Special Issue: Innovation and the Evolution of Human Behavior

Quantifying Novelty in the Archaeological Record

GORDON RUGG
School of Computing and Mathematics, Keele University, Keele, ST5 5BG Staffordshire, UNITED KINGDOM; g.rugg@cs.keele.ac.uk

NIKKI HOLLAND
School of Computing and Mathematics, Keele University, Keele, ST5 5BG Staffordshire, UNITED KINGDOM; nikki@cs.keele.ac.uk

ABSTRACT

This paper describes two methods for quantification of novelty, which are not well known in archaeology. These extend the methodological toolkit available to researchers into culture and innovation. The methods are:

- *Inverse frequency weighting*. This involves allocating a weighting of 1/n to instances of the chosen item, so that rarer items are given a heavier weighting than common ones, in a way which allows summing of novelty across components.
- *Minimum edit distance*. This involves calculating the minimum number of changes which need to be made to change one item into another specified item (e.g., the number of changes involved in changing from one manufacturing process to another).

Worked examples are given, showing how these can be used to quantify the novelty associated with specific changes in the archaeological record, such as the change from stone to copper and bronze. The article also shows how empirical approaches to measurement of novelty can be usefully linked to concepts from other literatures, such as the concept of the *honest indicator*, from evolutionary ecology, and to the literature on empirical aesthetics.

The "Innovation and Evolution" workshop was held at the Centre for the Archaeology of Human Origins, University of Southampton, United Kingdom; workshop papers guest edited by Hannah Fluck (University of Southampton; and, Landscape, Planning and Heritage, Hampshire County Council), Katharine MacDonald (Faculty of Archaeology, University of Leiden), and Natalie Uomini (School of Archaeology, Classics and Egyptology, University of Liverpool). This is article #6 of 7.

INTRODUCTION

This article consists of two main parts. The first is about things we can measure, and discusses ways of measuring novelty, with worked examples of two methods. These examples demonstrate that "novelty" is not a single unitary measure —instead, it is context-sensitive, and can be measured in different ways for different purposes. This gives more flexibility and power than trying to use a single unitary measure, even if that were possible.

The second part of this article discusses whether the perceived attractiveness of an innovation in material culture influences how likely it is to be adopted. This is another field in which empirical approaches have been developed in other disciplines which may be relevant to archaeological research. This part of the article describes the concept of empirical aesthetics, and discusses ways in which this concept can be applied to the study of innovation in archaeology. Although this concept offers considerable promise, it needs to be applied with caution, since there is a significant risk that it can lead to simplistic "Just So story" interpretations. We illustrate this point with examples from the ethology literature.

The concept of measuring novelty empirically is well established, often within the framework of empirical aesthetics or computational aesthetics (discussed in more detail later in this article). Some of this work has focused on the generic issue of measuring novelty. Other research has focused on specific domains and specific problems. Measuring novelty in images, for instance, has received considerable attention, and there is a well-established community of researchers in this field. Much of this research involves testing and comparing different approaches to assess empirically the novelty of an image or of components of an image. This typically involves some variant on Shannon's information theory (Shannon 1948), which is an extremely well established approach, generally viewed as one of the key underlying concepts in digital technology, and widely used in other fields. This use of information theory is often combined with software architecture based on some form of Artificial Neural Nets (a very well established field with an extensive and highly technical literature). There are numerous examples of this approach to measuring novelty computationally, for instance, Saunders and Gero (2001); Matas et al. (2006).

  ISSN 1545-0031
doi:10.4207/PA.2011.ART50

Empirical approaches have been used to assess perceived attractiveness of a range of artifact types, such as Web pages (e.g., Lavie and Tractinsky 2004). Research in this tradition has not only investigated the attributes of the object being aesthetically evaluated, but has also investigated the cognitive processing performed by the evaluator, to see whether there are correlates between perceived attractiveness of an artifact and the cognitive processing routes involved in processing it (e.g., Reberet al. 2004; Jacobsen et al. 2006).

This article describes two methods which fit within this broad tradition, but which appear not to have been widely used in archaeological research into innovation, namely inverse frequency weighting and minimal edit distance.

THINGS WE CAN MEASURE, EXAMPLE 1: INVERSE FREQUENCY WEIGHTING

Inverse frequency weighting is a concept widely used in information retrieval, where it is particularly useful for ranking the results in an online search. Online searches often produce very large numbers of hits and there has been a considerable amount of work in the online search research community into ways of automatically ranking the results so that the ones most likely to be relevant to the user are shown first in the list. Inverse frequency weighting is a widely used way of doing this. It works on the assumption that the more common a term is, the less likely it is to contain much useful information, for instance, the word "system" is ubiquitous, whereas "stochastic" is comparatively rare.

An online search is normally performed on a system where the records can be treated as a closed corpus and whose contents are routinely indexed automatically by software. This means that the search software can easily check how many instances there are of each search term in the corpus being searched, just by looking up the relevant entries in the online database. It might find, for example, that there are 9,260,000 instances of "stochastic" and 1,350,000,000 instances of "system" in the case of a Google search.

The usual approach in online search systems is to give highest ranking to any records which contain all of the terms in the search; for instance, both "stochastic" and "system." After that, the search system shows records which only contain a subset of the search terms—e.g., records containing only "stochastic" without "system"—in inverse frequency order. The usual way of calculating frequency order is to convert the frequency of each term into a fraction, whose value is 1/(number of occurrences of the term in the corpus). So, for instance, the value for "stochastic" would be 1/9,260,000 and the value for "system" would be 1/1,350,000,000. These values are instances of inverse frequency weighting.

This approach is particularly useful because it allows values for given combinations of terms to be calculated. For instance, if the user enters four search terms, and there are records containing various permutations of three but not all four search terms, then the system can calculate the relative weightings of each permutation of three, and present them to the user in the appropriate order.

The same underlying concept can be applied to entities other than search terms, for instance, physical artifacts in a corpus such as a set of finds from an archaeological site or a set of tools in use in a present-day village. Suppose, for instance, that a Neolithic community gathering at a ritual site happens to own 25 fully polished flint axe heads and 75 partially polished flint axe heads. The inverse frequency weightings would be 1/25 and 1/75 respectively. If one of the community turns up with a bronze headed axe, then that axe head would have an inverse frequency weighting of 1/1.

As with online search, the inverse frequency weighting starts to produce more interesting results when dealing with multiple items. Suppose, for instance, that we compare two axes in terms of the inverse frequency weightings of the individual components from which they are made. The first axe in our hypothetical example consists of an axe head and a handle, whereas the second axe consists of an axe head, an antler socket, and a handle. We can now sum the inverse frequency weightings for each component to give values for each axe as a whole. For instance, the first axe might consist of a partially polished flint head (1/75) plus a handle (1/100), whereas the second consists of a fully polished flint head (1/25) plus the innovation of an antler socket (1/1), and a wooden handle (1/100). Translating these fractions into decimals for ease of calculation, the first would therefore have an inverse frequency weighting of (0.0133 + 0.01 = 0.0233) and the second would have the much higher value of (0.04 + 1.00 + 0.01 = 1.05).

If one of the axes is an innovation, whether a new invention or a new trade import, then the inverse frequency value for that axe can be treated as a proxy for the degree of innovation. In the example just given, for instance, the axe with the antler socket would have a novelty value of 1.05, compared to the much lower novelty value of 0.0233 for the other axe.

There are various ways in which this approach can be made more sophisticated, and more suited to the issues involved in archaeology as opposed to information retrieval. One example is the level of granularity used in the analysis. Continuing the previous hypothetical example, we might compare the two axes at the level of "having a wooden handle" and decide that they both had a wooden handle and therefore should both have the same inverse frequency weighting for that component. We might instead, however, go down a level of granularity, and distinguish between different types of wooden handle. At this level of granularity, one of the axes might use a much rarer type of handle than the other—for instance, in terms of the type of wood involved (e.g., oak versus ash or yew) or in terms of using a knot in the wood as the point to receive the axe head as opposed to using a straight grained piece of wood. Similarly, a polished stone axe head might use a rare type of stone as opposed to a common one.

This can be handled by listing the novelty values for each level of analysis separately, which is simpler and clearer to the reader than trying to combine them into a single "boiled down" value, which is likely to be hard to understand, and of questionable usefulness. So, for instance, we might describe an axe as having a novelty value of 0.025 at the level of which components it contains and as having a novelty value of 0.85 at the level of the composition of the axe head.

This raises obvious questions about the choice of levels of analysis and about choice of features within each level of analysis. These choices are ones for the individual researcher to make, in light of the research questions which they are asking. For one researcher, for instance, the levels of analysis might involve the processes and sub-processes used in making the axe; for another, the levels might involve the materials used and the sources or varieties for each of those materials.

Inverse frequency weighting can give some useful measures of novelty and degree of innovation; however, it does not give much insight into the process of innovative change. If a culture decides to adopt an innovation, how many changes, and what sort of changes, will be involved? One way of measuring this is minimum edit distance, described in the next sub-section.

THINGS WE CAN MEASURE, EXAMPLE 2: MINIMUM EDIT DISTANCE

Minimum edit distance involves comparing two items and working out how many changes (i.e., edits) are required to change one item into another. There are various well established formalisms for doing this, such as Hamming distance. The core concept for these formalisms is much the same, and is usually illustrated via the example of changing one word into another, one operation at a time, where the set of permitted operations is specified in advance (e.g., adding a new letter, deleting a letter, and changing one letter into another). So, for instance, we could transform "CHAT" to "BEDS" as follows.

CHAT becomes CAT (deletion)
CAT becomes BAT (substitution)
BAT becomes BET (substitution)
BET becomes BED (substitution)
BED becomes BEDS (addition)

This example involves five operations to edit one word into the other. (Incidentally, there is no requirement for all the intermediate steps to involve acceptable English words rather than nonsense words, but we have chosen to do this for clarity.)

As in the previous example, i.e., of quantifying novelty via inverse frequency weighting, these changes can be at various levels of granularity, so the figure for minimum edit distance will depend on the level of granularity chosen for the analysis. This is a significant difference from traditional uses of minimum edit distances, which tend to operate with one-dimensional strings of text, rather than with hierarchically organized structures.

As a simple initial example, we can compare two axes, one of which has a bronze head and the other of which has a polished stone head, but which are otherwise identical in terms of handle and socket. Each axe consists of three components (head, socket, and handle); to change the stone-headed axe into a bronze-headed axe would involve changing only that one component, so the minimum edit distance would be 1 at this level of the hierarchy. One way of handling minimum edit distances in hierarchies is to show the minimum edit distance for each level of the hierarchy. An obvious refinement is to show the edit distance as a fraction of the total number of editable points—in this case, 1/3, since only one of the three points is being edited.

Describing the change from stone to bronze as "only" an edit distance of 1 top-level change has obvious limitations; the number of lower-level edits entailed by this change will clearly be substantial. Each of the lower levels can be assessed in turn, using the same approach. What may happen in some cases is that all of the lower-level materials and tools turn out to be ones already in use within that culture, so the changes involve only novelty in the uses of the materials and tools. In other cases, as with the transition from stone to bronze, the change will require the introduction of new materials and tools—copper ore, tin ore, tuyeres, moulds, etc. This can be modeled using graph theory (Euler 1741) to represent the production chain in graph theory format (Rugg this volume).

Figure 1 shows this schematically. This figure shows a graph theoretic representation demonstrating how an artifact (at level 0) is composed of two main components (at level 1), each of which is composed of two sub-components (at level 2); one of the level 2 sub-components is itself composed of three sub-sub-components (at level 3).

The graph in Figure 1 shows how an artifact can be logically decomposed into layers of lower-level components and/or materials. This approach is not limited to material composition. An obvious refinement is to take into account the number of processes affected by the change, as opposed to the components, materials, and tools. This can be done using approaches such as process models (e.g., Haidle & Bräuer this volume) to represent systematically the various processes, sub-processes, etc, involved in producing an artifact. The underlying concepts involved are very similar to those described for material components, and for brevity will not be further described here.

OTHER WAYS OF LOOKING AT NOVELTY

The previous sections suggested some ways of quantifying the concept of "novelty" in an innovation. The following sections of this article suggest some other empirical approaches to innovation which have been comparatively little addressed within archaeology and paleoanthropology. The underlying theme is the subjective attractiveness of an innovation—does this affect the likelihood that a particular physical innovation will be accepted and are there any ways of predicting the subjective perceived attractiveness of an innovation from objectively measurable features of that innovation?

There is an extensive literature on the diffusion of innovations, where a classic text is the book of that name (Rogers 2003). There is also a significant literature on this

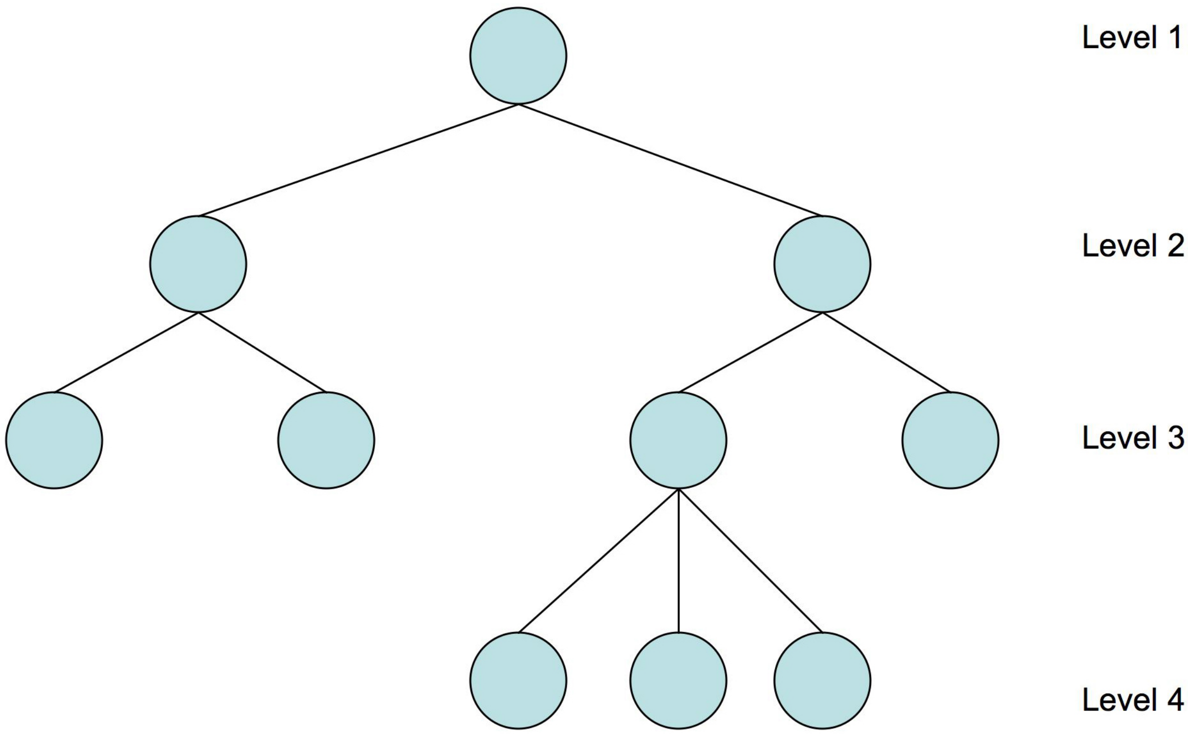

Figure 1. Graph theoretic representation of the layers of components in a hypothetical artifact.

topic with specific reference to archaeology (e.g., Renfrew 1978). Rogers identifies five key attributes of an innovation which affect the rate of adoption. Of these, two are functional: *compatibility*, i.e., the degree to which the innovation is compatible with other products currently in use; and, the *perceived advantage* conferred by adopting the innovation. Two relate to processes: *complexity*, i.e., the perceived ease of use of the innovation; and, the *rate of innovation*, i.e., the rate at which it spreads through the community. The fifth is *observability*, i.e., how conspicuously visible an innovation is to others in the community.

The factor of observability has obvious implications for fashion and for personal adornment. Non-functional personal adornments such as beads are well attested in the archaeological record for tens of thousands of years. Archaeologists and anthropologists are well aware of the significance of status displays, and of the role of rare and high-cost items in those displays. Status is singled out as an important variable by Rogers, not only for non-functional items but also for heavily functional ones; he cites, for instance, an example of farmers buying expensive storage silos as highly visible status symbols, instead of equally functional silos which would be significantly cheaper.

This raises the question of whether there may be consistent features in successful displays, beyond their novelty and their cost. Is it the case, for a given pair of innovations which are equal in visibility, novelty, and cost, that some other factor will consistently affect the likelihood of one being selected in preference to the other? If so, this has implications for interpreting the archaeological and paleontological record.

There is an extensive set of empirical literatures which deal with individuals' responses to observable displays. Much of this work is from ethology and deals with displays in the broader animal kingdom; some of it is from other fields, including neuropsychology. Taken together, these bodies of research imply that it is possible to predict to at least some extent what people's responses will be to a given observable innovation, across times and across cultures. In this part of the article, we briefly describe these literatures, and discuss how this approach can be applied to the archaeological record. We also highlight some areas for caution when using this general approach.

EMPIRICAL APPROACHES TO STUDYING DISPLAYS AND PHYSICAL INNOVATIONS

Although the distinction between functional and non-functional innovations might appear an obvious one, the reality is more complex. Some factors are clearly and uncontroversially functional, and many of these are easy to measure. In the case of an innovation involving an axe, for instance, one obvious objective factor is how long it takes to cut down a tree with the new type of axe as opposed to the old type of axe.

Another obvious objective factor is the production cost which has gone into an artifact. It is tempting to assume that an innovation which is lower-cost is more likely to be successful. The full story, though, is more complex (Rogers 2003). A high-cost item is typically associated with higher social status than a low-cost item. One reason for this is that the high-cost item signals that its owner has access to more resources than others in the community. A very similar concept has been extensively studied in ethology, where it is known as an *honest indicator*. In ethology and related fields, an honest indicator is an observable attribute which accurately reflects the reproductive fitness of its bearer, for instance, plumage quality in birds is generally an honest indicator. There is a well established body of research investigating the role of biological honest indicators in human society, such as whether or not human facial attractiveness is an honest indicator of health (e.g., Kalick et al. 1998; Rhodes et al. 2001). This concept has been applied in archaeology by, e.g., Mithen (2003), writing about handaxes, which he describes as "reliable indicators for four specific dimensions of fitness: resource location abilities, planning ability, good health, and capacity to monitor other individuals within the group." Similar cases have been made by other researchers (e.g., Kohn and Mithen 1999). Chamberlain applies a similar approach to a range of archaeological topics, including handaxe design and landscape preferences (Chamberlain 2000). There has, though, been debate about the applicability of such approaches within the archaeological literature (e.g., Machin 2008; Hodgson 2009a; Hodgson 2009b; Nowell and Chang 2009; Mellars 2010).

Ryan (1990) summed up the honest indicator approach succinctly: "It can suggest how females might evolve preferences in adaptive mate choice, how sensory biases could determine the direction of the runaway process, and how males might evolve traits that exploit pre-existing sensory biases of the female (sensory exploitation)." For brevity, we will ignore the considerable debate in evolutionary ecology about the respective importance of male and female choice in sexual selection (e.g., Buss 2006). Two key phrases for our purposes are "pre-existing sensory biases" and "the runaway process." The first of these implies that animal sensory systems are predisposed to have biases in particular directions. This is well supported by a considerable literature, and is discussed in more detail below. The second phrase refers to a particular form of sensory bias, known in ethology as "superstimulation" or "supernormal stimuli" (e.g., Ryan 1990). In brief, a superstimulus involves a sensory bias encountering something outside the usual range and being super-stimulated as a result. A textbook example is parent birds responding to the stimulus of the yellow skin around a nestling's mouth, by feeding the nestling. Cuckoo chicks have much bigger expanses of yellow skin than the parent birds' own chicks; this stimulates the parents to feed the cuckoo more than the other chicks. Superstimuli are a widespread feature in mating displays across the animal kingdom (e.g., Ryan 1990). Obvious potential cases in human culture include facial cosmetics, high heels, and shoulder pads in clothing. The issue of superstimuli in human culture is discussed in some detail in relation to art, under the name "peak shift effect" by Ramachandran and Hirstein (1999).

There are some robust findings in the literatures on human sensory biases and aesthetics, such as the tendency for people to rate symmetrical images as more attractive than asymmetric ones (e.g., Cardenas and Harris 2006). Other well-established findings include the following:

- Some preferences appear to be hard-wired, for instance, even very young babies prefer symmetrical faces to asymmetric ones.
- Prototypical faces, such as those produced in composite photos, tend to be viewed as more attractive than most of the individual photos which contributed to the composite (e.g., Rhodes et al. 2001).
- Novel things tend to be viewed as more attractive than familiar things, as long as they are not too novel (e.g., Forsythe, Mulhern and Sawey 2008).

At least some of these findings appear to apply across cultures; for example, consistent results have been found across cultures with regard to perceived attractiveness of human faces (e.g., Fink and Neaves 2005; Rhodes et al. 2001), and the innovation literature suggests that similar effects are found in regard to novelty across cultures (e.g., Rogers 2003).

We might therefore infer, extrapolating from biological to technological signals, that an innovation which sends out a clearly visible high-cost signal is more likely to be successful than one which does not, other things being equal. The first metal artifacts were honest indicators, being rare, and not easily counterfeited, so on this count they were likely to be successful innovations, even if they were not functionally more effective than their stone equivalents.

This extrapolation raises a broad set of research questions. Is it the case that some of the objective features involved in honest biological signalling carry across into subjective preferences for features of artifacts? For instance, do humans extend the underlying principle of preferring symmetry, not just to preferences for symmetrical faces, but also to preferences for objects in general, such as houses and handaxes? If so, this would be a source of systematic preferences affecting which physical innovations were perceived as desirable and which were not. There is suggestive evidence that this is the case, as discussed below. However, this approach needs to be treated with caution, since the full story is more complex than it might appear at first glance.

Another literature which has taken an empirical approach to the study of visual displays is *computational aesthetics*. In computational aesthetics, the usual approach is to produce computer-generated images and then to ask respondents to rate the images for subjective aesthetic attractiveness. This approach has the advantage of enabling the researcher to have full control of the objective features of the image. Another approach within computational aesthetics is to take existing objects or images, and to ask respondents to rate these for subjective aesthetic attractiveness, after which the researcher measures objective features in the object or image (e.g., an application of this approach to Web pages by Lavie and Tractinsky (2004); c.f. also Mosshagen and Thielsch [2010] for a typology of visual aesthetics intended for Web designs). This has the advantage of focusing on real objects and images. One subject which has been frequently studied within computational aesthetics is perceived attractiveness of fractals (e.g., Spehar et al. 2003); because the mathematical features of fractals can be specified in some detail, this subject has obvious attractions to researchers. Similarly, the information content of an image can be readily manipulated and correlated with aesthetic ratings (e.g., Avital and Cupchik 1998).

A third empirical approach is to start from the neurology literature, so as to identify attributes of objects and images which are likely to be significant. A prominent example of this approach is a widely-cited article by Ramachandran and Hirstein which outlines a neurological approach to explaining aesthetic experience (Ramachandran and Hirstein 1999). This literature includes use of brain scans to identify the underlying neurological processes involved in aesthetic judgments (e.g., Vartanian and Goel 2004; Jacobsen et al. 2006). A closely related approach involves the information processing literature. Information processing issues, such as reduction of cognitive load, have been proposed as a factor in aesthetics by various authors (e.g., Enquist et al. 2002; Forsytheet al. 2008; Reber et al. 2004; Jacobsen et al. 2006; Winkielman et al. 2006).

Although these literatures come from very different origins, they are generally in agreement about the key features which they are trying to explain. They would all suggest, for instance, that if other things are equal, an innovative object which is symmetrical and moderately novel, and which is honestly signalling high cost, is more likely to be perceived as attractive than one which is not. So, for instance, if two axe heads are equally efficient and equally costly, then the one most likely to be preferred by members of a community is the one which is the more symmetrical, moderately novel, and honestly signalling that it is high cost. From this point of view, the first copper axeheads had a lot going for them in non-functional terms. The same factors provide a plausible argument for why the degree of polish on polished stone axe heads often goes far beyond what is functionally necessary.

GROUNDS FOR CAUTION

There are, however, grounds for caution when attempting to apply these literatures to prediction of the success of apparently non-functional innovations. One obvious potential criticism is the role of culture in human behavior—to what extent are regularities in human aesthetic preferences, for instance, a reflection of ephemeral cultural preferences rather than biological biases hard-wired into the human brain? Ramachandran and Hirstein's 1999 article re-stoked the flames of this debate; tellingly, an editorial about this article in the *Journal of Consciousness Studies* (Anonymous 2001) was entitled "Another Front in the Science wars?" Meta-analyses, such as that by Langlois et al. (2000), who surveyed 1,800 articles on the topic, have concluded that there is solid evidence for cross-culture regularities in aesthetic preferences with regard to perceptions of human attractiveness. However, that is not the same as demonstrating that every aesthetic preference comes straight from biology rather than from culture, or that every claimed link between an objective feature and an underlying biological explanation is true.

Some of the standard findings described above, such as the preference for symmetry, appear to be solidly grounded in biology rather than culture. Symmetry preference is ubiquitous in the animal kingdom, and even insects exhibit a preference for symmetry (Rodriguez et al. 2004). The underlying neurophysiology is fairly simple, and there are plausible reasons for this preference arising, in terms of evolutionary biology, since symmetry is usually an honest indicator of genetic fitness. Given its ubiquity throughout the animal kingdom, we would therefore expect there to be a preference for symmetry among hominins as a biological "default setting." One archaeological implication is that innovative products which are symmetrical are more likely to be accepted than ones which are not; by the same reasoning, we would expect to find this preference from the earliest hominins onward.

There are, however, other issues which make the full story more complex. If we take handaxes as an example, it is tempting to argue that the symmetry in most handaxe designs could be caused by their makers having an inbuilt aesthetic preference for symmetrical shapes. However, it is also possible that the symmetry is present for other reasons, such as reduction of cognitive load—the process for making a symmetrical handaxe can be handled with fewer mental steps than a deliberately asymmetrical one. Other plausible explanations for handaxe symmetry can easily be generated, such as practical factors relating to how the handaxe was used. However, being plausible, and having some supporting illustrative examples, is not the same as being the best explanation, and researchers in this area are well aware of the risk of producing something which is no more than a "Just So" story—an entertaining and superficially plausible story which in reality is oversimplified, overgeneralized, or downright wrong.

A classic example is provided by the bowerbird, where the males of many species produce courtship areas decorated with colored objects. This behavior has been the topic of a considerable body of research, since it lends itself readily to a wide variety of experimental approaches, as described below, and since there are apparent similarities between the birds' elaborate visual displays and human art. We have used this as an example of how a set of findings can be taken out of context and be misinterpreted as demonstrating a simple set of hard-wired aesthetic principles common between humans and non-human animals.

The most widely known example is the satin bowerbird, whose males prefer to add smooth, shiny blue objects to the courtship area, and whose males have blue plumage. It is easy to speculate from this that the objects in the courship area function as superstimuli for the male's coloration—if female satin bowerbirds have a sensory bias towards blue coloration, as an honest indicator of the quality of the male's health, then the blue objects in the courtship area might act as superstimuli for this sensory bias. An analogous supposition for human behavior would be that smooth, shiny objects such as female jewellery are superstimuli for smooth, shiny eyes and/or smooth, shiny skin, both of which are honest indicators of physical health.

The full story, however, is more complex. Within a single bowerbird species, there are individual differences between birds in preferences for colors of objects; across species, there are differences in relation to whether or not smooth, round shiny objects are preferred over dull ones. Even within the classic example of the satin bowerbird, it appears that the decorative objects fulfil a different role in courtship displays from the plumage of the displaying bird (Borgia 2008) and are not a simple superstimulus.

Similar issues arise in relation to set of apparently robust findings relating to perceived attractiveness in humans. Tall men, for instance, are usually perceived as being of higher status than shorter men; women with facially neotenous features are typically perceived as more attractive than women with more mature facial features. At first sight, this suggests that innovations which enhance the appearance of the valued attributes will be more readily accepted than innovations which do not. Obvious examples include headwear, high heels, and raised seating, as ways of giving an individual the appearance of more height and therefore more status; similarly, cosmetics make the skin appear smoother and therefore youthful.

Findings such as those described in the preceding paragraphs give some interesting insights. There is, however, risk of overgeneralizing those findings, and of producing "Just So" stories which gloss over underlying complexities.

An example of more complex realities involves the difference between within-gender issues and between-gender issues, which may pull in conflicting directions. A classic early recognition of this occurs in Darwin's *Origin of Species* (Darwin 1859); this is still very much a live area of research (e.g., Buss 2006). An attribute which gives an individual an advantage in within-gender competition for status, etc, such as height, may also put them at a disadvantage in between-gender competition for mates—tall women are at an advantage over short women in terms of relative status within the community of women, but at a disadvantage in terms of stereotypical attractiveness to heterosexual males.

A related issue is highlighted by the phrase "stereotypical attractiveness"—it is easy to lump data into a single mean preferred value, rather than checking whether there are, e.g., multimodal preferences showing clusterings for several different preferred values. Ethological studies of mate preferences strongly suggest that there is a widespread tendency in the animal kingdom for there to be several different robust types of "attractiveness" within a species, rather than "one size fits all." An example of the temptation towards "Just So" stories involves the preference for smooth, round shiny things. Without the data from other species, there would be an easy Freudian just-so story relating this preference to the shape and visual texture of human breasts and buttocks; however, the results from species such as bowerbirds suggest that this preference was widespread in the animal kingdom (though not universal) long before humans evolved. Similarly, although it is tempting to assume that rare, costly displays will invariably be attractive items of sexual display, there is evidence that this is not always the case across species; Madden and Balmford (2004) found that spotted bowerbirds do not prefer rare or costly bower decoration.

It is also important to bear in mind that evolutionary explanations are not the only ones which can make sense of systematic biases in subjective aesthetic preferences. Information processing models offer a very different way of approaching the same issue, with a solidly empirical grounding, but could lead to some very different conclusions.

Despite these grounds for caution, we conclude that there is a strong case for at least some observable features of physical artifacts being reliable predictors of human subjective aesthetic responses to them in a way which is not culture-bound. This implies that these attributes can be factored in to innovation research, where they can be used to generate a new set of research questions.

CONCLUSION AND FURTHER WORK

For researchers wishing to use empirical measures of novelty, there are various available methods which are well established in other disciplines. Inverse frequency weighting and measurement of minimal edit distances are two approaches which can be readily integrated with other approaches, such as the use of graph theory to show levels of complexity. These methods are simple, and do not require complex equipment, but are powerful and rigorous; they can also identify fruitful new research questions.

The role of subjective factors in adoption of innovations is an area where approaches from other disciplines, in particular the literature on empirical aesthetics, also offer the prospect of interesting research questions which can be empirically investigated and which can offer fresh insights. For instance, is there a correlation between the costs involved in a given set of costly signalling and the degree of social stratification in a society? Is it the case that the degree of novelty in innovations is relatively constant across time, or does it behave in some other way? Are aesthetically attractive innovations more likely to succeed than non-attractive ones? Are innovations in the early archaeological record more consistent with ethological models of displays and are innovations later in the record more consistent with cognitive load models of aesthetics? The approaches we have described in this paper should make it possible to tackle these, and other, questions; we hope they will be of use to readers.

REFERENCES

Anonymous (editorial). 2001. Another front in the science wars? *Journal of Consciousness Studies* 8(1), 3–8.

Avital, T. and Cupchik, G.C. 1998. Perceiving hierarchical structures in nonrepresentational paintings. *Empirical Studies of the Arts* 16 (1), 59–70.

Borgia, G. 2008. Experimental blocking of UV reflectance does not influence use of off-body display elements by satin bowerbirds. *Behaviour Ecology*, Advance Access doi: 10.1093/beheco/arn010. Accessed 17.1.2011

Buss, D. 2006. Strategies of human mating. *Psychological Topics* 15(2), 239–260.

Cardenas, R.A. and Harris, L.J. 2006. Symmetrical decorations enhance the attractiveness of faces and abstract designs. *Evolution and Human Behavior* 27(6), 1–18.

Chamberlain, A.T. 2000. On the Evolution of Human Aesthetic Preferences. http://www.assemblage.group.shef.ac.uk/chamberl.html. Accessed 17.1.2011

Darwin, C. 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (1st ed.). London: John Murray.

Enquist, M., Ghirlanda, S., Lundqvist, D., and Wachtmeister, C.-A. 2002. In G. Rhodes and L. Zebrowitz (eds.), *Advances in Visual Cognition*, Volume 1: *Facial Attractiveness*. Ablex: Westport CT.

Euler, L. 1741 (facsimile of original paper). Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae* 8, 128–140. http://math.dartmouth.edu/~euler/docs/originals/E053.pdf. Accessed 28.3.2008.

Fink, B. and Neave, N. 2005. The biology of facial beauty. *International Journal of Cosmetic Science* 27, 317–325.

Forsythe, A., Mulhern, G., and Sawey, M. 2008. Confounds in pictorial sets: the role of complexity and familiarity in basic-level picture processing. *Behavior Research Methods* 40(1), 116–129.

Haidle, M.N. and Bräuer, J. 2011. From brainwave to tradition—how to detect innovations in tool behaviour. *PaleoAnthropology* 2011, 145–154 (this volume).

Hodgson, D. 2009a. Symmetry and humans: reply to Mithen's 'Sexy Handaxe Theory.' *Antiquity* 83, 195–198.

Hodgson, D. 2009b. Evolution of the visual cortex and the emergence of symmetry in the Acheulean techno-complex. *Comptes Rendus Palevol* 8, 93–97.

Jacobsen, T., Schubotz, R.I., Höfel, L. and v. Cramon, D.Y. 2006. Brain correlates of aesthetic judgment of beauty. *NeuroImage* 29, 276–285.

Kalick, S. M., Zebrowitz, L.A., Langlois, J.H., and Johnson, R.M. 1998. Does human facial attractiveness honestly advertise health? *Psychological Science* 9(1), 8–13.

Kohn, M. and Mithen, S.J. 1999. Handaxes: products of sexual selection? *Antiquity* 73, 518–526.

Langlois, J.H., Kalakanis, L., Rubenstein, A.J., Larson, A., Hallam, M., and Smoot, M. 2000. Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin* 126(3), 390–423.

Lavie, T. and Tractinsky, N. 2004. Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies* 60, 269–298.

Machin, A.J. 2008. Why handaxes just aren't that sexy: a response to Kohn and Mithen. *Antiquity* 82, 761–769.

Madden, J.R. and Balmford, A. 2004. Spotted bowerbirds *Chlamydera maculata* do not prefer rare or costly bower decorations. *Behavioral Ecology and Sociobiology* 55, 589–595.

Matas, M., Mancas-Thillou, C., Gosselin, B., and Macq, B.M. 2006. A rarity-based visual attention map - application to texture description. IEEE International Conference on Image Processing, Atlanta, Georgia, 2006, pp. 445–448.

Mellars, P. 2010. Neanderthal symbolism and ornament manufacture: The bursting of a bubble? *Proceedings of the National Academy of Sciences USA* 107, 20147–20148.

Mithen, S. 2003. Handaxes: the first aesthetic artefacts . In E. Voland and K. Grammer (eds.), *Evolutionary Aesthetics*, pp. 261–275. Berlin: Springer.

Mosshagen, M. and Thielsch, M.T. 2010. Facets of visual aesthetics. *International Journal of Human-Computer Studies*, 68(10), 689–709.

Nowell, A. and Chang. M.L. 2009. The case against sexual selection as an explanation of handaxe morphology. *PaleoAnthropology* 2009, 77–88.

Ramachandran, V.S. and Hirstein, W. 1999. The science of art: A neurological theory of aesthetic experience. *Journal of Consciousness Studies* 6, 15–51.

Reber, R., Schwarz, N. & Winkielman, P. 2004. Processing fluency and aesthetic pleasure: is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, 8(4), 364–382.

Renfrew, C. 1978. The anatomy of innovation. In D. Green, C. Haselgrove, and S.M. Spriggs, (eds.), *Social Organisation and Settlement*, pp. 89–117. Oxford: British Archaeological Reports Series 47(i).

Rhodes, G., Zebrowitz, L.A., Clark, A., Kalick, M., Hightower, A., and McKay, R. 2001.

Do facial averageness and symmetry signal health? *Evolution and Human Behavior* 22, 31–46.

Rodriguez, I., Gumbert, A., de Ibarra, N.H., Kunze, J., and Giurfa, M. 2004. Symmetry is in the eye of the 'beeholder': innate preference for bilateral symmetry in flower-naïve bumblebees. *Naturwissenschaften* 91, 374–377.

Rogers, E.M. 2003. *Diffusion of Innovations* (5th edition). New York: Free Press.

Rugg, G. 2011. Quantifying technological complexity. *PaleoAnthropology* 2011, 155–166 (this volume).

Ryan, M.J. 1990. Sexual selection, sensory systems and sensory exploitation. *Oxford Surveys in Evolutionary Biology* 7, 157–195.

Saunders R. and Gero, J.S. 2001. The digital clockwork muse: A computational model of aesthetic evaluation. AISB '01, University of York, York, UK, pp. 12–21.

Shannon, C.E. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal* 27, 379–423 and 623–656.

Spehar, B., Clifford, C.W.G., Newell, B.R., and Taylor, R.P. 2003. Chaos and graphics: universal aesthetic of fractals. *Computers and Graphics* 27, 813–820.

Vartanian, O. and Goel, V. 2004. Neuroanatomical correlates of aesthetic preference for paintings. *Cognitive Neuroscience and Neuropsychology* 15(5), 893–897.

Winkielman, P., Halberstadt, J., Fazendeiro, T., and Catty, S. 2006. Prototypes are attractive because they are easy on the mind. *Psychological Science* 17(9), 799–806.