[Review] The Origins and Development of Aesthetics

Daphne Maurer, PhD, FRSC
Distinguished University Professor,
Department of Psychology, Neuroscience & Behaviour,
McMaster University

and

Charles Maurer (independent scholar)

This paper is to be published in Philosophical Transactions B of the Royal Society on 15 July 2024.

This is the accepted manuscript in its state before layout and final proofing. It must not be quoted. Quote the published version only: https://doi.org/10.1098/rstb.2023.0246

The manuscript's figures are between pages 11 and 12, and the file of supplementary figures is appended. It begins after page 23.
Abstract

All people (and some other animals) have aesthetic responses to sensory stimulation, responses of emotional pleasure or displeasure. These emotions vary from one person and culture to another, yet they share a common mechanism. To survive, an adaptive animal (as opposed to a tropic animal) needs to become comfortable with normality and to have slight abnormalities draw attention to themselves. Walking through a jungle you need to notice a tiger from a single stripe: if you must wait to see the whole animal, you are unlikely to survive. In *Homo sapiens*, the brain’s adaptive neurochemistry does this naturally, partly because the brain’s neuronal networks are structured to react efficiently to fractal structures, structures that shape much of nature. Previous associations may turn a slight variation from normal into feelings of either pleasure or danger. The details of these responses—what is normal and what variations feel like—they will depend upon an individual’s experience, but the mechanism is the same, no matter whether a person is tasting a wine, seeing a face or landscape, or hearing a song.

*Keywords:* aesthetics, attractiveness, preferences, beauty, pleasure, displeasure

*Supplemental figures can be downloaded from https://.............*
Feelings, like other perceptions, start out simple then elaborate with experience. A newborn baby senses continua from normal sensation to abnormal, and feels more comfortable with normality.

Normality at this age is defined by experiences in the womb, experiences formed by built-in processes interacting with the chemicals of the mother’s amniotic fluid, the pressures of maternal and fetal movements, and the vibrations of the mother’s voice. If a stimulus is excessively abnormal, the nervous system responds by instigating low-level muscular and glandular reflexes. A newborn’s fingers may form fists, his arms may push, his legs may kick, and his muscles may contract throughout his trunk, neck and head: a crying baby will result. In contrast, the resumption of normality will come to feel pleasurable. Over months, years and decades the brain learns to control muscular reflexes, to process increasingly complex stimuli as expectable or normal, and, ultimately, to associate them with different feelings and labels.

Discriminating abnormal from normal is the fundamental capability of an adaptive organism, and discriminating this quickly with minimal information is essential for survival. If you do not notice a tiger by a mere hint of its stripes, then you may not live long enough to notice it at all. For this reason, the neurophysiology of the brain has adapted so that slight deviations from normal are particularly salient. The combination of these phenomena forms the mechanism of aesthetic pleasure and displeasure, of the attractiveness and unattractiveness of different sensory perceptions. We like normality, we dislike abnormality, and we especially like or especially dislike anything that is on the edge. With the processing of a sensation into the perception of either “normal” or “abnormal” comes a set of positive and/or negative feelings from associations formed during an individual’s development: those form a person’s aesthetic values.

Different personal histories cause different people to have different reactions to a similar situation—what is familiar to me may be novel to you—but the mechanism is constant. What any person comes to feel as normal, and whether someone interprets a variation as lovely or monstrous—that will depend upon individual experience, but the mechanism will be the same, no matter whether a person is tasting a wine, seeing a face, or hearing a song. In every case, “normality” begins to be learned during development and forms the basis of aesthetics.

A few other authors have presented similar ideas for specific aesthetic domains in adults, but we shall show that this mechanism develops from birth and forms the roots of aesthetic perceptions and preferences across all domains. Our discussion will begin with taste and smell, because they are relatively simple sensory systems, then continue with faces, music, and dance.
Taste and Smell

Babies are born with preferences for tastes and smells, preferences that they show through positive or negative reactions: either smiling and swallowing or pouting, frowning and spitting. They like tastes that are sweet and savoury and smells that are fruity or milky; they dislike tastes that are sour or bitter and smells that are fishy or rotten.7,8 Similar congenital preferences evolved in other species as well as man, probably because they help infants to survive by biasing them toward swallowing useful foodstuffs and spitting out poisonous ones. Useful foods tend to have sweet or savoury flavours; worrisome ones tend to be sour, bitter, or foul-smelling; and some that can go either way, depending on our physiological state, we taste as salty. However, aside from those congenital biases, children learn all of their flavour preferences. They come to like what smells and tastes familiar.

Learning to recognize flavours begins in the womb. Chemicals in the amniotic fluid vary with the mother’s diet. These chemicals stimulate the fetus’s olfactory bulbs and, when its mouth is open, its taste buds as well. After birth, a baby will usually frown and avoid a novel odour like anise, yet if the mother ate anise while pregnant, the baby is likely not to mind it.9 The same holds when a 6-month-old tastes carrot-flavoured cereal for the first time. A baby will normally dislike this, but not if the mother drank carrot juice before or shortly after birth, so that the flavour was in her amniotic fluid or breast milk.10 With development, more complex flavours come to be preferred but familiar ones remain comforting.

Through the entirety of a person’s life, familiarity with flavours forms preferences. For example, consider coffee. Coffee is too bitter for a baby to swallow, and no child nor adult will drink it naturally, yet societal pressures induce most of us to learn to drink it in adolescence, and the strength of coffee that comes to taste best is the strength of the coffee that we become used to. An approximation of this strength is the ratio of coffee beans per unit of water. From one market to another this ranges across an order of magnitude.11,12 Jus de chaussettes (juice of socks) to a Parisian is a Nebraskan farmer’s normal brew.

Yet although the average coffee you drink will taste good, a restaurant might make it taste even better with an unusual twist that you enjoy, perhaps serving it with a dollop of whipped cream instead of table cream. Or the twist can take an unpleasant turn, if the restaurant offers not cream but skim milk. The former provides unexpected pleasure and the latter unexpected displeasure: positive and negative aesthetic valuations.2

This is how we form all aesthetic preferences. Whether a twist of the average is pleasant or unpleasant—whether it evokes positive or negative emotions and aesthetic evaluations, and how strong they are—that depends upon the individual’s experience amplified or damped by congenital differences in reactivity.2 The averages and twists of aesthetic preferences are all learned,
and different people and peoples apply different words to their positive and negative experiences, but the mechanism is universal and is built into the neurochemistry of brains.

As we said, this evolved as a basic mechanism required for an adaptive organism to survive: recognizing abnormal conditions. Tropic organisms react automatically but adaptive organisms react selectively. For an adaptive organism to survive, normality needs to recede into the background so that anything abnormal can stand out quickly and be dealt with appropriately. If you are fishing in central Africa, you need to pay little attention to the surface of the river unless you notice a small variation: something sticking up. That variation you need to notice instantly and warily, because it might be the nose of a crocodile. Adaptation and sensitivity to abnormality is the mechanism of survival, and its spinoff is aesthetics.

**Vision and Averaging**

Repeated stimulation forms neuronal networks: this is adaptation. Networks facilitate the transmission of energy through some synapses and inhibit the transmission through adjacent synapses. This makes repeated stimulation easy to process and efficient, so that it takes little neurophysiological effort to be aware that a river is calm.

In a similar way, the brain forms networks of neurons representing faces we have seen. The central tendencies of those faces cause networks to emerge that represent our cognitive prototype of all faces. This cognitive prototype is easy to process, so that up to a point—we shall discuss that point later—people prefer the look of faces that approximate the prototype more closely.

Numerous experiments show that our judgments of facial attractiveness are heavily influenced by the cognitive prototype—the average over time—formed by faces we have seen before. For example, a composite of 60 faces appears more attractive than most of the individual faces forming it. Moreover, moving the features of an individual face closer to the average position will make the face more attractive than moving the features an equal distance away from average (see Supplementary Figure A). This is a generic phenomenon. It has been demonstrated for the attractiveness of birds, dogs, fish, and the faces of watches.

A comparable face preference is manifest by 12 months of age (or maybe 6 months). By that time babies look longer at faces made more average than those made less average. However, babies show this preference only for human faces. Since babies have no experience with monkeys, they have no prototype of simian faces and show no preference for the average simian face. In contrast, 3-month-old monkeys look longer at a more average simian face yet make no distinction among human faces.
Developing this preference requires months of experience, because babies are born with extremely poor vision. Acuity and contrast sensitivity are so poor at birth that a newborn sees a face 30 cm away as just an outline with some blurry patterning inside. At greater distances, only the outline remains visible.\textsuperscript{20} For this reason, newborns spend most of their time fixating spots on the external contours of faces,\textsuperscript{21,22} and while doing so they pick up little information from peripheral vision.\textsuperscript{23}

Despite these limitations, newborns will fixate contrasty internal features of some simple outlines.\textsuperscript{24} Also, when shown a “face” formed from a lightbulb-shaped outline with squares for the eyes and mouth, newborns will look longer when the squares are in a face-like position than when they are inverted.\textsuperscript{25} (See Supplementary Figure B.)

This attentional bias assures long fixations on human faces—or on any stimulus showing dark features against a light background and more features at the top than the bottom, especially if the widest part of the outline holds the greatest density of features.\textsuperscript{26,27,28,29} However, this is not an innate “face-processing module”, it is nothing more than a bias induced by structures of the physical world. The nose blocks vision toward the inside lower corner of each eye. In consequence, the overall field of view is broader on top, so that more features are visible toward the top. This matters more to newborns than to adults, because the newborn’s field of view is much narrower.\textsuperscript{30} In addition, since the visual field is larger at the top, when an object within that field is larger at the top, and when details within that object are larger at the top, there is more energy at the top across different scales, forming a fractal structure. As we shall explain below, this makes either a face or a face-like pattern easy to process.

To attract the newborn’s eyes, features need usually to be dark on light. This is an evolutionary adaptation to a statistical regularity of the visual world: most of the things we look at are darker than the background. To understand why, imagine yourself outside on a sunny day. If the sun is behind you, it will illuminate everything you see, and any one object may be lighter or darker than its background. However, the sun is less likely to be behind you than to be on your left or right or in front of you, or overhead. In any of those positions it will illuminate directly only some of what you see. Most of what you look at will be in shadow, where it will be illuminated indirectly by objects around it and thus darker than those objects. Hence, most of the time, the object you are looking at will be darker than its surroundings. (See Supplementary Figure C.)

For this reason the brain becomes more used to darks against lights. Adults see darks against lights more quickly than the opposite,\textsuperscript{31} and when 3-month-old babies are shown an assortment of light and dark bars against an intermediate background, they look toward the dark bars first.\textsuperscript{32} (See Supplementary Figure D.)
Those are primitive, structural, attentional biases that serve a useful function: they serve to pull a baby’s gaze toward the mother’s face and hold it there, so that her face becomes familiar. After a few hours of familiarity, a newborn prefers looking at his mother’s face to looking at another woman’s face, even when body odours are masked. A baby begins to learn a mother’s odour while still in the womb, from her amniotic fluid: this familiar sensation also attracts him.

A preference to look at the mother’s face continues after six or eight weeks, when the baby’s vision becomes clearer and he starts looking at the features of her face. It can be demonstrated with photographs by 3 months of age. However, this preference can be reversed—in photographs—by excessive familiarity. It can be reversed by showing to a baby a photo of his mother, and then showing the same photo repeatedly so many times that the baby becomes bored. This exemplifies the basic mechanism of aesthetic preferences: familiarity is comfortable but a pleasant change can be welcome.

Around 3 months of age, something fundamental changes: “familiar” becomes defined by statistical averages. Then, when a baby sees photographs of faces, the average of those faces is treated as more familiar than any individual face. The statistics of experience begin to shape what the baby expects to see and to form attentional biases. By 12 months, accumulating experience makes the baby sensitive to more subtle differences.

These developments parallel changes in the differentiation of emotions and their expression. A newborn’s smiles are reflexive. They involve only turning up the corners of the mouth, they are more common when the baby is sleeping, and they occur in bouts, as though excess energy is finding a channel to be emitted. A social smile, in contrast, occurs only when the baby is awake, involves both the mouth and the eyes, and is elicited most easily by a familiar adult interacting with the baby in familiar ways, through patterns of movement and sound that the baby has experienced before and, hence, can process easily. This begins around at 2 months. Then, if an adult suddenly stops interacting and stares at the baby impassively, the social smile vanishes and a distressed cry-face emerges. This small but unwelcome deviation from normal elicits distress. At first a social smile is most easily elicited by the most familiar face—the mother—but with experience other faces also become easily processed and can lead to smiles in familiar contexts.

Experience continues to play a role in emotional and aesthetic reactions to faces even in adulthood. We can see these effects by varying the height of a face’s features. Imagine three photographs of a woman’s face. In one of them her internal features are at an average location; in the two others her features are shifted 2 standard deviations downward and upward. Although the lowered features do not typify adults, they do typify children. Of the three, adults find the average location to be most attractive and the low location to be second-
best. This pattern matches the average adult’s experience seeing faces. However, tall adults usually look down on people, which makes foreheads look more prominent than chins: they prefer the picture with the features lowered. In contrast, short adults usually look upward, which makes chins more prominent than foreheads. They prefer the picture with the features raised. (See Supplementary Figure E.)

A baby’s experience is like a short adult’s except more so. Babies seldom see people from straight on or from above. Consequently, babies look and smile more at faces with high features than at faces with features at average or low locations. Children’s preferences begin to change only after they begin to play with peers and to sit with adults around a table. Thus, four-year-olds rate faces as equally attractive whether those faces have the low features of the children’s peers or the average features of adults across a table, and they rate faces with high features as unattractive. The behaviour of 3-year-olds is consistent with this. Whether or not they rate faces with low features as most attractive depends on whether they went to day care and so had daily experience looking at their peers.

At every age, quantitative integrations of our previous experience determine what is expected, form biases, and define what we perceive to be attractive. Those perceptions of attractiveness remain malleable even into adulthood. A striking example of their malleability is a two-part experiment we performed. The first part was a memory task. For 8 minutes we showed adults 40 faces. We showed them each face a different number of times. After each presentation, the subject had to key in how many times that face had already been seen. Each subject then participated in a “second experiment”. In this “second experiment”, the subject used a mouse to move the features of an average face to the most attractive location. Unbeknownst to the subject, all of the faces in the previous memory task had had features raised or lowered by 2 standard deviations. Those subjects who had just spent 8 minutes looking at faces with low features preferred lower features than those who had just been looking at faces with high features. Those who had been seeing features in an average position preferred features in between.

New experience incorporated into our running integration can change our biases. This is the mechanism of fashion. Designers lower waist-lines to a level that looks absurd at first, then a new style develops and we find ourselves disliking clothes we raved about a year before. The same mechanism we hear constantly in music. For example, archetypical western popular song starts with a melody 16 beats long, then repeats the melody, then plays a variation the same length, then repeats the original melody with a slight variation toward the end. The melody sounds good, the repetition makes it familiar, the variation prevents boredom, then the varied repetition brings back the familiar with a twist, sustaining the positive emotion and aesthetic evaluation. Familiarity juxtaposed
to variation seems to be the basis of every kind of music in every culture, even what must surely be the most ponderous music on earth, Japanese gagaku.²

**Around Average**

That is only part of the story, however. We can average many faces or we can average only the most attractive faces. The latter looks more attractive than the former. The difference in attractiveness between those two averages we can treat as a unit of preference or bias. It is an arbitrary unit but a unit none-theless. We can measure the physical instantiation of this unit by (in principle) converting the two images into vector graphics, then measuring the difference in each vector’s dimension and location. These measurements we can use to create new images with greater or lesser bias. If we define the original average as 0, and the average of attractive faces as +1, we can manipulate faces by units of that amount to create a scale of bias above and below zero. When shown faces manipulated like this, adults find a bias near +1 to be most attractive, with a gradual drop-off above that and a faster drop-off below zero.⁴⁵ (See Supplementary Figure F.)

That illustrates the general nature of aesthetic preferences: the more average a stimulus, the better, but only up to a point. Too average can be boring, so we develop a preference for a flourish, for a bias toward some variation that our experience has taught us to enjoy.

Of course, boredom takes time to develop. Five-year-olds prefer average faces quite as adults do, but 5-year-olds do not show the adult preference for a flourish. For faces, the shift does not occur until adolescence.⁴¹

A similar mechanism works with body shape. A mere minute of exposure to an extremely thin or extremely fat body will shift in that direction the shape that a subject deems to be most normal and the shape judged to be ideal.⁴⁶,⁴⁷ When the prevalence of thin bodies shifts over the course of an experiment, normally sized bodies are more likely to look overweight. When the prevalence of thin bodies shifts over centuries—well, compare life-sized paintings of a nude Venus by Peter Paul Rubens and by William-Adolph Bouguereau, both erotica dressed as culture (see Supplementary Figure G). Rubens was an international celebrity of the 17th century. His customers were aristocrats who saw portliness as indicating wealth and virtue, and leanness as indicating poverty and untrustworthiness:⁴⁸

‘A bears him like a portly gentleman;
And to say truth, Verona brags of him
To be a virtuous and well-govern’d youth.”

*(Shakespeare, *Romeo and Juliet)*
Another lean unwashed artificer...
(Shakespeare, King John)
Yond Cassius has a lean and hungry look...
(Shakespeare, Julius Caesar)

Rubens’s ideal Venus was fleshly, but Bouguereau lived two centuries later in Paris, a recognizably modern city of 2 million where prototypes of female sexuality came from risqué young women working in theatres, cabarets, and dance halls, much as they come today from their counterparts in film, television, and music videos. Bouguereau captured this aesthetic so well and so often that he became one of the most respected artists in France, and the (all-male) Government of France bought his slim Venus for the nation. Over those two centuries normal experiences changed, so people learned to prefer different body shapes. The two paintings make this clear. They also incorporate subtle variations reflecting preferences of their times, so they illustrate not just historical norms but historical ideals.

Facilitation: Fractal Patterns

A shoot grows up from the ground and branches into two. Each branch grows awhile then branches into two. Each of the four new ones does the same. Eventually you will see a tree or a bush.

Nature builds trees and bushes by this kind of repetition, and many other objects as well. Leaves, snowflakes, sea shells, clouds, lightning, mountains, watersheds—the list can is infinite and includes the human circulatory and nervous systems. Each of these objects looks similar at any scale of measurement. They form a kind of symmetry (“same measure”) across scales. They are physical instantiations of what mathematicians call fractals. The redundant structure of fractal patterns makes them easier to perceive and remember, which facilitates feeling familiarity and noticing deviation.

Fractal patterns form the structures of music. Melodies are fractal structures unfolding in time, and so are individual notes. The archetypical musical note is a plucked string. The string vibrates over its whole length and simultaneously over half its length, one-third of its length, one-quarter, one-fifth, etc. This is called an harmonic series. Although it is an arithmetic progression, it can also be viewed as a set of fractals.

Musicians are taught that the first harmonic determines the pitch of a note and that higher harmonics determine its timbre (although they commonly use the terms “fundamental frequency” and either “overtones” or “partials”). Higher harmonics do determine timbre, but if the first harmonic determined the pitch of a note, you would never hear the pitches of bass instruments heard through a small radio, because the radio’s loudspeaker cannot reproduce the first harmonic or two of the lowest octaves in music.
In fact, we do not hear a note’s pitch through its first harmonic, we hear it through the fractal pattern formed by its harmonics. Fractal structures repeat themselves across scales, so they contain redundant information. When the first harmonic is missing because it is too low for a radio to reproduce, we still hear its pitch because the rest of the fractal pattern fills in the gaps: the higher harmonics’ redundancy limns the first harmonic well enough that we hear it. (See Figure 1.) Babies can do this from around 3 months of age.53

---

INSERT FIGURE 1 ABOUT HERE

Waveforms of an harmonic series. A shows the first harmonic; B shows the first 10 harmonics in their theoretically ideal proportions; C shows harmonics 2—10; D shows harmonics 4—10. In each case, the overall wavelength is the same. This shows that the overall pattern of harmonics determines the pitch we hear, not the first harmonic.

---

Redundancy facilitates cranial processing, so fractal structures are easy to process and, hence, tend to be pleasing. Chords formed of the simplest harmonics are consonant; chords formed of more complex harmonics are less likely to be. Newborns’ brains already hear that difference.54 Like adults, newborns also prefer consonant chords to dissonant.55 However when music is completely consonant, adults become used to it and bored. Mozart and Beethoven wrote largely consonant music but filled it with dissonance to provide contrasts that wake people up.

All of this holds with pictures as well. Images can be analyzed by spatial frequency as sounds can be analyzed by temporal frequency. Images do not have a properly harmonic structure, but they do contain a multitude of frequencies and filtering out many of them still leaves the subject recognizable. (See Figure 2.) A perfectly fractal structure like the golden rectangle looks attractive but soon seems repetitive and tedious, so a modest variation looks more interesting. (See Figure 3.) We prefer partially fractal pictures, pictures with a fractal dimension around 1.4. This holds whether looking at clouds, landscapes, or drip-and-splash paintings like those by Jackson Pollock.56,57

---

INSERT FIGURE 2 ABOUT HERE

Images 1 to 8 show a series of spatial frequencies taken from image 9. The spatial frequency doubles with each number.
The golden rectangle is a diminishing series of identical rectangles. Its fractal structure makes it easy to process but repetitive. Varying the structure adds interest.

Multimodal Perception

We have shown that our experience interacts with our perceptions to shape our preferences and emotional reactions, but we have talked only about individual senses. This is somewhat simplistic because from birth, our experience of almost everything is multimodal.

More sensory modes provide more information so, for example, newborn babies will learn to recognize their mother’s face more quickly if they hear her as well as see her, and shortly thereafter, her body odour will enhance brain functions involved in treating faces as a distinct category. Indeed, babies appear to be born intermixing all of their senses. We gradually learn to sort out our several sensory systems, but only up to a point. Ten to 15% of us notice specific intermixtures, reporting perhaps that calendrical sequences (January, February March) form spatial patterns or that words have tastes, but these are only the end of a continuum of synaesthesia. All of us intermix sensory systems in ways that we are usually unaware of.

Smell and taste mix synaesthetically in all of us, as we notice when a clogged nose kills flavours, but they interact with sight, too. For example, mint-flavoured water smells stronger if it is coloured yellow rather than clear. The vestibular system also interacts synaesthetically with sight, so that it is difficult to balance on one foot with both eyes closed. The sense of gravity interacts with sight as well: when standing up, a vertical line looks more attractive than a horizontal line, yet we see no difference between them when lying down rolled over on one side.

Touch and seeing also interact. The rubber-hand illusion shows this nicely. Imagine you are sitting at a table with one of your hands hidden beneath it, while a rubber hand is sitting above it on the table. If an experimenter strokes your visible hand with a brush, and also the rubber hand, you will feel that your hand beneath the table is being brushed. (See Supplementary Figure H.) Some people also feel touched while watching somebody else being touched.

Some adults hear music in colour, with idiosyncratic but specific colours evoked by specific pitches and timbres. Others do not perceive those colours, yet when asked to match colour samples to pitches, they do tend to show some
comparable correspondences: systematic variations of chroma with timbre, and higher brightness corresponding to higher pitch.\textsuperscript{67}

People usually think of these sensory intermixtures as noisy imperfections of a compartmentalized brain, but this is wrong. The intermixtures do not add noise to perceptual signals, they clarify signals. No sensory system can respond to everything all at once, so every sensory system must sample the information available to it. Much if not most of what we perceive, we do not sense directly, we infer from experience. Cars seldom turn into your street, and you saw no car a few seconds ago, so you perceive it safe to cross. This is an intuitive calculation of a Bayesian probability: few-cars-ever times no-car-seconds-ago equals a low probability of danger. Intermixing sensory dimensions provides additional information to be multiplied this way into perceptions. (See Figure 4) Multiplying probabilities facilitates forming accurate perceptual prototypes, so a side-effect of multimodal perception is helping to form aesthetic preferences and pleasure.\textsuperscript{2}

\begin{center}
\textbf{Multimodality: Music and Dance}
\end{center}

Listening to music, like all aesthetic experiences, is also multimodal, even for people who do not hear music in colour. You may not tap a foot while listening yet you are still likely to synchronize micro-movements to the beat,\textsuperscript{68} and of course you cannot make music of any kind without moving to the music in some way. Moving the whole body to music we usually call dancing, but movement and music are so intertwined that many languages do not distinguish between them—nor does English always do so. “A minuet” or “a waltz” or “a tango” can refer to the music, the movement, or the music and movement combined.

Moving with sounds begins in infancy. Even if there is no regular beat, just the rhythmic lilt of a mother speaking to her baby, still newborns will match their micro-movements to the ebb and flow of phonemes.\textsuperscript{69,70,71}

Beginning as early as 7 months, moving the head helps us to hear rhythms. Phillips-Silver and Trainor\textsuperscript{72} played a drum beat that could be inter-
preted as either in 2 (a march) or in 3 (a waltz). Simultaneously they bounced babies in time with every second beat or with every third beat. Afterwards they played the drumming through two different loudspeakers, accented as a march in one and a waltz in the other. The babies turned toward the sound that matched their bouncing. Follow-up experiments in adults teased apart the effect to find that in this situation, the vestibular system is key.\textsuperscript{73,74}

Rhythmic movement also affects vision and memory: we notice and remember more about people who move with us in synchrony. This comes out in an experiment by Mathew Woolhouse.\textsuperscript{75} Undergraduates danced in a room 10 at a time. To mix them up in a controlled way, they moved through a randomized set of assigned spots on the floor, dancing an equal time on each. Each dancer wore a sash of a different colour, and half of the sashes were decorated with a black cat. The music came through headphones and each dancer heard one of two different songs, one fast and the other slow. Afterwards a surprise quiz asked each dancer which person wore which sash. They remembered better than chance only when the other person had been dancing to the same song.

Simultaneous movement not only enhances one person’s perception of another, it can tie people together perceptually. One variation of the rubber hand illusion has somebody rhythmically open and close his hidden hand while watching an experimenter open and close her hand either at the same frequency or randomly. When they do this at the same frequency, most participants report themselves “causing” or “controlling” the movement of the experimenter’s hand.\textsuperscript{76}

While you dance, watching other people dance to the same beat strengthens the perceptual prototype of that dance, and unless you are an expert dancing with amateurs, it will also show you some pleasant variations. This generates aesthetic pleasure. On the other hand, if you are an expert dancing with amateurs, the other dancers will demonstrate unpleasant variations. When Mozart was 21, he wrote of a wedding, “There was dancing, but I only danced four minuets, and was in my own room again by eleven o’clock, for, out of fifty young ladies, there was only one who danced in time”.\textsuperscript{77}

We understand little of the physiological processes that we feel as aesthetic pleasures, but we know that dancing can involve the endogenous opioid system, a system that also appears to generate “runner’s high”.\textsuperscript{78} On the other hand, pleasurable dancing clearly need not generate runner’s high. For 150 years the most popular ballroom dances were minuets. These were not energetic—they were not the quick waltzes in three that most musicians play them as now—they were stately dances in six with complex but gentle evolutions and steps.\textsuperscript{79} It seems that opioid response is likely one physiological element that can be involved in aesthetic pleasures, but not necessarily. Physiological systems are so complex that other elements are surely involved as well.

The pleasure we take in dancing seems particularly germane to human evolution. It has become a truism that mankind developed so far technologically
and intellectually partly because our ancestors evolved the neurological means
to work together efficiently. Before industrialization, working together often
meant moving together, moving together to till a field, to harvest a crop, or to
paddle a canoe. Moving together forms a perceptual prototype of movement
comparable to our prototypes of sights and sounds, so efficient partners feel that
they work together beautifully. Indeed, that is one reason they are efficient:
their efficiency and beauty form a virtuous circle. They come to feel, synestheti-
cally, their partner’s movement as involved with their own.

These effects have had a major impact on the history of the world. Ancient
Romans discovered that drilling soldiers—moving them synchronously in time—
turns a mob of individual fighters into unified army with greater power to kill.
That is why infantrymen are marched and drilled to chants and drums.80

A more beneficent form of moving together in time is dancing. Today in
Europe and North America, few people dance, but before television took over
brains and lives, dancing was a more common way to socialize. (In 1938, a
Gallup poll showed that 13% of Americans chose dancing as their favourite way
to spend a evening; in 1986 only 3% did, and 33% chose watching television.81)
Wealthy people used frequently to attend fancy balls and ordinary people went
to dance halls82 or just cleared away the furniture in someone’s home. (See
Supplementary Figure I.) Dancing helped bind individuals into couples and
couples into social classes. A form of dance even helped bind together clergy and
their congregations: processionals. Dancing was so commonplace that if you went
to the cinema in1929, you might have seen the latest step demonstrated on the
weekly newsreel.83

In short, dancing and music—they are two sides of one coin—are not a
societal side-show, they are a human universal. In different societies they serve
different purposes, and they seldom look and sound similar, but they have been
a fundamental factor in human evolution. Dancing, like music, creates pre-
dictable patterns, prototypical patterns that become attractive as an average
face becomes attractive.84

Conclusion

Our ancestors needed to notice the nose of a crocodile, the rumbling of an
elephant, and the slightly off flavours of food starting to spoil. This required
adapting to normal circumstances so that unusual stimulation stood out. This is
a basic adaptation.

That is the fundamental mechanism behind aesthetics. When something
familiar is encountered, previous experience with it may make it comforting or
worrisome. When something slightly different is encountered, those perceptions
amplify. Some of the amplifications we feel as aesthetic preferences. Because we
each have different developmental histories, individuals will react differently in
different circumstances, but the underlying mechanism will always be the same.
Although many of our examples are from the WEIRD world (western, educated, industrial, rich, and democratic), the mechanisms described are universally human. People growing up in different cultures can have radically different experiences, so they can develop radically different aesthetic preferences, but they develop their preferences the same way.

Indeed, the mechanism applies to non-human animals as well as humans. Other adaptive animals have similar requirements to be especially sensitive to small deviations from normality, so we would expect other animals to develop aesthetic preferences too; and so they do. For example, primatologist Anne Russon observed, “Orangutans are all so meticulous. I remember one...was sitting there and picked up sticks from the ground and very systematically broke them all down so that they were about the same length and then laid them all out parallel to each other. They like things to be neat”.

Nor do aesthetic judgments seem restricted to primates. Any animal will prefer one food to another, and watch a dog sniff out new terrain: how fast its tail wags will vary from one spot to another.

Finally, consider Lucy, a chimpanzee raised by Maurice and Jane Temerlin as though she were their human daughter. Maurice wrote:

Lucy loves to dress up. She will adorn herself with a blanket or any interesting article of our clothing, wrap it around herself, take it to a mirror and watch herself in the mirror as she preens and postures in Jane's scarf or blouse, or my hat or shirt.

Roger Fouts used to tutor her at her home in American Sign Language. He wrote:

At 8:30 a.m....Lucy, would greet me at the front door, give me a hug, and show me into the house. While I sat in the kitchen, six-year-old [pubescent] Lucy would go to the stove, grab the teakettle, and fill it with water from the kitchen sink. She did all this chimpanzee-style, by jumping from counter to counter. After getting two cups and tea bags out of the cupboard, she would brew the tea and serve it like a perfect hostess. Then her ASL lesson would begin. It was all very civilized.

Lucy makes clear that, to the extent their anatomy and experience are similar, animals can develop aesthetic preferences much as we do. Other animals are built differently than we and experience the world differently, so they develop different expectations. Lucy’s tea service was suitable for her but not for the Queen. Nevertheless, their neuronal mechanism will have been comparable.
Or consider another example. Just as babies can learn to like carrots from their mother’s amniotic fluid and milk, so kittens can learn to like bananas. Cats are born without taste buds that react to sugar, so no cat will develop a sweet tooth. Cats enjoy meat and dislike bananas. However, if a pregnant cat is forced to eat a diet of bananas, her kittens will prefer bananas to meat.  

Since at least the days of ancient Greece, philosophers, artists and poets have argued about aesthetic endeavours, each claiming that one form of art is more sublime than another. We would like to have seen such a discussion over tea served by Lucy in her robe du jour, with a kitten nibbling bananas by their feet. Empirical evidence shows that aesthetic preferences are neither sublime nor even uniquely human. Aesthetic preferences flow from the basic neurophysiology of brains. They form a general-purpose mechanism that helps an animal adapt to its environment. This mechanism may cause a person to experience some songs and paintings as glorious and others as grotesque, but those experiences are incidental. The origin and function of a brain’s aesthetic mechanism is to help the body to survive.


This is the published file of supplementary figures, which identifies them by number. Since the manuscript above refers to them by letter, we have modified the captions to match.
Figure A

**Figure 1.** Moving the features of an individual face closer to their average positions (right) makes the face more attractive than moving the features the same distance away from average (left).

Figure B

**Figure 2.** Newborns look longer at the “face” on the left because its layout mirrors the shape of the paddle, which mirrors in turn the shape of the baby’s visual field.

Figure C

**Figure 3.** If the sun is behind you, the odds are even that any one object will be lighter or darker than an object behind it. However, the sun will seldom be behind you. If it is to the top, side or back, then it will illuminate directly little of what you look at, leaving most things illuminated indirectly —i.e., in shadow. For example, in this photograph the sun is on the left and illuminates directly the left side of the bird and branch. From every other direction the sun illuminates the objects indirectly, by reflecting off the surroundings—in this case by reflecting off the sky. ©Charles Maurer
Figure 4. Since darks against lights are a statistical regularity of the environment, the nervous system has evolved to work with them more readily than lights against darks. Adults see darks against lights more quickly than the opposite, and when 3-month-old babies are shown a pattern like this one, they look toward the dark bars first (Dannemiller JL, Stephens BR. 2001 Asymmetries in contrast polarity processing in young human infants. J. Vis. 1, 112–125.).

Figure 5. The middle photo shows the model’s features in an average position, the others show them lowered and raised 2 standard deviations. Most adults prefer the middle faces, tall adults prefer faces more like the left one—they are used to looking down on people and seeing more forehead and less chin—but babies, who look upward, prefer the right (except for 3-year-olds who have been going to day-care: they are used to seeing other children en face, so they prefer the middle photo).

Figure 6. The lines show the mean rating of attractiveness (±1 standard error) for male and female faces containing different levels of bias. The average of 60 faces is bias 0. The average of the 15 most attractive faces is more attractive, so we call it a unit of bias (bias = 1, described by the yellow bar). Adults find faces with 1-2 units of bias to be the most attractive.
Figure 7. Portions of life-sized paintings of Venus by Peter Paul Rubens (left) and William-Adolphe Bouguereau show ideal European body shapes in 1635 and 1879. *Venus and Adonis* courtesy Metropolitan Museum of Art, *Birth of Venus* courtesy Musée d’Orsay.

Figure 8. The subject (in blue) is holding his right hand beneath the table, yet when the experimenter strokes both the subject’s left hand and a rubber hand, the subject feels both of his hands being stroked. Adapted from a drawing by Norimichi Kitagawa in Norimichi Kitagawa, "Link Between Hearing and Bodily Sensations." *NTT Technical Review* 11, no. 12 (2013).
Figure 9. *Country Dance* (1883) by Martin Edgar Ferrill, courtesy (U.S.) National Gallery of Art.

Figure I