no doubt incorporate an unconscious bias towards patterns of a particular type. To be sure that our results remain valid over the full range of possible dot patterns, it will be necessary to adopt a more principled approach to the selection of the patterns, e.g., using a randomised procedure to generate the patterns. It will also be necessary to investigate larger dot patterns, but the inherent intractability of any algorithm to generate all the hulls for a given pattern would make this impractical for patterns much larger than those already considered. Alternative approaches, involving sampling from the full set of hulls, may have to be considered instead.

2. *Choice of experimental procedures.* Instead of asking subjects to draw hulls for the dot patterns presented, other tasks may also yield useful information. Examples are
   (a) Subjects are presented with a selection of possible outlines for a dot pattern and asked to choose the one which, for them, best represents the perceived shape of the pattern.
   (b) Subjects are presented with pairs of outlines for a given dot pattern and asked to select the preferred outline.
   (c) Subjects are presented with a selection of possible outlines for a dot pattern and asked to rank them in order of acceptability.
   (d) Free-form commentary: in any of the above situations, subjects are invited to explain why they judge one outline to be more acceptable than another.

3. *Application context.* A possible concern with any of the above procedures is that they are assumed to be conducted in the absence of any proposed application context. Subjects are not being asked to rate the outlines as good *for* anything in particular, but merely what looks 'right' to them. A priori, one might suppose that this would prove problematic for some subjects, although in the pilot study it was found that subjects were very willing to treat the task as an abstract exercise without reference to any application. However, most of the subjects in the pilot study were university-educated, many of them actually working in the university, and if a wider-ranging set of subjects is used, this may become a more serious consideration, and it may be appropriate to embed the tasks in some 'real-world' problem context (e.g., map generalisation) in order to provide better motivation.

4. *Other objective criteria.* The results of the pilot study were only examined from the point of view of the area/perimeter minimisation hypothesis. But no doubt other factors are involved: in particular, once it is established that preferred outlines tend to lie on or close to the Pareto front of the area/perimeter

graph, the obvious question is what further factors influence exactly whereabouts on the Pareto front the preferred solutions will be found. As the examples in the pilot study show, the Pareto front may take various forms. The point of maximum curvature sometimes assumes the form of a well-marked 'knee', to the right of which the slope is quite gentle, representing a series of hulls with increasing area but similar perimeter. A priori one might expect the preferred hulls to lie towards the left of this series, near the knee, but the experimental results do not really bear this out. Further investigation is needed to determine what factors influence the location of the optimal hulls along the front. Factors that might be considered include *sinuosity* (a measure of which is the number of times the outline changes from convex to concave or vice versa as it is traversed), or the number of vertices in the hull. Both of these are to some extent correlated with perimeter, although the correlation is far from exact. One might also wish to investigate other factors such as symmetry, which undoubtedly affect visual salience.

5. *Evaluation of algorithms.* Having established an appropriate set of criteria for evaluating polygonal hulls, one can then begin experimenting with different algorithms. Many of the published algorithms for producing outlines of dot-patterns yield polygonal hulls in the sense defined in this paper, and an obvious first step would be to investigate to what extent these algorithms tend to produce outlines that are optimal according to the criteria that have been established. In particular, most of the existing algorithms involve a parameter — typically a real-valued length parameter, but in the case of the k-nearest neighbour algorithm of [6], it is a positive integer. It would therefore be interesting to investigate how the objective evaluation criteria vary as the parameter is varied: one could, for example, trace the path followed by an algorithm's output in area-perimeter space as the parameter runs through the full range of its possible values, and hence find which parameter settings optimise the quality of the output. For the hulls shown in Figure 4, for example, the number of dominators in area-perimeter space are 0, 5, 4, 0, 5, and 0 respectively, suggesting that this algorithm, like our human subjects, is very good at finding hulls on or near the Pareto front.

6. *Algorithm design.* Going beyond this, one might also ask whether it is possible to *design* an algorithm with those criteria in mind, that is, to tailor an algorithm to produce hulls which are optimal with respect to the criteria. With larger point sets, one can only expect to identify the Pareto-optimal hulls to some degree of approximation, suggesting that a fruitful approach here might be to use some form of evolutionary algorithm.

7. *Extension to three dimensions.* Many of the ideas discussed here could probably be generalised to apply to three-dimensional dot patterns. A hull must now be a volume of space bounded by a polyhedral surface rather than an area bounded by a polygonal outline: a 'polyhedral hull'. Some, but not all, of the algorithms that have been used for generating outlines of two-dimensional dot patterns readily generalise to three dimensions; little work has been done on this, though the Power Crust algorithm of [12, 13] is not unrelated. There would be obvious practical difficulties in asking experimen-

tal subjects to construct polyhedra in space rather than drawing outlines on a piece of paper, but no doubt some suitable experiments could be devised. For the time being, however, the two-dimensional case already offers ample scope for further investigation.

Acknowledgments

References

1. H. Edelsbrunner, D. G. Kirkpatrick, and R. Seidel. On the shape of a set of points in the plane. *IEEE Transactions on Information Theory*, IT-29(4):551–559, 1983.
2. Gautam Garai and B. B. Chaudhuri. A split and merge procedure for polygonal border detection of dot pattern. *Image and Vision Computing*, 17:75–82, 1999.
3. M. Melkemi and M. Djebali. Computing the shape of a planar points set. *Pattern Recognition*, 33:1423–1436, 2000.
4. A. R. Chaudhuri, B. B. Chaudhuri, and S. K. Parui. A novel approach to computation of the shape of a dot pattern and extraction of its perceptual border. *Computer Vision and Image Understanding*, 68(3):257–275, 1997.
5. A. P. Galton and M. Duckham. What is the region occupied by a set of points? In M. Raubal, H. J. Miller, A. U. Frank, and M. F. Goodchild, editors, *Geographic Information Science: Proceedings of the 4th International Conference, GIScience 2006*, pages 81–98. Springer, 2006.
6. Adriano Moreira and Maribel Santos. Concave hull: a k-nearest neighbours approach for the computation of the region occupied by a set of points. In *Proceedings of the 2nd International Conference on Computer Graphics Theory and Applications (GRAPP2007), Barcelona, Spain, 8-11 March, 2007*, 2007.
7. H. Alani, C. B. Jones, and D. Tudhope. Voronoi-based region approximation for geographical information retrieval with gazetteers. *International Journal of Geographical Information Science*, 15(4):287–306, 2001.
8. A. Arampatzis, M. van Kreveld, I. Reinbacher, C. B. Jones, S. Vaid, P. Clough, H. Joho, and M. Sanderson. Web-based delineation of imprecise regions. *Computers, Environment and Urban Systems*, 30:436–459, 2006.
9. A. P. Galton. Dynamic collectives and their collective dynamics. In D. M. Mark and A. G. Cohn, editors, *Spatial Information Theory*. Springer, 2005.
10. M. Duckham, L. Kulik, M. Worboys, and A. Galton. Efficient generation of simple polygons for characterizing the shape of a set of points in the plane. *Pattern Recognition*, in press. Accepted for publication, March 2008.
11. Kalyanmoy Deb. *Multi-objective Optimization Using Evolutionary Algorithms*. John Wiley and Sons, 2001.
12. Nina Amenta, Sunghee Choi, and Ravi Kolluri. The power crust. In *Sixth ACM Symposium on Solid Modeling and Applications*, pages 249–260, 2001.
13. Nina Amenta, Sunghee Choi, and Ravi Kolluri. The power crust, unions of balls, and the medial axis transform. *Computational Geometry: Theory and Applications*, 19(2-3):127–153, 2001.
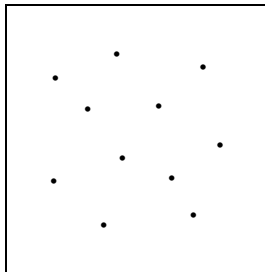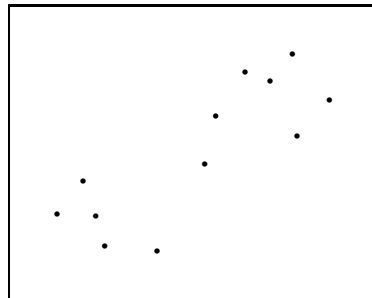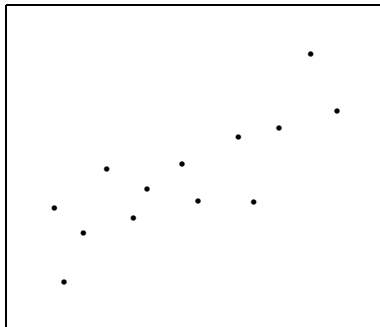
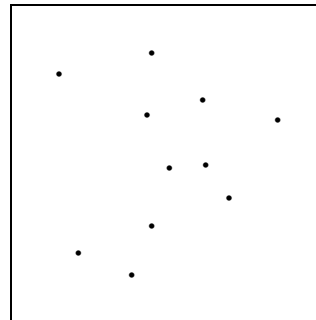# A   The dot patterns used in the pilot study
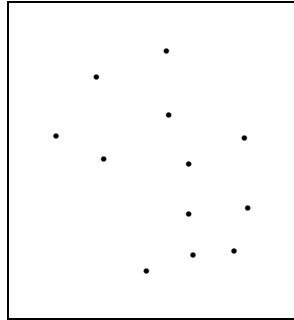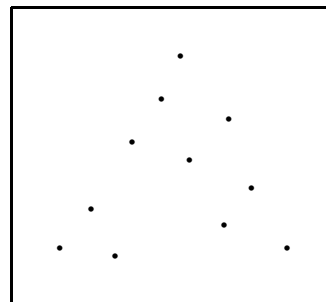
Pattern 1

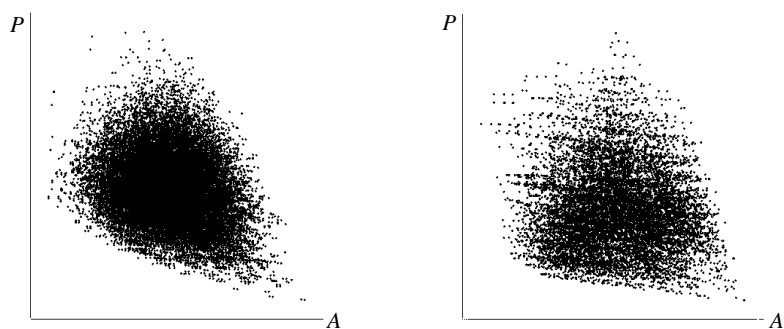Pattern 2

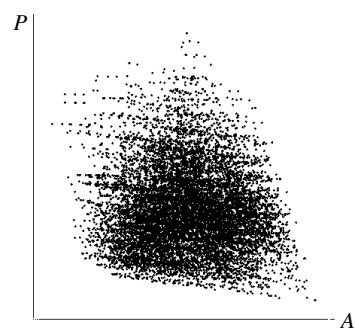Pattern 3

Pattern 4

Pattern 5

Pattern 6

Pattern 7

Pattern 8

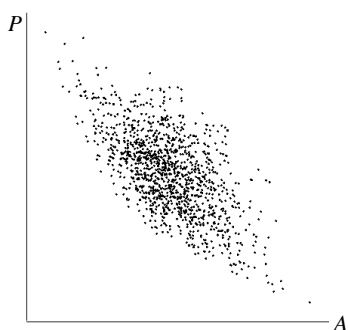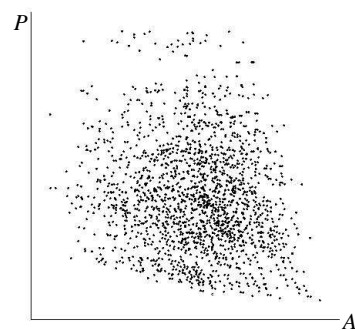# B   Area-perimeter plots for pilot study dot patterns
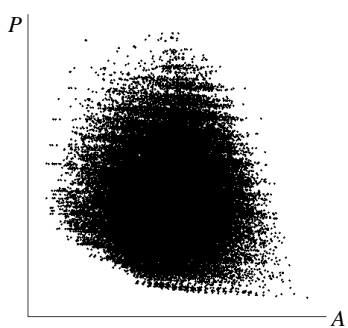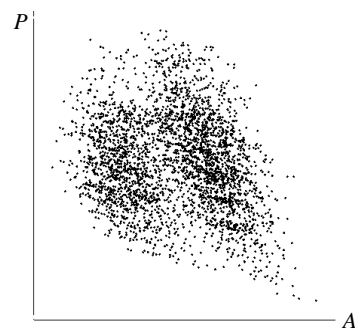

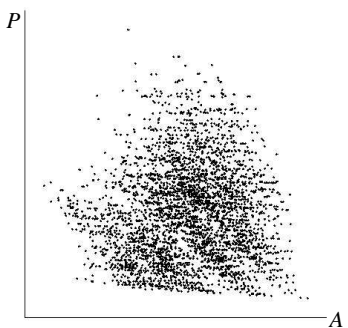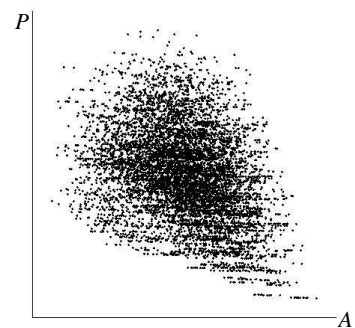Pattern 1

Pattern 2

Pattern 3

Pattern 4

Pattern 5

Pattern 6

Pattern 7

Pattern 8

C Pareto fronts, with pilot study responses

Pattern 1

Pattern 2

Pattern 3

Pattern 4

Pattern 5

Pattern 6

Pattern 7

Pattern 8