Alexis D.J. Makin

The Gap Between Aesthetic Science and Aesthetic Experience

Abstract: *For over a century we have attempted to understand human aesthetic experience using scientific methods. A typical experiment could be described as reductive and quasi-psychophysical. We vary some aspect of the stimulus and systematically measure some aspect of the aesthetic response. The limitations of this approach can be categorized as problems on the Y axis (what we measure) and the X axis (what we manipulate). The most enigmatic components of aesthetic experience include inclination to cry, aesthetic rapture, a sense of the sublime, and intense fascination. However, we cannot evoke these 'hot' aesthetic emotions in the lab, at least not with well controlled stimuli on multiple trials. We thus resort to measuring cold, cognitive preference ratings. There are also problems on the X axis. The reductive psychophysical approach explicitly assumes that there are lawful relations between different stimulus dimensions and preferences. It also tacitly assumes that these dimensions are independent and orthogonal. The second assumption is implausible. Whatever stimulus-preference laws we discover are likely to be twisted and modulated when another dimension is added to the stimuli. This 'gestalt nightmare' has long been recognized, but never resolved. This matters, because human aesthetic faculties are probably tuned to the balance and relationship of parts which make up a whole and are indifferent to the parts presented in isolation. I conclude that the future of scientific aesthetics depends on how successfully we can transcend reductive, quasi-psychophysical approaches.*

Correspondence:
Alexis D.J. Makin, University of Liverpool, UK.
Email: Alexis.Makin@liverpool.ac.uk

Journal of Consciousness Studies, **24**, No. 1–2, 2017, pp. 184–213

1. Introduction

Empirical aesthetics seeks to explain human preferences and aesthetic experience by using scientific methods (Palmer, Schloss and Sammartino, 2013). Nobody doubts that near-universal human preferences exist. For example, every sane person likes eating cakes more than gravel or poison. Everyone prefers clean drinking water and comfortable room temperatures, and we all want peers to admire us. These mundane preferences can easily be explained by natural selection: brains are partly organized by genes, and the genes that code for brains with these preferences were replicated. Other genes may have coded for brains inclined towards self-destruction, but these genes were never replicated. Certain adaptive preferences are bound to become universal over evolutionary time.

However, aesthetic preferences seem to be a different category because they do not have an *immediately obvious* evolutionary explanation (although they may have a less obvious one upon careful consideration). For instance, most people have a strong positive reaction to their favourite music. But why? How is listening to music relevant for survival and reproduction? From our genes' point of view, indulging our aesthetic faculties seems like a waste of time. Nevertheless, people do spend a lot of time listening to music, reading literature and watching films, looking at artworks and admiring nature, and numerous other practices which engage our conscious aesthetic faculties. Preference without obvious evolutionary explanation is an intriguing mystery for psychologists. This is just one of the reasons that empirical aesthetics has expanded dramatically in recent years, with new journals (e.g. *Art and Perception*), books (e.g. Shimamura and Palmer, 2012), international conferences (e.g. Visual Science of Art), and even institutions (e.g. Max Plank Institute for Empirical Aesthetics in Frankfurt: http://www.aesthetics.mpg.de/).

Here I argue that there are deep limitations with contemporary aesthetic science. More specifically, there are problems with what we might call the 'psychophysical approach' or the 'reductive approach'. Psychophysicists vary some objective property of the stimuli, and measure subjective experience of that property. For example, we can vary the brightness of a light (plotted on the X axis), and measure subjective brightness evaluations (plotted on the Y axis). Then we can discover the function that relates the objective stimulus properties to subjective experience (Figure 1A). Perhaps it is possible to discover truths about human aesthetic preferences in a similar way? Can we vary some objective property of the stimuli (X axis) and measure subjective preferences (Y axis)? Early pioneers of scientific aesthetics, such as Fechner (1876), Birkhoff (1932), Eysenck (1941), and Berlyne (1971), all proposed lawful functions relating stimulus properties to subjective preferences. This is still a very common approach (reviewed in Palmer, Schloss and Sammartino, 2013).

Criticism of empirical aesthetics is also old. Dickie (1962) claimed that psychology experiments that measure average preferences in lay people are 'irrelevant' to understanding aesthetics. For one thing, no serious critic or artist would ever change their practice based on such results. Instead, expert aestheticians are intuitively tuned to subtle factors like intensity, completeness, and coherence, which are difficult to study experimentally. But this was 54 years ago — can we now say Dickie was too pessimistic about the future of the field? Following other contemporary authors (e.g. Kubovy, 2000; Holmes and Zanker, 2012), I argue that many problems remain. My aim is to organize some of these valid and interesting criticisms in a novel way. As we will see, *there are problems with both axes*.

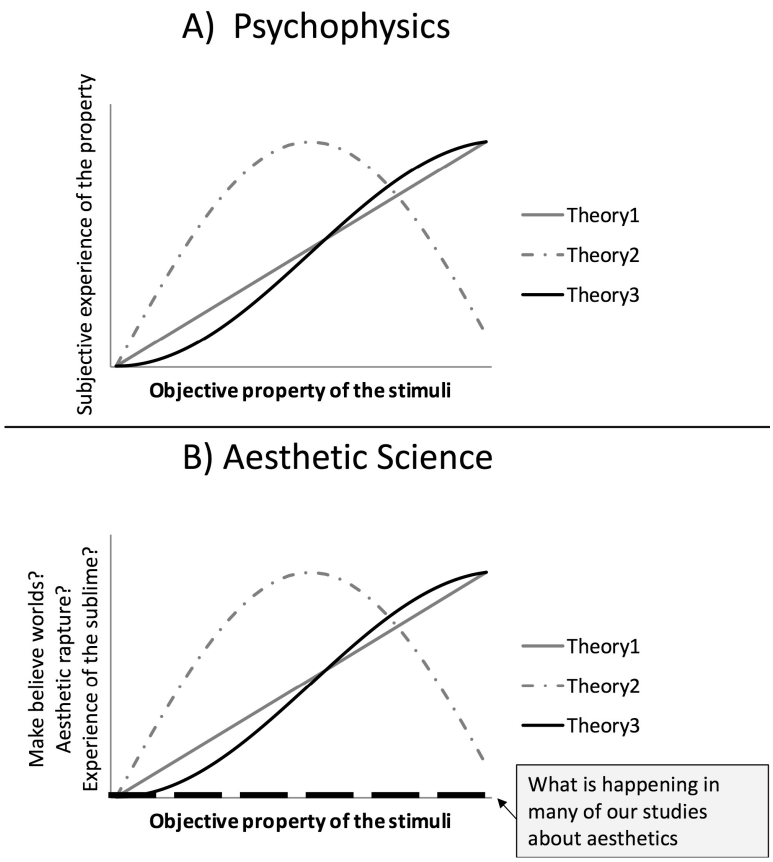

Figure 1. Psychophysics (A) and the quasi-psychophysical approach to aesthetic science (B).

2. What Goes on the Y Axis?

Words like 'aesthetics' and 'aesthetic experience' could be described as 'grandiose'. Aesthetic experience is supposedly something rare and wonderful. Art, beauty, and truth are esteemed as the 'higher pinnacles' of the human mind. Aesthetic experience is one of those mental jewels that supposedly separates us from the beasts (see Jacobsen *et al.*, 2006, for evidence that this is not just anthropocentric bias). Religions allude to deep connections between beauty, perfection, morality, and the divine. Philosophers have also considered the nature of beauty for millennia. Emanuel Kant (1724–1804) identified aesthetic experience as involving a 'sense of the sublime' and 'disinterested fascination' (as described in Shimamura, 2012). According to Kant, aesthetic emotions are rare and qualitatively different to ordinary ones. While most emotions are about satisfying homeostatic drives, aesthetic emotions do not compel action so urgently (although they certainly inspire us to create and understand art).

Modern accounts in psychological science do not always shy away from the richness of aesthetic experience. For example, Markovic (2012) describes the state of *aesthetic rapture* as an intense fascination with the object of appreciation, in a rare and special transcendent moment. He links aesthetic rapture to other hard-to-describe mental states like *flow experience* (being so engrossed we blissfully forget about time and the narrative self — Csikszentmihalyi, 1991), *mindfulness* (a simultaneously wakeful and peaceful state sometimes attained though mediation — Langer and Moldoveanu, 2000), or *peak experienc*e (a vivid, enthusiastic participation in life associated with personal fulfilment — Maslow, 1972).

Let's look at another attempt to put aesthetic experience into words. Kubovy (2000) lists the following six components of a prototypical aesthetic response to art:

1) Attention is firmly fixed upon heterogeneous but interrelated components of a pattern.
2) Where there is little interference from potentially distracting stimuli like environmental noise or bodily sensations.
3) Where the viewer is conscious of the relationship between the object and artist or object and culture which created it.
4) Is dominated by intense feelings or emotions.
5) All aspects of the experience are transparently related to each other.

6) May have a sense of make believe, in which the viewer is immersed to the extent that they do not think of the artwork as a material object.

These valuable accounts describe something about the human mind and experience, and they are of interest to psychologists, even though they are not stereotypically scientific. I think it is uncontroversial to say aesthetic rapture cannot be studied with the reductive quasi-psychophysical approach. For one thing, it is virtually impossible to evoke these kinds of intense emotions in the lab on repeated trials with well controlled stimuli. In all likelihood, the participants experience *absolutely zero* aesthetic rapture at any time during such experiments (Figure 1B).

Superficially, this conclusion seems to be in stark contrast to a claim in Palmer, Schloss and Sammartino (2013), who proposed that we have 'some kind of aesthetic experience to nearly everything we see'. However, Palmer makes the vital distinction between 'aesthetics with a capital A' and 'aesthetics with a lower case a'. The latter refers to more ordinary preferences and reactions, which need not be intensely emotional. If we can study anything about aesthetic experience using the quasi-psychophysical approach, it is only aesthetics with a lower case 'a'.

Let's consider a final example. Pelowski and Akiba (2011) propose that aesthetic experience is NOT always about blissfully losing ourselves in moments of sensory perfection or cognitive mastery, but about challenging the ego. They propose that art can disrupt and attack our self-schema. This is followed by denial and resistance (experienced as negative tensions and anxiety) and then surrender to a new worldview (sometimes experienced as a tearful epiphany). Pelowski (2015) explored this model empirically, and found that 36–43% of visitors to Rothko exhibitions reported some inclination to cry. However, a Rothko painting is the opposite of a controlled stimulus — it is a labyrinth of interacting visual and semantic dimensions which cannot easily be listed, isolated, and quantified. Furthermore, the participants probably entered the 'experiment' with prior expectations and opinions about Rothko. It is instructive that Pelowski (2015) had to move a long way from the reductive psychophysical approach in order to record aesthetic emotions of any intensity.

3. Difficulty with Studying Ordinary Emotions

As well as rare and special aesthetic emotions, the full aesthetic response also involves perceptual processing, cognitive appraisal, cultural knowledge, and also more ordinary emotions, like happiness or surprise (Leder *et al.*, 2004; Markovic, 2012). So, even if aesthetic rapture cannot be reliably evoked in the lab, perhaps we can study the other emotional responses instead? For example, can we vary some aspect of the stimuli in a controlled way (X axis) and measure happiness (Y axis)? Next I will argue that studying ordinary emotions is also very difficult with the psychophysical approach.

Mauss and Robinson (2009) focus on theoretical and methodological issues in emotion research. One idea is that there are *core emotions* that are culturally universal in humans and evident in other species. Core emotions are supposedly triggered by specific circumstances which were common to our ancestors (such as loss, achievement, frustration, or assault). Furthermore, core emotions activate a unique set of neural, hormonal, and cardiovascular responses (e.g. amygdala activation, adrenaline release, heart rate increase), they produce distinct facial expressions (e.g. smiling, frowning, snarling), and finally, they have stereotypical behavioural consequences (e.g. running, shouting, hiding). There may be 15 core human emotions: *Amusement, Anger, Contempt, Contentment, Disgust, Embarrassment, Excitement, Fear, Guilt, Pride, Relief, Sadness, Satisfaction, Sensory Pleasure, and Shame* (Ekman, 1999). These are supposedly distinct, brief, and automatic reactions to certain antecedent events. Core emotions are thus different from moods, which are emotion-laden but more persistent (see Saarimäki *et al.*, 2015, and Sauter *et al.*, 2010, for recent work on basic emotions).

An alternative to the core emotion account is the *dimensional account*. We can imagine a 3D graph: on the respective axes we have *arousal* (low to high), *valence* (negative to positive), and *approach-avoid inclination* (tendency to move towards or away). Each emotional state can be described as a point in that 3D space. For example, fear is characterized by high arousal, negative valence, and avoidance (i.e. running from an enemy). Anger is also high arousal and negative valence, but with high approach inclination (i.e. attacking an enemy). Contentment would involve low arousal, positive valence, and no strong tendency to approach or avoid anything. Mauss and Robinson (2009) claim that there is more evidence for the dimensional account than for the core emotion account. Perhaps the two can be reconciled if emotion space is uneven and heterogeneous: there could be certain regions which are frequently revisited, and these regions are signposted by the core emotion words.

We need not have a final classification scheme in place before we can study emotions scientifically. There are various methodologies for diagnosing or 'reading out' a person's emotional state. These include questionnaires, self-report, heart rate measurement, galvanic skin responses, pupil dilation, and facial expression analysis. These measures all have strengths and weaknesses. However, these different measures of emotion are poorly correlated: for example, pupil dilation is not strongly correlated with self-reported arousal. Even worse, they are not diagnostic: I cannot conclusively claim that you are not happy because you are *not* smiling, or that you must be happy because you *are* smiling.

Researchers studying emotion often use very potent stimuli, such as the International Affective Pictures (IAPs) or fear conditioning where sounds are paired with very aversive electric shocks (LeDoux, 1998). Emotion research which uses these stimuli is difficult enough. However, in aesthetic science we have an additional problem: we want to use precisely controlled visual stimuli, which differ from each other in specific and quantifiable ways. This means the emotional difference between our conditions is miniscule. The kinds of images used in empirical aesthetics probably produce only sub-threshold and sporadic emotional responses at most. For example, it is not likely that they would elicit strong states of arousal, with increase in heart rate, adrenaline release, and sweating, etc. We are still very much in the realms of Figure 1B, where the Y value is zero on virtually every trial. In other words, we cannot evoke a 'hot' emotional reaction with carefully controlled visual stimuli, on multiple trials, across many participants.

Before moving on, it is worth briefly mentioning another account, which is highly relevant to scientific aesthetics. Barrett (2012) asks whether 'Emotions are real'. The reflexive answer is 'yes', but this seems less obvious after deeper inspection. She begins with the ontological analysis of the categories 'flower' and 'weed': there is no objective, physical difference between plants that distinguishes a flower from a weed. Should we say that flowers and weeds are illusory? No: surely our culturally shared knowledge about gardens makes sense of the words 'flower' and 'weed'. For one thing, these words suggest different appropriate actions, weeds are to be discarded, flowers admired and cherished. Changes in heart rate or adrenaline release can also be objectively measured (like plants) and, in this case, internally discriminated and perceived. However, Barrett proposes that we understand and interpret these events using our learned categories and our full semantic knowledge (*ibid.*). Like 'flower' and 'weed', conceptual categories such as 'fear' and 'anger' are complex, social constructions, learned and elaborated through countless conversations and cultural interchanges. An emotion is properly understood as more than a physiological response or fixed action sequence, it is also the full range of semantic and conceptual knowledge which we use to *make sense* of these bodily events. Barrett uses the term 'emotional gestalt' to refer to this cognitive and affective whole (*ibid.*). This theoretical position allows for near infinite flexibility in emotional experience. People in different cultures could have quite different emotional gestalts, because they have different emotional words and categories, a different history of ideas about the kind of actions appropriate in certain situations, and so on. It would be interesting to analyse aesthetic experience using the concept of emotional gestalts, although that is beyond the scope of this paper.

4. Cold Evaluations and Preferences

We have seen that it is very difficult to evoke either special aesthetic emotions or *hot* emotions using well controlled stimuli on repeated trials. This leaves us with studying *cold* evaluations or preferences. Fortunately, people can easily evaluate things without feeling a hot emotional reaction. I can judge that winning a prize is positive, and losing my wallet is negative, without actually feeling triumph and frustration. This is presumably what happens in the vast majority of experiments in empirical aesthetics that use the psychophysical approach, even if we sometimes interpret results in terms of 'affective processing'.

It is easy to find examples in modern scientific aesthetics. We know that there are reliable preferences for particular colours. People typically like blue and greenish-blue and dislike brownish-yellows (Palmer and Schloss, 2010). Certain colour combinations are harmonious, and some people have a preference for harmony (Schloss and Palmer, 2011). People rate curved edges as more attractive than angular, spiky edges (Bar and Neta, 2007; Bertamini *et al.*, 2016). People prefer symmetrical arrangements to random arrangements (Eisenman, 1967; Jacobsen and Höfel, 2002; Makin, Pecchinenda and Bertamini, 2012), and have a preference for mid-range fractals

(Spehar *et al.*, 2003). In all these experiments, the authors varied a particular aspect of the simple stimuli, and measured preference in one way or another. Often, participants simply rated preference for each presentation on a scale (e.g. unattractive to attractive, or pleasant to unpleasant). I presume participants were making cold evaluations in all these experiments. When Palmer talked of 'aesthetics with a lower case a', he was partly referring to the fact that we are studying cold, cognitive evaluations, not hot emotional reactions.

So, we have found something *workable* for the Y axis: namely cold evaluations. This may be an underwhelming and mundane part of the aesthetic experience when contrasted with glamourous things like 'aesthetic rapture', 'immersion in make believe worlds', or 'sense of the sublime', but at least it is something we can precisely measure in the lab with controlled stimuli![1]

5. What Goes on the X Axis? Facing Up to the 'Gestalt Nightmare'

So far we have concentrated on the Y axis, and asked which aspects of the aesthetic experience can be evoked in the lab, quantified, and measured. We now move onto the X axis, and consider issues with stimuli which we sometimes use in empirical aesthetics.

Can we learn anything about real-world preferences by measuring preference for one visual dimension at a time? What if features fail to summate in a predictable way? For example, we know that people like blue more than brown (Palmer and Schloss, 2010). But surely findings like this can always be turned upside down when blue and brown are properties of other objects, not just patches of colour? After all, nobody wants a blue banana, and everybody likes brown chocolate. Holmes and Zanker (2012) state the problem like this:

> [The reductionist approach] allows more precise questions about the relationship between some feature of the stimulus and the consciously articulated preferences to be explored in a highly controlled manner. However, as observed by Arnheim (1974), the aesthetic experience is rather fragile and typically relates to the gestalt, or whole, rather than the sum of any isolated parts meaning that a reductionist approach will

---

1 Of course, some researchers would object here that aesthetics-with-a-lower-case-a is not *their* plan b — it is interesting in its own right to understand the ubiquitous preferences that enhance everyday life.

> never be able to explain the effects of the complex interaction of many individual features on the aesthetic evaluation of the stimulus. (p. 426)

This problem could be called the '*gestalt nightmare*'.[2] The reductive psychophysical approach assumes that there are many stimulus dimensions which alter preference in a lawful way. We can do experiments to discover these laws. However, the approach tacitly assumes *that all these dimensions are orthogonal, and all effects on preference are independent*. This second assumption is totally unrealistic. It is more likely that stimulus-preference laws are modulated, turned, and twisted whenever another dimension is added to the stimuli.

This is particularly important, because the human aesthetic faculty *is tuned to gestalts*: we are impressed by the precise balance and interplay of elements, and comparatively uninterested in the elements in isolation. For example, the aesthetic appeal of music comes from the relationship between the individual notes, melodies, and lyrics, etc. It would be totally meaningless to play each component sound separately, obtain a preference rating, then sum all the preferences. However, the reductive approach sometimes implies that studying the isolated parts is somehow step one on an ambitious research project. Dickie (1962), Kubovy (2000), and many others have talked about the gestalt nightmare in one way or another. However, as a community, we have still not found an unproblematic way to transcend it.

I have conducted a short experiment which illustrates the gestalt nightmare. Participants first evaluated patterns which varied only in one dimension, 1) symmetry or random, 2) blue or brown, 3) curved or angular (Figure 2A). They were then shown patterns which included all three dimensions (Figure 2B). Do findings from the one-dimensional experiments generalize to the three-dimensional experiment? For instance, if people like blue more than brown coloured patches, will this translate into liking blue in complex patterns as well? In the same spirit, we can test whether individual differences transfer from 1D to 3D experiments. Does a participant who really liked blue when colour was the only dimension remain a blue-lover when there are other features available as well? Although this experiment alone does not conclusively tell us how much to worry about the gestalt nightmare, it does help illustrate the nature of the problem, and suggests some recurring variants.

2 Professor Stephen Palmer (personal communication) suggested the phrase 'gestalt nightmare'.

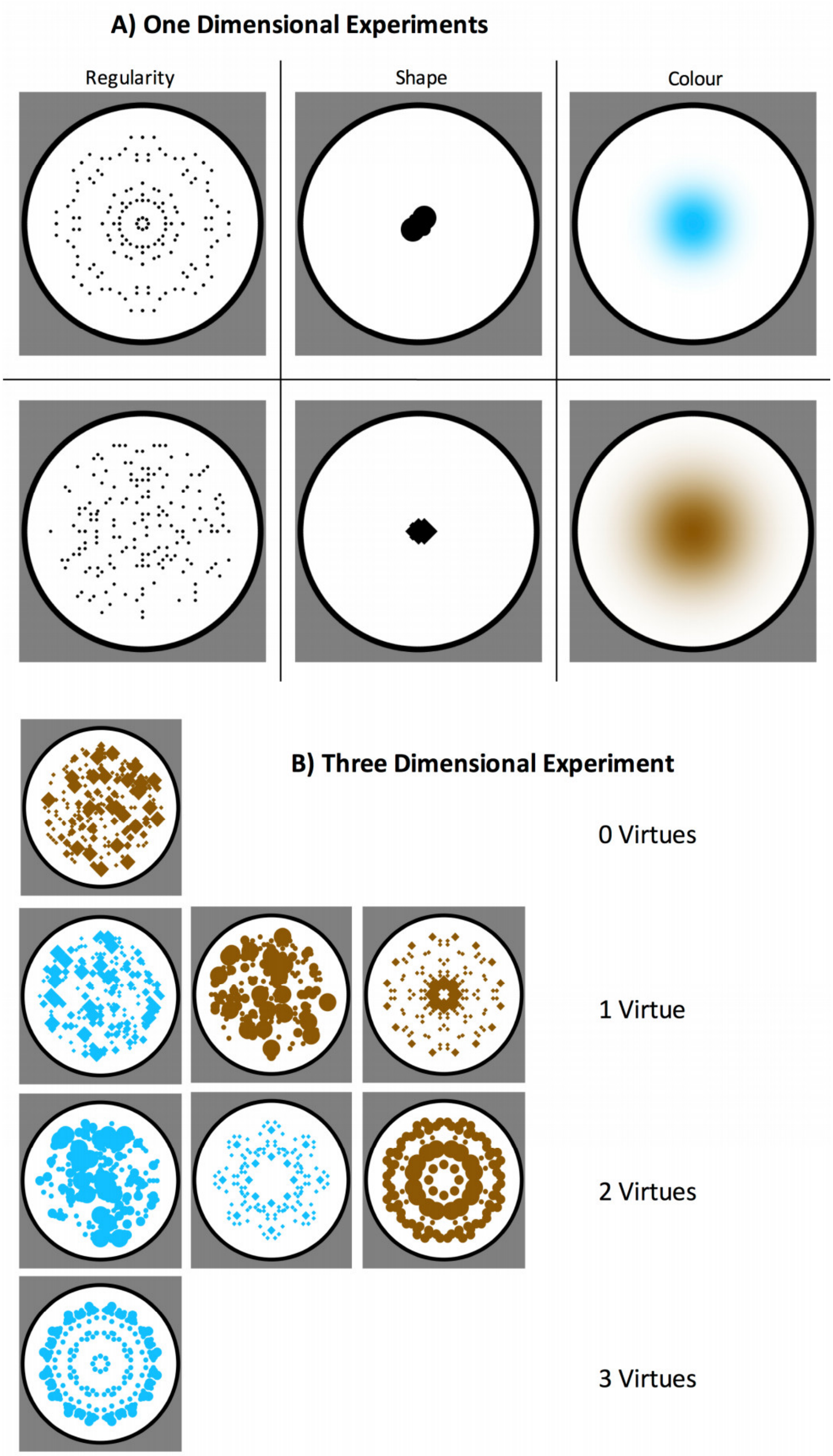

Figure 2. A) There were three separate 1D experiments, where stimuli varied on a single dimension only: Regularity, Shape or Colour. B) Stimuli from the subsequent 3D experiment. Here all three dimensions were available, and all could contribute to preference.

6. Method

6.1. Participants and Apparatus

42 participants from the University of Liverpool (age 18–25, 4 males, 7 right-handed) were involved in the study. All had normal or corrected to normal vision. The study had local ethics committee approval and was conducted in accordance with the declaration of Helsinki (revised 2008). The session lasted about 25 minutes. The experiment was programmed in Python using open source PsychoPy libraries (Peirce, 2007).

6.2. Procedure and Stimuli

The experiment had four parts. First there were three separate *one-dimensional experiments* (1D). All participants then completed the *three-dimensional experiment* (3D). Trial structure is shown in Figure 3. The order of the 1D experiments was counterbalanced across participants, then all participants completed the 3D experiment last.

6.3. 1D Experiments

The **regularity experiment** varied the arrangement of the dots (symmetrical or random). Following numerous previous studies, we predicted that symmetry would be preferred (Makin, Pecchinenda and Bertamini, 2012). The symmetrical patterns had four folds. The random patterns had the same average number of dots, but with no systematic spatial relationship between elements. The program generated the patterns afresh on every trial using the same algorithm. No pattern was ever presented twice and no two participants saw the same set of patterns.

The **shape experiment** varied the angularity or roundness of a small black solid shape. This was composed of several overlapping squares or circles, which varied in size. Previous work suggests that people prefer rounded edges to squared, angular edges (e.g. Bar and Neta, 2007), either because the angles signify threat and risk, or because the curved edges are positively evaluated (Bertamini *et al.*, 2016). Again, the program generated these shapes afresh on every trial.

Finally, the **colour experiment** varied the colour of a central patch, either light blue or light brown. These two colours were chosen to be approximately similar to the most and least preferred colours found by

Palmer and Schloss (2010). The size of the patch was varied randomly from trial to trial to provide some extra variety.

On each of the 1D experiments, 20 trials were shown, with 10 examples from the hypothetically positive category, and 10 from the hypothetically negative category. Patterns were presented for 3 secs. All patterns were framed by an identical white circle with a black border, 11.25 cm in diameter. This was followed by a rating scale, which varied from 0 to 100. The screen read: 'How much did you like the pattern?' The low end was marked 'not at all', the high end was marked 'extremely' (Figure 3). For each participant, the mean rating scores of the 10 positive patterns and the 10 negative patterns was obtained.

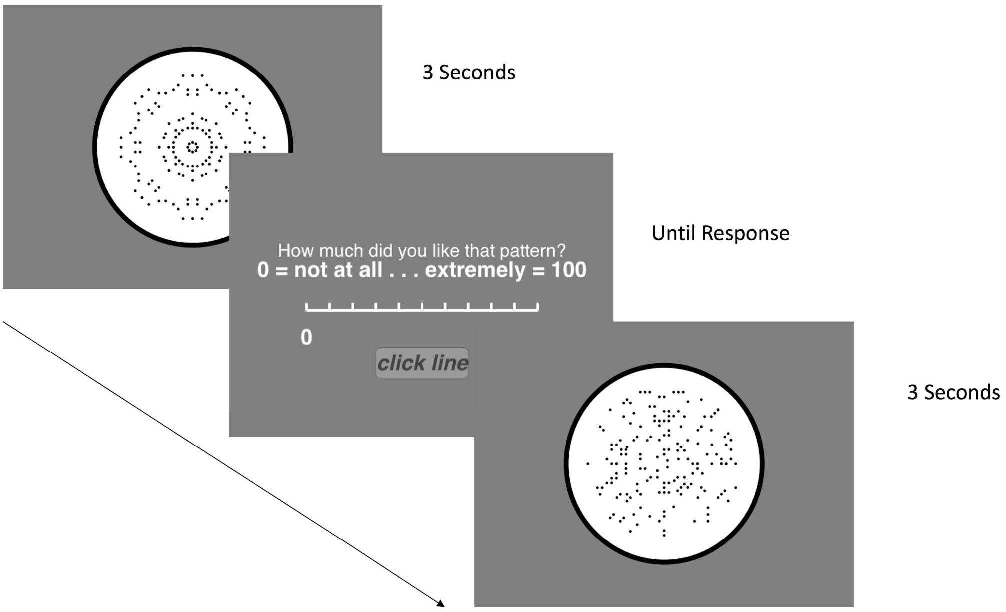

Figure 3. Trial structure and design.

6.4. 3D Experiment

The 3D experiment was always completed after the 1D experiments. Stimuli were novel, but represented a combination of the features used in the 1D experiments: patterns were symmetrical or random, made of rounded or angular sub-units, and were either blue or brown. The same basic algorithms used in the 1D experiments were used again to arrange symmetry and generate element shape. The same blue and brown colours were used. There were 80 trials in this experiment, with 10 repeats of each condition (Figure 2B). The mean preference ratings from each condition were obtained.

We can think of the stimuli in the 3D experiment as including a 0–3 of aesthetic virtues (Figure 2B). Even ignoring the results of the 1D experiments, it is interesting to consider how these virtues might interact with each other. Will this always be some kind of summation, or will there be complicated interactions?

7. Results

7.1. 3D Experiment

It is instructive to begin with the results of the 3D experiment. Preferences in all 8 conditions of the 2 (Regularity) x 2 (Element shape) x 2 (Colour) design are shown in Figure 4A. The other panels in Figure 4 show the same data, but collapsed over one or more factors because this makes some results easier to visualize. Participants preferred the symmetrical to the random patterns ($F(1,41) = 97.091$, $p < 0.001$, partial $\eta^2 = 0.703$, Figure 4D), and the blue patterns to the brown patterns ($F(1,41) = 25.531$, $p < 0.001$, partial $\eta^2 = 0.384$, Figure 4E). There was no main effect of element shape ($F(1,41) < 1$, Figure 4F).

There was no three-way interaction between all factors ($F(1,41) < 1$). The Regularity x Colour interaction approached significance ($F(1,41) = 3.618$, $p = 0.064$, partial $\eta^2 = 0.081$, Figure 4B), while the Regularity x Element shape interaction was clear ($F(1,41) = 17.024$, $p < 0.001$, partial $\eta^2 = 0.293$, Figure 4C). When patterns were random, people preferred spiky, angular edges ($t(41) = -3.114$, $p = 0.003$), but when patterns were symmetrical, there was a trend in the other direction ($t(41) = 1.637$, $p = 0.109$).

7.2. 1D Experiments

Results of the 1D experiments are shown in Figure 5. As expected, participants again preferred symmetry to random ($t(41) = 11.828$, $p < 0.001$, Figure 5A) and blue to brown ($t(41) = 8.925$, $p < 0.001$, Figure 5B). Numerically, people preferred spiky to smooth (the opposite of the expected results), although this was not significant ($t(41) = -1.725$, $p = 0.092$, Figure 5C).

We can now compare effects from these 1D experiments with the equivalent main effects from the 3D experiment. Does each 1D result tell us something about preference for richer stimuli, where there are multiple dimensions which could potentially be considered?

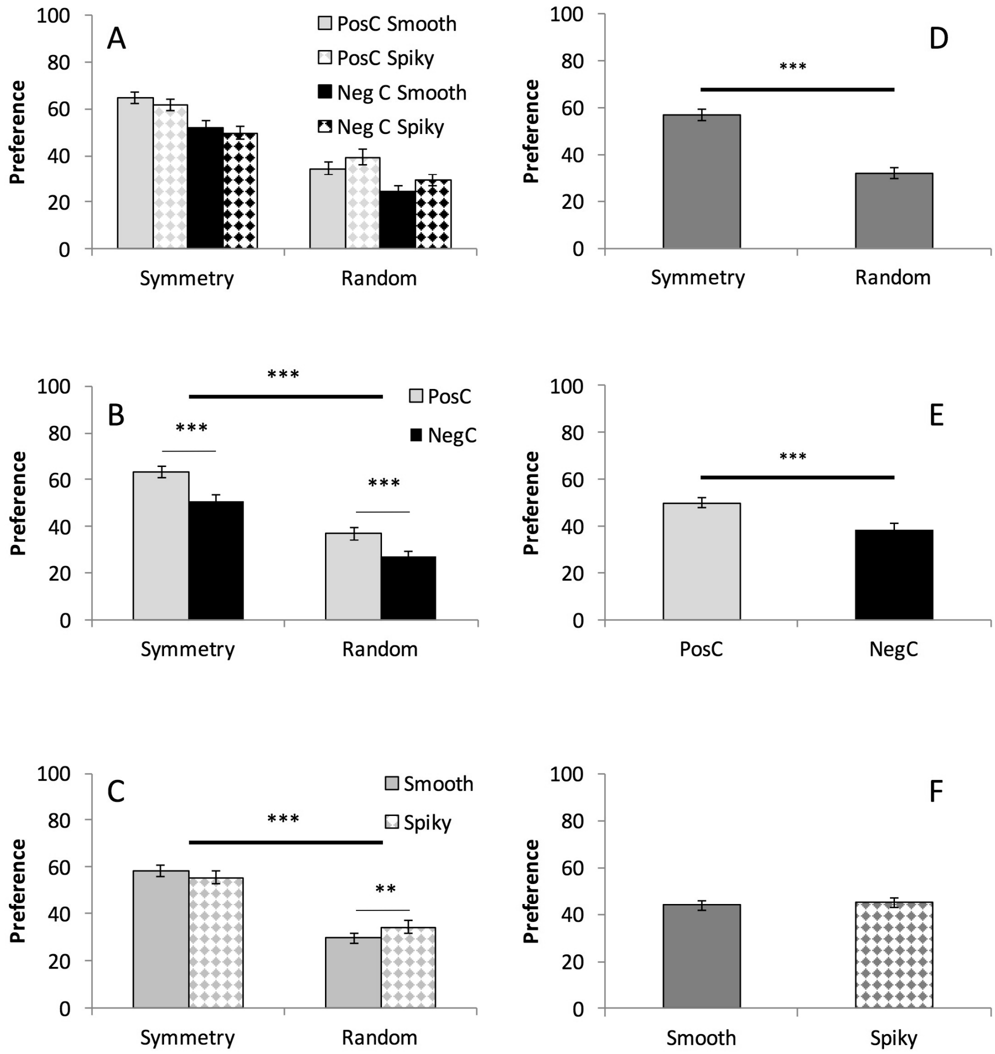

Figure 4. Results of the 3D experiment. A, preference in all conditions. B–F, same results averaged over one on more factors. Error bars = +/- 1 S.E.M. ** $p < 0.01$, *** $p < 0.001$.

First, preference for symmetry over random arrangements was explored. Results are shown in Figure 5D. These data were analysed with a Regularity (Symmetry, Random) x Context (1D, 3D) repeated measures ANOVA. Although people always preferred symmetry, this was slightly larger in the 1D experiment than the 3D experiment, resulting in a Regularity x Context interaction ($F(1,41) = 8.438$, $p = 0.006$, partial $\eta^2 = 0.171$).

Next, preference for blue over brown was analysed in the same way. Results are shown in Figure 5E. Preference for blue was found in both experiments, but was greatly reduced in the 3D experiment. This was confirmed with a Colour (Blue, Brown) x Context (1D, 3D) repeated measures ANOVA. There was a main effect of Context ($F(1,41) = 14.245$, $p < 0.001$, partial $\eta^2 = 0.258$), and a Context x Colour interaction ($F(1,41) = 22.273$, $p < 0.001$, partial $\eta^2 = 0.352$).

Finally, we examined the effect of context on element shape (Figure 5F). The only significant result was a main effect of Context, with higher ratings in the 3D experiment ($F(1,41) = 44.437$, $p < 0.001$, partial $\eta^2 = 0.520$).

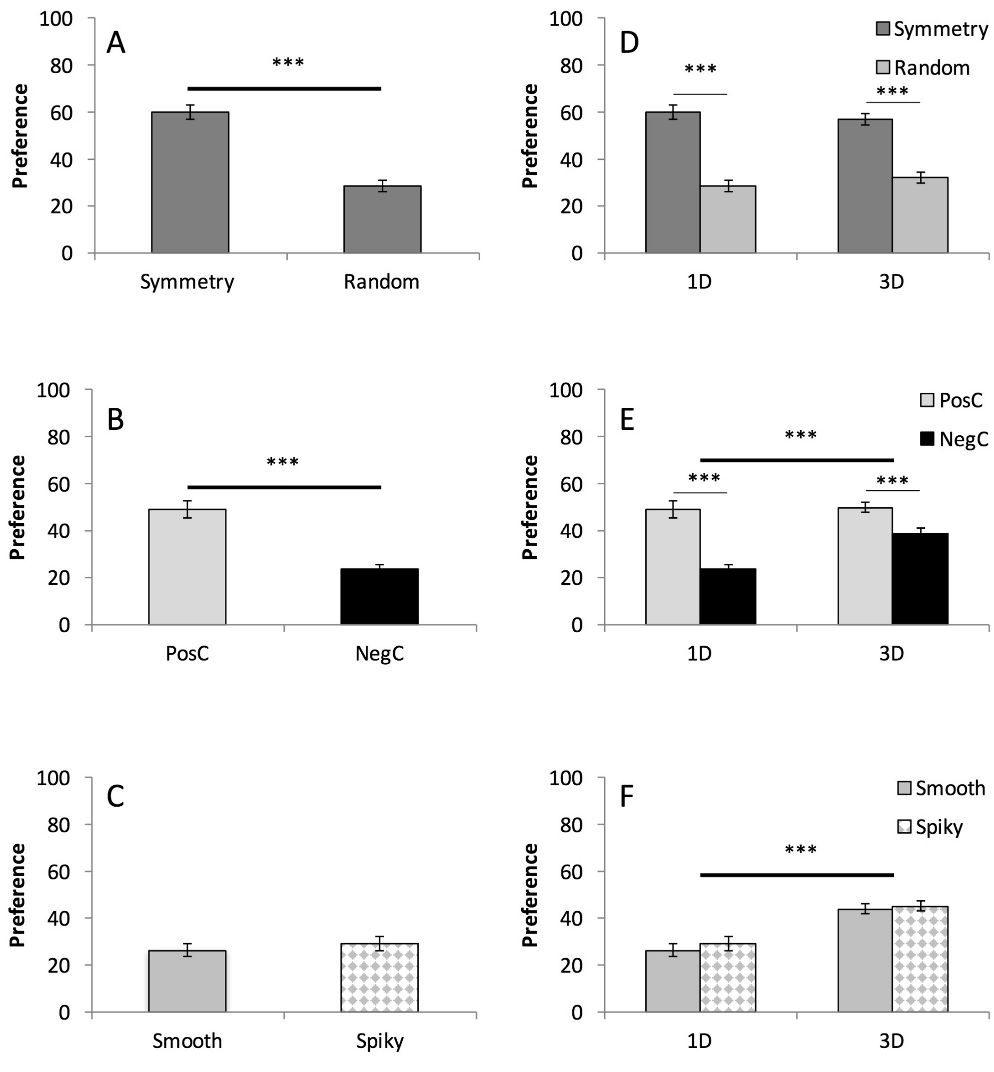

Figure 5. A–C show results of the 1D experiments. D–F show results of 1D experiments and the equivalent main effect from the 3D experiment. Error bars = +/- 1 S.E.M. *** p < 0.001.

7.3. Individual Differences

People have different preferences. The generalizability of individual differences can be estimated by correlating preferences in the 1D and 3D experiments. Results are shown in Figure 6. People who rated symmetry highly in the 1D experiment also tended to give a high rating for symmetry in the 3D experiment ($r = 0.655$, $p < 0.001$, Figure 6A). There was an apparent transfer of individual differences in colour preference as well, although this did not reach significance with a two-tailed test ($r = 0.291$, $p = 0.061$, Figure 6B). There was no correlation for element shape, although the relationship was positive ($r = 0.227$, $p = 0.148$, Figure 6C).

8. Discussion

I presume that most people working on empirical aesthetics would agree that we can never predict how much someone will like a piece of art by breaking it into separate visual parts, and then measuring their preference for each part in isolation. However, the reductive psychophysical approach to aesthetic science is still very common. So how scary is the gestalt nightmare? The experiments reported here provide some preliminary insights.

Perhaps there are common variations of the gestalt nightmare, which keep recurring? In the 1D experiments, preference for symmetry was stronger than preference for blue. The preference for symmetry remained at about the same magnitude in the 3D experiment, but the preference for blue was substantially reduced. *Here the stronger factor dominated preferences when there are many factors available.*

This kind of *masking* is probably a very common variant of the gestalt nightmare. If we have very impoverished stimuli, then we may find quite substantial and statistically significant effects on preference. However, these effects may vanish if more potent dimensions are added, but would still be found if less potent dimensions were added.

For multidimensional stimuli, preference judgments will likely be dominated by a few salient dimensions, and otherwise important factors that may significantly affect preference when tested in isolation will be drowned out. For example, if we added extreme positive or negative semantic content as a second dimension, the aesthetic effect of colour or symmetry might disappear completely (because the participant's aesthetic faculties are totally absorbed with the obvious valence difference). This is a particularly important consideration for empirical aesthetics, because we typically measure low magnitude effects on preference (symmetry, colour, curvature), while real art often includes very potent factors that are hard to quantify (ideological messages, extreme positive or negative valence, challenge to stylistic norms, and so on).

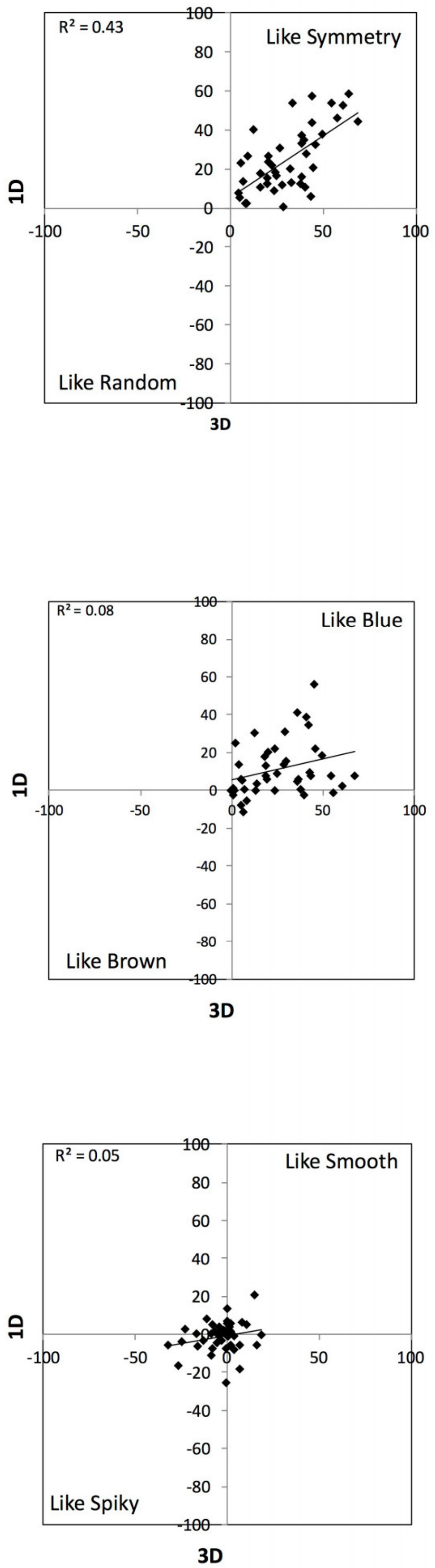

Figure 6. Correlations between individual preference magnitudes in 1D and 3D experiments. Each data point is a participant.

There was also an interesting interaction between regularity and element shape in the 3D experiment. When the patterns were symmetrical, there was a trend in favour of curved over angular shapes, but when patterns were random, people significantly preferred angular shapes to curved. This points to another potential variant of the nightmare. What mattered was not the absolute valence, but the agreement between valence of different dimensions of the stimuli. People do not just look at an image and award a point for every part with positive valence, and subtract a point for every part that has negative valence. Instead, consistency between different parts of the image is good, even if that 'consistency' means double negative valence. This relates to the idea of *representational fit*: for instance, if an artist wants to paint a battle scene, they might use disharmonious, clashing colours. Colour clash on its own may be unattractive, but here aesthetic merit comes from deliberate congruence between dimensions of the work at different levels (Palmer, Scloss and Sammartino, 2013). This could be another common variant of the gestalt nightmare. It should be noted that this congruence wasn't found with the Colour-Symmetry condition. Here the effects of colour and symmetry were approximately additive (participants did not like random, brown patterns for instance). This shows again that any apparent aesthetic law, such as the value of congruence, is potentially plastic.

In summary, the hyper-dimensional human preference space may not be so tightly tangled and folded that we should abandon the reductive approach completely, but there are substantial problems.

9. Gestalt Nightmare and Different Research Designs

The gestalt nightmare still haunts other studies which use slight variations of the reductive psychophysical approach. For example, the X axis is not always an objective dimension of the stimuli. Sometimes we measure the relationship between two kinds of subjective rating. For example, Berlyne proposed a lawful relationship between subjective arousal and subjective preference, where people like mid-level arousal (Berlyne, 1971). This is still an example of a reductive psychophysical approach. We are still varying just one dimension at a time, albeit a subjective dimension. Again, there is no guarantee that addition of other dimensions will leave the lawful relationship unchanged.

A nice recent example of this interaction was reported by Albrecht and Carbon (2014). They reviewed the fluency model, which suggests that people are sensitive to the speed and efficiency of their own perceptual and cognitive processing. Generally, fluent processing engenders positive emotions, which are sometimes attributed to the merit of the stimuli (Reber, 2012). However, fluency was found to amplify pre-existing valence, so visual fluency made already-negative images more negative, and already-positive images more positive.

The gestalt nightmare is still a problem if the experiment has more than one factor and an interaction is hypothesized *a priori*. We never know how this interaction will be further twisted by the addition of another factor. For example, we cannot be sure that the fluency by valence interaction found by Albrecht and Carbon (2014) would remain if something else was included as a third factor. The mere use of a multifactorial design does not immediately save us from the gestalt nightmare.

Despite these considerations, the gestalt nightmare is probably most acute when we plot lawful relationships between discrete, objective stimulus properties and subjective preferences (Figure 1B). It perhaps fades somewhat when plotting the relationship between basic psychological states (Berlyne, 1971). Finally, the gestalt nightmare may not be relevant at all to research that examines correlations between higher-level components of the aesthetic experience, like psychological tension and the inclination to cry (e.g. Pelowski, 2015). These 'top-level' constructs can be conceptualized as emotional gestalts (Barrett, 2012). Emotional gestalts are viable entities for empirical measurement, even though they have a very complicated internal structure and many unconscious components. Aesthetic science can possibly avoid the gestalt nightmare by measuring psychodynamics at this level.

Palmer and Schloss (2010) reported a mini-gestalt nightmare. The lawful relations between colour and preference were modulated by brightness. Specifically, preference for yellows was affected by brightness, but this was not found for other colours. Palmer and Schloss provided evidence that colour preferences are caused by association with objects that have a pre-existing valence. Dark yellow resembles faeces or vomit, while light yellow often reminds us of objects with positive valence, like the sun. They found that colour preferences were dependent on the strength of the colour–object associations. Objects are combinations of features, and preferences are targeted towards familiar objects. By studying objects, empirical aesthetics is partly relieved from the impossible burden of measuring preference for every conceivable dimension and every combination. This may reduce the gestalt nightmare. However, I would tentatively argue that it does not eliminate it. Preferences for objects can be documented, but any systematic result is likely to be modulated when another feature is added to the display, such as composition, familiarity, or artistic intentions.

Indeed, we should also note that ‘additional dimensions’ need not be properties of the objects in the strictest sense. For example, mere familiarity alters preference (Zajonc, 1968), while the dynamics of prototype formation and innovation play out over decades (Carbon, 2010; Cutting, 2003). History and fashion can strongly influence preference for a particular design, and the effects of history on preference may interact with the intrinsic properties in unpredictable ways. The gestalt nightmare has a temporal dimension.

There is growing interest in understanding the neural basis of aesthetic experience (Chatterjee, 2011). Leading contemporary researchers such as Leder and Nadal (2014) are optimistic that neuroscience can be increasingly integrated with empirical aesthetics in future. Others are more sceptical (Conway and Rehding, 2013). I will sidestep the broader debate to focus on a specific point. The gestalt nightmare arises whenever there are very few *independent variables*, and it is *not* remedied be recording high-dimensional data sets with a lot of *dependent variables* (although this may be desirable for other reasons). In neuroaesthetics we get an explosion of dependent variables, but experiments may still have relatively few independent variables. The gestalt nightmare remains as it was in old-fashioned preference rating experiments.

We can illustrate this with research that has measured activation of the facial muscle responsible for smiling (Zygomaticus Major, ZM). Research has found weak ZM activation when participants observe symmetrical patterns (Makin *et al.*, 2012, Experiment 1), prototypical dot patterns (Winkielman *et al.*, 2006), and attractive faces (Gerger *et al.*, 2011). ZM recordings tell us about the latency of the emotional response, and therefore contain more information than verbal preference reports. However, the extra time-dimension in our data set says nothing about the consequences of including a new independent variable in our experimental design (indeed, Experiment 2 of Makin *et al.*, 2012, found that the effect could be reversed if response-key labels were switched).

Of course, all psychophysics is reductionist in a sense: we study motion, colour, and contrast perception by putting participants in a chin rest, in a dark room in front of a screen at a fixed distance, and isolating and varying one aspect of the stimuli at a time. Would it be legitimate to complain that these results tell us nothing about complex real-world stimuli? No, that would miss the point: the aim of psychophysics is to probe the discriminatory limits of the visual system, and relate this to facts about optics and neurobiology of the eye and visual brain. In scientific aesthetics there is some aspiration to generalize beyond the dimensions tested in the lab, and find laws of aesthetics which help us understand the appeal of art and nature. We are not trying to push the system to make the smallest and most precise aesthetic discrimination possible!

10. Attractive Faces

Human faces are gestalts that provoke strong emotional responses, and they are occasionally very beautiful. What lessons can facial attractiveness research provide for scientific aesthetics? Facial attraction is about truthful indicators of health and fertility, with some fine-tuning depending on our current reproductive strategy (Little, Jones and DeBruine, 2011). Amongst other things, we like smooth skin, reddish colouration (from carotenoids), approximate symmetry, and prototypicality. Males are attracted to feminine features (shaped by oestrogen), females sometimes like masculinized faces (shaped by testosterone), although this is modulated by the menstrual cycle and current partnerships. Aesthetic responses to other objects are less obviously linked to natural selection.

Tsao and Livingstone (2008) review the neuroscience of face perception. They claim that faces are processed holistically and immediately, and this is explained by an early, automatic detection stage, like a template that finds upright eye-nose-mouth arrangements in the retinal image. Other shapes are not captured by innate detectors that gate dedicated upstream circuitry. This is another reason why facial beauty might be a separate category.

I tentatively suggest that facial attractiveness is tractable with the reductive approach. Faces vary on (relatively) few discrete dimensions, so we can work through multidimensional space exhaustively (although no doubt this is still a huge challenge). Natural selection has rationalized our feelings about faces (from the genes point of view). Consequently, face preferences are probably more uniform and predictable, so effects (and new hypotheses) can be discovered more easily. Studying other kinds of beauty is a different game altogether.

11. Summary and Conclusions

Empirical aesthetics is popular and growing. The reductive psychophysical approach remains common. We vary a property of the stimuli, and measure some aesthetic reaction. We have seen that there are very significant problems with this approach, which can be categorized as problems on the X and Y dimensions. Although the paradigmatic aesthetic experience involves strong emotions, fascination, and aesthetic rapture, it is nearly impossible to evoke these feelings in the lab, especially on multiple trials with well controlled stimuli. Most of this applies to studying 'hot' emotions generally. We thus have to resort to 'aesthetics with a lower case a', and study cold preferences and evaluations. Now we have problems with the X axis. We cannot expect to discover lawful relations between stimulus attributes and preference, and then generalize these findings far beyond the initial stimulus set (at least not without further justification). The gestalt nightmare is real: the vast number of stimulus dimensions which systematically alter preference are not orthogonal and independent, and the reductive approach tacitly assumes that they are.

To avoid any misunderstanding at this point, we should quickly acknowledge the reductive approach as *only one part* of a broader research programme. There have been very interesting attempts to apply general theories from cognitive neuroscience to understand art (Ramachandran and Hirstein, 1999; Redies, 2007; Van de Cruys and Wagemans, 2011; Zeki, 2002). Others have pointed out that artists are experts on vision. They have learned which cues carry compelling visual impressions (e.g. contours, occlusions), and what can be ignored (e.g. precise optics of shadows and reflections). Cavanagh (2005) claimed that 'paintings and drawings are a 40,000-year record of experiments in visual neuroscience'. There is also valuable cross-talk between science and modern artists (http://www.gestaltrevision.be/en/).

The reductive approach will sometimes find a basic law which is fairly robust. For example, there are reliable preferences for compositional arrangements (Palmer, Gardner and Wickens, 2008) and colours (Palmer and Schloss, 2010). People like symmetry in abstract patterns (Jacobsen and Höfel, 2002), smoothed rather than sharp edges (Bar and Neta, 2007), and certain types of fractal structure (Spehar *et* *al.*, 2003). Sometimes people like stimuli which are fluently processed (Reber, 2012) or have the medium arousal potential (Berlyne, 1971), or which have been targets rather than distractors on previous visual search tasks (Fenske and Raymond, 2006). These regularities have been discovered with the reductive psychophysical approach, and they represent genuine insights. It would be too extreme to say that these studies tell us nothing about aesthetics: they do, but only in a rather limited sense.

We can use the *Turing test* rationale to get a handle on how important these limitations are. If we really understand aesthetics, we should be able to program a computer to make art which humans find appealing, using explicit *production rules* inspired by empirical findings. Computers can indeed run pattern-generation algorithms with surprising and appealing visual results. For instance, dynamic screensavers are sometimes fascinating. But do we credit the programmer or the program? At present, we are a long way from simulating the judgments of a human artist, even if computers can mediate between the artist/programmer and artwork in increasingly complex ways (Chamberlain, Mullin and Wagemans, 2015).

Compositional rules validated in empirical experiments may help a photographer produce desirable photographs, and knowledge of colour preferences may help a web designer make an attractive website. However, we are certainly not at a stage where empirical aesthetics provides a recipe book which can lead to successful art and design. Empirical aesthetics can provide tips to help a complete amateur, but not much more. In fact, these tips may lead in completely the wrong direction, because they will always interact unpredictably, and any trained artist or designer will probably be able see these aesthetic blunders immediately (e.g. Dickie, 1962).

This was illustrated in an amusing way by the artists Komar and Melamid: they telephoned a sample of around 1,000 Americans and surveyed their preferences on many dimensions (favourite colour, second favourite colour, prefer domestic or wild animals, and so on), and then proceeded to paint 'America's favourite painting'. The resulting 'art by democracy' is not highly rated (http://awp.diaart.org/km/).[3] Good art and design requires the careful and precise inter-

[3] The 2016 Eurovision song contest revisited the humour of Komar and Melamid (https://www.youtube.com/watch?v=aMgW54HBOS0).

play of every visual and semantic element, and this precise balance cannot be dismantled and understood bit by bit (e.g. Kubovy, 2000).

Is all this just a straw-man argument — am I overstating the case of my opponent? Not really, because there *is no vocal advocate* of reductive aesthetic science who denies the existence of its limitations. At the same time, the research community continues to publish new papers using the reductive, quasi-psychophysical approach, as if the problems are small and solving them can be postponed indefinitely. I claim here that the problems are substantial, and need to be faced.

12. Could a New Aesthetic Science Say More About Aesthetic Experience?

We can put the conclusions of this paper into a single sentence: *aesthetic experience is fundamentally about hot emotional reactions to wholes, but empirical aesthetics is stuck measuring cold evaluation of parts*. I propose that the future of scientific aesthetics depends on how successfully we can transcend the reductive, quasi-psychophysical approach.

Furthermore, dwelling on these problems is a worthwhile exercise because it sharpens our sense of what an ideal psychological theory would look like. It would give a detailed neural model of why some gestalts evoke aesthetic rapture when viewed in exactly the right conditions at exactly the right moment in our personal history. It would explain why aesthetic emotions sometimes occur unexpectedly, while we often feel flat in almost exactly the same situation. It would give us precise rules to produce art that evoked these feelings reliably. Models would make specific, falsifiable predictions, and be transparently connected to the rest of the natural sciences.

Although this is science fiction, there have been admirable recent attempts to transcend the narrow reductive approach while retaining some insights and methods from cognitive science. As mentioned, Pelowski (2015) elaborated on the schema shift ideas in Pelowski and Akiba (2011), and measured inclination to cry in response to Rothko paintings. This circumvented many problems on the X and Y axis. However, the approach has other weaknesses instead. First, there are *still* no precise production rules which could be programmed into a computer. Second, the background theory draws very heavily on introspections of artists and critics. This yields insights which cannot be obtained in other ways (e.g. from neuroscience). But introspection is problematic: people will readily tell stories about their own feelings and behaviours. But are these stories informed by a private, internal epistemic channel? Nisbett and Wilson (1977) argued that 'intro-spection' is just the application of our own folk-psychological theories to ourselves. Third, there are problems with falsifiability. For example, the cognitive flow model in Pelowski and Akiba (2011) makes some novel predictions, but it could not be falsified by empirical data (e.g. the finding that most visitors did not feel tearful in art galleries was easily accommodated). Rich and multifaceted accounts cover more aspects of aesthetic experience, but at a cost: it becomes difficult to adjudicate between alternative accounts experimentally. These are old frustrations with philosophical aesthetics, and they make the reductive scientific approach seem bold and pioneering again!

Perhaps we can understand the gap between aesthetic science and aesthetic experience by comparing it to other gaps in human knowledge. The philosopher Colin McGinn (e.g. 1994) divides the realm of the unknown into *problems* that can be solved by scientists, and the insoluble *mysteries* which eternally vex philosophers. McGinn proposes that mysteries exist because intelligent minds have specific abilities and, consequently, specific limits. Rats will never understand prime numbers, so perhaps the human ability to understand natural phenomena is also limited?

The origin of the universe is a mystery in the McGinn sense. Saying 'there was the Big Bang' is tantamount to saying 'it just happened', while any advancement crashes into our unshakable intuitions about the nature of causality and time. According to McGinn the failure is a consequence of our uniquely human cognitive architecture. McGinn states: 'That reason is flummoxed by a certain class of problems is thus no proof that those problems possess any inherent refractoriness, nor that there are no other conceivable epistemic systems that might take these problems in stride' (*ibid.*, p. 137).

McGinn proposes a way of diagnosing mysteries, called the 'DIME shape'. When trying (in vain) to explain a mysterious natural phenomenon, we first try to *Demystify and Domesticate* the object of our puzzlement by proposing a reductive explanatory theory (**D**), when this fails, we *declare it Irreducible* (**I**) and sometimes resort to *Magical pseudo-explanations* (**M**), or *Eliminate the object from our ontology completely* (**E**). These attempts are antagonistic, 'so we hop unhappily from one unsatisfactory option to the next; or dig our heels (squintingly) into a position that seems the least intellectually unconscionable of the bunch' (*ibid.*, p. 145). The DIME shape is the telltale signature of human intelligence labouring under delusional aspirations to exceed innate limits. The recurring philosophical debates about the fit between consciousness and neurobiology can be attributed to the DIME dynamics. Of course, aesthetic experience is a conscious experience, and thus a mystery. But are there additional domain-specific mysteries for scientific aesthetics besides this big one?

When considering scientific aesthetics, we oscillate between preferring a reductive research programme (that says little about hot aesthetic experience) and preferring rich philosophical discourses (that are unfalsifiable). This grass-is-always-greener business loosely resembles the DIME dynamics discussed by McGinn. However, the DIME resemblance is not perfect. There is NO sheer cognitive cliff, which makes intuition seem painfully impotent. Aesthetic science would progress nicely *if only* we could reliably evoke and measure hot aesthetic emotions under controlled conditions. But aesthetic emotions are too fleeting and idiosyncratic, they are too minutely sensitive to thousands of internal and external details. We are like scientists who would love to measure a very rare whirlpool in a chaotic system, but cannot reliably recreate it in an artificial fluid tank. The DIME-like dynamics in aesthetic science reflect deliberation about how to proceed given this practical obstacle, not innate cognitive limitations of the type that make it impossible to intuitively understand consciousness or the origin of the universe.

References

Albrecht, S. & Carbon, C.C. (2014) The Fluency Amplification Model: Fluent stimuli show more intense but not evidently more positive evaluations, *Acta Psychologica*, **148**, pp. 195–203.

Arnheim, R. (1974) *Art and Visual Perception: A Psychology of the Creative Eye*, Berkeley, CA: University of California Press.

Bar, M. & Neta, M. (2007) Visual elements of subjective preference modulate amygdala activation, *Neuropsychologia*, **45**, pp. 2191–2200.

Barrett, L.F. (2012) Emotions are real, *Emotion*, **12**, pp. 413–429.

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

Bertamini, M., Palumbo, L., Gheorghes, T.N. & Galatsidas, M. (2016) Do observers like curvature or do they dislike angularity?, *British Journal of Psychology*, **107**, pp. 154–178.

Birkhoff, G (1932) *Aesthetic Measure*, Cambridge, MA: Harvard University Press.

Carbon, C.C. (2010) The cycle of preference: Long-term dynamics of aesthetic appreciation, *Acta Psychologica*, **134**, pp. 233–244.

Cavanagh, P. (2005) The artist as neuroscientist, *Nature*, **434** (7031), pp. 301–307.

Chamberlain, R., Mullin, C. & Wagemans, J. (2015) The artistic Turing test: An exploration of perceptions of computer-generated and man-made art, *Journal of Vision, 15. VSS Annual Meeting*, Florida, 15–20 May 2015, 112.

Chatterjee, A. (2011) Neuroaesthetics: A coming of age story, *Journal of Cognitive Neuroscience*, **23**, pp. 53–62.

Conway, B.R. & Rehding, A. (2013) Neuroaesthetics and the trouble with beauty, *PLoS Biology*, **11**, e1001504.

Csikszentmihalyi, M. (1991) *Flow: The Psychology of Optimal Experience*, New York: Harper Perennial.

Cutting, J.E. (2003) Gustave Caillebotte, French impressionism, and mere exposure, *Psychonomic Bulletin & Review*, **10**, pp. 319–343.

Dickie, G. (1962) Is psychology relevant to aesthetics?, *The Philosophical Review*, **71**, pp. 285–302.

Eisenman, R. (1967) Complexity-simplicity: I. Preference for symmetry and rejection of complexity, *Psychonomic Science*, **8**, pp. 170–196.

Ekman, P. (1999) Facial expressions, in Dalgleish, T. & Power, M. (eds.) *Handbook of Cognition and Emotion*, pp. 301–320, New York: Wiley.

Eysenk, H. (1941) The empirical determination of an aesthetic formula, *Psychological Review*, **48**, pp. 83–92.

Fechner, G.T. (1876) *Vorschule der Ästhetik*, Leipzig: Breitkopf und Härtel.

Fenske, M.J. & Raymond, J.E. (2006) Affective influences of selective attention, *Current Directions in Psychological Science*, **15**, pp. 312–316.

Gerger, G., Leder, H., Tinio, P.P.L. & Schacht, A. (2011) Faces versus patterns: Exploring aesthetic reactions using facial EMG, *Psychology of Aesthetics, Creativity, and the Arts*, **5**, pp. 241–250.

Holmes, T. & Zanker, J.M. (2012) Using an oculomotor signature as an indicator of aesthetic preference, *I-Perception*, **3**, pp. 426–439.

Jacobsen, T. & Höfel, L. (2002) Aesthetic judgments of novel graphic patterns: Analyses of individual judgments, *Perceptual and Motor Skills*, **95**, pp. 755–766.

Jacobsen, T., Schubotz, R.I., Höfel, L. & Cramon, D.Y. (2006) Brain correlates of aesthetic judgment of beauty, *Neuroimage*, **29**, pp. 276–285.

Kubovy, M. (2000) Visual aesthetics, in Kardazin, A.E. (ed.) *Encyclopedia of Psychology*, New York: Oxford University Press.

Langer, E.J. & Moldoveanu, M. (2000) The construct of mindfulness, *Journal of Social Issues*, **56**, pp. 1–9.

Leder, H., Belke, B., Oeberst, A. & Augustin, M.D. (2004) A model of aesthetic appreciation and aesthetic judgments, *British Journal of Psychology*, **95**, pp. 489–508.

Leder, H. & Nadal, M. (2014) Ten years of a model of aesthetic appreciation and aesthetic judgments: The aesthetic episode — developments and challenges in empirical aesthetics, *British Journal of Psychology*, **105**, pp. 443–464.

LeDoux, J. (1998) *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*, London: Weidenfeld and Nicolson.

Little, A.C., Jones, B.C. & DeBruine, L.M. (2011) Facial attractiveness: Evolutionary based research, *Philosophical Transactions of the Royal Society B: Biological Sciences*, **366**, pp. 1638–1659.

Makin, A.D.J., Pecchinenda, A. & Bertamini, M. (2012) Implicit affective evaluation of visual symmetry, *Emotion*, **12**, pp. 1021–1030.

Makin, A.D.J., Wilton, M.M., Pecchinenda, A. & Bertamini, M. (2012) Symmetry perception and affective responses: A combined EEG/EMG study, *Neuropsychologia*, **50**, pp. 3250–3261.

Markovic, S. (2012) Components of aesthetic experience: Aesthetic fascination, aesthetic appraisal, and aesthetic emotion, *I Perception*, **3**, pp. 1–17.

Maslow, A. (1972) *The Farther Reaches of Human Nature*, Chapel Hill, NC: Maurice Bassett.

Mauss, I. & Robinson, M. (2009) Measures of emotion: A review, *Cognition & Emotion*, **23**, pp. 209–237.

McGinn, C. (1994) The problem of philosophy, *Philosophical Studies*, **76**, pp. 133–156.

Nisbett, R.E. & Wilson, T.D. (1977) Telling more than we can know — verbal reports on mental processes, *Psychological Review*, **84**, pp. 231–259.

Palmer, S.E., Gardner, J.S. & Wickens, T.D. (2008) Aesthetic issues in spatial composition: Effects of position and direction on framing single objects, *Spatial Vision*, **3**, pp. 421–449.

Palmer, S.E. & Schloss, K.B. (2010) An ecological valence theory of human color preference, *Proceedings of the National Academy of Sciences USA*, **107**, pp. 8877–8882.

Palmer, S.E., Schloss, K.B. & Sammartino, J. (2013) Visual aesthetics and human preference, *Annual Review of Psychology*, **64**, pp. 77–107.

Peirce, J.W. (2007) PsychoPy — psychophysics software in Python, *Journal of Neuroscience Methods*, **162**, pp. 8–13.

Pelowski, M. (2015) Tears and transformation: Feeling like crying as an indicator of insightful or 'aesthetic' experience with art, *Frontiers in Psychology*, **6**, art. 1006.

Pelowski, M. & Akiba, F. (2011) A model of art perception, evaluation and emotion in transformative aesthetic experience, *New Ideas in Psychology*, **29**, pp. 80–97.

Ramachandran, V.S. & Hirstein, W. (1999) The science of art: A neurological theory of aesthetic experience, *Journal of Consciousness Studies*, **6** (6–7), pp. 15–31.

Reber, R. (2012) Processing fluency, aesthetic pleasure, and culturally shared taste, in Shimamura, A.P. & Palmer, S.E. (eds.) *Aesthetic Science: Connecting Mind, Brain, and Experience*, pp. 223–249, New York: Oxford University Press.

Redies, C. (2007) A universal model of esthetic perception based on the sensory coding of natural stimuli, *Spatial Vision*, **21**, pp. 97–117.

Saarimäki, H., *et al.* (2015) Discrete neural signatures of basic emotions, *Cerebral Cortex*, **26**, pp. 2563–2573.

Sauter, D.A., Eisner, F., Ekman, P. & Scott, S.K. (2010) Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations, *Proceedings of the National Academy of Sciences USA*, **107**, pp. 2408–2412.

Schloss, K.B. & Palmer, S.E. (2011) Aesthetic response to color combinations: Preference, harmony and similarity, *Attention, Perception and Psychophysics*, **73**, pp. 551–571.

Shimamura, A.P. (2012) Towards a science of aesthetics, in Shimamura, A.P. & Palmer, S.E. (eds.) *Aesthetic Science: Connecting Minds, Brains and Experiences*, pp. 3–28, New York: Oxford University Press.

Shimamura, A.P. & Palmer, S.E. (2012) *Aesthetic Science: Connecting Minds, Brains and Experience*, New York: Oxford University Press.

Spehar, B., Clifford, C.W.G., Newell, B.R. & Taylor, R.P. (2003) Universal aesthetic of fractals, *Computers & Graphics-UK*, **27**, pp. 813–820.

Tsao, D.Y. & Livingstone, M.S. (2008) Mechanisms of face perception, *Annual Review of Neuroscience*, **31**, pp. 411–437.

Van de Cruys, S. & Wagemans, J. (2011) Putting reward in art: A tentative prediction error account of visual art, *I Perception*, **2**, pp. 1035–1062.

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

Zajonc, R.B. (1968) Attitudinal effects of mere exposure, *Journal of Personality and Social Psychology Monographs*, **9**, pp. 1–27.

Zeki, S. (2002) Neural concept formation and art: Dante, Michelangelo, Wagner, *Journal of Consciousness Studies*, **9** (3), pp. 53–76.

Paper received February 2016; revised July 2016.