New Prospects for Information Theory in Arts Research

Alan Marsden, Lancaster University, Lancaster Institute for the Contemporary Arts, Lancaster LA1 4YW, UK. Email: <a.marsden@lancaster.ac.uk>. ORCID: 0000-0001-9822-4317

Abstract

Information Theory provoked the interest of arts researchers from its inception in the mid-20th century, but failed to produce the expected impact partly because the data and computing systems required were not available. With the modern availability of data from public collections and sophisticated software, there is renewed interest in Information Theory. Successful application in the analysis of music implies potential success in other artforms also. I give an illustrative example, applying the information-theoretic similarity measure Normalized Compression Distance with the aim of ranking paintings in a large collection by their conventionality.

The Early Period of Information Theory and Arts Research

"Information Theory has, in the last few years, become something of a scientific bandwagon," wrote Claude Shannon in 1956 [1]. The paper which founded the theory had been published eight years earlier, and republished in book form with an explanatory and discursive essay by Warren Weaver [2] just a year after that. Shannon could have omitted the word "scientific" since, as he noted himself, Information Theory had been applied in other domains also. Examples from the time include groundbreaking work by Leonard Meyer on music [3] and Abraham Moles' *Théorie de l'information et perception ésthetique* [4]. However, by the time the English translation of Moles' book was published in 1966 [5], a number of reviewers noted that enthusiasm for Information Theory had declined, some apparently taking this as evidence that it was not applicable to the arts. Indeed there are sound arguments that some have confused aesthetic information with the information which Information Theory measures [6]. I do not have space to consider these arguments here; suffice that I aim to concentrate on studies which are not guilty of this confusion. In this paper I re-examine Information-Theoretic research in the arts, and argue for the renewed

1

potential of the theory at this juncture. A bibliography on Information Theory and the arts is given in the supplementary material.

The revolutionary aspect of Information Theory is a method of quantifying information. Shannon's original, explicit concern was how to communicate information efficiently through a channel of limited capacity in the presence of noise. The essential insight of the theory is that the quantity of information carried by an item, sometimes called its *surprisal*, is inversely related to the probability of occurrence of that item: if we know for certain what a person is going to say next, their saying it conveys to us no information. The mathematically convenient formulation of this is as the negative logarithm of the probability. The rate of information expected in sending a message -- the information content per item -- is then the average of the surprisal of all items weighted by the probability of their occurrence:

$$H = -\sum p_i \log(p_i)$$

where p_i is the probability of item i. Because of its similarity to the formula for entropy in statistical mechanics, Shannon chose to use the same term for this quantity in Information Theory.

Information Theory immediately appealed to arts researchers because it spoke to shared concerns with the expectedness of an event or item within a work or in relation to other works in a genre. It also gave a possible explanation for moderation, noted as a valuable quality in aesthetic theories from the "Aristotelean mean" of classical and medieval theory to Berlyne's theory of "optimal complexity" [7]. Arts researchers were as fascinated as others with the notion of entropy, which appeared to have the power to unite science with art. Weaver wrote excitedly of finding that entropy had a role in communication theory and quoted with approval Arthur Eddington's joining of entropy with beauty and melody [8].

Despite their excitement at the prospect of a scientific approach to aesthetics, the work of Moles and others contained little actual calculation. A 1971 book on algorithmic aesthetics [9] contained only toy examples rather than discoveries about genuine works of art. The most substantial musical study, by Hiller and Bean [10], found an expected increase in entropy from Mozart through Beethoven and Berg to Hindemith. A study by Bailey from the same period [11] used Information Theory in the analysis of phonemes in poetic and prose texts, but his positive results came in fact from simple comparisons of the relative frequency of

phonemes. Nake refers to a number of quantitative studies carried out by students of Max Bense in the 1960s [12].

Data was not easy to come by. Bailey used a handful of phonemic transcriptions made by other researchers and made a few of his own. Hiller & Bean gathered their own data (apparently by hand) and only came to use a computer in a subsequent study. The situation now is completely different. Large quantities of data are relatively easy to obtain, and software tools and expertise are also available.

Recent Applications of Information Theory in Music Research

Music has the linear and temporal form which Shannon took as axiomatic of communication, so Information Theory is more readily applicable to music than, say, visual art. It is in the field of Music Information Retrieval in particular (already a conjunction of engineering and aesthetic concerns) that Information Theory has had its greatest revival.

One example is a test of "compression distance" in the analysis of music. Data compression typically relies on predictability, and compressing a sequence is ideally equivalent to encoding it in a way which allows it to be sent through a communication channel using the minimum number of tokens. If using the information from one sequence allows a second sequence to be substantially more compressed, the compression distance between the two sequences is low. The content of the second sequence becomes more predictable (less surprising; containing less information) when it follows the first sequence. Louboutin and Meredith [13] applied this measure to two sets of musical data where similarity was already known: a set of 163 Dutch folksongs which had been assigned by experts to "tune families," and the fugues from J.S. Bach's *The Well-Tempered Klavier, book 1*. The authors found that both general purpose compression algorithms and a specialized music-compression algorithm were effective in distinguishing automatically between similar and non-similar items.

A more direct application of Information Theory, and the most extensive in recent music research, is found in the system called IDyOM (Information Dynamics of Music) [14]. IDyOM is based on variable-order Markov models (ones which take into account a varying number of previous items) and it aims explicitly to model a listener's changing expectations as they listen to a piece of music. Its essential behavior is to take a sequence of musical symbols as input and determine the probabilities for the next symbol. From this, one can infer what musical event is most expected, and also calculate how surprising the actual next event

is. These are the outputs one would expect from any Markov model, but IDyOM has some additional features designed to model aspects of both the process of listening and the process of musical acculturation. The model has separate "long-term" and "short-term" components (illustrated in Fig. 1), modeling the expectations derived from knowledge of music in general ("long-term") and expectations derived from what has previously occurred in the piece currently being listened to ("short-term"). Crucially, the system determines which "viewpoints" of the data, or combinations of viewpoints, result in the lowest entropy, i.e., which allow for the most efficient modeling of the data, and uses those.
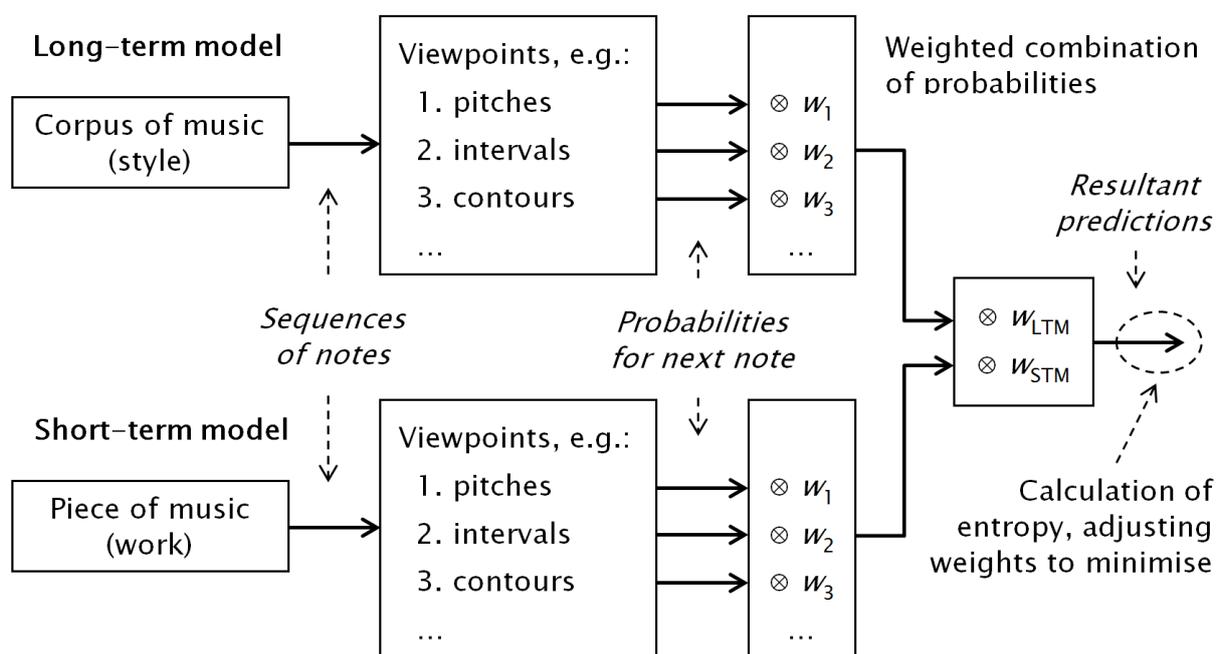

Figure 1. Schematic diagram of IDyOM model.

Since the first article about what was to become IDyOM was published in 2004 [15], nearly thirty papers have been published which make use of IDyOM in one way or another. Six are concerned with the theory of the model and of models like it [16]. Seven test the model's predictions of expectation against experimental data on listeners' expectations of melodic continuation, measured by various means including direct ratings, behavioral measurements, and EEG [17]. More interesting perhaps are the applications of IDyOM which reveal something about other aspects of music, constituting half of the published papers. Five of these concern melodic grouping, i.e., dividing a melody into phrases and sub-phrases [18]. Other applications, represented by eight publications, concern emotional responses, expressive performance and tension, meter induction, comparisons between styles, the

influence of expertise, and similarity [19]. The model has also found applications in neuroscience and linguistics [20].

Information Theory and Arts Research Questions

The achievements of IDyOM are impressive, but it has rarely been applied to the kind of question musicologists typically ask. Only one of the published applications of IDyOM includes an extended study of a specific piece of music, an analysis of the expectation and surprise at each moment of a piece by Philip Glass [21].

In visual art, there was early interest and debate provoked by Moles' book referred to above, and by an essay by Rudolf Arnheim [22]. This too was largely general. Many applications of technology to research in visual art have used theoretical tools other than Information Theory, and they have often focused on predicting the level of beauty in an image [23]. There have been some recent applications of Information Theory (see the bibliography in the supplementary material) but, showing the same kind of distance from the concerns of arts researchers as in the case of music research, these have often concerned generalized beauty rather than specific aesthetic qualities. A notable exception is the work of Jaume Rigau and his co-workers, who have used Information-Theoretic measures to examine color palate and differences and developments of style [24].

In the case of literature, too, the focus has been more on studies of bodies of work than on individual works. Information Theory is readily applicable to language (indeed Shannon used language to illustrate his theory), but studies have often addressed general problems such as author attribution (see the supplementary bibliography). Language and literature researchers have been quick to take the opportunities afforded by this age of Big Data. In the field of corpus studies, computers extract summary data, for example, word counts or collocations, drawing conclusions from large corpora. But despite this being called "distant reading" by some, such an approach does not allow the conclusions which can be drawn from actually reading those works. Information Theory, with its calculation of expectation and surprise which form such an important part of reading (and viewing and listening) affords a possible basis on which a computer might "read" for us, and the study of the piece by Philip Glass demonstrates that conclusions about individual works are possible.

An Example: Finding "Non-Conventional" Paintings

As an illustration of what is possible in visual art, and how Information Theory can be used to investigate aesthetic qualities, I present a simple application of Information Theory to investigating conventionality in a large collection of paintings. Art UK [25] makes available images of over 200,000 artworks from collections in the UK. This study presents a way of ranking works within categories from that collection according to their putative conventionality. It is not intended as a conclusive study but merely as an illustration of what can be achieved by a relatively simple application of Information Theory.

Conventional works are assumed to be those which are similar to other works in their category. As a measure of similarity this study employed the Information-Theoretic measure normalized compression distance (NCD), as used in the musical study referred to above. Strictly speaking NCD uses Algorithmic Information Theory or Kolmogorov Complexity rather than Shannon's Information Theory, but the two kinds are closely related [26]. As mentioned above, the compression distance between two items measures how much the information from one item allows greater compression of the other item. Its basic formula is

$$NCD = \frac{Z(xy) - \min(Z(x), Z(y))}{\max(Z(x), Z(y))}$$

where $Z(x)$ is the size of the result of compressing x and $Z(xy)$ is the size of the result of compressing the composition of x and y together.

NCD has been used successfully as a measure of the similarity between images in prior research but usually applied to photographs and often to faces or other restricted types of image [27]. Rigau used NCD to compare parts of images with each other as a measure of the composition of artworks [28]. Several caveats about the use of NCD for images are evident from these works. Different measures of similarity arise from using different image-compression algorithms. Strictly, the compressor needs to be "normal" (meaning roughly that the relation of compression to information content is well behaved), but common image-compression algorithms are not. Nevertheless, Rigau was able to get useful results using JPG compression. Other researchers found that the best results came from specially designed compressors and/or complex methods of composing x and y into xy. Both of these are beyond the scope of this simple illustration.

For this study, the following method was chosen for its simplicity and because, among other simple options, it produced the smallest NCD measures between two copies of the same

image (one of the indications of normality). Searches of the Art UK website employing several different tags were made, delivering large numbers of images scaled to a uniform width of 300 pixels. Each was initially converted to a bitmap from its original JPG format. The images were then rotated through 90 degrees and concatenated left-right so that each scan line contained pixels from both images and there were no blank regions. The single images and concatenated images were compressed into PNG format and the resultant file sizes recorded, from which the NCD was computed. Having calculated the NCD between every pair of images in each set of images with each tag, a putative measure of an image's conventionality within each set was calculated as the average NCD between an image and all others in the set.

Figure 2 shows, among 1,432 images tagged "horses" downloaded from the Art UK website, six of the nine images with the lowest average NCD to other images tagged "horses", six of the nine with middling average NCD, and six of the nine with largest average NCD [29]. If NCD usefully measures the similarity between images, these should be, respectively, examples of the most conventional, middling and least conventional paintings of horses. Proper testing of this would require some other credible measure of conventionality or some empirical data for comparison, but this is not readily available. These results therefore remain only suggestive, but it is worth noting some evident similarities between the "most conventional" images. All of them have a horizon near the middle of the painting (either lacking or less prominent in the "less conventional" images) and they also mostly have cloudy skies and shades of brown.
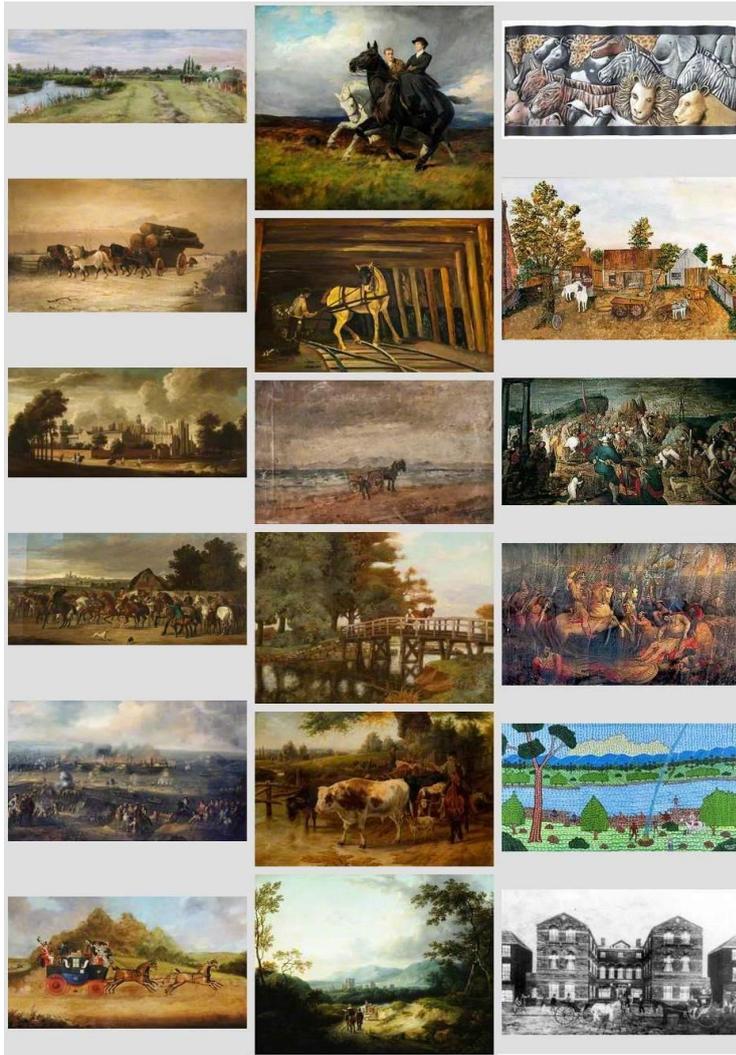
Figure 2. Left column: six images with smallest average NCD to other images tagged "Horses"; center: six images with middling average NCD; right: six images with largest average NCD. (For details, credits and copyright information, see supplementary materials).

Similar results come from analysis of 264 paintings tagged "fruit and vegetables" (Fig. 3). All the "conventional" paintings have fruit and vegetables arranged on a table top seen from an angle above. The middle-ranking paintings are much more varied, and the most distant ones are in some cases barely of fruit and vegetables at all.
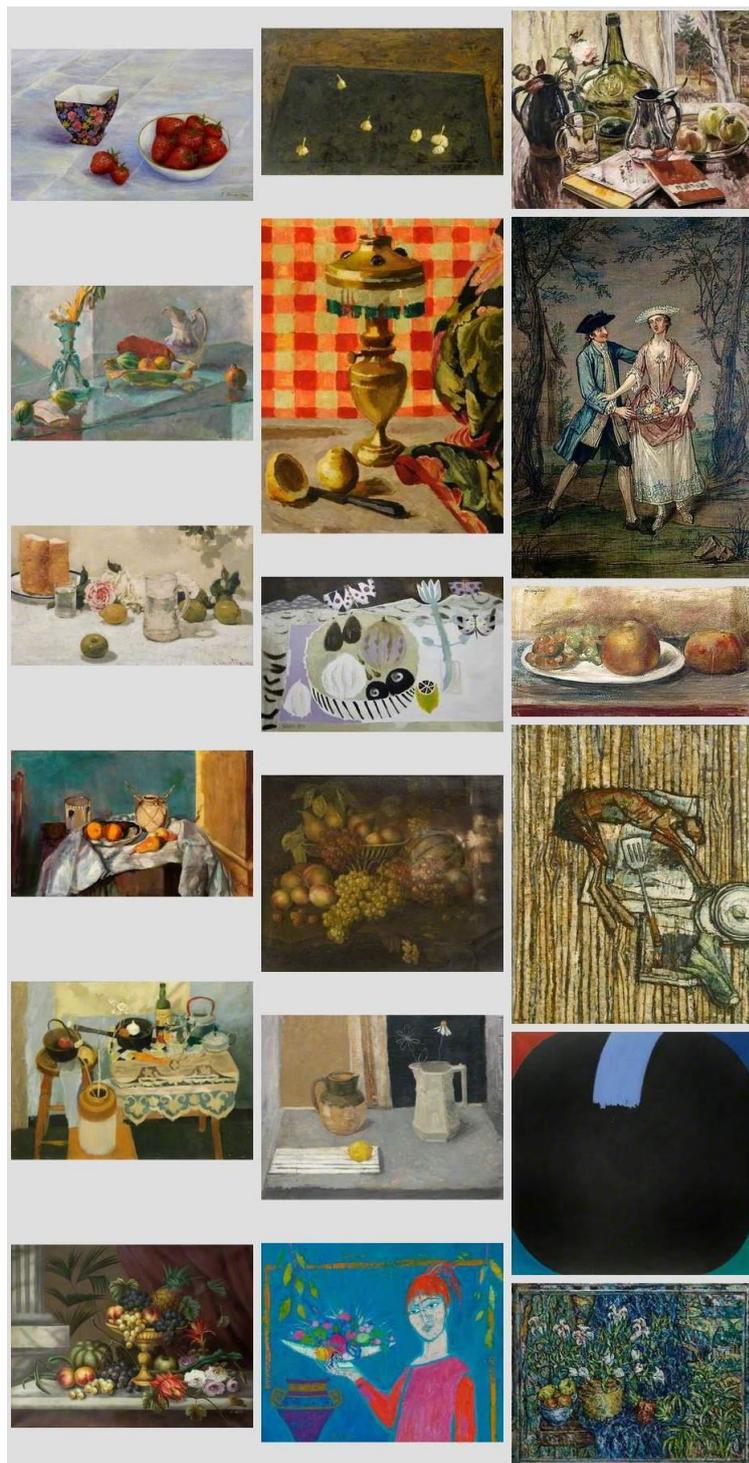
Figure 3. Left column: six images with smallest average NCD to other images tagged "Fruit and vegetables"; center: six images with middling average NCD; right: six images with largest average NCD. (For details, credits and copyright information, see supplementary materials).

Results similar to those presented in Figs 2 and 3 could have been derived by first finding out what makes each type of painting conventional (perhaps central horizons for horses, table tops for fruit and vegetables, etc.) and then finding an image-processing method to recognize these things. However, one advantage of the approach employed here is that no such prior assumptions about conventionality are required. The method can be applied to any set of images, and, in theory at least, the similarities discovered arise from each set according to its common features. In practice, the similarities will depend on the assumptions inherent in the choice of compressor and method of composition. Further research is thus required before any claim for a general method of measuring conventionality can be made. It should also be noted that the method presented here requires calculation of the NCD between every pair of images in a collection, and so is not scalable to truly massive collections.

Nevertheless, the study does illustrate what is possible in applying Information Theory, even in a manner much less sophisticated than an explicit predictive model like IDyOM. To develop and apply a tool of the sophistication of IDyOM to visual art would require several hurdles to be overcome. One problem is how to account for spatial data using a theory designed to work with sequences of symbols. Eye-tracking studies give information about how viewers scan the salient parts of a painting and there are computational tools for determining the salient regions of an image. These two lines of research might yield a way of sequentializing the data in an image in a more meaningful way than simple rasterization. A second problem is that IDyOM uses notes, units of information at a higher level than the audio signal. It is far from self-evident what level of information a predictive model like IDyOM should use for paintings or how that information should be derived from the image. Whatever the method for applying Information Theory, given the deluge of data facing researchers of the future it is likely to become an essential tool for research.

Acknowledgements

References and Notes

1. Claude E. Shannon, "The bandwagon," *IRE Transactions on Information Theory* **2**, No. 1, 3 (1956).

2. Claude E. Shannon and Warren Weaver, *The Mathematical Theory of Comunication* (Urbana: University of Illinois Press, 1949).

3. Leonard Meyer, "Meaning in music and information theory," *Journal of Aesthetics and Art Criticism* **15**, No. 4, 412--424 (1957).

4. Abraham Moles, *Théorie de l'information et perception ésthetique* (Paris: Flammarion, 1958).

5. Abraham Moles, *Information Theory and Esthetic Perception* (Urbana: University of Illinois Press, 1966).

6. For a recent example, see Frieder Nake, "Information Aesthetics: An heroic experiment," *Journal of Mathematics and the Arts* **6**, No. 2--3, 65--75 (2012).

7. Daniel Berlyne, *Aesthetics and Psychobiology* (New York: Appleton-Century-Crofts, 1971).

8. See [2], p.117.

9. George Stiny and James Gips, *Algorithmic Aesthetics: Computer Models for Criticism and Design in the Arts* (Berkeley: University of California Press, 1971).

10. Lejaren Hiller and Calvert Bean, "Information theory analyses of four sonata expositions," *Journal of Music Theory* **10**, No. 1, 96--137 (1966).

11. Richard Bailey, "Statistics and the sound of poetry," *Poetics: International Review for the Theory of Literature* **1**, No. 1, 16--37 (1971).

12. See [6].

13. Corentin Louboutin and David Meredith, "Using general-purpose compression algorithms for music analysis," *Journal of New Music Research* **45**, No. 1, 1--16 (2016).

14. Marcus Pearce, *The construction and evaluation of statistical models of melodic structure in music perception and composition*. PhD thesis, City University, London, UK (2005).

15. Marcus Pearce and Geraint Wiggins, "Improved methods for statistical modelling of monophonic music," *Journal of New Music Research* **33**, No. 4, 367--385 (2004).

16. Items M14, M27, M30, M32, M40 and M50 in the supplementary bibliography.

17. Items M15, M25, M26, M29, M31, M34 and M37.

18. Items M16, M17, M19, M23 and M24.

19. Items M33, M39, M43, M45, M46, M51, M53 and M54.

20. Items M41, P8 and P9.

21. Keith Potter, Geraint Wiggins and Marcus Pearce, "Towards greater objectivity in music theory: information-dynamic analysis of minimalist music," *Musicae Scientiae* **11**, No. 2, 295--324 (2007).

22. Rudolf Arnheim, *Entropy and Art: An Essay on Disorder and Order*, (Berkeley: University Of California Press, 1971).

23. For a recent review, see Stephen E. Palmer, Karen B. Schloss and Jonathan Sammartino, "Visual aesthetics and human preference," *Annual Review of Psychology* **64** (2013), pp. 77--107.

24. Jaume Rigau, Miquel Feixas, and Mateu Sbert, "Informational aesthetics measures," *IEEE Computer Graphics and Applications* **28**, No. 2, 24--34 (2008). Other works are listed in the supplementary bibliography.

25. <artuk.org>.

26. Ming Li and Paul Vitanyi, *An Introduction to Kolmogorov Complexity and its Applications*, (New York: Springer-Verlag, 2008). See also T14, T15 and T24 in the supplementary bibliography.

27. See items A9, A16, A18, A19 and A20 in the supplementary bibliography.

28. See [24].

29. The selection of six was made on the basis of images which would fit in the figure, without reference to other image characteristics. The full list of works, ranked by the average NCD value and with links to higher resolution images on the Art UK website, is available in supplementary material.

Alan Marsden is a lecturer in the Lancaster Institute of Contemporary Arts at Lancaster University, UK and editor of *Journal of New Music Research*.