Synergistic Combination of Visual Features in Vision–Taste Crossmodal Correspondences

Byron P. Lee* and Charles Spence

Crossmodal Research Laboratory, University of Oxford, Oxford, OX1 2JD, UK

*Corresponding author; e-mail: byron.lee@psy.ox.ac.uk

ORCID iDs: Lee: 0000-0002-3773-4513; Spence: 0000-0003-2111-072X

Received 15 March 2023; accepted 20 July 2023; published online 14 August 2023

Abstract
There has been a rapid recent growth in academic attempts to summarise, understand, and predict the taste profile matching complex images that incorporate multiple visual design features. While there is now ample research to document the patterns of vision–taste correspondences involving individual visual features (such as colour and shape curvilinearity in isolation), little is known about the taste associations that may be primed when multiple visual features are presented simultaneously. This narrative historical review therefore presents an overview of the research that has examined, or provided insights into, the interaction of graphic elements in taste correspondences involving colour, shape attributes, texture, and other visual features. The empirical evidence is largely in line with the predictions derived from the proposed theories concerning the origins of crossmodal correspondences; the component features of a visual stimulus are observed to contribute substantially to its taste expectations. However, the taste associated with a visual stimulus may sometimes deviate from the taste correspondences primed by its constituent parts. This may occur when a new semantic meaning emerges as multiple features are displayed together. Some visual features may even provide contextual cues for observers, thus altering the gustatory information that they associate with an image. A theoretical framework is constructed to help more intuitively predict and conceptualise the overall influence on taste correspondences when visual features are processed together as a combined image.

Keywords
crossmodal correspondences, taste, visual features, interaction, synergy
1. Introduction

Crossmodal correspondences refer to people’s tendency to voluntarily associate sensory attributes in one modality with those in another modality. The existence and ubiquity of such connections between the senses has fascinated both researchers and the consumer industry because of their theoretical and practical implications (Spence, 2011; Spence and Deroy, 2013). The vast literature on vision–taste crossmodal correspondences now provides compelling evidence for a robust, and sometimes surprising, tendency of human observers to associate visual features or attributes with gustatory qualities (see Lee and Spence, 2022, for a review). When questioned, people tend to connect specific taste qualities with certain visual properties, such as colour attributes (Spence et al., 2015), geometric features (Velasco et al., 2016a), and visual textures (Barbosa Escobar et al., 2022) in systematic ways that are also generally consistent (or consensual, given that there is no objectively ‘correct’ answer; Koriat, 2008). Perhaps more surprisingly, presenting these visual stimuli has, on occasion at least, been shown to influence people’s taste expectations (Saluja and Stevenson, 2018; Velasco et al., 2018a) and sometimes even their taste/flavour experiences after tasting the food accompanying or decorated by the said features (Rolschau et al., 2020). With these well-documented implications in mind, designers and researchers alike have understandably become increasingly interested in harnessing the effects of crossmodal correspondences in the development of more engaging consumer/user experiences (Elliott, 2012; Piqueras-Fiszman et al., 2012; Spence and Van Doorn, 2022; Velasco and Spence, 2019).

In the case of food and beverage packaging (and indeed any other forms of visual presentation in product and service delivery), design solutions can convey meanings by deliberately configuring visual elements such as colour, shape, and texture (Gómez et al., 2015; Underwood and Klein, 2002). Considering how graphic elements are typically intertwined when presented, it is presumably in the interests of marketing practitioners and those researchers studying crossmodal correspondences alike to try and better understand how multiple visual features collectively influence taste expectations/experiences. Currently, a substantial body of empirical research has documented the bidirectional matches that exist between taste (gustation) and specific visual features, such as the hue–taste (see Spence and Levitan, 2021, for a review), curvilinearity–taste (Spence and Ngo, 2012), and visual texture–taste correspondences (Barbosa Escobar et al., 2022). However, relatively few studies have attempted to systematically examine the combined influence of individual visual elements, which requires an experimental paradigm that simultaneously manipulates more than one visual stimulus attribute. In this review, we follow the evidence available to explore the dynamics of co-presented features.
in vision–taste correspondences (e.g., Salgado-Montejo et al., 2015; Wang et al., 2019).

The conflicts and synergies between different visual features are inevitable topics for the researchers studying vision–taste correspondences. Regrettably, the majority of the published literature on crossmodal correspondences has examined the associations of taste qualities with visual features such as colour and shape individually (i.e., by explicitly manipulating only a single feature at a time — see Spence, 2019; Velasco et al., 2016b, for reviews). The simultaneous manipulation of multiple visual features could hypothetically have enabled the researchers concerned to examine the respective roles and collective effect of different visual features in the process of priming taste information. A growing trend amongst researchers has been to engage the various visual attributes in a stimulus as a coherent, integrated ensemble (e.g., Liang et al., 2021; van Rompay et al., 2018; H. Wang et al., 2022 — see Moutoussis, 2015; Spence, 2015, for a review). Such an approach can address some of the issues in the previous studies, for example: (1) the observers may be voluntarily or involuntarily combining all accessible visual cues when establishing a connection with specific gustatory qualities (Alais et al., 2010; Deroy and Spence, 2016; Treisman and Gelade, 1980); (2) a typical scene/image/object almost always incorporates visual attributes of what people describe as ‘different dimensions’. Consider, for instance, the paradox of having an object without any colour or having a surface with no texture (Grossberg, 1984). By understanding the collective and respective contributions of visual features from different dimensions in the vision–taste correspondences, designers will be better informed to coordinate the graphic elements in their work when attempting to deliver or shape–taste expectations (e.g., Chitturi et al., 2019; Fairhurst et al., 2015). Ultimately, examining the interaction of different visual design elements may help to shed light on the hierarchy of visual cues when priming the associated gustatory qualities (Lee and Spence, 2022; Spence and Levitan, 2021).

Considering the simultaneous effects of different elements incorporated in an image, a paradigm that involves the manipulation of multiple visual features could better reflect real-world scenarios, thus offering more realistic approximations of the vision–taste correspondences in practical contexts (Motoki and Velasco, 2021). For example, when asking participants to match tastes with a given shape, Salgado-Montejo et al. (2015) manipulated the curvilinearity, symmetry, and segment complexity of the shapes presented to participants (see Fig. 1). Such an experimental design enabled the researchers concerned to demonstrate several interaction effects between the attributes of different visual dimensions.

Due to the scarcity of empirical evidence concerning the taste expectations that may be associated with combined visual design features, we also reference
the theories that have been put forward to try and explain crossmodal correspondences (specifically, those concerning the origins of intermodal relations) and attempt to apply them when predicting the taste correspondences associated with more complex images. Intriguingly, an emerging body of empirical research has now started to uncover the reasons behind the connections that people make between taste and specific visual design features (see Spence and Levitan, 2021, for a review). For example, some theories have highlighted the role of the connotative and semantic meaning that may be attached to the visual features (Spence and Van Doorn, 2022), such as the values and personalities that people deem as attached to specific colours (e.g., green is perceived as associated with the concepts of calming, nature, and healthiness; Clarke and Costall, 2008; Kunz et al., 2020). Meanwhile, certain associations would appear to reflect the emotions triggered by innate responses (Gómez-Puerto et al., 2016; LoBue, 2014; Nairne et al., 2009) that can nevertheless be influenced by contextual cues (Elliot and Maier, 2012; Motoki and Velasco, 2021). Additionally, many associations are believed to result from the statistical correlations of objects/concepts encountered in the environment at large (Schloss et al., 2018; Spence, 2021). Take, for instance, the highly consensual associations of red with sweet and green with sour, which may be putatively attributed to the regular pairing of these sensory attributes in experiences with fruits.
taste mappings of stimuli incorporating multiple visual design features could ideally reflect the competition between different underlying factors accounting for vision–taste associations when multiple visual features are present.

This narrative historical review (see Furley and Goldschmied, 2021, for a discussion on this style of review) focuses on the gustatory information that is primed, or modulated, by presenting graphic elements from different visual dimensions. We draw upon the empirical evidence to review vision–taste correspondences involving one or more visual attributes (e.g., hue, curvilinearity, symmetry, texture, etc.) and assess the combined effect of multiple visual features when they happen to be matched with taste qualities. The review focuses specifically on the vision–taste matching studies that have probed the effects of interactions between visual features (e.g., between two different colours, Woods et al., 2016 — or between colour and shape; Stewart and Goss, 2013).

Based on the theories accounting for the origins of crossmodal correspondences, a theoretical framework is proposed as a more intuitive approach to conceptualise the interactions between visual features, grouping the latter by the mediator/basis of correspondence instead of by the visual dimensions involved. In relation to the Gestalt experience (Köhler, 1929; Wagemans et al., 2012), we also explore the idea of having new connotative, semantic, and/or semiotic meanings emerge when visual features are displayed together (Spence and Van Doorn, 2022). It is suggested that the new meanings attached to a compound image could also modify or even alter the taste expectations that would otherwise be determined by the constituting features.

2. An Overview of Vision–Taste Correspondences

The assessment of interactions between visual features across different dimensions necessitates a clear definition of visual dimensions, which may be regarded as an umbrella term that encapsulates elements such as colour, curvilinearity, symmetry, geometric positioning, and visual texture. At times, such an approach to categorising visual features is inevitably (and intentionally) liberal and ambiguous. For example, think only about how visual texture is derived from more basic visual features, often seen as a product of contour, orientation, lightness, and depth working together to create the impression of a surface (Caelli, 1985; Norcia et al., 2005). For the purpose of the current review, two criteria were considered as defining a visual dimension: (1) if the neural signals for processing a group of visual features involve dedicated pathways in the brain, such as hue, curvilinearity, and symmetry (Desimone et al., 1985; Goodale et al., 1994); (2) if there exist a distinct pattern of crossmodal correspondences between a type/group of visual features and taste qualities. For example, visual textures are shown to match taste qualities in a way that is noticeably independent of their finer, constituting geometric properties (e.g.,
Di Stefano and Spence, 2022 — see Spence, 2022a, for a review on different kinds of crossmodal correspondence). Contrary to visual textures, typeface–taste correspondences appear to be predominantly determined by stroke curvilinearity, which is why the typeface effects are generally regarded as a mere extension of curvilinearity effects but not an independent dimension in terms of vision–taste correspondences.

Over the past few decades, a high degree of consensuality has been observed across a wide range of studies that have attempted to document the taste qualities that are matched to colour, curvilinearity, symmetry, and visual texture. The taste qualities in this case are the four or five basic tastes of sweet, sour, salty, bitter, and umami (Halpern, 2002; McBurney and Gent, 1979; Trivedi, 2012 — though it should be noted that some researchers have questioned the validity of the definition of ‘basic taste’: Beauchamp, 2019). People of different ages and cultural backgrounds seem to broadly agree on what visual features match each of the basic taste qualities (see Table 1). Notably, the visual associations for the quality known as ‘umami’ or ‘savoury’ have not been conclusively investigated among the relevant literature, potentially due to

Table 1.
Visual features from different dimensions associated with each of the four basic tastes

<table>
<thead>
<tr>
<th>Study</th>
<th>Visual dimension</th>
<th>Sweet</th>
<th>Sour</th>
<th>Salty</th>
<th>Bitter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Déribére (1978)</td>
<td>Colour (hue)</td>
<td>Red</td>
<td>Yellow</td>
<td>Blue</td>
<td>Brown</td>
</tr>
<tr>
<td></td>
<td></td>
<td>White</td>
<td>Green</td>
<td>White</td>
<td>Green</td>
</tr>
<tr>
<td>O’Mahony (1983)</td>
<td>Colour (hue)</td>
<td>Red</td>
<td>Yellow</td>
<td>White</td>
<td></td>
</tr>
<tr>
<td>Raevskiy et al. (2022)</td>
<td>Colour (hue)</td>
<td>Pink</td>
<td>Yellow</td>
<td>White</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Light red</td>
<td>Green</td>
<td>Blue</td>
<td></td>
</tr>
<tr>
<td>Saluja and Stevenson (2018)</td>
<td>Colour (hue)</td>
<td>Pink/Red</td>
<td>Yellow</td>
<td>Blue</td>
<td></td>
</tr>
<tr>
<td>Wan et al. (2014)</td>
<td>Colour (hue)</td>
<td>Pink</td>
<td>Green</td>
<td>White</td>
<td>Black</td>
</tr>
<tr>
<td>Motoki and Velasco (2021)</td>
<td>Curvilinearity</td>
<td>Rounded</td>
<td>Angular</td>
<td>Angular</td>
<td>Angular</td>
</tr>
<tr>
<td>Salgado-Montejo et al. (2015)</td>
<td>Curvilinearity</td>
<td>Rounded</td>
<td>Symmetrical</td>
<td>Asymmetrical</td>
<td>Complex</td>
</tr>
<tr>
<td></td>
<td>Symmetry</td>
<td>Simple</td>
<td>Asymmetrical</td>
<td>Simple</td>
<td>Complex</td>
</tr>
<tr>
<td>Barbosa Escobar et al. (2022)</td>
<td>Visual texture</td>
<td>Fluffy</td>
<td>Crunchy</td>
<td>Rough</td>
<td></td>
</tr>
</tbody>
</table>
the limited familiarity in non-Asian cultures (Cecchini et al., 2019; Raevskiy et al., 2022). This section therefore provides an overview of the vision–taste associations that have been documented by researchers who have assessed the taste correspondences of visual features from each dimension.

2.1. Taste Associated With Different Constituent Attributes of Colour

Colour (hue) is historically one of the first visual features to have been investigated in terms of crossmodal correspondences involving visual features and taste qualities (Déribéré, 1978 — see Lee and Spence, 2022). Researchers (and practitioners) have long been interested in the links between hue and basic taste (e.g., Déribéré, 1978; Favre and November, 1979; Guinard et al., 1996; O’Mahony, 1983). As presented in Table 1, the field of crossmodal correspondences research has now established a fairly extensive record of crossmodal mappings between colour hues and basic taste qualities. It is worth noting that the associations between taste qualities and colour hues have been investigated using a diverse range of experimental paradigms (see Lee and Spence, 2022; Spence, 2021). These range from O’Mahony’s (1983) study, which examined within-participant consistency in the pairing of colour and taste words, through Koch and Koch’s (2003) exploration of the expected taste potency associated with specific colours, to the open-ended approach used by Saluja and Stevenson (2018), in which participants tasted pure solutions of each basic taste and were allowed to pick any possible colour from a colour wheel that they felt best matched the given taste.

It should be recognised that while the most fundamental component of gustatory experience is likely its quality (i.e., the basic categories such as sweet and bitter, which is supported by subjective reports and physiological evidence; Chaudhari and Roper, 2010; Yarmolinsky et al., 2009), the perception of taste also involves attributes such as intensity, duration, and the variation/trajectory of these attributes before, during, and after consumption/tasting (Kelling and Halpern, 1983; Lenfant et al., 2009). Similarly, colour experiences can also be broken down into constituent attributes and examined in terms of chroma (hue) and saliency. The term ‘colour saliency’, as used here, refers to the prothetic dimensions of colour appearance, such as saturation and brightness (Panek and Stevens, 1966). Although taste/flavour intensity is speculated to correspond with colour saliency (e.g., Knöferle and Spence, 2012; Shermer and Levitan, 2014 — cf. Guinard et al., 1996; Ueda et al., 2020; Wang et al., 2022), the taste associations of colour saturation and brightness have not been studied as extensively as the taste associations of colour hue (Spence et al., 2015).

Other attributes of taste (gustation), such as the perceptual differences induced by different chemical compounds (McDowell, 2017; Wilk et al., 2022), could also play a role in colour–taste associations. For example,
Guinard et al. (1996) noted that their participants used more than one colour to describe those chemical compounds that elicited multiple taste qualities, such as the bittersweet-tasting saccharin. People also appear capable of consistently designating specific colours to different types of sweetening agents, but they do not associate different colours with different bitterants (Higgins and Hayes, 2019; Wardle et al., 2007). Interestingly, it happens to be common practice amongst marketers to colour code the packages of artificial sweeteners based on their chemical constituents (Builder, 2018; Elliott, 2012; Shmerling, 2019; Wardy et al., 2017). Perhaps as a result of repeated exposure, specific colour hues may come to cue the presence of particular subqualities of taste (e.g., Simmonds and Spence, 2019), which, in this case, would be determined by the onset, intensity, duration, and extinction of the sweet sensation created by different types of sweetener (Ketelsen et al., 1993; Shallenberger, 1993, pp. 5–46; Stuckey, 2012, p. 216; Walsh et al., 2014).

### 2.2. Tastes Associated With Geometric Features

The perception of visual forms is constructed by the human brain from retinotopic signals such as segment, contour, contrast, connection, depth, size, and other elements (Loffler, 2008). Amongst a number of geometric features, curvilinearity (roundedness/angularity) has been shown to be an influential feature of vision–taste crossmodal correspondences (Velasco et al., 2016a). Researchers have reported that taste expectations can be influenced by the curvilinearity of abstract shape (Velasco et al., 2015a), package elements (Motoki and Velasco, 2021), plate shape (Fairhurst et al., 2015), typeface stems (Velasco et al., 2014a), and even the shape of the food itself (Baptista et al., 2022; Wang et al., 2017). However, the taste associations of curvilinearity would appear to involve less mapping variety than those of colour: that is, rounded shapes are matched with sweet taste and angular shapes with all other basic taste qualities (Velasco et al., 2015a — see also Lee and Spence, 2022). Although some three-dimensional objects can be exclusively matched to a non-sweet quality (Isbister et al., 2006; Obrist et al., 2014 — cf. Gottlieb et al., 2007; Juravle et al., 2022), the mapping patterns do not appear to be as consensual as when matching taste with two-dimensional (2D) angularity/roundness (Cytowic and Wood, 1982; Deroy and Valentin, 2011).

In addition to curvilinearity, several other geometric properties have been found to be systematically associated with taste qualities (see Table 2). Symmetry and segmentation complexity (i.e., the number of points in a concave polygon), for example, were confirmed by Salgado-Montejo et al. (2015) to be consensually mapped in terms of taste correspondences. That is, people from different cultures associate symmetric, simple shapes with a sweet taste, while asymmetric, complex shapes were associated with non-sweet tastes (e.g., sour and bitter) instead. There is also weak evidence for associations between taste
Table 2.
An overview of the associations established between taste qualities and various geometric properties other than curvilinearity

<table>
<thead>
<tr>
<th>Geometric properties</th>
<th>Study</th>
<th>Taste quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segmentation complexity</td>
<td>Salgado-Montejo et al. (2015)</td>
<td>Simple Complex</td>
</tr>
<tr>
<td>Symmetry</td>
<td>Salgado-Montejo et al. (2015)</td>
<td>Symmetrical Asymmetrical</td>
</tr>
<tr>
<td>Reflectional and rotational symmetry</td>
<td>Turoman et al. (2018)</td>
<td>Symmetrical Asymmetrical Asymmetrical</td>
</tr>
<tr>
<td>Volume (of 3D objects)</td>
<td>Deroy and Valentin (2011)</td>
<td>Voluminousness (larger volume)</td>
</tr>
<tr>
<td>Vertical location</td>
<td>Velasco et al. (2018b)</td>
<td>Relatively higher Relatively lower</td>
</tr>
</tbody>
</table>

and vertical locations. When mapping taste qualities in vertical space, people tend to place sweet higher than bitter (Velasco et al., 2018b). Meanwhile, Deroy and Valentin (2011) documented an association between voluminousness and sweetness when matching a set of 2D and 3D images with three beers that had different flavour profiles. However, to date, not much is known about the correspondences between taste and size, as evidence remains scarce for how size would be associated with specific taste qualities (cf. Woods et al., 2013). As a side note, there is a fair prospect that the temporal profile of each taste quality can be represented by (and thus connected to) various geometric shapes. For example, the sweetness induced by tasting sucrose tends to build up gradually and dissipate slowly, and sour taste typically exhibits a short but intense pulse with rapid onset and sudden finish (Obrist et al., 2014). Although each taste quality was perceived in a unique pattern of temporal characteristics, no research has looked into this link as a potential origin of crossmodal correspondence thus far.

2.3. Tastes Associated With Visual Textures

Visual texture, at least the graphical features that represent the tactile impression of a surface (Whitaker et al., 2008), may be regarded as a somewhat peculiar visual attribute (Di Stefano and Spence, 2022; Henson et al., 2006; Picard, 2007). When inspected closely, what people perceive as visual textures are essentially repeating patterns of contour, shadow, and lighting arranged in a manner that can create specific impressions about the surface properties (Groissboeck et al., 2010). In this respect, it would be reasonable to expect
Figure 2. The visual texture stimuli used by (a–c) Barbosa Escobar et al. (2022), (d, e) Wan et al. (2014), and (f, g) Spence et al. (2019) that were reported to be associated consensually with basic taste qualities. In Barbosa Escobar et al.’s (2022) study, the visual texture designated as (a) was found to associate with sweetness; while the visual textures indexed (b) and (c) were found to associate with saltiness. In Wan et al.’s (2014) study, two visual patterns of different roughness levels (d, e) were associated with saltiness. The findings present a unique mapping between visual texture and taste quality that agrees with the effect of tactile texture on taste/flavour (Biggs et al., 2016; van Rompay and Groothedde, 2019). Among the various taste qualities that were examined by Barbosa Escobar et al. (2022) and Wan et al. (2014), only sweet and salty appeared to be systematically associated with some of the visual textures presented in their studies. Similarly, when Spence et al. (2019; N = 339) investigated the taste mappings of four differently-textured amuses bouche (each designed to cue one of the four basic taste qualities, in part based on prior research by Turoman et al., 2018), only sweet and salty were reported to have a notable association with the textures/food forms that were presented. Spence et al. (2019) documented consensual crossmodal associations between sweet taste and candy floss texture (Fig. 2f; n = 276; 81% of all participants questioned), even more so
than the element resembling a sugar cube \((n = 250; 74\% \text{ of all participants questioned})\). The only other taste quality that had a close-to-majority correspondence with any visual texture was saltiness, which was associated with a ‘web-clotted’, asymmetric, triangular piece of thin starch (Fig. 2g; \(n = 148; 44\% \text{ of all participants questioned}\). As Spence et al. (2019) noted, further research efforts are still needed to help identify those textures that can be more consensually associated with sourness, bitterness, and saltiness.

It is also worth pointing out that many stimuli used in the texture–taste matching studies were not prepared or administered in a consistent manner. That is, there is no clear definition concerning what amounts to the repeating patterns of a texture and what might be just an extraction of an object/scene with repetitive visual features (though see Cimpoi et al., 2014, for a project to develop a universal dataset for identifying textual attributes with image and vocabulary). Consider here only how the ‘fluffy’ texture (see Fig. 2a) used in Barbosa Escobar et al.’s (2022) study could very well be seen as an image of a cluster of bubbles; the latter would likely carry much more semantic weight, or meaning, than the feeling of fluffy (be it visual or tactile). Relevant here, there is an ongoing discussion on the labels used to describe textures, such as the confusion surrounding the term ‘roughness’ (Di Stefano and Spence, 2022). Although researchers have yet to agree on which level of visual texture should be investigated, the empirical evidence suggests both the properties (e.g., roughness and blurriness) and specific patterns (e.g., stripes, cracks and waves) of visual texture to be influential when enabling crossmodal associations (Barbosa Escobar et al., 2022; Wan et al., 2014).

3. Predicting Interactions Based on Theories Concerning the Origins of Crossmodal Correspondences

In recent years, the field of crossmodal correspondences research has witnessed a noticeable growth of interest in the reasons that motivate (or cause) people to connect specific visual features with (gustatory) taste qualities (Spence and Levitan, 2021, 2022). As confirmed by how the taste matching patterns deviate (in terms of their diversity and consistency) from one type of visual feature to another (Velasco et al., 2015a), people appear to be guided by (or fall back on) different principles/justifications when associating taste qualities with visual features from the different dimensions. Among the various theories that have been discussed (Spence, 2011; Wang et al., 2016), the statistical and hedonic accounts stand out as being most promising, in that they have adequately explained a substantial proportion of the variance in the empirical data. As proposed in Spence’s (2022a) recent review, the crossmodal correspondences between different features can perhaps be fruitfully grouped by the laws ‘governing’ these intermodal relations (Walker-Andrews, 1994).
For those matches involving visual features and taste, the most relevant principle would appear to be the internalisation of the statistics in the environment and the mediation of emotions. The following section focuses on how these principles would operate together when visual features are combined. A novel theoretical framework is also proposed to help understand and thus predict the possible outcomes of the taste mappings when visual features are combined.

People’s experience of the world contributes to the belief that certain perceptual features should accompany, or predict, the presence of other features, such as their repeated encounters with physical laws and man-made products (Ernst, 2006; Marks, 1978, pp. 11–48; Spence, 2011, 2018; Walker-Andrews, 1994). Those researchers who have studied this principle have designated the term *statistical correspondences* when referring to those crossmodal associations that are thought to be acquired from the internalisation of the statistics of the general environment (Spence and Levitan, 2021). The term ‘internalisation’ describes the process of registering the regularities in the world, which are consequently used to inform how attributes are connected across the senses (Barlow, 2001; Ernst, 2007; Parise et al., 2014; Shankar et al., 2010). Some researchers have even reintroduced the idea that different types of crossmodal mapping can be organised and better understood by referring to what those regularities are based on (Walker-Andrews, 1994 — see Spence, 2022a, for a review).

Exploring the concept further through specific examples, one might consider the taste correspondences of colour and visual texture in the context of Walker-Andrews’s (1994) approach to intermodal relations. The taste correspondences of colour and visual texture may be based on the ‘arbitrary’ combinations of visual and gustatory features (Ernst, 2007; Spence, 2022a), particularly given the appearance of processed foods can theoretically take any colour. The idiosyncratic nature of arbitrary correspondences, or rather the fact that the source objects behind correspondences are ‘man-made’ (not just artificial, but as a result of human activities; Spence, 2021), could potentially explain why the vision–taste associations involving hue and visual texture are subject to a certain degree of cultural variation (e.g., Raevskiy et al., 2022; Wan et al., 2014). On the other hand, there may well be a systematic and predictable relationship between the chemical properties of the source object and the crossmodal correspondences derived. Ward et al. (2022) recently found the physicochemical features of olfactory stimuli to reliably correlate with the crossmodal associations made with shape curvilinearity, texture roughness, and, to a smaller extent, colour hue. At the same time, ripening fruits typically go through chemical changes that result in the appearance becoming warmer in tone (Foroni et al., 2016). These findings imply the supposedly arbitrary correspondences to not only be influenced by idiosyncratic, cultural differences,
but may also be subject to physicochemical, and potentially predictable, properties.

When used to explain the acquisition of crossmodal correspondences, the statistical account provides a model for ‘how’ and ‘why’ the connections are established but not ‘which’ connections are internalised. To a certain extent, the consensuality of crossmodal correspondences appears to correlate with the frequency of co-occurrence of the stimuli-in-association in the environment (Spence and Levitan, 2021). However, it remains undetermined what may drive people to internalise certain environmental regularities while pushing for the alternatives to be rejected. Current knowledge on this matter is insufficient to confidently answer why, for example, it should be saltiness that is matched to blue rather than, say, sweetness (as in the case of blueberries and energy drinks, for example, Spence, 2021 — cf. Shankar et al., 2010; Velasco et al., 2016c), or why sweetness is matched with the fluffy texture of cotton-like strands of floss instead of a cube made of coarse grains that resembles a sugar cube (e.g., Spence et al., 2019). Meanwhile, the statistical account currently fails to provide an explanation for why the regularities of most geometric features are not internalised and used as a basis for taste correspondences.

Curiously, when comparing the patterns of colour–taste associations across different cultures (Wan et al., 2014), there would seem to be occasional cases of variation supposedly linked to specific marketing conventions (Piquerias-Fiszman and Spence, 2011), agricultural practices (Osorio et al., 2016; Raevskiy et al., 2022), and wider human activities. While these cultural differences can be indicative of how certain associations are influenced by the environment, it requires further findings to suggest how internalised statistics would compete to drive people to make the documented connections (see also Ernst, 2006; Ho et al., 2014; Schoenlein and Schloss, 2022; Spence, 2011, for discussions on the Bayesian decision theory and other models that may help to explain crossmodal matches).

Contrary to those correspondences that are based on the internalisation of the statistics of the environment, some vision–taste associations have been shown to be mediated or facilitated by the emotions that are attached to the sensory stimuli. Here, the term affective (also known as emotionally mediated) correspondences has been used to describe those associations that are supposedly formed due to the corresponding stimuli happening to be individually associated with a similar emotional tone (Schifferstein and Tanudjaja, 2004; Spence, 2020, 2022b). The idea is that people tend to group concepts of similar hedonic value (or tone) together and draw associations accordingly, which is putatively why rounded contours are associated with ‘sweet’ while angular is associated with ‘non-sweet’ (Spence and Levitan, 2021; Velasco et al., 2016b). Interestingly, the shape correspondences involving salt are generally found to be less robust than other qualities, albeit a systematic association with angular
shapes can still be observed (Spence, 2023; Velasco et al., 2015a, 2016a). The peculiar pattern of preference for saltiness, which peaks at a certain range of perceived intensity (i.e., is a function of tastant concentration; Hayes et al., 2010), may confound the mediating role of emotion and thus contribute to the discounted consensuality documented in salt-shape correspondences (Wang et al., 2016). Yet, contrary to saltiness, despite the documented ‘sweet spot’ in taste intensity that correlates with a similar peak in hedonic liking (Cheung et al., 2022; Jayasinghe et al., 2017), the visual correspondences concerning sweetness are among the most consensual associations (Wan et al., 2014).

It has previously been suggested that emotional mediation may function as a secondary route when it comes to forming associations between sensory properties, stepping in when no other meaningful source object/concept is available (Lee and Spence, 2022; Spence, 2020; Wang et al., 2016 — see also Spence, 2022c, for perceptual similarity as an alternative route to matching stimuli). Indeed, emotion is observed to have contributed to the mappings of cross-modal correspondences involving more ‘abstract’ stimuli that lack access to internalised statistics or alternate routes to establish crossmodal connections, such as neural/structural similarity (e.g., Lindborg and Friberg, 2015; Velasco et al., 2016d — see Spence, 2020, for a review). If the visual features, either when combined or else assessed individually, fail to establish crossmodal correspondence with taste by connecting a source object/concept, it is worth entertaining the notion that people might fall back and rely on the hedonic value (e.g., Schifferstein and Tanudjaja, 2004 — see Lee and Spence, 2022) or perceptual similarities (see Spence, 2022c) of the stimuli and make their inferences accordingly.

The affective account is especially appealing to researchers when it comes to trying to explain the existence of consensual crossmodal correspondences between tastes and geometric features/properties, such as curvilinearity and symmetry (Deroy et al., 2013; Spence, 2022b; Spence and Levitan, 2021; Velasco et al., 2016d). That said, the emotion evoked by (or associated with) these geometrical features might have been the result of different mechanisms of feature perception and evaluation. For example, the preference for curvature (and thus for rounded shapes) might originate from the approach-avoidance decision when visually evaluating unfamiliar objects (Bar and Neta, 2006, 2007; Bertamini et al., 2016 — see Palumbo et al., 2015 for a review). In the context of taste quality, the motivation to approach (prefer) or avoid (dislike) may pertain to the need to determine the nutritious value or toxicity of foods (i.e., their palatability; Katz and Sadacca, 2011, p. 130). In the case of symmetry, the preference for symmetrical shapes could be explained by how
Figure 3. A possible hierarchy of how vision–taste correspondences may be established when people infer the gustatory information that seems most appropriate to associate with a given visual feature (or vice versa).
symmetry is considered to be a sign of health and normality (Shepherd and Bar, 2011), which respectively pertain to reproductive advantages (i.e., evolutionary fitness; Etcoff, 1955/2000; Thornhill and Gangestad, 1999 — see also Motoki et al., 2019) and processing fluency (Reber et al., 2004; Winkielman et al., 2006).

Based on our current understanding of the origins of intermodal relations, the mediators/bases (or what some called the ‘governing laws’; Walker-Andrews, 1994) underpinning crossmodal correspondences would determine how the patterns are established between taste and visual features. For the purpose of predicting the taste information that happens to be associated with a given visual feature, which dimension that feature belongs to may not be as important as how that feature comes to be involved in crossmodal connections. Imagine, for example, an asymmetric object featuring angular elements. Since both curvilinearity and symmetry are facilitated by emotions when people establish their taste correspondences, the collective taste profile of the said object can be explained and predicted using the same model. Thus, a more streamlined approach is suggested: as an alternative to referring to a visual dimension and defining interactions as either ‘between dimensions’ or ‘within dimension’, it would be more intuitive to conceptualise the dynamics of co-presented visual features by why/how their taste correspondences were formed when assessed in isolation. Table 3 presents an overview of possible modes of interaction, categorised by how the visual features involved in the interaction may be associated with taste qualities. Regardless of which visual dimension a feature belongs to, the taste correspondence of visual features involved in the interaction could be attributed to either the same or different putative cause(s).

3.1. Interactions of Vision–Taste Associations That May Be Attributed to the Statistics of the Environment

It is intriguing to envisage the combined effect on vision–taste associations when hue and visual texture are presented together, because both visual dimensions appear to be involved in the so-called arbitrary type of crossmodal correspondence (i.e., connections that are based on the statistics of the environment). There is a possibility that putting certain combinations of colour and texture together may create a new meaning in the mind of the observer, in such a way that people would associate the compounded image with a different taste that could not be implied unless both colour and texture are present. In the language of the statistical account, observers may rely on an otherwise omitted regularity to connect the compounded image with another taste quality (Barlow, 2001). Hypothetically, then, putting together the colour red (associated with sweet, and/or ripeness) and the texture of the membranes (also known as the pulp) in citrus fruit would create the taste expectation of ‘bitter and
<table>
<thead>
<tr>
<th>Modes of interaction</th>
<th>Examples of combined stimuli</th>
<th>Corresponding dimensions</th>
<th>(Possible) taste mappings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical only</td>
<td>Green foreground on red background</td>
<td>Colour–Colour</td>
<td>Presenting a coloured patch with green in front (as body) and red behind (as outline) led to an even stronger association with sour taste than displaying green or red colour alone (Woods et al., 2016).</td>
</tr>
<tr>
<td></td>
<td>Translucent liquid having a blue tint</td>
<td>Colour–Transparency</td>
<td>Blue is found to be the second sweetest colour when people assess the taste of liquid in different colours (Wan et al., 2014) — see Spence, 2021, for how recently such a link may have been established).</td>
</tr>
<tr>
<td>Affective only</td>
<td>Rounded typeface printed on curvy package</td>
<td>Curvilinearity–(Typeface)</td>
<td>Putting ‘sweet-tasting’ typeface and packaging shape together makes the product seem sweeter than the alternatives arrangements (Velasco et al., 2014a).</td>
</tr>
<tr>
<td></td>
<td>Angular and asymmetric shape</td>
<td>Curvilinearity–Symmetry</td>
<td>Shapes with angular and asymmetric features were regarded as the most unpleasant and ‘sour-tasting’ stimulus among other combinations of curvilinearity and symmetry features (Salgado-Montejo et al., 2015).</td>
</tr>
<tr>
<td>Statistical–affective</td>
<td>Black and square plateware</td>
<td>Colour–Curvilinearity</td>
<td>Food served on a square black plate was rated sweeter than on round black plate, but the food on a round white plate was rated to be sweeter than on a square white plate (Stewart and Goss, 2013).</td>
</tr>
</tbody>
</table>
sour’ due to the resemblance to grapefruit (Hayes et al., 2011). This speculated association could potentially be demonstrated by allowing assessors to report associations for each taste quality, rather than forcing them to choose the best match (e.g., O’Mahony, 1983; Wan et al., 2014). Similarly, since blue colourant is increasingly used in sweet sports beverages (Spence, 2021), any visual property cueing the context of translucent liquid may trigger the association with sweetness when configured together with blue (e.g., Wan et al., 2014, see also Lee and Spence, 2022).

Having established that people use the statistics of the environment in order to infer the taste associations of both colour and visual texture, one might be understandably curious about the situations in which colour and texture of the different taste correspondences are put together and assessed by the observers (i.e., when the visual features involved are individually matched with different taste qualities). According to the consensuality principle (Koriat, 2008), when the assessors are not certain (or sceptical) about the link between a visual feature and a potential source object (or concept), the crossmodal association derived from this link is not likely to be highly consensual. Here, assuming all accessible source objects of a given image are mutually exclusive (as a good fit would deny the probability of alternatives being endorsed), there will either be a prevailing source object or there will be none. If the visual cues presented in an image can be potentially linked to different source objects or concepts in order to match with taste qualities, the observers would have to pick a ‘winner’ from the competing source objects instead of computing an averaged midpoint between all possible regularities (Deroy et al., 2016; Schmitz et al., 2021; Wang et al., 2021). Drawing on the previous example with grapefruit, one may consider the simultaneous presentation of colour and visual texture has allowed novel meanings to emerge. If assessors recognise the taste profile represented by these new meanings, these taste-altering visual cues may be considered when deciding on the taste correspondences. That said, if the assessors fail to extract any meaningful representation from the combined stimuli, they might instead select a dominant visual feature and refer to the taste associated with that feature (e.g., Woods et al., 2016). The model is, however, likely different if the features presented are similar enough to be visually averaged, such as fusing multiple colours by blending the hues (Spence and Levitan, 2022; Webster et al., 2014).

Based on the framework of crossmodal correspondences between colour and taste, Woods and Spence (2016) tested the conflict and cooperation between different colour cues in the process of conveying taste information and shaping taste expectations. Woods and Spence were interested in the possible effect of congruent visual stimuli on the robustness of the connections, that is, if simultaneously showing two colours of the same taste correspondence will lead to stronger and more consistent associations. The participants
were presented with two colour patches placed side-by-side in the matching task, they had to decide which of the four basic tastes best represented the pair of colours. In addition to the nature of the associations that their participants made, Woods and Spence also assessed the reaction time for every matching decision made by the participants. The results demonstrated that pairing two congruent colours (i.e., both corresponding to the same taste quality) led to more consistent colour–taste associations. Presenting a pair of colours with congruent taste profiles was found to be more strongly associated with the corresponding taste than its component colours when assessed in isolation.

In Woods and Spence’s (2016) study, the participants were significantly slower to assign a basic taste quality to the visual stimuli when viewing colour patches of equal size side-by-side than when responding to individual colour patches. Putatively, a slower decision in crossmodal matching tasks could imply less certainty about the decision (Beck et al., 2008; Churchland et al., 2008) and, therefore, presumably less consensuality in terms of the association reported. In this case, though, the slower reaction time may hint at how the two colours placed side-by-side were processed serially instead of as a unified stimulus (Kiani et al., 2014). Committing to the theory that a coherent stimulus would lead to a faster reaction time, a follow-up study by Woods et al. (2016) redesigned the stimuli. Woods et al. combined visual stimuli by filling in the foreground (body) and background (outline) of the patch with two different colours (see Fig. 4). The results yielded an intriguing contrast to Woods and Spence’s (2016) study: when two colours of conflicting taste associations were applied to a stimulus as a foreground–background colour scheme, the participants no longer found the stimuli cognitively overwhelming. Judging by the reported mappings of vision–taste associations, the foreground colour

![Figure 4](https://example.com/f4.png)

**Figure 4.** Comparing stimuli from different studies: side-by-side setting (Woods and Spence, 2016) on the left and foreground–background setting (Woods et al., 2016) on the right. (a) Sample of the side-by-side matching test from Woods and Spence (2016); (b) sample of the foreground–background colour combination matching test from Woods et al. (2016). The proportion of the two colours was intentionally kept to such ratio as shown here in order to create a hierarchy with the main colour at the centre.
Figure 5. Comparing the mappings of colour–taste correspondences documented by Woods and Spence, 2016) on the left and Woods et al. (2016) on the right. The colour stimuli are ranked in the order of decreasing association strength from left to right for both studies. Note that the split squares displayed in Woods and Spence’s (2016) study represent the two colours placed as two side-by-side patches, the taste associations of co-appearing patches were assessed together in a single trial.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sweet</td>
<td><img src="image1" alt="Sweet" /></td>
<td><img src="image2" alt="Sweet" /></td>
</tr>
<tr>
<td>Sour</td>
<td><img src="image3" alt="Sour" /></td>
<td><img src="image4" alt="Sour" /></td>
</tr>
<tr>
<td>Salty</td>
<td><img src="image5" alt="Salty" /></td>
<td><img src="image6" alt="Salty" /></td>
</tr>
<tr>
<td>Bitter</td>
<td><img src="image7" alt="Bitter" /></td>
<td><img src="image8" alt="Bitter" /></td>
</tr>
</tbody>
</table>

appeared to be the primary cue as it largely determined the taste profile of the combined stimulus (see Fig. 5). Woods et al.’s (2016) study also revealed that when presented with two colours together in a foreground–background configuration, participants \(N = 100\) completed the taste matching task just as rapidly as when seeing either constituent colour as a single-colour patch.

As Woods et al. (2016) noted, the crossmodal correspondence between simultaneously-presented colours and taste does not appear to reflect a simple case of summing the individual association strength of all colours displayed. The patterns established by their study imply a cognitive process of computing taste associations based on a much more complex model. By presenting multiple colours simultaneously, the taste association observed appears to be built on the computed ensemble of multiple visual cues. The approach adopted by Woods et al. (2016) successfully demonstrates how taste associations can be categorically changed by combining multiple colours, which are visual features that count on the learnt statistics of the environment as bases to match with taste qualities. Concurrently, by featuring two colours in one stimulus and creating a visual hierarchy with clearly differentiated foreground and background, Woods et al.’s (2016) presentation of stimuli has putatively produced
a unitary Gestalt (Wagemans et al., 2012). In this case, not only do individual elements convey taste information, but the stimulus configuration as a whole could exert an influence beyond simply aggregating the individual elements together.

Suppose that presenting multiple visual stimuli provides additional context for matching/accessing internalised statistics, such that inferring the taste associated with a compounded image is comparable to finding a source object with clues given about the context.¹ Certain combinations of colour and texture, presumably by resembling the visual appearance of a specific food, happen to cue taste associations that are seemingly independent from their incorporating visual features. Velasco et al. (2016c) presented a scenario in which colour–taste associations deviated from the established mappings when translucent drinks were coloured. A notable proportion of participants \( n = 1480; 28\% \) regarded the drink dyed blue as the sweetest, only after those who chose red instead \( n = 2164; 41\% \). The findings are rather remarkable given that blue colour is typically associated with saltiness when assessed as an abstract visual feature (i.e., in a non-food context; Spence et al., 2015). A recent review by Spence (2021) on the peculiar case of ‘blue foods’ mentioned the link between blue-sweet impression and the widespread application of blue colourants, especially in soft drinks. The review also noted the popularity of artificial food dyes in Western countries (Aslam, 2006; Hisano, 2016), which has supposedly contributed to the cross-cultural differences in flavour and taste associations depending on the market success of blue foods and beverages (Shankar et al., 2010; Velasco et al., 2014b; Wan et al., 2014). In a previous review, Lee and Spence (2022) endorsed the possibility of turning typically non-sweet visual features in the food (e.g., blue and green) into contextual cues for sweetness, which can be potentially achieved by highlighting the artificiality/ultra-processed nature with translucent appearance (Spence, 2021; Vignola et al., 2021). Perhaps, then, certain textures may cue the characteristics of artificial food; pairing such visual texture with blue might produce a compounded image that conveys the impression of a higher sugar level (Monteiro et al., 2019).

3.2. Interactions Between Crossmodal Associations That Are Mediated by Emotion

For those visual features associated with taste because of sharing a similar valence when appraised, it is worth noting how these hedonic values can be systematically measured using the Semantic Differential Technique (Osgood et

¹ In the Bayesian integration model, the stimuli that are matched together due to the intermediate priors (e.g., knowledge learnt from the environmental statistics) are described as coupling priors (Ernst, 2007; Parise and Spence 2009).
Initially proposed as part of the effort to capture and map the connotative meanings of a wide range of items, the technique is increasingly being favoured by researchers in this field as a better way to conceptualise people’s feelings concerning sensory experience/stimuli (e.g., Dalton et al., 2008; Spence, 2020; Spence and Levitan, 2021; Velasco and Spence, 2022; Velasco et al., 2015a). The method requires assessors to rate the stimulus in question on a scale with its two ends anchored by a pair of bipolar adjectives. To measure valence, the *evaluation* (as in the dimension of meaning that identifies emotion) of the stimulus is collected from the scales with a set of adjectives describing good/bad feelings (e.g., happy–unhappy, optimistic–pessimistic, elegant–vulgar; see Heise, 1969, for an overview). In addition to evaluation, the technique has been demonstrated to reliably survey the feelings in other dimensions of affective meaning, such as *activity* (with ‘active–passive’ adjectives) and *potency* (with ‘strong–weak’ adjectives). Since evaluation appears to be the only semantic dimension closely pertaining to preference, fewer efforts have gone into analysing other connotative meanings of stimuli in vision–taste correspondences research (Weierich et al., 2010). With that being the case, more recent studies have started to examine the role of other semantic dimensions when the observers were asked to make inferences from perceptual stimuli (Motoki et al., 2022 — see Spence, 2023 for a review).

Depending on how the assessors arrive at the vision–taste correspondences of the constituting components, the outcome of taste correspondences for a combined visual stimulus can vary drastically. For example, when assessing the taste of abstract colours, there is no certain ‘scale’ of valence value which a colour needs to be mapped onto in order to establish crossmodal associations. In this regard, one might say the emotionally mediated associations are less arbitrary when compared to those that happen to be based on the internalised statistics of the environment. At the very least, it is possible to conceptualise the mediating emotions as a prothetic value (i.e., hedonic value) by measuring how much a feature is liked by the observers. That said, it is also possible to record the confidence in the observers when they report a cross-modal association (Deroy et al., 2016; Woods and Spence, 2016), which can be used to conceptualise the strength of statistical connections (Koriat, 2008).

For a stimulus incorporating multiple features, all having the possibility of being affectively associated with taste qualities when assessed individually, the hedonic value (and thus taste association) of this stimulus would likely be a computed mid-point accounting for the weight of all the valence values presented (Wilson and Brewster, 2017). So, for instance, the level of sourness associated with a rounded and asymmetric shape was found to lie somewhere in-between an angular, asymmetric shape and a rounded, symmetric shape (Salgado-Montejo et al., 2015).
For the crossmodal correspondences involving taste and mediated by emotions, the hedonic value that people attribute to a combined stimulus would appear to depend on its component members. Not only does this notion apply to visual features such as curvilinearity, symmetry, and segment complexity (Salgado-Montejo et al., 2015), but also auditory cues when presented alongside visual features. In a rare attempt to analyse the taste impression of multisensory packaging designs, Velasco et al. (2014a) documented a four-way interaction between the curvilinearity of typeface, of package shape, and the sound accompanying the product (both auditory pitch and verbal presentation) when rating the expectation for sour and sweet taste (see Fig. 6). Velasco et al. (2014a) concluded that packaging elements such as rounded shapes, curvy typefaces, soft-sounding names, and low-pitched sounds are associated with sweetness, especially when these features are combined. The findings provide another level of empirical support for the notion that each emotionally mediated feature in a combined stimulus contributes to the final hedonic value when evaluated altogether (see Wang et al., 2016, for the links between taste, emotion, and auditory pitch).

Studies examining the mediating role of emotion or valence in crossmodal correspondences have provided a robust approach to systematically explaining the patterns in vision–taste correspondences with linear regression models (Salgado-Montejo et al., 2015; Turoman et al., 2018; Velasco et al., 2015a; Wang et al., 2016). For example, by measuring the emotional ratings and the strength of taste correspondences for the visual stimuli assessed, the correlation coefficient then offers an approximate measure of the variance in crossmodal associations that can be accounted for by valence. However, this approach offers only a partial predictive model, as further explorations have started to highlight the influence of familiarity (Carvalho and Spence, 2019; Chuquichambi et al., 2021; Motoki et al., 2023; Westerman et al., 2013), arousal (Blazhenkova and Kumar, 2018; Marin et al., 2012; Wang et al., 2016), and even cultural notions (Bremner et al., 2013 — cf. Ćwiek et al.,
2022) in facilitating the crossmodal associations that are typically regarded as emotionally mediated. Should one consider examining the contribution of other potential mediating factors, then the aforementioned Semantic Differential Technique should also serve as an adequate tool to capture a wide spectrum of feelings beyond hedonic values (e.g., Velasco et al., 2016b).

3.3. Interactions Between Crossmodal Associations of Different Origins/Bases

Theoretical understandings of the crossmodal correspondences allow predictions about the dynamics between certain visual features in their link with taste. As an example, it would be reasonable to presume that the taste correspondence of an asymmetric yellow triangle is substantially modulated, if not determined, by a combination of the internalised statistics and the hedonic values linked to the said shape. However, these speculations, albeit informed, remain superficial due to the scarcity of empirical evidence on this point. Only a handful of research efforts have explored the collective taste impressions of multiple visual features of different mediating routes (see Lee and Spence, 2022; Spence, 2019; Velasco et al., 2016a, for reviews). For vision–taste associations, not many studies had the explicit goal of examining the interaction of visual features. The curvilinearity–symmetry–complexity study by Salgado-Montejo et al. (2015) and the similar study by Velasco et al. (2014a) are among the few to have explored the combined visual effect with a factorial design, allowing them to inspect the interaction of geometric components in the shape–taste correspondences. When the interaction involves visual features that rely on different ways of establishing taste correspondences, the current understanding of the vision–taste association is insufficient for predicting the collective taste impression of these visual features. This section covers the past literature and extracts relevant evidence that could help to provide insight into such a mode of interaction (i.e., between visual features that typically depend on statistics and on emotional valence/hedonics). Table 4 provides a summary of these prominent studies covered by the review, in the order that they are introduced, to serve as a quick reference guide to the currently available empirical evidence.

Stewart and Goss (2013) examined the combined influence of colour and shape on taste expectation and perception. Their participants (N = 48) tasted a cheesecake from either a white round, white square, black round, or black square plate, then reported the strength of sweetness perceived. Affirming previous findings (Piqueras-Fiszman et al., 2012), Stewart and Goss (2013) found a main effect of plate colour but not of plate shape. Importantly, however, they also found a significant interaction between the plate colour and plate shape, the plate having a white colour and round shape combined creates the strongest synergy for cheesecakes to be perceived as sweet. That said,
### Table 4.
Studies that have allowed the investigation of combined taste associations by assessing multiple visual features simultaneously

<table>
<thead>
<tr>
<th>Study</th>
<th>Visual features assessed</th>
<th>Mode of interaction involved</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salgado-Montejo et al. (2015)</td>
<td>Curvilinearity–Symmetry–Segment Complexity</td>
<td>Affective only (between dimensions)</td>
<td>Documented the expectations of sweet and sour for shapes varying at two levels in each of the three visual dimensions</td>
</tr>
<tr>
<td>Woods and Spence (2016)</td>
<td>Colour–Colour</td>
<td>Statistical only (within colour dimension)</td>
<td>Assessed the combined taste associations of two colour patches placed side-by-side</td>
</tr>
<tr>
<td>Woods et al. (2016)</td>
<td>Colour–Colour</td>
<td>Statistical only (within colour dimension)</td>
<td>Featuring two colours joined as foreground–background colour of the same visual stimulus</td>
</tr>
<tr>
<td>Velasco et al. (2014a)</td>
<td>(Package) Curvilinearity–(Typeface) Curvilinearity</td>
<td>Affective only (within curvilinearity dimension)</td>
<td>Examined the interaction between the curvilinearity and sound (as auditory pitch and verbal names) when shaping taste expectations</td>
</tr>
<tr>
<td>Stewart and Goss (2013)</td>
<td>Curvilinearity–Colour</td>
<td>Statistical–affective</td>
<td>Taste experience of the food was found subject to the interaction effect between plate shape and plate colour</td>
</tr>
<tr>
<td>Rolschau et al. (2020)</td>
<td>Curvilinearity</td>
<td>–</td>
<td>Manipulated only the typeface curvilinearity, but potentially encountered the conflict between colour and typeface curvilinearity</td>
</tr>
<tr>
<td>Lee and Spence (in press)</td>
<td>Colour–Curvilinearity</td>
<td>Statistical–affective</td>
<td>Noted a diminished effect of curvilinearity when joining the so-called ‘sour-tasting’ colours</td>
</tr>
</tbody>
</table>
Stewart and Goss’s (2013) quasi-experimental design has involved numerous uncontrolled factors potentially tangled in their tasting sessions, such as the interaction between plate colour and food colour (see Woods et al., 2016). Of course, when simulating the real dining experience, taste perception is likely to be influenced by a range of other factors that may have been unanticipated by the researchers, such as the texture of the food (Saint-Eve et al., 2011), or of the plate (Biggs et al., 2016), or even their combination.

In addition to demonstrating a collaborative relationship between colour and shape in taste correspondences, the study by Stewart and Goss (2013) also provided a potential case of processing fluency in combining the black colour with a square shape. In this context, processing fluency suggests people would prefer seeing stimuli possessing similar quality in proximity or being grouped together (Reber et al., 1998, 2004). Essentially, the preference for features of similar property could explain why a square plate in black colour had received a higher liking rating than the square white plate and even came close to the level of the round white plate. For the taste correspondence of visual features, the positive evaluation for an image could be potentially translated (by the observers) to a stronger association between the image-being-assessed and sweetness (Salgado-Montejo et al., 2015; Spence, 2022c; Velasco et al., 2016d). It is worth noting that this synergy between the co-presented visual features has not been studied (and arguably could not be studied) in the literature concerning the taste associations of visual features in isolation. Put simply, processing fluency is one of the unanticipated effects in crossmodal correspondences when stimuli are assessed together.

The unanticipated interaction of visual features can also be observed in other studies of vision–taste correspondences. Rolschau et al. (2020) tested the typeface–taste congruency in their quasi-experimental field study in a naturalistic setting. They changed the typeface used on the beer menu at a bar to investigate if typefaces could have any effect on taste expectations and therefore influence consumers’ choices. As demonstrated in the literature on typeface–taste crossmodal correspondence (Velasco et al., 2015b, 2018c), typeface could prime people with gustatory information about the product, largely in accordance with the curvilinearity instances of the strokes/stems. For example, rounded typeface could subtly indicate the presence of sweetness and invite more people to choose the associated product (see also Wang et al., 2020). That being said, if a rounded typeface causes the product it decorates to appear less sour, consumers might be convinced to try such product that they would not consider otherwise. Effectively, by alleviating the unpleasantness, rounded typeface could lower the threshold and turn would-be-shunned products into viable or possibly even desirable option. This counterintuitive effect was exactly what Rolschau et al. (2020) observed: contrary to the previously documented taste profile of typeface curvilinearity, the sweet-inducing
rounded typeface was found to increase their participants’ likelihood of choosing the sour beers. It should be recognised, however, that it is not known what aspects of product assessment (e.g., sweetness, acidity, and Bouba/Kiki effect in the beer names — see Fairhurst et al., 2015) were reflected by the purchase decision, which has not been a well-understood topic in the field of crossmodal correspondences (Biswas and Szocs, 2019; Spence and Van Doorn, 2022).

Relevant to the interaction between visual features, there is a compelling yet undiscussed explanation accounting for the unintuitive behaviour of the pubgoers as observed by Rolschau et al.’s (2020) field study. In the beer menu prepared by Rolschau et al., a blackboard was used as the only available menu for the consumers wanting to order beer (see Fig. 7). Important here, while the typeface displayed was changed between the angular and the round condition, the name of each beer remained written in the specific colour. Ostensibly, if the colour had remained the same for each menu item throughout the experiment, the only variable should indeed be the typeface (including the weight and stem filling of the font). However, the interaction between typeface and colour might have created unforeseen synergies when conveying information about taste expectations.

Using a vision–taste crossmodal matching task, Lee and Spence (in press) investigated the interaction of colour and typeface in an online study. The participants \((N = 102)\) provided their taste evaluations for a series of text samples. The text stimuli were rendered with either round or angular typeface and outlined by a thick coloured stroke, the stem body of the text is also coloured, albeit it could be a different colour than the outline. As expected, a main effect was found for colour and for typeface in taste expectations. Except for salt ratings, there was a significant interaction effect between colour and typeface for all basic tastes. For the effect of colour on taste expectations, painting texts with body–outline colouring schemes operates almost identically to painting patches with foreground–background patterns (e.g., Woods et al., 2016 — see
On top of that, a rounder typeface could significantly enhance the expectation of sweetness and discount that of sourness on most occasions.

In their study of the interaction between typeface and colour, Lee and Spence (in press) also demonstrated a significant effect of colour when people associate sour taste with angular and round typefaces. In particular, when the text was in the colours that are generally matched with sourness (e.g., green, yellow, and their combination), the respective sourness ratings were not statistically different between the angular and round typefaces. For instance, no curvilinearity effect was found for texts in the ‘sour-tasting’ colours (see Fig. 8). Critically, this would suggest that colour could affect the taste association of typeface curvilinearity (and vice versa). Such an effect of inhibition, which researchers have only just begun to understand, may have confounded Rolschau et al.’s (2020) field study, making it even harder to translate vision–taste correspondences into consumers’ choice behaviour in a meaningful setting. Given the evidence presented above, the taste correspondence of Fig. 5).
statistically-based colour and emotionally mediated curvilinearity appear to work in synergy when incorporated into a combined stimulus. In certain circumstances, however, the effect of curvilinearity would appear to have been inhibited. It is not clear what might have caused the situational conflict/inhibition that led the curvilinearity to be less influential. Possible reasons include the domination of colour effects over the relatively less influential curvilinearity, as presumably observed in the interaction of plate colour and plate shape (Stewart and Goss, 2013).

4. Conclusions

Among the documented cases of perceptual and semantic connection between taste qualities and visual features, not much research has been directed towards studying the gustatory message conveyed by combining visual features from different dimensions. This is not just surprising because the field of crossmodal correspondences research has grown rapidly in recent years, but also due to how powerful those synergies can be. As highlighted by the corresponding sections, the existing theoretical framework available to the researchers did not anticipate some instances of conflicts. It takes revisions to the current model (i.e., theories on the acquisition and mappings of crossmodal correspondences involving visual features in isolation) to explain the effects resulting from presenting multiple visual features in vision–taste associations. At the same time, though, it is reassuring that we can generally rely on the established accounts to predict the direction of cooperation/conflict as a result of visual interaction.

The present review assessed the gustatory information collectively (if not coherently) conveyed by multiple visual features. This has been accomplished by examining studies that have simultaneously manipulated more than one dimension of visual elements in vision–taste correspondences. For the visual features matched with taste because they share a similar hedonic tone, such as curvilinearity and symmetry, the taste association is adjusted according to the additive (when features are congruent) or averaged (when in conflict) value of the constituting emotions (Salgado-Montejo et al., 2015; Velasco et al., 2014a). If the combined visual stimulus (which is composed of multiple visual features) reminds people of an internalised regularity concerning how a specific taste accompanies certain visual imagery, the combination of visual features might be matched with a different taste quality accordingly (Wan et al., 2014; Woods et al., 2016). In this case, the altered taste association could be completely detached from the taste correspondence of the individual visual components making up the ensemble imagery. Given the evidence that has been evaluated here, this would suggest new semantic meanings being made accessible as the result of combining visual elements (see also Lee and Spence,
Moreover, there is also a good prospect that some visual features may be regarded by the observers as an expected or preferred combination when presented together (Reber et al., 2004; Sundar and Noseworthy, 2016). Drawing upon what we have learnt from the research on processing fluency and the affective (emotional mediation) account, the preference for certain ways to group visual features could ultimately lead to a higher chance of associating these configurations with sweetness. On top of these observations, the evidence reviewed points to a domination of colour over curvilinearity when the two visual dimensions work together to shape–taste expectations (Lee and Spence, 2023; Stewart and Goss, 2013).

Due to the scarcity of evidence, many claims and predictions made in this review must necessarily remain speculative at this stage. One might suggest that the current model of emotional mediation, as an explanation for certain types of crossmodal correspondences, should be revised to account for the influence of familiarity/typicality, personal differences (Chen et al., 2021), and non-valence emotions. Meanwhile, the parameters in the internalisation of regularities, such as the statistics of motivation, reward, and sensory inputs, have not been sufficiently investigated to accurately predict the taste correspondences associated with combined, multi-dimensional visual features (Aleem et al., 2020). Further research is warranted to verify our hypotheses and fill in the gaps identified here. For example, to clarify if the effect of processing fluency enhances the preference for combinations of interconnected visual features. It is also worth investigating what could have led certain visual features to be regarded as interconnected in the first place, such as the notion known as the Kandinsky correspondences (Kandinsky, 1914; Kharkhurin, 2012 — see Dreksler, 2020; Dreksler and Spence, 2019). Further, to confirm the limitation of synergy when presenting visual elements with a congruent profile of taste correspondence (e.g., the domination of colour over shape curvilinearity). Eventually, as studies progressively depart from matching taste qualities with abstract visual stimuli, they will inevitably engage with more complex stimuli in both quantity and quality. The shift will bring experimental conditions closer to real-world settings, and, in turn, the methodologies will be more relevant in terms of addressing commercial interests (e.g., Chitturi et al., 2019; Rolschau et al., 2020 — see also Motoki and Velasco, 2021; Velasco et al., 2016a). With the evidence reviewed in mind, the current field of research is notably unprepared to unpack the cognitive process of translating taste expectations/preferences into purchase decisions.

References


Piqueras-Fiszman, B., Alcaide, J., Roura, E. and Spence, C. (2012). Is it the plate or is it the food? Assessing the influence of the color (black or white) and shape of the plate on the perception of the food placed on it, *Food Qual. Pref.* **24**, 205–208. DOI:10.1016/j.foodqual.2011.08.011.


