

Exploring Group Differences in the Crossmodal Correspondences

Charles Spence*

Crossmodal Research Laboratory, Department of Experimental Psychology, University of Oxford, New Radcliffe House, Walton Street, Oxford, OX2 6BW, UK

*E-mail: charles.spence@psy.ox.ac.uk

ORCID iD: Spence: 0000-0003-2111-072X

Received 9 March 2022; accepted 22 July 2022; published online 9 August 2022;
published in print 22 August 2022

Abstract

There has been a rapid growth of interest amongst researchers in the cross-modal correspondences in recent years. In part, this has resulted from the emerging realization of the important role that the correspondences can sometimes play in multisensory integration. In turn, this has led to an interest in the nature of any differences between individuals, or rather, between groups of individuals, in the strength and/or consensuality of cross-modal correspondences that may be observed in both neurotypically normal groups cross-culturally, developmentally, and across various special populations (including those who have lost a sense, as well as those with autistic tendencies). The hope is that our emerging understanding of such group differences may one day provide grounds for supporting the reality of the various different types of correspondence that have so far been proposed, namely structural, statistical, semantic, and hedonic (or emotionally mediated).

Keywords

cross-modal correspondences, group differences, autism, development, sensory loss, synaesthesia

1. Introduction

The body of evidence supporting the existence of a wide range of cross-modal correspondences between various different attributes, features, and dimensions of sensory experience has expanded rapidly in recent years (e.g., Marks, 1974, 1978, 2004; Sathian and Ramachandran, 2020; Spence, 2011). Furthermore, the research that has been published to date has also highlighted the influence that a number of such correspondences can sometimes exert over various

aspects of multisensory perception (e.g., Parise and Spence, 2009; Spence and Sathian, 2020). It is interesting, therefore, to consider whether there are also any meaningful differences between particular groups of individuals in terms of the strength and/or consensuality of the cross-modal correspondences that have been documented to date. A growing body of empirical research has recently started to highlight various differences between groups of individuals attributable to everything from neurotypical development (e.g., Marks *et al.*, 1987) through to cultural influences (Walker, 1987), and from the consequences of sensory loss (e.g., Hamilton-Fletcher *et al.*, 2018; Sourav *et al.*, 2019; though see also Barilari *et al.*, 2018) through to autism/autistic tendencies (e.g., Chen *et al.*, 2021).

Importantly, however, the question of differences between groups of individuals can potentially be further broken down as a function of the type of cross-modal correspondence under investigation, and perhaps even the particular pair of sensory modalities that happen to be involved. For while the majority of research published to date has tended to focus on audiovisual cross-modal correspondences (e.g., Bonetti and Costa, 2018; Guzman-Martinez *et al.*, 2012; Orchard-Mills *et al.*, 2013; see Spence and Sathian, 2020, for a review), there has also been interest from researchers in those correspondences involving the sense of touch (see Spence, 2020a, for a review) and even the chemical senses (namely olfaction and gustation; see Belkin *et al.*, 1997; Gilbert *et al.*, 1996; Kemp and Gilbert, 1997; Spence, 2020b; Spence and Levitan, 2022).

Spence (2011) argued for the existence of four putatively distinct kinds of cross-modal correspondence, namely structural (e.g., Stevens, 1957; Walsh, 2003; cf. Pinel *et al.*, 2004; Yates *et al.*, 2012), statistical (e.g., Wagner *et al.*, 1981), semantic (Gallace and Spence, 2006; Martino and Marks, 1999, 2000, 2001; Osgood, 1960; Walker and Walker, 2016), and emotional or hedonic (e.g., Palmer *et al.*, 2013; Spence, 2020c; see Table 1). Structural correspondences are thought to be based on common neural encoding (e.g., of magnitude or intensity; Stevens, 1957). Statistical correspondences are based on associative learning (i.e., they are based on the internalization of the statistical regularities of the multisensory environment such as, for example, the physical correlation between pitch and size). Semantic (or lexical; Spence and Di Stefano, 2022; Walker, 2016) correspondences are based on the same term being used to describe different dimensions of experience (as in the terms ‘high’ and ‘low’ being used to describe pitch of sound and elevation; see Martino and Marks, 1999). Finally, mood-based or emotionally-mediated correspondences are thought to be based on matching of the mood, emotion, or hedonic valence associated with each of the component stimuli being similar (see Palmer *et al.*, 2013; Schifferstein and Tanudjaja, 2004; Spence, 2020c).

Table 1.

Different kinds of crossmodal relation that have been proposed in the literature to date, including four types of crossmodal correspondence, along with amodal, metaphorical, and crossmodal semantic congruency. Relevant here, Walker-Andrews (1994) distinguished between four kinds of crossmodal mapping: namely those based on amodal specification, such as (in her view) size, shape, texture, and substance (see also Wagner *et al.*, 1981, on the notion of amodal specification); Arbitrary and Artificial correspondences (i.e., the ringtone on one's phone; man-made objects, she notes, mostly fall into this category); Arbitrary but Natural (such as the sound of a person's voice); and Typical intermodal relations, “describing those that are typically in the environment but are only partially specified by amodal invariants – such as heavy objects typically producing deeper and louder sounds on impact than do lighter objects)” (Walker-Andrews, 1994, p. 48). One might wonder whether the latter two differ only in terms of the precision of the crossmodal match, rather than reflecting a difference in kind

Structural (physiological)	Based on neural organization/coding (e.g., of stimulus intensity/magnitude; Stevens, 1957; Walsh, 2003). Spence and Di Stefano (2022) have argued that physiological may be better than structural, given that the latter term has already been used to describe alignment based on physical similarity of the perceptual dimensions (e.g., hue and pitch both being circular dimensions; e.g., see Sebba, 1991).
Statistical	Based on the internalization of the statistics of the environment (e.g., pitch-size mapping; e.g., Ernst, 2007; Spence, 2011). That said, Walker-Andrews (1994) distinguishes between physical laws (or Typical intermodal relations) and man-made objects (Arbitrary and Artificial in her terminology).
Semantic (lexical)	Based on common linguistic terms used to differentiate aspects of sensory experience (e.g., Gallace and Spence, 2006; Martino and Marks, 1999, 2000, 2001; Osgood, 1960). Walker and Walker (2012) suggest that it is the lexical, rather than linguistic meaning of the terms that is key to the effects reported in this case (cf. Shayan <i>et al.</i> , 2014).
Emotionally-mediated (hedonic)	Based on emotional, or affective associations of component stimuli, such as when paintings and musical selections matched because they are associated with the same emotion (e.g., Palmer <i>et al.</i> , 2013; Schifferstein and Tanudjaja, 2004; Spence, 2020c)
Amodal	Pick-up of the same environmental information by different senses (e.g., shape information detected by vision and touch; e.g., Gibson, 1969; cf. Lewkowicz and Turkewitz, 1980; Smith, 1994).
Metaphorical	Based on the more abstract mapping of (sensory) features (such as the matching of auditory duration with visual length (Gardner, 1974; Marks <i>et al.</i> , 1987; Wagner <i>et al.</i> , 1981 – cf. Jalal and Ramachandran, 2014).
Semantic	Based on different unisensory properties that are associated with semantically-meaningful objects or concepts, such as the mapping of a barking sound with the image of a dog (e.g., Barenholtz <i>et al.</i> , 2014; Chen and Spence, 2010, 2011). Notice how the terminology conflicts with the so-called semantic category (terminology introduced by Martino and Marks, 1999, 2000) above.

One might, for example, expect there to be more scope for individual differences in the case of semantic (or lexical) and/or emotional cross-modal correspondences than for those correspondences that happen to be structural, or statistical in origin (at least when the latter happen to be based on the physical regularities of the environment). The hope amongst some researchers is therefore that any regularities in the patterns of differences that may be observed in various groups of participants might one day help to support the reality of the distinctions that have been suggested between the various different classes of cross-modal correspondence (see Smith, 1994). At the same time, however, it should also be acknowledged that there is something of an unresolved question in the literature as to whether amodal attributes and metaphorical mappings should be listed alongside the other cross-modal correspondences (e.g., Smith, 1994; Walker, 1987; Walker-Andrews, 1994), or whether instead they should simply be subsumed within the other categories of cross-modal relation (see Spence *et al.*, 2013, on the problematic notion of amodality in multisensory research).

Structural (or perhaps better said, physiological; Spence and Di Stefano, 2022, on this point), and some statistical correspondences (namely those based on the physical regularities of the environment), such as the pitch–size association, might well be expected to be universal (e.g., Gallace and Spence, 2006; Parise and Spence, 2009; see also Peters *et al.*, 2015; Pisanski *et al.*, 2017). That said, those statistical correspondences that happen to be based on more ‘arbitrary’ combinations of features (what Walker-Andrews, 1994, has termed ‘Arbitrary and Artificial’ correspondences), such as, for example, those that have been documented to exist between smell and taste stimuli (e.g., Blank and Mattes, 1990; Spence, 2008), and/or between colour and taste (e.g., Shankar *et al.*, 2010; Spence *et al.*, 2015), or between tastes and visual textures (Barbosa Escobar *et al.*, 2022) are presumably more likely to show robust cultural variation. Consistent with such a view, Wan *et al.* (2014) conducted an online study ($N = 452$) of the nature of any cross-modal correspondences between basic tastes and abstract visual features (specifically hue category and form) in groups of participants from four countries (China, India, Malaysia, and the USA). While the results showed a high degree of cross-cultural similarity, some salient cross-cultural differences were also observed. Indeed, according to the authors: “In terms of the whole picture of color-shape-texture-taste associations, we found the four countries, China, India, Malaysia, and the USA, are quite different from each other” (Wan *et al.*, 2014, p. 9). That being said, the authors were unable to establish any plausible factors underpinning the differences that they reported (such as Eastern or Western culture), or meaningful differences in terms of the particular correspondences under investigation.

Levitan *et al.* (2014) also documented a number of cross-cultural similarities and differences in terms of the odour–colour correspondences in a study

that was conducted simultaneously in six different countries/cultural groups (Dutch, Netherlands-residing-Chinese, German, Malay, Malaysian–Chinese, and a group of people living in the USA). The researchers used representational dissimilarity analysis to assess the degree of cross-cultural similarity and difference in the patterns of colour–odour associations that were reported. These researchers found that the German and US participants were most similar in terms of their specific cross-modal correspondences, followed by German and Malay participants. By contrast, the largest group differences in terms of colour–odour cross-modal correspondences that were reported were between the Malay and Netherlands-resident Chinese groups and also between the Dutch and Malaysian–Chinese. Given this pattern of results, the researchers were led to the conclusion that culture (i.e., in terms of different prior patterns of multisensory perceptual experience) played an important role in determining the particular colour–odour cross-modal associations that they observed.

Meanwhile, in a sample of more than 5000 participants from around the world, Velasco and his colleagues were able to demonstrate that, of the six drinks that were shown to participants, the dark red one was rated as looking sweetest by participants (the other drinks had a blue, yellow, green, orangey, and bright red appearance), no matter which continent the participants happened to come from (see Velasco *et al.*, 2016). That said, it should also be noted that the correspondence was far from perfectly consensual, with only 44% of the participants choosing the dark-red drink as the sweetest of the six available options (see Note 1). Furthermore, as soon as colour happens to be present in terms of the appearance properties of a particular food or drink, then the likelihood is that it may well be associated with a specific product or brand (see Shankar *et al.*, 2010). Consequently, in such cases, it can be hard to rule out the possibility that semantic knowledge (regarding the particular configuration of sensory properties that normally co-occur together in a given food product – such as the colour brown, a sweet taste, and vanilla–cinnamon–citrus scent of a cola drink) may play in driving any cross-modal associations that are documented (i.e., rather than necessarily a more abstract, and direct, correspondence between colour and taste/smell; e.g., Spence, 2020b; Spence and Levitan, 2022; see also Shankar *et al.*, 2010; Speed *et al.*, 2021). This may also be expected to introduce a cultural element to the cross-modal correspondences that have been reported (see Walker-Andrews, 1994, on those correspondences that are associated with, or based on, man-made objects). In such cases, the correspondences may be better considered to be semantic in nature (e.g., as in the literature on cross-modal semantic congruency; see Chen and Spence, 2010, 2011; rather than semantic cross-modal correspondences; see Martino and Marks, 1999). Here, it is important not to confuse this with semantic cross-modal correspondences suggested by Martino and Marks

(1999, 2000; see also Spence, 2011). This is one of the reasons why it may make more sense to label the latter as lexical correspondences (see Spence and Di Stefano, 2022; and Walker, 2012, on this point).

One of the intriguing questions about the cross-modal correspondences (especially for some of the more surprising cross-modal mappings) for which it may be difficult to come up with a statistical explanation (e.g., Sidhu and Pexman, 2018; Spector and Maurer, 2012; Stevenson *et al.*, 2012; Wagner *et al.*, 1981) concerns the consensuality of the mappings that have so far been obtained (Chen *et al.*, 2019; Spence *et al.*, 2015; Woods *et al.*, 2013). According to certain of the research that has been published to date, those cross-modal correspondences that people are themselves more confident about also tend to be the ones that they believe will be shared by others (e.g., see Spence *et al.*, 2015, fig. 4c). Potentially relevant here, Koriath (2008) has introduced the notion of ‘consensuality’ (to describe the feeling that we often know how other people are likely to respond, and this may be more important to performance than the correct answer, assuming there is one). Others, meanwhile, talk of people’s ‘metacognitive awareness’ (see Chen *et al.*, 2019). That being said (and as we will see later), an individual’s feeling regarding the consensuality of a given cross-modal correspondence need not always accurately predict its strength within the general population (see Parise, 2016).

At the same time, however, it is easy to imagine how any semantic or rather lexical correspondences (Spence and Di Stefano, 2022; Walker, 2016) will likely depend on, or at the very least be influenced by, the differing linguistic (Wagner *et al.*, 1981) or lexical (see Walker and Smith, 1984; Walker and Walker, 2012; cf. Glaser and Glaser, 1989) terms that are sometimes used to describe sensory experiences in different languages. One might think here only of the way in which different terms, such as ‘thickness’ or ‘height’, are sometimes used to describe the pitch of sounds in different languages (e.g., Dolscheid *et al.*, 2014; Eitan and Timmers, 2010; Parkinson *et al.*, 2012; Shayan *et al.*, 2011). At the same time, however, it is worth considering how there may also be a relevant historical component to a few such semantic correspondences, given the changing meanings, uses, and associations that terms such as ‘thick’ and ‘bright’ have had over the decades/centuries in different languages when referring, for example, to the description of sounds (e.g., see Cohen, 1934; Mudge, 1920; Schiller, 1935, on this point).

Beyond this, there is also a sense in which affective, or emotionally-mediated, correspondences might well be expected to show more individual, or cultural, variation (e.g., Palmer *et al.*, 2013; Spence, 2020c; Velasco *et al.*, 2015; Wang and Spence, 2017; Wang *et al.*, 2016), given the cultural differences that have been reported in terms of people’s affective responses to unisensory stimuli (such as colours and odours; e.g., see Ayabe-Kanamura *et*

et al., 1998; Blazhenkova and Kumar, 2018; Jonauskaitė *et al.*, 2019, for a couple of representative examples). By contrast, it is easy to imagine how those cross-modal correspondences that happen to be based on the physical regularities of the environment (Evans and Treisman, 2010), such as, for example, the correspondence between pitch and size (Gallace and Spence, 2006; Parise and Spence, 2009), would tend to be universal, given the regularity with which people will presumably be exposed to them. They might perhaps even be shared across species (e.g., see Korzeniowska *et al.*, 2022; Ratcliffe *et al.*, 2016). The auditory pitch–elevation correspondence is presumably also likely to be universal (Parise *et al.*, 2014; see also Rusconi *et al.*, 2006), as would any putatively innate correspondences, such as possibly between visual brightness and auditory pitch (Haryu and Kajikawa, 2012; Ludwig *et al.*, 2011; though see Spence and Deroy, 2012a) (Note 2). Intriguingly, researchers have also raised the possibility that certain of the cross-modal correspondences between auditory pitch (low vs high) and specific taste qualities (namely, bitter and sweet) might be shared universally. This suggestion is based on the different stereotypical orofacial gestures that these basic tastants evoke across species at birth, and the presumed sonic differences in the quality of any utterances that would be made (see Knöferle and Spence, 2012; Spence, 2012).

Interestingly, when Parise and Spence (2012) assessed five different audio-visual cross-modal correspondences (including examples of both verbal and non-verbal sound symbolism), their strength, as measured by the d' score (using a version of the Implicit Association Test; IAT), turned out to be essentially equivalent, regardless of the correspondence being tested (see Fig. 1). By contrast, however, when Parise (2016) took a wide range of different putative correspondences (described verbally) that those he interrogated thought would not exist, he actually found (in an informal survey of 62 people) that many of the mappings did, in fact, show a consensual response mapping (in terms of aligning the ends of the scale). Taken together, therefore, Parise's informal results suggest that the strength, or consensuality, of the mapping between a wide range of polar attributes/dimensions (see Parise and Spence, 2013; Smith and Sera, 1992) varies from chance-level responding (i.e., ca 50%) up to near-perfect agreement on the most appropriate polarity alignment (see Fig. 2). That said, it should be noted that only a subset of the correspondences assessed by Parise are actually cross-modal.

2. On Different Kinds of Individual Difference Relevant to the Cross-Modal Correspondences

It is important to recognize how individual differences in the cross-modal correspondences might reveal themselves in a number of different ways: for

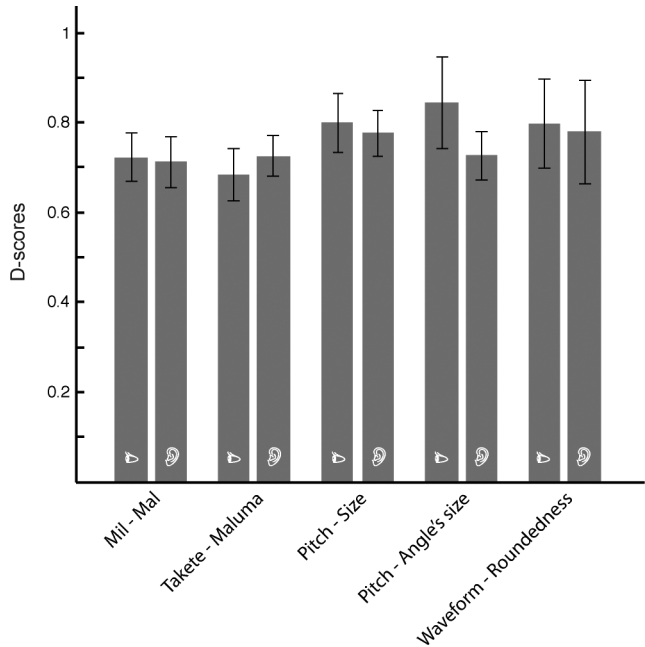


Figure 1. Comparison of the effect size (d score) for vision and audition between the five Implicit Association Test (IAT) experiments assessing various different audiovisual cross-modal correspondences reported by as Parise and Spence (2012). The participants had to respond to a random sequence of auditory and visual stimuli. When the corresponding stimuli were associated with the same response key performance was better (i.e., faster and more accurate) than when non-corresponding stimuli were associated with each of the two response keys. Note that all of the cross-modal correspondences had a very similar effect size. Error bars represent the standard error of the mean. Figure reprinted from Parise and Spence (2012).

instance, as a difference in the absolute number of cross-modal correspondences that are exhibited (i.e., consider here only how while adults have been reported to exhibit a robust cross-modal correspondence between visual size and auditory loudness, younger children of 2–3 years of age apparently do not; Smith and Sera, 1992). More precisely, this claim might be taken to suggest that the strength of different cross-modal mappings varies from chance-level responding to near-perfect agreement across a sampled population (e.g., Dai-ley *et al.*, 1997; Parise and Spence, 2013). It is, though, important to highlight the fact that ‘strength’ might have one of several different meanings here: it might, for example, be taken to refer to an individual’s own belief in the appropriateness, or feeling of rightness, of the cross-modal mapping. Alternatively, however, it might also be taken to refer to the degree on consensuality across a population (Koriat, 1975, 2008). The latter is presumably what Rader and Tellegen (1987) had in mind when they referred to the ‘collective norm’ response (cf. Gardner, 1974). Alternatively, however, strength might instead

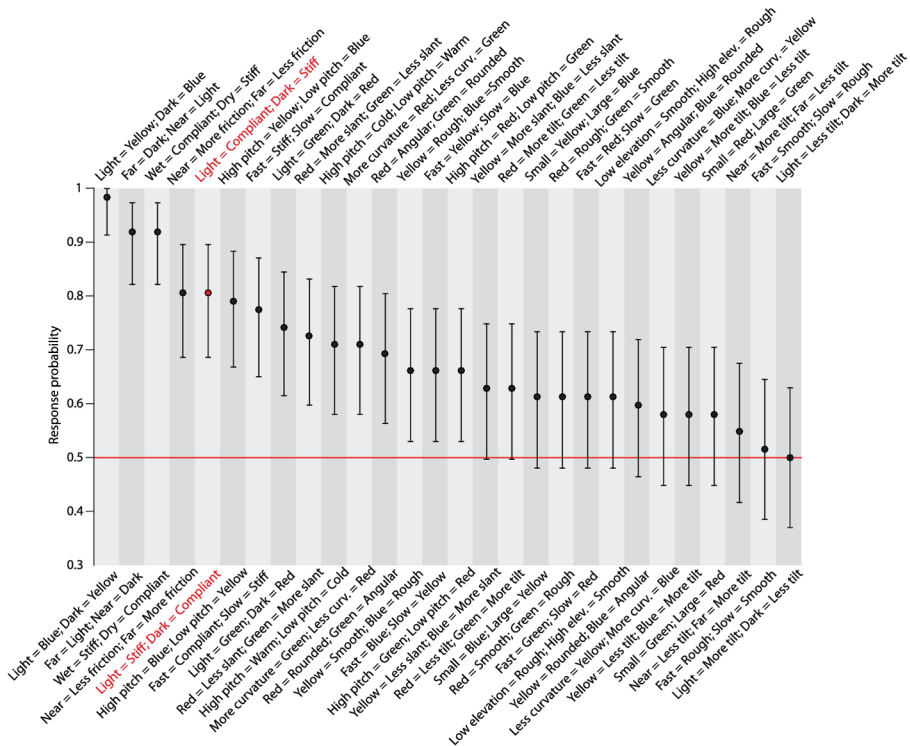


Figure 2. Results of an informal survey ($N = 62$) reported by Parise (2016) highlighting the polarity alignment for various dimensions that people had suggested did not correspond.

be taken to refer to a variation in the vividness of the association (Rader and Tellegen, 1987), the latter possibly linked to individual differences in the vividness of mental imagery (cf. Nanay, 2017, 2018; Spence and Deroy, 2013a). This would seem to be what Rader and Tellegen described as the ‘experiential component’ of synaesthesia. Considering how synaesthesia has so often been defined (e.g., with synaesthetes being required to pass the ‘Test of Consistency’; Baron-Cohen *et al.*, 1993), one might also wonder whether the consistency of people’s cross-modal mappings over time should be included here (cf. Sabaneev and Pring, 1929, p. 266). That said, while a few researchers have explicitly tested for such consistency (e.g., over a period of years in Gilbert *et al.*, 1996; and over a matter of weeks in O’Mahony, 1983), this is by no means a common practice currently.

In their study, Rader and Tellegen (1987) had 374 participants (undergraduates enrolled in an introductory psychology course at a North American university) match each of three tones (200 Hz, 1000 Hz, and 4000 Hz; as well as a range of other musical and vocal speech stimuli) with each of 11 colours (white, pink, yellow, red, orange, green, blue, purple, brown, grey,

and black) identified verbally. The participants were given five points to distribute between the colours as they saw fit in each matching task. White was most frequently matched with the 4000 Hz tone, blue (closely followed by green) with the 1000 Hz tone, and black (closely followed by brown and grey) was the most popular choice for the 200 Hz tone, suggesting a pitch–brightness/lightness mapping. The participants were also given a short questionnaire in order to assess the frequency of any visual–auditory synaesthetic experiences that they had outside of the experimental setting. While the terminology used by Rader and Tellegen is somewhat confusing, they do report “generalized and continuously distributed individual differences in the tendency to represent sounds as colors in accordance with empirically derived norms” (Rader and Tellegen, 1987, p. 981; see also Odberet *et al.*, 1942), the latter referring to what they term a *general translation tendency*, or variation in the strength of a correspondence. Of course, any correspondences that involve auditory pitch as one of the corresponding dimensions might also be influenced by individual differences in holistic vs spectral listening (Schneider *et al.*, 2005).

Indeed, over the years, a number of studies that have attempted to address individual differences in the strength of cross-modal correspondences have been confusingly described in terms of synaesthetic tendencies (e.g., Rader and Tellegen, 1987; Simpson *et al.*, 1956; Walker and Smith, 1984). It should, however, be noted here that there is also a separate question (and experimental literature) concerning whether there may be a meaningful continuum in terms of any individual differences in synaesthetic tendencies (e.g., Deroy and Spence, 2016a, 2017; Lehman, 1972; Wicker, 1968; Wicker and Holahan, 1978; cf. Rader and Tellegen, 1987). However, once cross-modal correspondences are clearly distinguished from the phenomenon of synaesthesia (as argued for, at length, by Deroy and Spence, 2013a, b), the answer to the continuity question can, and presumably should, be addressed separately in the two cases (see also Cohen, 2017; Deroy and Spence, 2017).

Taken together, therefore, the research that has been reported in this section highlights how any individual differences in the strength of cross-modal correspondences may be exhibited in one (or more) of several different ways. Namely, the strength of a particular cross-modal correspondence may be related to the general agreement within a tested population, what Rader and Tellegen (1987) refer to as the “general translation tendency” and which may vary between chance-level mapping and perfect agreement (sometimes referred to as “empirically-derived norms” or the “modal response”; Gardner, 1974). Second, researchers have occasionally considered the strength of a correspondence in terms of the vividness of the “experiential component” (Dailey *et al.*, 1997; Rader and Tellegen, 1987). Third, the strength of cross-modal correspondences has also been conceptualized by researchers in terms of an

individual's belief, or confidence, that other people will share the same mapping (Koriat, 2008; Wan *et al.*, 2014; Woods *et al.*, 2013). This is sometimes referred to as metacognitive awareness (Chen *et al.*, 2019). Fourth, one could potentially also consider the consistency of a particular cross-modal mapping over time to provide another measure of the strength of a given cross-modal correspondence (Gilbert *et al.*, 1996; O'Mahony, 1983). Finally, one might consider whether the automaticity of a cross-modal correspondence (Fumarola *et al.*, 2014; Getz and Kubovy, 2018; Spence and Deroy, 2013b, or its resistance to interference or perceptual load manipulations; see Evans, 2020) might provide another means of quantifying the strength of a correspondence. As we will see below however, the majority of the literature on individual differences in relation to the cross-modal correspondences that has been published to date has tended to focus on the first of these measures, namely on differences in the general translation tendency as a function of either normal or atypical development and as a result of cultural factors.

3. Differences Between Groups of Individuals in Cross-Modal Correspondences

The last few years has seen an explosive growth of interest in any between-groups differences in people's sensitivity to a wide range of cross-modal correspondences. One of the first studies to have assessed such differences (i.e., without confounding it with synaesthesia) was reported by Walker (1987). The latter researcher conducted a study in which 838 people were invited to pick the most appropriate (metaphorically speaking) of four pairs of visual stimuli for a range of pairs of auditory stimuli. Walker's results revealed that the consistency of the choices for four cross-modal correspondences/metaphorical cross-modal mappings (namely, auditory frequency with vertical placement, loudness with visual size, duration with length of horizontal line, and auditory waveform with visual pattern) were determined primarily by the participants' level of musical training, thus suggesting a cultural influence over the cross-modal correspondences that were exhibited (cf. Weinberger *et al.*, 2022).

3.1. Cross-Cultural Similarities and Differences in the Cross-Modal Correspondences

As has been mentioned already, a growing number of cross-cultural studies now highlight the existence of both similarities as well as some intriguing cross-cultural differences in terms of the cross-modal correspondences that have been documented to date. Early research has, for example, demonstrated the cross-cultural, and cross-linguistic, generalizability of a number of sound-symbolic correspondences (Köhler, 1929, 1947; Sapir, 1929; see

also Chen *et al.*, 2016a) (Note 3). Similarly, Bremner *et al.* (2013) documented a robust bouba–kiki effect amongst the Himba tribe of Kaokoland in Northern Namibia. This remote group of hunter-gatherers, without either schools or written language, made similar shape–sound matches, but different shape–taste matches, to what had been documented in Western consumers previously. In particular, Western consumers have often been shown to exhibit a robust cross-modal correspondence between angularity and carbonation levels in water samples, as well as between increasing bitterness and angularity in chocolate samples varying in terms of their cocoa content (cf. Ngo *et al.*, 2011, 2012). By contrast, the Himba exhibited no such correspondence between angularity and carbonation while choosing to associate bitterness with roundness more than with angularity (i.e., the opposite pattern of results to that documented in Western participants). As yet, there is no obvious explanation for this striking difference [though it may argue against the hedonic or emotional-mediation account of cross-modal correspondences, given the suggestion that sweet is universally and innately pleasant sensation (Drewnowski and Greenwood, 1983), whereas bitter is unpleasant, and potentially poisonous (see Spence and Deroy, 2013c, for a review)].

It is perhaps worth stressing here the fact that while the cross-modal correspondences that have been documented between colour and taste/odour are likely based on the internalization of the relevant statistics of the environment (i.e., of the natural environment, see Foroni *et al.*, 2016; or supermarket), drinks do not have a distinctive shape (Spence and Deroy, 2012b, 2013c). As such, the wide range of shape–taste cross-modal correspondences that have been documented (e.g., see Cytowic and Wood, 1982; Deroy and Valentin, 2011) may be based on matching the temporal qualities of tasting (meaning the way in which sweet tastes onset and offset gradually in experience, whereas sour tastes appear to onset and offset far more suddenly; see Obrist *et al.*, 2014). At the same time, however, they may also be hedonically mediated, meaning that pleasant shapes (Larson *et al.*, 2012) and tastes are matched (e.g., sweet and round) while unpleasant stimuli, such as bitter and angular, are also paired (see Velasco *et al.*, 2015). Potentially relevant here, similar cross-modal mappings between angularity and basic taste have now been reported in populations from both China and India (e.g., see Liang *et al.*, 2016).

A subtle modulation of the sound-symbolic bouba–kiki effect has also been reported between Western and Eastern participants – from UK and Taiwan, respectively (see Chen *et al.*, 2016a; and see Blasi *et al.*, 2016 and Ćwiek *et al.*, 2021, for the latest findings on sound symbolism). The difference was in terms of the point at which the majority of responses switched from one sound-symbolic response to the other (i.e., from rounded to angular). Crucially, the basic polarity of the cross-modal association was the same in both groups

of participants. However, as with a number of other studies that have documented a high degree of concordance in terms of the cross-modal matching in the presence of subtle cultural differences, it is hard to derive any straightforwardly meaningful conclusions regarding the appropriateness of discriminating between the various different classes of correspondence that have been proposed (see Table 1) (Note 4). Meanwhile, Shang and Styles (2017) reported that speakers of different languages match Mandarin Chinese tones of speech sounds/vowels [i] and [u] to the angularity of visual shapes differently. In particular, according to the authors, speakers of Mandarin Chinese classified as Chinese-dominant systematically matched a high, steady tone to the curvy shape and the falling tone to the pointy shape, whereas the English speakers with no knowledge of Chinese matched the high, steady tone to the pointy shape and a low, dipping tone to the curvy shape. The existence of such cross-cultural variation is presumably most consistent with a statistical (and possibly also emotional-mediation) account of these particular cross-modal correspondences.

Over the years, a number of striking cross-cultural differences have been reported in terms of olfactory–gustatory cross-modal associations (Blank and Mattes, 1990; see also Spence, 2022). For instance, Blank and Mattes reported how the association of anise and nutmeg aroma/flavour with sweetness (i.e., a basic taste quality) differed somewhat between the groups of white North American and non-white participants whom they happened to test. It has also been suggested that the almond smell of benzaldehyde might be more strongly associated with sweetness amongst Western consumers and with a savoury taste amongst Japanese consumers (see Spence, 2008). In both cases, note, the suggestion is that the particular taste associated with the odorant can be traced back to people's exposure to the co-occurrence in food (i.e., supporting a statistical account of such cross-modal correspondences). This explanation also appeared earlier to help explain the cross-cultural consistencies and differences that have been documented in terms of colour–taste correspondences (e.g., see Velasco *et al.*, 2016). While such cross-modal correspondences are based on cultural statistics, it is worth noting that the ingredients typically combined in recipes differ by region, or culture (e.g., Jain *et al.*, 2015; Rozin, 1983). The specific colours (hues) given to various processed food products is fairly arbitrary – the latter constitute an example of what Walker-Andrews (1994) calls “Arbitrary and Artificial” correspondences. That said, the ripening of many fruits and vegetables tends to be associated with a shift from green towards warmer hues, and that might be expected to be a universal environmental statistic (Foroni *et al.*, 2016).

Taken together, the research that has been reviewed in this section demonstrates, as one might have expected, both broad agreement in terms of the cross-modal correspondences that have been documented between cultures

along with some modest cultural differences, especially when it comes to the case of man-made products (e.g., such as food and drink). In this regard, it would be interesting to have more data explicitly comparing those cross-modal correspondences cross-culturally that are hypothesized to be based on “typical intermodal relations” (in Walker-Andrews, 1994, terminology) and/or putatively amodal properties, the prediction being that a high degree of cross-cultural agreement would be obtained in the latter case.

3.2. *On the Development of Cross-Modal Correspondences*

According to the developmental literature, very young infants are sensitive to a number of cross-modal correspondences, including those that are sound-symbolic in nature (e.g., Haryu and Kajikawa, 2012; Lewkowicz and Turkewitz, 1980; Maurer *et al.*, 2006; Mondlock and Maurer, 2004; Ozturk *et al.*, 2013; Pejovic and Molnar, 2017; Peña *et al.*, 2011; Walker *et al.*, 2010). The literature that has been published to date shows that many of the correspondences appear to develop in a seemingly predictable sequence over the first decade or so of life – meaning that a sensitivity to additional cross-modal correspondences is acquired (i.e., rather than lost) with increasing age. At the same time, however, it is important to highlight the fact that some inconsistencies have also been reported in the cross-modal correspondences that have been demonstrated at various ages. The latter may, however, in part, simply be explainable in terms of differences in whether explicit or implicit response measures have been used, as well as by various other methodological features, and modality combinations that have been studied by the researchers concerned. (The use of dynamic vs static stimuli might also be relevant here; Jeschonek *et al.*, 2013.) One of the challenges with assessing pre-linguistic infants’ abilities has been to avoid falling into the trap of simply assuming that any cross-modal effects that are observed in this group should necessarily be interpreted in terms of synaesthesia (e.g., see Dailey *et al.*, 1997; Deroy and Spence, 2013a, b; Domino, 1989; Maurer, 1997; Rader and Tellegen, 1987; Wagner and Dobkins, 2011). Another challenge has been to distinguish between cross-modal correspondences and other cross-modal mappings that may be more metaphorical in nature (Gardner, 1974; Marks *et al.*, 1987; Wagner *et al.*, 1981; Walker, 1987) (Note 5).

Walker *et al.* (2010) examined preferential looking behaviour in 16 three- to four-month-old preverbal infants. These researchers demonstrated a cross-modal correspondence linking auditory pitch (a sound changing continuously between 300 and 1700 Hz) to oscillating height (i.e., elevation) of a visual stimulus and sharpness of a dynamically-changing visual stimulus (in 12 out of the 16 participants). The magnitude of the looking preference was equivalent for the two stimuli, leading the authors to argue that this reflects both

an unlearned aspect of perception and that neonatal perception is fundamentally synaesthetic in nature (Note 6). Meanwhile, research from Haryu and Kajikawa (2012) in Japan suggests that 10-month-old infants ($N = 16$) are implicitly sensitive to the cross-modal correspondence between auditory pitch and visual brightness (as assessed by a violations of expectation task). However, no cross-modal correspondence was observed between auditory pitch and visual size in this particular age group ($N = 16$). Consistent with earlier findings reported by Mondloch and Maurer (2004) in a group of 12 30–36-month-old infants, all exhibited a preference for high-frequency sounds and brighter objects, while only 80% of the sample were sensitive to the pitch–size correspondence. Such a developmental pattern was taken to support the innate basis of the pitch–brightness correspondence and the statistical account of the pitch–size mapping.

One might have expected that any putatively structural correspondences, such as between auditory loudness and visual brightness, would emerge prior to more arbitrary and artificial statistical correspondences (cf. Loconsole *et al.*, 2022; Walker-Andrews, 1994). According to Marks *et al.* (1987), certain cross-modal correspondences likely strengthen with age. The latter researchers also suggest that the first cross-modal correspondences to develop are likely to involve prothetic (e.g., magnitude) scaling, and thereafter, the mapping of polar dimensions (Smith and Sera, 1992). There is also a sense that certain correspondences cannot be established until the underlying concept, or perceptual dimension, is understood by the developing child. In particular, Marks *et al.* tested 500 children aged 3.5–13.5 years of age in response to a range of audiovisual cross-modal correspondences. Importantly, pitch–loudness and pitch–brightness cross-modal correspondences were observed by all age groups (including adults). By contrast, the children only became sensitive to the auditory pitch–visual size correspondences when they reached around 11 years of age (see Marks *et al.*, 1987). Marks *et al.* were led to conclude that the latter correspondence might only be acquired as the result of multisensory experience.

According to Lewkowicz and Turkewitz (1980), the auditory loudness to visual brightness correspondence can be considered as an example of the implicit amodal matching of stimulus intensity. The latter researchers claimed to have shown that at 30 days of age, infants already treated loud sounds and bright lights as cross-modally equivalent. That said, not everyone necessarily agrees on the existence of amodal qualities (see Spence *et al.*, 2013; Spence and Di Stefano, submitted). Whatever the nature of the relationship between loudness and brightness, the developmental research that has been published to date would appear to show that it is one of the first correspondences that very young children exhibit a sensitivity to (see Table 2). Meanwhile, Wagner

et al. (1981) found no evidence of preferential looking for visual size and auditory loudness among a sample of 6–14-month-old infants though a number of them showed a sensitivity to broken/continuous line–pulsing/continuous tone, and jagged/smooth circle–pulsing/continuous tone, up/down pointing arrow, and tone ascending or descending in pitch (though the pitch range was not specified; cf. Parise *et al.*, 2014). Intriguingly, these researchers described their findings in terms of metaphoric matches rather than necessarily as cross-modal correspondences.

Nava *et al.* (2016) demonstrated the existence of cross-modal correspondences between increasing/decreasing auditory pitch and the vertical motion (i.e., upwards vs downwards) of a visual stimulus (as in the Barber pole illusion) in 4–5-years-old children. The researchers also documented weak correspondences for the vertical movement of a tactile stimulus delivered to the participant's back (audiotactile condition) and the spatial motion of visual objects to touch (visuotactile condition). According to Nava *et al.*'s results, pitch–height correspondences for audiovisual and audiotactile combinations are weak in pre-school-aged children. The authors suggested that the limited sensitivity of their young participants to such correspondences may have reflected limited language development and also perhaps the immature development of the auditory system. By contrast, the adults tested in this study exhibited near-perfect agreement in matching the direction of vertical motion with the direction of auditory pitch change, regardless of the pairing of modalities that was assessed, suggesting that, whatever the cause, the strength of cross-modal correspondences varies with age.

Recently, Speed *et al.* (2021) assessed a relatively wide range of different cross-modal associations with olfactory, auditory, and tactile stimuli in both children and adults. These researchers assessed cross-modal correspondences between auditory pitch, tactile (shape/texture) properties, colours, and a number of odours associated with familiar everyday stimuli (such as the smell of caramel, lemon, menthol, onion, and raspberry). Their results demonstrated little evidence of consensual cross-modal associations in the youngest age group (4–5-years-old) for any of the modality combinations that were tested. At the same time, however, Speed *et al.* also reported that the correspondences tended to strengthen markedly in the older age groups (6–9 years, 10–18 years, and 19+ years old). The authors argued that this pattern of results was consistent with the importance of exposure to statistical associations in the environment to the establishment of cross-modal correspondences. They also argued that it highlighted the important role played by language and semantics (e.g., in the case of matching colours to odours having a familiar smells – such as the smell of caramel being associated with the colour brown; see also Metatla *et al.*, 2019; Spence, 2020b).

Table 2.

Individual differences in different kinds of crossmodal correspondence that have been documented to date

Crossmodal mapping	Correspondence and relevant individual differences that have been reported
Statistical correspondence	
Typical intermodal relation	<p>Pitch–brightness: Demonstrated in 10-month-olds by Hidaka and Kajikawa (2018), and shown by all 30–36-month-olds by Mondloch and Maurer (2004; see also Marks <i>et al.</i>, 1987)</p> <p>Loudness–brightness: Available at 30 days of age (Lewkowicz and Turkewitz, 1980), though weaker in those with higher Autism Quotient (Hidaka and Yaguchi, 2018)</p> <p>Pitch–loudness: Demonstrated at 3.5 years upwards by Marks <i>et al.</i> (1987)</p> <p>Pitch–visual size: Develops to adult-like levels only at 11 years of age (Marks <i>et al.</i>, 1987). Only 80% of 30–36-month-olds show this correspondence in Mondloch and Maurer (2004), while Hidaka and Kajikawa (2018) find that 10-month-olds are not sensitive to this correspondence</p> <p>Pitch–shape (angularity): Available very early (in 75% of 3.5-month-olds; Walker <i>et al.</i>, 2010)</p> <p>Pitch–elevation: Available very early according to Walker <i>et al.</i> (2010; i.e., in 75% of 3.5-month-olds, though see also Marks <i>et al.</i>, 1987; Nava <i>et al.</i>, 2016)</p> <p>No evidence for loudness–visual size correspondence in 6–14-month-olds</p>
Arbitrary and Artificial	<p>Colour–taste and colour–smell correspondences show cross-cultural variation (Levitan <i>et al.</i>, 2014; Spence <i>et al.</i>, 2015; Velasco <i>et al.</i>, 2016; Wan <i>et al.</i>, 2014) and consensuality of mapping within culture increases with age (Speed <i>et al.</i>, 2021)</p>
Metaphorical crossmodal mapping	<p>Develops to adult-like levels of performance during childhood (e.g., Wagner <i>et al.</i>, 1981); may also be influenced by musical training (Walker, 1987)</p>
Emotionally-mediated (hedonic) correspondence	<p>Shape–taste and shape–smell correspondences acquired during development (Chen <i>et al.</i>, 2021; Metatla <i>et al.</i>, 2019; Speed <i>et al.</i>, 2021), and weaker in high-Autism-Quotient individuals (Chen <i>et al.</i>, 2021)</p>
Analogical	<p>Individual differences in second-order reasoning between senses linked to general intelligence factor g (Weinberger <i>et al.</i>, 2022; see also Jewanski, 2010)</p>

It is interesting to note that the emergence of multiple correspondences in the 6–9-years-old age group in Speed *et al.*'s (2021) study seems consistent with a general inflection point in terms of multisensory development that

appears to occur at around nine years of age (e.g., Gori *et al.*, 2008; Jüttner *et al.*, 2006; Vaught *et al.*, 1975). For example, according to research from Gori *et al.*, visuohaptic integration becomes optimal at around age 8–10 years; Meanwhile, according to Vaught *et al.*, haptic-visual performance reaches a steady-state around age nine years. Jüttner *et al.* report that 9–10-year-olds but not 8–9-year-olds have been reported to show view independence. In future research, it would be especially interesting to determine whether particular classes of cross-modal correspondence are more likely to come on line following this inflection point than others.

In conclusion, the developmental research that has been reviewed in this section would appear to demonstrate something of a predictable developmental trajectory concerning the acquisition of different (kinds of) cross-modal correspondences. While the development of a sensitivity to certain correspondences undoubtedly depends on a child's acquisition of the relevant underlying concept, or perceptual dimension (e.g., 'density', Marks *et al.*, 1987), others may be based on structural (e.g., Lewkowicz and Turkewitz, 1980; Spence, 2011), or physiologically-based (Spence and Di Stefano, 2022), correspondences instead. Often, as has been mentioned already, the suggestion has been made by the researchers concerned that those correspondences evidenced in the first few months of life are unlikely to have been learnt from the statistics of the environment (Lewkowicz and Turkewitz, 1980; Spector and Maurer, 2012). This has led to the suggestion that such correspondences may be innate. Of course, such claims should not be accepted without a careful evaluation of the supporting evidence. At the same time, however, the gradual development of a sensitivity to other cross-modal correspondences (Speed *et al.*, 2021), as well as the developmental trend documented amongst certain audiovisual size correspondences to develop from a coarse to a more fine-grained mapping during primary-school age (Cuturi *et al.*, 2019), suggests a role for experience and/or sensory/intellectual/language development (see also Goubet *et al.*, 2018) (see Table 2 once again).

3.3. *Autism Spectrum and Individual Differences in Cross-Modal Correspondences*

The research that has been published to date shows that those scoring high on the autism spectrum appear to be less sensitive to certain of the correspondences that have been demonstrated in the normal population, such as, for example, the so-called bouba–kiki effect (e.g., Gold and Segal, 2017; Król and Ferenc, 2020; Stewart *et al.*, 2016). In one of the first studies of its kind to document such individual differences, Oberman and Ramachandran (2008) provided some preliminary evidence for a reduced strength of cross-modal correspondences in a group of 10 individuals with autism spectrum disorder (ASD). In particular, their neurotypical group exhibited 88% consensuality on

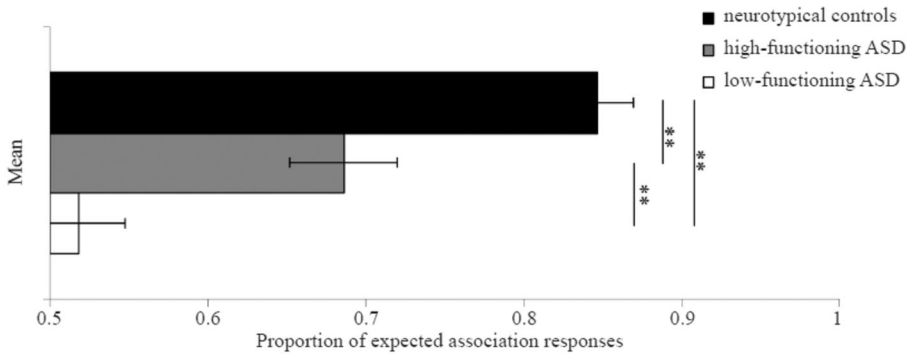


Figure 3. Mean proportion of participants in the groups of children with low-functioning autism spectrum disorders (ASDs) ($N = 20$), high-functioning ASDs ($N = 15$), and neurotypical controls ($N = 37$) who chose the corresponding name for each pair of stimuli. The error bars represent the standard error of the means. The significant comparisons are indicated by asterisks (**, $p < 0.001$). Figure reprinted from Ocelli *et al.* (2013).

the boubá–kiki effect, as compared to just 56% for those with ASD (chance being 50%). Subsequently, Ocelli *et al.* (2013) conducted a larger study of the takete–maluma phenomenon in a group with autism spectrum disorders. The latter researchers compared the performance of a group of 37 neurotypical children/adolescents (7–19 years of age) with that of both low- and high-functioning ASD children ($N = 20$ and 15, respectively) on matching 14 pairs of meaningless angular vs rounded line drawings (rectilinear vs curvilinear, respectively) with 14 meaningless pairs of phoneme clusters. The results (see Fig. 3) revealed no evidence for a sound-symbolic correspondence in the low-functioning autistic group with an intermediate level of performance being documented in the high-functioning group (Note 7).

In another study, Hidaka and Yaguchi (2018) investigated the relationship between autistic traits and cross-modal correspondences in a typically developing adult Japanese population ($N = 87$). These researchers found that, among three types of cross-modal correspondences tested (namely, brightness–loudness, visual size–pitch, and visual elevation–pitch), the brightness–loudness correspondence was related with total ASD traits and a sub-trait (social skill) when the participants were split into higher and lower ASD groups, respectively. Specifically, those in the latter group exhibited a significantly larger cross-modal congruency effect in the speeded visual classification task, thus suggesting that they experienced more of a correspondence between the component auditory and visual stimuli than did those in the group with higher ASD traits (see Fig. 4). Relevant to the earlier discussion, it is worth noting here that brightness–loudness might reflect a structural (or, according to some, a possibly amodal) correspondence (Lewkowicz and Turkewitz, 1980; Spence, 2011), whereas visual size–pitch and visual

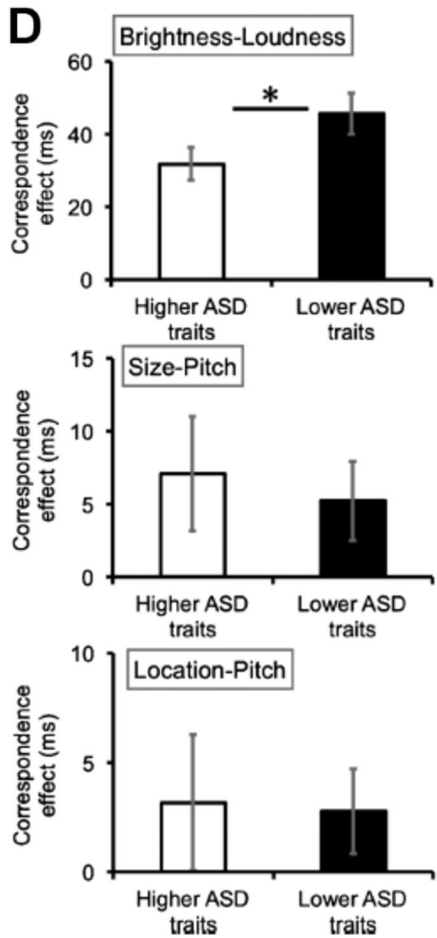


Figure 4. (D) The correspondence effect in the higher and lower total Autism-Spectrum Quotient (AQ) score groups for each session. (*, $p < 0.05$). ASD, autism spectrum disorder. Figure reprinted from Hidaka and Yaguchi (2018).

elevation–pitch are both often discussed as statistical correspondences based on the internalization of robust environmental regularities (e.g., Bernstein and Edelstein, 1971; Evans and Treisman, 2010; Gallace and Spence, 2006; Parise *et al.*, 2014; see also Stewart *et al.*, 2016).

Those individuals with high autistic traits have recently been reported to exhibit less consensual cross-modal correspondences between visual features and basic tastes (Chen *et al.*, 2021). In particular, the latter researchers conducted an online study in Japan with a convenience sample of 85 adults whom they divided into three groups (with a low, medium, or high Autism Quotient). They assessed colour–taste, shape–taste, and shape–colour correspondences

(Note 8). The results revealed that the likelihood of choosing the consensual response was significantly associated with an individual's autistic traits. In particular, those participants in the higher autistic quotient group (according to the tripartite division) tended to choose fewer consensual colour–taste and shape–colour correspondences, while no difference was reported in the case of shape–taste mappings (see Fig. 5). Chen *et al.* put forward a statistical account of their findings. Furthermore, it is perhaps also worth repeating the point made earlier that flavourful stimuli nearly always have a statistically associated colour, whereas shape (e.g., in the case of a drink) is typically determined by the receptacle in which a drink happens to be presented. Shape, in other words, is not necessarily so closely linked to taste as is colour, and the link may be more emotional in origin.

Taken together, the results reported in this section would appear to suggest that autistic individuals and those normal participants with a high Autism Quotient show weaker cross-modal correspondences (Note 9). That is, perhaps autistic tendencies might interfere with an individual's feeling concerning the consensuality of the mapping/response) The fact that (potentially amodal) loudness–brightness correspondence is impaired in those with autistic tendencies is a little surprising, though, given that this has been postulated to be an amodal correspondence (see Lewkowicz and Turkewitz, 1980; though see Spence and Di Stefano, submitted).

3.4. *Sensory Loss and Cross-Modal Correspondences*

The last few years have seen an explosion of interest concerning the status of sound-symbolic and other cross-modal correspondences in those who are blind (e.g., Bottini *et al.*, 2019; Graven and Desebrock, 2018; Hamilton-Fletcher *et al.*, 2018) (Note 10). For example, Hamilton-Fletcher *et al.* examined the role of visual experience in the emergence of various audiotactile cross-modal correspondences. Intriguingly, they observed both differences and similarities in terms of the correspondences documented in the blind and sighted groups of participants whom they tested ($N = 59$ early and late blind, and 63 sighted controls). In particular, the congenitally blind participants who they tested ($N = 32$) were found to be insensitive to the pitch–shape correspondence that has so often been documented in sighted participants (e.g., Marks, 1987; O'Boyle and Tarte, 1980; Parise and Spence, 2012; Walker, 2012) (Note 11). Meanwhile, the pitch–size, and to a lesser extent pitch–weight, cross-modal correspondences were unaffected by an individual's prior visual experience (namely normally sighted, early or late blind, $N = 27$). Furthermore, and somewhat surprisingly, the pitch–texture and pitch–softness cross-modal correspondences were only documented in the early-blind participants.

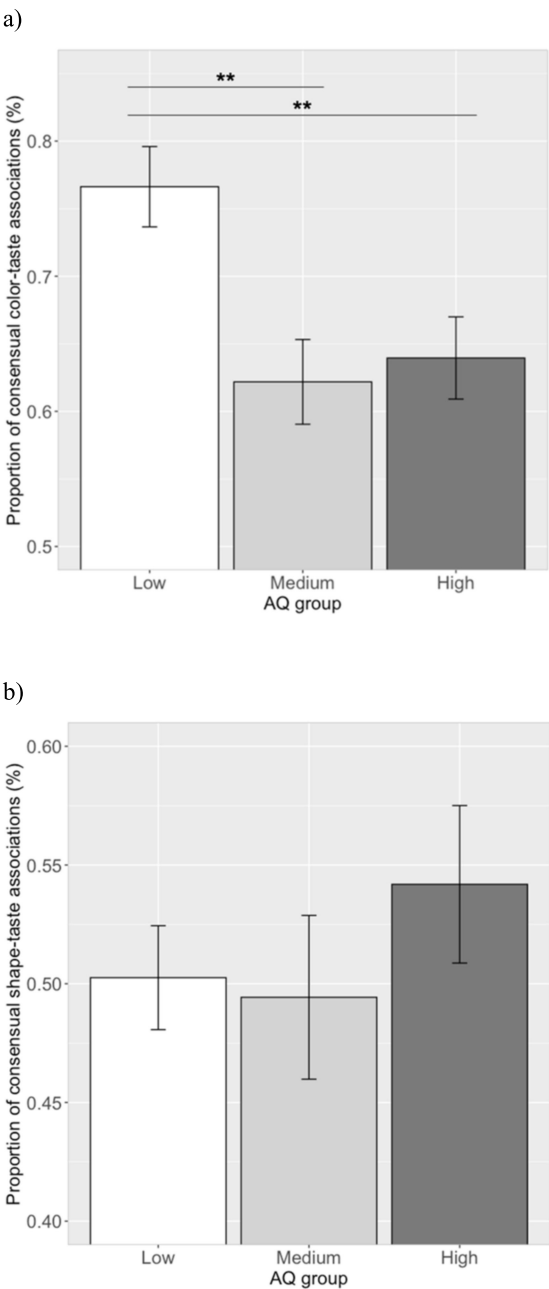


Figure 5. Proportion of consensual colour–taste associations (a) and shape–taste associations (b) in the three Autism-Spectrum Quotient (AQ) groups. Error bar represents the standard mean error. Asterisks indicate significant differences (**, $p < 0.01$). Figures reprinted from Chen *et al.* (2021).

Intriguingly, mixed data have been published concerning pitch–elevation cross-modal correspondences in sighted, late, and congenitally blind participants (e.g., see Deroy *et al.*, 2016; Eitan *et al.*, 2012; Occelli *et al.*, 2009). For instance, Deroy *et al.* failed to find any evidence for increasing pitch being associated with tactile motion along the finger in either early- or late-blind participants (≤ 3 or > 5 years, respectively) using an IAT, while demonstrating an effect in the sighted participants. By contrast, Eitan *et al.* reported that the majority of cross-modal mappings involving musical auditory stimuli were preserved in the blind participants whom they tested. Elsewhere, Barilari *et al.* (2018) conducted a study on a group of 46 early-blind individuals (who had all become totally blind before the age of three years) and a control group of 46 normally-sighted individuals. Their results showed that the early blind also thought that red is heavier than yellow, though the strength of this particular cross-modal correspondence was somewhat weaker than that observed in the sighted control participants.

According to research from Sourav *et al.* (2019), there may be a protracted sensitive period regulating the development of cross-modal sound–shape associations in humans. The latter researchers investigated the role of visual experience in the development of both audiohaptic and audiovisual (where possible) cross-modal correspondences using five groups of participants. The typically sighted ($N = 70$), and the late permanent blind groups of participants ($N = 12$), showed highly robust cross-modal correspondences between sound (the spoken words ‘bouba’ and ‘kiki’) and angular vs rounded felt (or seen) shape (haptics or vision, respectively). Meanwhile, those in the congenital permanent blindness group ($N = 15$), as well as those individuals who had suffered a transient period of congenital blindness as a result of congenital bilateral dense cataracts before undergoing cataract-reversal surgeries ($N = 30$), and those individuals with a history of developmental cataracts groups did not ($N = 24$). Taken together, therefore, these results provide support for the existence of a protracted sensitive (or critical) period during which aberrant vision would appear to prevent the possibility of acquiring sound-symbolic–shape cross-modal correspondences later in life (see also Fryer *et al.*, 2014). However, while the stimuli used by Sourev *et al.* are described as assessing sound–shape associations, a closer look at two of the pairs of tactile stimuli used suggests that they might equally well be described as differing in terms of their surface texture as much as they do in terms of their shape. It would be interesting to know whether this critical period for visual experience also affects the internalization of other basic perceptual correspondences. Even late permanent blindness does not appear to eliminate cross-modal correspondences. In particular, ontogenetically early visual input (defined here as less than 12 years of age) would appear to drive the acquisition of certain sound-symbolic associations. It would be intriguing to know, therefore, about

the role of early vision in the establishment of other cross-modal correspondences, including both structural and ubiquitous statistical associations.

3.5. *Synaesthetic Tendencies and Sensitivity to Cross-Modal Correspondences*

One could imagine that more creative artists/designers might be more sensitive to cross-modal correspondences than those who are less creative. That said, while creativity has often been associated with synaesthesia, it need not be (cf. Weinberger *et al.*, 2022). Indeed, many of the studies would appear to have confounded cross-modal correspondences with synaesthesia. Intriguingly, Dailey *et al.* (1997) conducted an intriguing study of 52 male undergraduate North American psychology students in which they assessed the strength of what they describe as “synaesthesia-like correspondences” (Dailey *et al.*, 1997, p. 3). A median split on participants’ performance on the Remote Associates Test (Mednick and Mednick, 1967) gave rise to two groups, differing in terms of their creativity. All of the participants were then presented with six pure tones (150, 300, 600, 1200, 2400, and 4800 Hz), and separately with 12 vowel sounds. They were shown each of six colour patches (blue, purple, green, red, orange, and yellow) and requested to rate the similarity of the auditory and visual stimuli on a 7-point scale (from not at all, 1, to very much, 7).

Although the results for the two groups on the two cross-modal matching tasks looked pretty similar, statistical analysis nevertheless revealed a three-way interaction between Creativity, Pitch (or Vowel sound), and Hue. The authors interpreted this finding as showing that: “[m]ore creative participants differed significantly from less creative participants on their ratings, with more creative participants exhibiting stronger associations between colors and pure tones, vowels, and emotional terms” (Dailey *et al.*, 1997, p. 1). That is, more creative individuals made stronger associations between all low-pitched sounds and shorter-wavelength colours, and between all high-pitched tones and longer-wavelength colours as compared to less creative participants. The individual difference in the strength of associations was specifically in terms of rated similarity. Importantly, however, it should be noted that prototypical colours were used in Dailey *et al.*’s study (here one might refer again to Table 2), meaning that the relative importance of hue, saturation, and brightness cannot easily be disentangled based on the results of this study.

While the question of whether synaesthetes show enhanced multisensory integration relative to non-synaesthetes, and if so, under what conditions, has received mixed support from researchers (e.g., Brang, Williams and Ramachandran, 2012; Neufeld *et al.*, 2012; Whittingham *et al.*, 2014; see also Mulvenna and Walsh, 2006; Ward *et al.*, 2006), one of the few studies to explicitly assess the strength of cross-modal correspondences was reported

by Lacey *et al.* (2016). According to the latter researchers, those individuals with synaesthesia ($N = 17$) exhibited stronger sound-symbolic cross-modal correspondences (between auditory pseudo-words and visual shapes). In particular, these researchers tested the strength of cross-modal correspondences using a simplified version of the IAT (cf. Parise and Spence, 2012). Their results demonstrated that synaesthetes were more sensitive than non-synaesthetes ($N = 18$) to sound-symbolic cross-modal correspondence, but there was no difference between the groups in terms of low-level sensory associations between auditory pitch and visual size and auditory pitch and visual elevation.

4. Experimentally-Induced Correspondences

It is interesting to note how several laboratory studies have presented their participants with statistical correlations between pairs of otherwise unrelated unisensory dimensions, such as, for example, luminance and haptic stiffness (Ernst, 2007; see also Baier *et al.*, 2006; Flanagan *et al.*, 2008; Zangenehpour and Zatorre, 2010, for a few other examples of associative learning of arbitrary combinations of auditory and visual stimuli). For example, Ernst (2007) exposed participants to an arbitrary statistical mapping between the luminance of a visual object and its felt stiffness, a haptically ascertained stimulus property that is not correlated with luminance in the natural environment. The participants were trained with multisensory stimuli where an artificial correlation had been introduced between the stimulus dimensions: For some of the participants, the stiffer the object, the brighter it appeared, while this mapping was reversed for others. Ernst's results highlighted a significant change in participants' discrimination performance when their responses to congruent and incongruent pairs of haptic stimuli were compared before and after training. These changes were attributable to changes in the distribution of the coupling prior. The training took place in a single testing session lasting for between 1.5 and 2.5 hours. Presumably the coupling priors for those stimulus dimensions that have been correlated over the course of a person's lifetime are likely to be correspondingly higher.

Given that relatively short associative training regimes have been shown to bias multisensory integration across a range of behavioural/perceptual tasks (see also Brunel *et al.*, 2015; Connolly, 2014; Tong *et al.*, 2021) (Note 12), it would seem that such training protocols may be said to induce differences between individuals in terms of their sensitivity to the specific (i.e., trained) correspondence, which is presumably internalized as a coupling prior in terms of Bayesian decision theory (Parise and Spence, 2013; see also Körding *et al.*, 2007, on the causal inference model of multisensory integration). That said, it is not clear how long such cross-modal correspondences last after the study has

been completed. As such, the researchers behind these studies might best be considered to have induced a *phasic* cross-modal correspondence in contrast to the *tonic* cross-modal correspondences that have typically been documented elsewhere in the literature (Note 13). Such short-lasting, or temporary, correspondences can presumably also be considered weak. Alternatively, however, one might want to distinguish in some way between such temporarily-induced and other seemingly more long-lasting correspondences in terms of a difference in kind.

In the context of multisensory flavour perception, Stevenson and his colleagues have also conducted a number of studies demonstrating that unfamiliar odorants can take on the taste properties they happen to be associated with (see Stevenson and Boakes, 2004; Stevenson *et al.*, 1998). Importantly, these researchers demonstrated that this can occur after only a surprisingly small number of co-exposures (i.e., no more than three), and even without the participants being necessarily aware of the tastant's presence in the combined stimulus. Taken together, the research that has been reviewed in this subsection clearly demonstrates that new cross-modal correspondences can be acquired rapidly given exposure to the appropriate environmental statistics. However, as yet there is little evidence concerning how long-lasting such correspondences may be (this possibly considered as one element of the strength of a cross-modal correspondence). There may also be fundamental differences in the rate of learning across different senses (with olfactory–taste correspondences seemingly learnt much more rapidly than other modality combinations) that may override the differences that one might want to study in terms of the kind of correspondence (e.g., when comparing statistical with structural).

Nevertheless, it is easy to imagine that it would be much more difficult to change, or update, a structural cross-modal mapping, such as the putative correspondence between stimulus intensity (i.e., if participants were exposed to an artificial environment in which auditory loudness was inversely correlated with visual brightness, say), than it would be to update a statistical correspondence that is based on a weak statistical relationship between two stimulus dimensions. Here, one might consider putatively amodal stimulus dimensions, such as shape, as providing highly correlated visual and tactile information.

5. Conclusions

As the literature on the cross-modal correspondences continues to flourish (e.g., see Spence, 2011; Spence and Sathian, 2020, for reviews), the interest of several groups of researchers has increasingly started to move beyond merely documenting the existence of an ever-expanding range of (more-or-less surprising) correspondences between attributes, features, or dimensions of experience in the different senses. Increasingly, researchers are starting

to assess the ubiquity of such cross-modal correspondences across different groups of individuals. This rapidly-emerging body of research has undoubtedly helped to highlight the existence of a number of intriguing developmental (and neuro-atypical) trends in the appearance of various cross-modal correspondences (e.g., Gold *et al.*, 2022; Marks *et al.*, 1987; Nava *et al.*, 2016; Spector and Maurer, 2012; Speed *et al.*, 2021). According to Marks *et al.*, a number of the cross-modal correspondences based on more complex perceptual dimensions (or acquired as a result of experience) simply do not reach adult-like levels of consistency until the second decade of life (see also Cutietta and Haggerty, 1987). The evidence is currently rather mixed though, with the results depending on the method used to study the correspondence, the nature of the correspondence under investigation, and the nature of the sensory loss (e.g., Hamilton-Fletcher *et al.*, 2018; Sourav *et al.*, 2019). In particular, intriguing research from Sourav *et al.* has convincingly demonstrated the existence of a critical period for visual experience in the establishment of certain sound-symbolic pseudoword–shape correspondences. Autistic tendencies also appear to influence the strength of various cross-modal correspondences (e.g., Chen *et al.*, 2021; Hidaka and Yaguchi, 2018; Oberman and Ramachandran, 2008; Occelli *et al.*, 2013). Individual differences in synaesthetic tendencies (or creativity) and well as musical experience (i.e., a cultural factor) have also been shown to influence the strength of various cross-modal correspondences (e.g., Dailey *et al.*, 1997; Rader and Tellegen, 1987; Walker, 1987; see also Chun and Hupé, 2016; Weinberger *et al.*, 2022).

If one accepts the distinction between multiple different causes (or types) of cross-modal correspondences (see Spence, 2011; Spence and Di Stefano, 2022; Spence and Sathian, 2020; and see Table 1 for a summary), then it is interesting to consider whether certain kinds of cross-modal correspondences (namely, structural or physiological, statistical, semantic or lexical, and emotional or hedonic) are more vs less subject to different kinds of individual difference (see Chen *et al.*, 2021; Hidaka and Yaguchi, 2018; Parise, 2016, on this theme; and see Table 2). As mentioned earlier, it is also important to stress the various ways in which such individual differences might express themselves – namely, as differences in the strength of particular correspondences, differences in vividness or naturalness, and/or differences in whether a particular correspondence is experienced or not. Intriguingly, the evidence is currently rather mixed with regard to the extent to which consensuality of various cross-modal correspondences tracks (or is correlated with) the strength of the correspondence; e.g., see Chen *et al.*, 2016a; Parise, 2016; Wan *et al.*, 2014, on this theme).

At the same time, however, one of the challenges currently is the relatively limited amount of data that have been collected in relation to group-level differences between individuals to date. Moreover, there would appear to be

an assumption amongst researchers that various correspondences are universal/amodal and hence the country in which the data is collected is treated as irrelevant (e.g., as when brightness–loudness correspondences are studied in different countries). Another challenge concerns the continuing uncertainty over the very existence of various of the different types of cross-modal correspondence that have been proposed (see Spence, 2011; Spence and Sathian, 2020) (Note 14). Potentially relevant here, the latest evidence would appear to support a potential distinction between sound-symbolic cross-modal correspondences and those that are based on more basic perceptual features (e.g., Bottini *et al.*, 2019; Lacey *et al.*, 2016; Margiotoudi *et al.*, 2022; see also Hung *et al.*, 2017).

Unfortunately, the confusion introduced by a number of earlier researchers around the theme of so-called individual differences in synaesthetic tendencies has not helped to clarify matters in this area (see Dailey *et al.*, 1997; Domino, 1989; Rader and Tellegen, 1987). Nevertheless, as the importance of the cross-modal correspondences to those interested in trying to understand multisensory integration/perception becomes ever clearer (see Miller *et al.*, 1958; Parise and Spence, 2009), it is likely that the various causes, and consequences, of such individual/group differences will likely only grow in relevance. In the years ahead, it is to be hoped that progress will both help to clarify which types of cross-modal correspondence are involved and in which way such group-level differences between individuals express themselves (Parise, 2016).

Notes

1. Chance level responding would be expected to give 1/6, or c. 17% agreement in this case.
2. That said, universal should certainly not be treated as being synonymous with innate.
3. The only exception here is for those languages in which pseudoword phonemes are either absent or else violate phonotactic rules (see Rogers and Ross, 1975; Styles and Gawne, 2017).
4. Though, once again, there is still no plausible explanation for the cross-cultural differences that have been reported (e.g., Bremner *et al.*, 2013).
5. Oftentimes, though by no means always, either label has been given to one and the same cross-modal association.
6. Though note that Walker *et al.*'s (2010) findings have not gone unchallenged (see Lewkowicz and Minar, 2014; and Walker *et al.*, 2014).

7. Taken together, it is unclear whether the lack of consensual mapping in these autistic individuals is specifically linked to the use of verbal stimuli, given the suggestion of reduced neural activity in Broca's area in autistic individuals.
8. Note that there were 11 colours, 15 shapes, and five basic tastes to choose from.
9. One might only wonder how weak 'theory of mind' (Baron-Cohen, 1995) would influence/impact cross-modal correspondences based on consensuality of the mapping.
10. Cf. Chen *et al.* (2016b) on colour–shape associations in both deaf and hearing people.
11. And which may also be available in the statistics of the environment according to Walker *et al.* (2010), given that harder objects tend to shatter into more angular pieces and also make a higher-pitched sound when struck.
12. Note that Hidaka and Yaguchi (2018) also demonstrate that those with a higher Autism-Spectrum Quotient (AQ), specifically attention switching component/trait of ASD, were slower to associatively learn a new cross-modal correspondence between horizontal visual apparent motion direction and pitch pairs (i.e., low-high, or vice versa).
13. That said, no one has yet, at least not as far as I am aware, suggested a 'Test of Consistency' for the cross-modal correspondences equivalent to that used by those researchers interested in synaesthesia (e.g., see Baron-Cohen *et al.*, 1993).
14. See, for example, Sadaghiani *et al.* (2009) for one attempt to use neuroimaging to try and distinguish between natural, metaphorical, and linguistic correspondences.

References

- Ayabe-Kanamura, S., Schicker, I., Laska, M., Hudson, R., Distel, H., Kobayakawa, T. and Saito, S. (1998). Differences in perception of everyday odors: a Japanese–German cross-cultural study, *Chem. Senses* **23**, 31–38. DOI:10.1093/chemse/23.1.31.
- Baier, B., Kleinschmidt, A. and Müller, N. G. (2006). Cross-modal processing in early visual and auditory cortices depends on the statistical relation of multisensory information, *J. Neurosci.* **26**, 12260–12265. DOI:10.1523/JNEUROSCI.1457-06.2006.
- Barbosa Escobar, F., Wang, Q. J., Corredor, A. and Velasco, C. (2022). The taste of visual textures, *Food Qual. Pref.* **100**, 104602. DOI:10.1016/j.foodqual.2022.104602.

- Barenholtz, E., Lewkowicz, D. J., Davidson, M. and Mavica, L. (2014). Categorical congruence facilitates multisensory associative learning, *Psychon. Bull. Rev.* **21**, 1346–1352. DOI:10.3758/s13423-014-0612-7.
- Barilari, M., de Heering, A., Crollen, V., Collignon, O. and Bottini, R. (2018). Is red heavier than yellow even for blind?, *i-Perception* **9**, 2041669518759123. DOI:10.1177/2041669518759123.
- Baron-Cohen, S. (1995). *Mindblindness: an Essay on Autism and Theory of Mind*. MIT Press, Cambridge, MA, USA.
- Baron-Cohen, S., Harrison, J., Goldstein, L. H. and Wyke, M. (1993). Coloured speech perception: is synaesthesia what happens when modularity breaks down?, *Perception* **22**, 419–426. DOI:10.1068/p220419.
- Belkin, K., Martin, R., Kemp, S. E. and Gilbert, A. N. (1997). Auditory pitch as a perceptual analogue to odor quality, *Psychol. Sci.* **8**, 340–342. DOI:10.1111/j.1467-9280.1997.tb00450.x.
- Bernstein, I. H. and Edelstein, B. A. (1971). Effects of some variations in auditory input upon visual choice reaction time, *J. Exp. Psychol.* **87**, 241–247. DOI:10.1037/h0030524.
- Blank, D. M. and Mattes, R. D. (1990). Sugar and spice: similarities and sensory attributes, *Nursing Res.* **39**, 290–293.
- Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F. and Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages, *Proc. Natl Acad. Sci. U S A* **113**, 10818–10823. DOI:10.1073/pnas.1605782113.
- Blazhenkova, O. and Kumar, M. M. (2018). Angular versus curved shapes: correspondences and emotional processing perception, *Perception* **47**, 67–89. DOI:10.1177/0301006617731048.
- Bonetti, L. and Costa, M. (2018). Pitch-verticality and pitch-size cross-modal interactions, *Psychol. Music* **46**, 340–356. DOI:10.1177/0305735617710734.
- Bottini, R., Barilari, M. and Collignon, O. (2019). Sound symbolism in sighted and blind: the role of vision and orthography in sound–shape correspondences, *Cognition* **185**, 62–70. DOI:10.1016/j.cognition.2019.01.006.
- Brang, D., Williams, L. E. and Ramachandran, V. S. (2012). Grapheme-color synesthetes show enhanced crossmodal processing between auditory and visual modalities, *Cortex* **48**, 630–637. DOI:10.1016/j.cortex.2011.06.008.
- Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J. and Spence, C. (2013). “bouba” and “Kiki” in Namibia? A remote culture make similar shape–sound matches, but different shape–taste matches to westerners, *Cognition* **126**, 165–172. DOI:10.1016/j.cognition.2012.09.007.
- Brunel, L., Carvalho, P. F. and Goldstone, R. L. (2015). It does belong together: cross-modal correspondences influence cross-modal integration during perceptual learning, *Front. Psychol.* **6**, 358.8. DOI:10.3389/fpsyg.2015.00358.
- Chen, N., Tanaka, K., Namatame, M. and Watanabe, K. (2016b). Color-shape associations in deaf and hearing people, *Front. Psychol.* **7**, 355. DOI:10.3389/fpsyg.2016.00355.
- Chen, N., Watanabe, K. and Wada, M. (2021). People with high autistic traits show fewer consensual crossmodal correspondences between visual features and tastes, *Front. Psychol.* **12**, 714277. DOI:10.3389/fpsyg.2021.714277.

- Chen, Y.-C. and Spence, C. (2010). When hearing the bark helps to identify the dog: semantically-congruent sounds modulate the identification of masked pictures, *Cognition* **114**, 389–404. DOI:10.1016/j.cognition.2009.10.012.
- Chen, Y.-C. and Spence, C. (2011). Crossmodal semantic priming by naturalistic sounds and spoken words enhances visual sensitivity, *J. Exp. Psychol. Hum. Percept. Perform.* **37**, 1554–1568. DOI:10.1037/a0024329.
- Chen, Y.-C., Huang, P.-C., Woods, A. and Spence, C. (2016a). When “Bouba” equals “Kiki”: cultural commonalities and cultural differences in sound–shape correspondences, *Sci. Rep.* **6**, 26681. DOI:10.1038/srep26681.
- Chen, Y.-C., Huang, P.-C., Woods, A. and Spence, C. (2019). I know that “Kiki” is angular: the metacognition underlying sound–shape correspondences, *Psychon. Bull. Rev.* **26**, 261–268. DOI:10.3758/s13423-018-1516-8.
- Chun, C. A. and Hupé, J.-M. (2016). Are synesthetes exceptional beyond their synesthetic associations? A systematic comparison of creativity, personality, cognition, and mental imagery in synesthetes and controls, *Br. J. Psychol.* **107**, 397–418. DOI:10.1111/bjop.12146.
- Cohen, J. (2017). Synesthetic perception as continuous with ordinary perception, or: we’re all synesthetes now, in: *Sensory Blending: New Essays on Synaesthesia*, O. Deroy (Ed.), pp. 59–83. Oxford University Press, Oxford, UK.
- Cohen, N. E. (1934). Equivalence of brightness across modalities, *Am. J. Psychol.* **46**, 117–119. DOI:10.2307/1416240.
- Connolly, K. (2014). Multisensory perception as an associative learning process, *Front. Psychol.* **5**, 1095. DOI:10.3389/fpsyg.2014.01095.
- Cutiitta, R. A. and Haggerty, K. J. (1987). A comparative study of color association with music at various age levels, *J. Res. Music Educ.* **35**, 78–91. DOI:10.2307/3344984.
- Cuturi, L. F., Tonelli, A., Cappagli, G. and Gori, M. (2019). Coarse to fine audio-visual size correspondences develop during primary school age, *Front. Psychol.* **10**, 2068. DOI:10.3389/fpsyg.2019.02068.
- Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., Kawahara, S., Koutalidis, S., Krifka, M., Lippus, P., Lupyan, G., Oh, G. E., Paul, J., Petrone, C., Ridouane, R., Reiter, S., Schümchen, N., Szalontai, Á., Ünal-Logacev, Ö., Zeller, J., Winter, B. and Perlman, M. (2021). Novel vocalizations are understood across cultures, *Sci. Rep.* **11**, 10108. DOI:10.1038/s41598-021-89445-4.
- Cytowic, R. E. and Wood, F. B. (1982). Synaesthesia II: psychophysical relations in the synesthesia of geometrically shaped taste and colored hearing, *Brain Cogn.* **1**, 36–49. DOI:10.1016/0278-2626(82)90005-7.
- Dailey, A., Martindale, C. and Borkum, J. (1997). Creativity, synesthesia and physiognomic perception, *Creat. Res. J.* **10**, 1–8. DOI:10.1207/s15326934crj1001_1.
- Deroy, O. and Spence, C. (2013a). Are we all born synaesthetic? Examining the neonatal synaesthesia hypothesis, *Neurosci. Biobehav. Rev.* **37**, 1240–1253. DOI:10.1016/j.neubiorev.2013.04.001.
- Deroy, O. and Spence, C. (2013b). Why we are not all synesthetes (not even weakly so), *Psychon. Bull. Rev.* **20**, 643–664. DOI:10.3758/s13423-013-0387-2.
- Deroy, O. and Spence, C. (2016a). Lessons of synaesthesia for consciousness: learning from the exception, rather than the general, *Neuropsychologia* **88**, 49–57. DOI:10.1016/j.neuropsychologia.2015.08.005.

- Deroy, O. and Spence, C. (2016b). Crossmodal correspondences: four challenges, *Multisens. Res.* **30**, 29–48. DOI:10.1163/22134808-00002488.
- Deroy, O. and Spence, C. (2017). Questioning the continuity claim: what difference does consciousness make?, in: *Sensory Blending: New Essays on Synaesthesia*, O. Deroy (Ed.), pp. 191–214. Oxford University Press, Oxford, UK.
- Deroy, O. and Valentin, D. (2011). Tasting liquid shapes: investigating the sensory basis of cross-modal correspondences, *Chemosens. Percept.* **4**, 80–90. DOI:10.1007/s12078-011-9097-1.
- Deroy, O., Fasiello, I., Hayward, V. and Auvray, M. (2016). Differentiated audio-tactile correspondences in sighted and blind individuals, *J. Exp. Psychol. Hum. Percept. Perform.* **42**, 1204–1214. DOI:10.1037/xhp0000152.
- Dolscheid, S., Hunnius, S., Casasanto, D. and Majid, A. (2014). Prelinguistic infants are sensitive to space-pitch associations found across cultures, *Psychol. Sci.* **25**, 1256–1261. DOI:10.1177/0956797614528521.
- Domino, G. (1989). Synesthesia and creativity in fine arts students. An empirical look, *Creat. Res. J.* **2**, 17–29. DOI:10.1080/10400418909534297.
- Drewnowski, A. and Greenwood, M. R. C. (1983). Cream and sugar: human preferences for high-fat foods, *Physiol. Behav.* **30**, 629–633. DOI:10.1016/0031-9384(83)90232-9.
- Eitan, Z. and Timmers, R. (2010). Beethoven's last piano sonata and those who follow crocodiles: cross-domain mappings of auditory pitch in a musical context, *Cognition* **114**, 405–422. DOI:10.1016/j.cognition.2009.10.013.
- Eitan, Z., Ornoy, E. and Granot, R. Y. (2012). Listening in the dark: congenital and early blindness and cross-domain mapping in music, *Psychomusicology* **22**, 33–45. DOI:10.1037/a0028939.
- Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch, *J. Vis.* **7**, 7. DOI:10.1167/7.5.7.
- Evans, K. K. (2020). The role of selective attention in cross-modal interactions between auditory and visual features, *Cognition* **196**, 104119. DOI:10.1016/j.cognition.2019.104119.
- Evans, K. K. and Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features, *J. Vis.* **10**, 6. DOI:10.1167/10.1.6.
- Flanagan, J. R., Bittner, J. P. and Johansson, R. S. (2008). Experience can change distinct size-weight priors engaged in lifting objects and judging their weights, *Curr. Biol.* **18**, 1742–1747. DOI:10.1016/j.cub.2008.09.042.
- Foroni, F., Pergola, G. and Rumiati, R. I. (2016). Food color is in the eye of the beholder: the role of human trichromatic vision in food evaluation, *Sci. Rep.* **6**, 37034. DOI:10.1038/srep37034.
- Fryer, L., Freeman, J. and Pring, L. (2014). Touching words is not enough: how visual experience influences haptic-auditory associations in the “Bouba–Kiki” effect, *Cognition* **132**, 164–173. DOI:10.1016/j.cognition.2014.03.015.
- Fumarola, A., Prpic, V., Da Pos, O., Murgia, M., Umiltà, C. and Agostini, T. (2014). Automatic spatial association for luminance, *Atten. Percept. Psychophys.* **76**, 759–765. DOI:10.3758/s13414-013-0614-y.
- Gallace, A. and Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size, *Percept. Psychophys.* **68**, 1191–1203. DOI:10.3758/bf03193720.

- Gardner, H. (1974). Metaphors and modalities: how children project polar adjectives onto diverse domains, *Child Dev.* **45**, 84–91. DOI:10.2307/1127753.
- Getz, L. M. and Kubovy, M. (2018). Questioning the automaticity of audiovisual correspondences, *Cognition* **175**, 101–108. DOI:10.1016/j.cognition.2018.02.015.
- Gibson, E. J. (1969). *Principles of Perceptual Learning and Development*. Appleton, New York, NY, USA.
- Gilbert, A. N., Martin, R. and Kemp, S. E. (1996). Cross-modal correspondence between vision and olfaction: the color of smells, *Am. J. Psychol.* **109**, 335–351.
- Glaser, W. R. and Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing, *J. Exp. Psychol. Gen.* **118**, 13–42. DOI:10.1037/0096-3445.118.1.13.
- Gold, R. and Segal, O. (2017). The Bouba–Kiki effect and its relation to the Autism Quotient (AQ) in autistic adolescents, *Res. Dev. Disabil.* **71**, 11–17. DOI:10.1016/j.ridd.2017.09.017.
- Gold, R., Klein, D. and Segal, O. (2022). The Bouba–Kiki effect in children with childhood apraxia of speech, *J. Speech, Lang. Hear. Res.* **65**, 43–52. DOI:10.1044/2021_JSLHR-21-00070.
- Gori, M., Del Viva, M., Sandini, G. and Burr, D. C. (2008). Young children do not integrate visual and haptic form information, *Curr. Biol.* **18**, 694–698. DOI:10.1016/j.cub.2008.04.036.
- Goubet, N., Durand, K., Schaal, B. and McCall, D. D. (2018). Seeing odors in color: cross-modal associations in children and adults from two cultural environments, *J. Exp. Child Psychol.* **166**, 380–399. DOI:10.1016/j.jecp.2017.09.007.
- Graven, T. and Desebrock, C. (2018). Bouba or Kiki with and without vision: shape-audio regularities and mental images, *Acta Psychol.* **188**, 200–212. DOI:10.1016/j.actpsy.2018.05.011.
- Guzman-Martinez, E., Ortega, L., Grabowecky, M., Mossbridge, J. and Suzuki, S. (2012). Interactive coding of visual spatial frequency and auditory amplitude–modulation rate, *Curr. Biol.* **22**, 383–388. DOI:10.1016/j.cub.2012.01.004.
- Hamilton-Fletcher, G., Pisanski, K., Reby, D., Stefańczyk, M., Ward, J. and Sorokowska, A. (2018). The role of visual experience in the emergence of cross-modal correspondences, *Cognition* **175**, 114–121. DOI:10.1016/j.cognition.2018.02.023.
- Haryu, E. and Kajikawa, S. (2012). Are higher-frequency sounds brighter in color and smaller in size? Auditory–visual correspondences in 10-month-old infants, *Infant Behav. Dev.* **35**, 727–732. DOI:10.1016/j.infbeh.2012.07.015.
- Hidaka, S. and Yaguchi, A. (2018). An investigation of the relationships between autistic traits and crossmodal correspondences in typically developing adults, *Multisens. Res.* **31**, 729–751. DOI:10.1163/22134808-20181304.
- Hung, S.-M., Styles, S. J. and Hsieh, P.-J. (2017). Can a word sound like a shape before you have seen it? Sound–shape mapping prior to conscious awareness, *Psychol. Sci.* **28**, 263–275. DOI:10.1177/0956797616677313.
- Jain, A., Rakhi, N. K. and Bagler, G. (2015). Analysis of food pairing in regional cuisines of India, *PLoS ONE* **10**, e0139539. DOI:10.1371/journal.pone.0139539.
- Jalal, B. and Ramachandran, V. S. (2014). A pilot investigation of “metaphor blindness” in a college student population, *Med. Hypotheses* **82**, 648–651. DOI:10.1016/j.mehy.2014.01.033.

- Jeschonek, S., Pauen, S. and Babocsai, L. (2013). Cross-modal mapping of visual and acoustic displays in infants: the effect of dynamic and static components, *Eur. J. Dev. Psychol.* **10**, 337–358. DOI:10.1080/17405629.2012.681590.
- Jewanski, J. (2010). Color-tone analogies: a systematic presentation of the principles of correspondence, in: *Audiovisiology: a Multidisciplinary Survey of Audiovisual Culture*, D. Daniels, S. Naumann and J. Thoben (Eds), pp. 338–347. König, Cologne, Germany.
- Jonauskaite, D., Wicker, J., Mohr, C., Dael, N., Havelka, J., Papadatou-Pastou, M., Zhang, M. and Oberfeld, D. (2019). A machine learning approach to quantify the specificity of colour–emotion associations and their cultural differences, *R. Soc. Open Sci.* **6**, 190741. DOI:10.1098/rsos.190741.
- Jüttner, M., Müller, A. and Rentschler, I. (2006). A developmental dissociation of view-dependent and view-invariant object recognition in adolescence, *Behav. Brain Res.* **175**, 420–424. DOI:10.1016/j.bbr.2006.09.005.
- Kemp, S. E. and Gilbert, A. N. (1997). Odor intensity and color lightness are correlated sensory dimensions, *Am. J. Psychol.* **110**, 35–46.
- Knöferle, K. and Spence, C. (2012). Crossmodal correspondences between sounds and tastes, *Psychon. Bull. Rev.* **19**, 992–1006. DOI:10.3758/s13423-012-0321-z.
- Köhler, W. (1929). *Gestalt Psychology*. Liveright, New York, NY, USA.
- Köhler, W. (1947). *Gestalt Psychology: an Introduction to New Concepts in Modern Psychology*. Liveright, New York, NY, USA.
- Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B. and Shams, L. (2007). Causal inference in multisensory perception, *PLoS ONE* **2**, e943. DOI:10.1371/journal.pone.0000943.
- Koriat, A. (1975). Phonetic symbolism and feeling of knowing, *Mem. Cognit.* **3**, 545–548. DOI:10.3758/BF03197529.
- Koriat, A. (2008). Subjective confidence in one's answers: the consensuality principle, *J. Exp. Psychol. Learn. Mem. Cogn.* **34**, 945–959. DOI:10.1037/0278-7393.34.4.945.
- Korzeniowska, A. T., Simner, J., Root-Gutteridge, H. and Reby, D. (2022). High-pitch sounds small for domestic dogs: abstract crossmodal correspondences between auditory pitch and visual size, *R. Soc. Open Sci.* **9**, 211647. DOI:10.1098/rsos.211647.
- Król, M. E. and Ferenc, K. (2020). Silent shapes and shapeless sounds: the robustness of the diminished crossmodal correspondences effect in autism spectrum conditions, *Psychol. Res.* **84**, 1507–1516. DOI:10.1007/s00426-019-01163-9.
- Lacey, S., Martinez, M., McCormick, K. and Sathian, K. (2016). Synesthesia strengthens sound-symbolic cross-modal correspondences, *Eur. J. Neurosci.* **44**, 2716–2721.
- Larson, C. L., Aronoff, J. and Steuer, E. L. (2012). Simple geometric shapes are implicitly associated with affective value, *Motiv. Emot.* **36**, 404–413. DOI:10.1007/s11031-011-9249-2.
- Lehman, R. S. (1972). A multivariate model of synesthesia, *Multivariate Behav. Res.* **7**, 403–439. DOI:10.1207/s15327906mbr0704_1.
- Levitan, C. A., Ren, J., Woods, A. T., Boesveldt, S., Chan, J. S., McKenzie, K. J., Dodson, M., Levin, J. A., Leong, C. X. R. and van den Bosch, J. J. F. (2014). Cross-cultural color-odor associations, *PLoS ONE* **9**, e101651. DOI:10.1371/journal.pone.0101651.
- Lewkowicz, D. J. and Minar, N. J. (2014). Infants are not sensitive to synaesthetic cross-modality correspondences: a comment on Walker *et al.* (2010), *Psychol. Sci.* **25**, 832–834. DOI:10.1177/0956797613516011.

- Lewkowicz, D. J. and Turkewitz, G. (1980). Cross-modal equivalence in early infancy: auditory–visual intensity matching, *Dev. Psychol.* **16**, 597–607. DOI:10.1037/0012-1649.16.6.597.
- Liang, P., Biswas, P., Vinnakota, S., Fu, L., Chen, M., Quan, Y., Zhan, Y., Zhang, G. and Roy, S. (2016). Invariant effect of vision on taste across two Asian cultures: India and China, *J. Sens. Stud.* **31**, 416–422. DOI:10.1111/joss.12225.
- Loconsole, M., Gasparini, A. and Regolin, L. (2022). Pitch–luminance crossmodal correspondence in the baby chick: an investigation on predisposed and learned processes, *Vision* **6**, 24. DOI:10.3390/vision6020024.
- Ludwig, V. U., Adachi, I. and Matzuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (*Pan troglodytes*) and humans, *Proc. Natl Acad. Sci. U S A* **108**, 20661–20665. DOI:10.1073/pnas.1112605108.
- Margiotoudi, K., Bohn, M., Schwob, N., Taglialatela, J., Pulvermüller, F., Epping, A., Schweller, K. and Allritz, M. (2022). Bo-NO-Bouba-Kiki: picture-word mapping but no spontaneous sound symbolic speech-shape mapping in a language trained bonobo, *Proc. R. Soc. B Biol. Sci.* **289**, 20211717. DOI:10.1098/rspb.2021.1717.
- Marks, L. (1978). *The Unity of the Senses: Interrelations Among the Modalities*. Academic Press, New York, NY, USA.
- Marks, L. E. (1974). On associations of light and sound: the mediation of brightness, pitch, and loudness, *Am. J. Psychol.* **87**, 173–188.
- Marks, L. E. (1987). On cross-modal similarity: auditory–visual interactions in speeded discrimination, *J. Exp. Psychol. Hum. Percept. Perform.* **13**, 384–394. DOI:10.1037/0096-1523.13.3.384.
- Marks, L. E. (2004). Cross-modal interactions in speeded classification, in: *The Handbook of Multisensory Processes*, G. A. Calvert, C. Spence and B. E. Stein (Eds), pp. 85–105. MIT Press, Cambridge, MA, USA.
- Marks, L. E., Hammeal, R. J. and Bornstein, M. H. (1987). Perceiving similarity and comprehending metaphor, *Monogr. Soc. Res. Child Dev.* **52**, 1–102. DOI:10.2307/1166084.
- Martino, G. and Marks, L. E. (1999). Perceptual and linguistic interactions in speeded classification: tests of the semantic coding hypothesis, *Perception* **28**, 903–923. DOI:10.1068/p2866.
- Martino, G. and Marks, L. E. (2000). Cross-modal interaction between vision and touch: the role of synesthetic correspondence, *Perception* **29**, 745–754. DOI:10.1068/p2984.
- Martino, G. and Marks, L. E. (2001). Synesthesia: strong and weak, *Curr. Dir. Psychol. Sci.* **10**, 61–65. DOI:10.1111/1467-8721.00116.
- Maurer, D. (1997). Neonatal synaesthesia: implications for the processing of speech and faces, in: *Synaesthesia: Classic and Contemporary Readings*, S. Baron-Cohen and J. E. Harrison (Eds), pp. 224–242. Blackwell, Oxford, UK.
- Maurer, D., Pathman, T. and Mondloch, C. J. (2006). The shape of boubas: sound–shape correspondences in toddlers and adults, *Dev. Sci.* **9**, 316–322. DOI:10.1111/j.1467-7687.2006.00495.x.
- Mednick, S. A. and Mednick, M. T. (1967). *Remote Associates Test: Examiners' Manual*. Houghton Mifflin, Boston, MA, USA.
- Metatla, O., Maggioni, E., Cullen, C. and Obrist, M. (2019). “Like popcorn”: crossmodal correspondences between scents, 3D shapes and emotions in children, in: *CHI’19: Proceedings of*

- the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–13. Association for Computing Machinery, New York, NY, USA. DOI:10.1145/3290605.3300689.
- Miller, A., Werner, H. and Wapner, S. (1958). Studies in physiognomic perception: V. Effect of ascending and descending gliding tones on autokinetic motion, *J. Psychol.* **46**, 101–105. DOI:10.1080/00223980.1958.9916273.
- Mondloch, C. J. and Maurer, D. (2004). Do small white balls squeak? Pitch–object correspondences in young children, *Cogn. Affect. Behav. Neurosci.* **4**, 133–136. DOI:10.3758/cabn.4.2.133.
- Mudge, E. L. (1920). The common synaesthesia of music, *J. Appl. Psychol.* **4**, 342–345. DOI:10.1037/h0072596.
- Mulvenna, C. M. and Walsh, V. (2006). Synaesthesia: supernormal integration?, *Trends Cogn. Sci.* **10**, 350–352. DOI:10.1016/j.tics.2006.06.004.
- Nanay, B. (2017). Sensory substitution and multimodal mental imagery, *Perception* **46**, 1014–1026. DOI:10.1177/0301006617699225.
- Nanay, B. (2018). Multimodal mental imagery, *Cortex* **105**, 125–134. DOI:10.1016/j.cortex.2017.07.006.
- Nava, E., Grassi, M. and Turati, C. (2016). Audio-visual, visuo-tactile and audio-tactile correspondences in preschoolers, *Multisens. Res.* **29**, 93–111. DOI:10.1163/22134808-00002493.
- Neufeld, J., Sinke, C., Zedler, M., Emrich, H. M. and Szykic, G. R. (2012). Reduced audio-visual integration in synaesthetes indicated by the double-flash illusion, *Brain Res.* **1473**, 78–86. DOI:10.1016/j.brainres.2012.07.011.
- Ngo, M. K., Misra, R. and Spence, C. (2011). Assessing the shapes and speech sounds that people associate with chocolate samples varying in cocoa content, *Food Qual. Pref.* **22**, 567–572. DOI:10.1016/j.foodqual.2011.03.009.
- Ngo, M. K., Piqueras-Fiszman, B. and Spence, C. (2012). On the colour and shape of still and sparkling water: insights from online and laboratory-based testing, *Food Qual. Pref.* **24**, 260–268. DOI:10.1016/j.foodqual.2011.11.004.
- Oberman, L. M. and Ramachandran, V. S. (2008). Preliminary evidence for deficits in multisensory integration in autism spectrum disorders: the mirror neuron hypothesis, *Soc. Neurosci.* **3**, 348–355. DOI:10.1080/17470910701563681.
- O’Boyle, M. W. and Tarte, R. D. (1980). Implications for phonetic symbolism: the relationship between pure tones and geometric figures, *J. Psycholinguist. Res.* **9**, 535–544. DOI:10.1007/BF01068115.
- Obrist, M., Comber, R., Subramanian, S., Piqueras-Fiszman, B., Velasco, C. and Spence, C. (2014). Temporal, affective, and embodied characteristics of taste experiences: a framework for design, in: *CHI’14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2853–2862. Association for Computing Machinery, New York, NY, USA. DOI:10.1145/2556288.2557007.
- Occelli, V., Spence, C. and Zampini, M. (2009). Compatibility effects between sound frequency and tactile elevation, *Neuroreport* **20**, 793–797. DOI:10.1097/WNR.0b013e32832b8069.
- Occelli, V., Esposito, G., Venuti, P., Arduino, G. M. and Zampini, M. (2013). The takete–maluma phenomenon in autism spectrum disorders, *Perception* **42**, 233–241. DOI:10.1068/p7357.

- Odbert, H. S., Karwoski, T. F. and Eckerson, A. B. (1942). Studies in synesthetic thinking: I. musical and verbal associations of colour and mood, *J. Gen. Psychol.* **26**, 153–173. DOI:10.1080/00221309.1942.10544721.
- O'Mahony, M. (1983). Gustatory responses to nongustatory stimuli, *Perception* **12**, 627–633. DOI:10.1068/p120627.
- Orchard-Mills, E., Van der Burg, E. and Alais, D. (2013). Amplitude-modulated auditory stimuli influence selection of visual spatial frequencies, *J. Vis.* **13**, 6. DOI:10.1167/13.3.6.
- Osgood, C. E. (1960). The cross-cultural generality of visual-verbal synesthetic tendencies, *Behav. Sci.* **5**, 146–169. DOI:10.1002/bs.3830050204.
- Ozturk, O., Krehm, M. and Vouloumanos, A. (2013). Sound symbolism in infancy: evidence for sound–shape cross-modal correspondences in 4-month-olds, *J. Exp. Child Psychol.* **114**, 173–186. DOI:10.1016/j.jecp.2012.05.004.
- Palmer, S. E., Schloss, K. B., Xu, Z. and Prado-León, L. R. (2013). Music–color associations are mediated by emotion, *Proc. Natl Acad. Sci. U S A* **110**, 8836–8841. DOI:10.1073/pnas.1212562110.
- Parise, C. and Spence, C. (2013). Audiovisual cross-modal correspondences in the general population, in: *The Oxford Handbook of Synesthesia*, J. Simner and E. M. Hubbard (Eds), pp. 790–815. Oxford University Press, Oxford, UK.
- Parise, C. V. (2016). Crossmodal correspondences: standing issues and experimental guidelines, *Multisens. Res.* **29**, 7–28. DOI:10.1163/22134808-00002502.
- Parise, C. V. and Spence, C. (2009). ‘When birds of a feather flock together’: synesthetic correspondences modulate audiovisual integration in non-synesthetes, *PLoS ONE* **4**, e5664. DOI:10.1371/journal.pone.0005664.
- Parise, C. V. and Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: a study using the implicit association test, *Exp. Brain Res.* **220**, 319–333. DOI:10.1007/s00221-012-3140-6.
- Parise, C. V., Knorre, K. and Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing, *Proc. Natl Acad. Sci. U S A* **111**, 6104–6108. DOI:10.1073/pnas.1322705111.
- Parkinson, C., Kohler, P. J., Sievers, B. and Wheatley, T. (2012). Associations between auditory pitch and visual elevation do not depend on language: evidence from a remote population, *Perception* **41**, 854–861. DOI:10.1068/p7225.
- Pejovic, J. and Molnar, M. (2017). The development of spontaneous sound-shape matching in monolingual and bilingual infants during the first year, *Dev. Psychol.* **53**, 581–586. DOI:10.1037/dev0000237.
- Peña, M., Mehler, P. and Nespore, M. (2011). The role of audiovisual processing in early conceptual development, *Psychol. Sci.* **22**, 1419–1421. DOI:10.1177/0956797611421791.
- Peters, M. A. K., Balzer, J. and Shams, L. (2015). Smaller = denser, and the brain knows it: natural statistics of object density shape weight expectations, *PLoS ONE* **10**, e0119794. DOI:10.1371/journal.pone.0119794.
- Pinel, P., Piazza, M., Le Bihan, D. and Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments, *Neuron* **41**, 983–993. DOI:10.1016/S0896-6273(04)00107-2.

- Pisanski, K., Isenstein, S. G. E., Montano, K. J., O'Connor, J. J. M. and Feinberg, D. R. (2017). Low is large: spatial location and pitch interact in voice-based body size estimation, *Atten. Percept. Psychophys.* **79**, 1239–1251. DOI:10.3758/s13414-016-1273-6.
- Rader, C. M. and Tellegen, A. (1987). An investigation of synesthesia, *J. Pers. Soc. Psychol.* **52**, 981–987. DOI:10.1037/0022-3514.52.5.981.
- Ratcliffe, V. F., Taylor, A. M. and Reby, D. (2016). Cross-modal correspondences in non-human mammal communication, *Multisens. Res.* **29**, 49–91. DOI:10.1163/22134808-00002509.
- Rogers, S. K. and Ross, A. S. (1975). A cross-cultural test of the Maluma–Takeke phenomenon, *Perception* **4**, 105–106. DOI:10.1068/p040105.
- Rozin, E. (1983). *Ethnic Cuisine: the Flavor-Principle Cookbook*. The Stephen Greene Press, Brattleboro, VT, USA.
- Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C. and Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect, *Cognition* **99**, 113–129. DOI:10.1016/j.cognition.2005.01.004.
- Sabanev, L. and Pring, S. W. (1929). The relation between sound and colour, *Music Lett.* **10**, 266–277.
- Sadaghiani, S., Maier, J. X. and Noppeney, U. (2009). Natural, metaphoric, and linguistic auditory direction signals have distinct influences on visual motion processing, *J. Neurosci.* **29**, 6490–6499. DOI:10.1523/JNEUROSCI.5437-08.2009.
- Sapir, E. (1929). A study in phonetic symbolism, *J. Exp. Psychol.* **12**, 225–239. DOI:10.1037/h0070931.
- Sathian, K. and Ramachandran, V. S. (2020). *Multisensory Perception: from Laboratory to Clinic*. Elsevier, San Diego, CA, USA.
- Schifferstein, H. N. J. and Tanudjaja, I. (2004). Visualizing fragrances through colors: the mediating role of emotions, *Perception* **33**, 1249–1266. DOI:10.1068/p5132.
- Schiller, P. (1935). Interrelation of different senses in perception, *Br. J. Psychol.* **25**, 465–469.
- Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H. J., Dosch, H. G., Bleeck, S., Stippich, C. and Rupp, A. (2005). Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference, *Nat. Neurosci.* **8**, 1241–1247. DOI:10.1038/nn1530.
- Sebba, R. (1991). Structural correspondence between music and color, *Color Res. Appl.* **16**, 81–88. DOI:10.1002/col.5080160206.
- Shang, N. and Styles, S. J. (2017). Is a high tone pointy? Speakers of different languages match mandarin Chinese tones to visual shapes differently, *Front. Psychol.* **8**, 2139. DOI:10.3389/fpsyg.2017.02139.
- Shankar, M. U., Levitan, C. and Spence, C. (2010). Grape expectations: the role of cognitive influences in color–flavor interactions, *Conscious. Cogn.* **19**, 380–390. DOI:10.1016/j.concog.2009.08.008.
- Shayan, S., Ozturk, O. and Sicoli, M. A. (2011). The thickness of pitch: crossmodal metaphors in Farsi, Turkish, and Zapotec, *Senses Soc.* **6**, 96–105. DOI:10.2752/174589311X12893982233911.
- Shayan, S., Ozturk, O., Bowerman, M. and Majid, A. (2014). Spatial metaphor in language can promote the development of cross-modal mappings in children, *Dev. Sci.* **17**, 636–643. DOI:10.1111/desc.12157.
- Sidhu, D. M. and Pexman, P. M. (2018). Five mechanisms of sound symbolic association, *Psychon. Bull. Rev.* **25**, 1619–1643. DOI:10.3758/s13423-017-1361-1.

- Simpson, R. H., Quinn, M. and Ausubel, D. P. (1956). Synaesthesia in children: association of colors with pure tone frequencies, *J. Genet. Psychol.* **89**, 95–103. DOI:10.1080/00221325.1956.10532990.
- Smith, L. B. (1994). Foreword, in: *The Development of Intersensory Perception: Comparative Perspectives*, D. J. Lewkowicz and R. Lickliter (Eds), pp. ix–xix. Lawrence Erlbaum, Hillsdale, NJ, USA.
- Smith, L. B. and Sera, M. D. (1992). A developmental analysis of the polar structure of dimensions, *Cogn. Psychol.* **24**, 99–142. DOI:10.1016/0010-0285(92)90004-L.
- Sourav, S., Kekunnaya, R., Shareef, I., Banerjee, S., Bottari, S. and Röder, B. (2019). A protracted sensitive period regulates the development of cross-modal sound–shape associations in humans, *Psychol. Sci.* **30**, 1473–1482. DOI:10.1177/0956797619866625.
- Spector, F. and Maurer, D. (2012). Making sense of scents: the colour and texture of odours, *Seeing Perceiving* **25**, 655–677. DOI:10.1163/187847612x648800.
- Speed, L. J., Croijmans, I., Dolscheid, S. and Majid, A. (2021). Crossmodal associations with olfactory, auditory, and tactile stimuli in children and adults, *i-Perception* **12**, 1–34. DOI:10.1177/20416695211048513.
- Spence, C. (2008). Multisensory perception, in: *The Big Fat Duck Cook Book*, H. Blumenthal, pp. 484–485. Bloomsbury, London, UK.
- Spence, C. (2011). Crossmodal correspondences: a tutorial review, *Atten. Percept. Psychophys.* **73**, 971–995. DOI:10.3758/s13414-010-0073-7.
- Spence, C. (2012). Managing sensory expectations concerning products and brands: capitalizing on the potential of sound and shape symbolism, *J. Consum. Psychol.* **22**, 37–54. DOI:10.1016/j.jcps.2011.09.004.
- Spence, C. (2020a). Temperature-based crossmodal correspondences: causes and consequences, *Multisens. Res.* **33**, 645–682. DOI:10.1163/22134808-20191494.
- Spence, C. (2020b). Olfactory-colour crossmodal correspondences in art, science and design, *Cogn. Res.* **5**, 52. DOI:10.1186/s41235-020-00246-1.
- Spence, C. (2020c). Assessing the role of emotional mediation in explaining crossmodal correspondences involving musical stimuli, *Multisens. Res.* **33**, 1–29. DOI:10.1163/22134808-20191469.
- Spence, C. (2022). Factors affecting odour-induced taste enhancement, *Food Qual. Pref.* **96**, 104393. DOI:10.1016/j.foodqual.2021.104393.
- Spence, C. and Deroy, O. (2012a). Crossmodal correspondences: innate or learned?, *i-Perception* **3**, 316–318. DOI:10.1068/i0526ic.
- Spence, C. and Deroy, O. (2012b). On the shapes of tastes and flavours, *Petits Propos Culinaires* **97**, 75–108.
- Spence, C. and Deroy, O. (2013a). Crossmodal mental imagery, in: *Multisensory Imagery: Theory and Applications*, S. Lacey and R. Lawson (Eds), pp. 157–183. Springer, New York, NY, USA. DOI:10.1007/978-1-4614-5879-1_9.
- Spence, C. and Deroy, O. (2013b). How automatic are crossmodal correspondences?, *Conscious. Cogn.* **22**, 245–260. DOI:10.1016/j.concog.2012.12.006.
- Spence, C. and Deroy, O. (2013c). On the shapes of flavours: a review of four hypotheses, *Theoria Hist. Sci.* **10**, 207–238. DOI:10.12775/ths-2013-0011.
- Spence, C. and Di Stefano, N. (2022). Coloured hearing, colour music, colour organs, and the search for perceptually meaningful correspondences between colour and pitch, *i-Perception* **13**, 20416695221092802. DOI:10.1177/20416695221092802.

- Spence, C. and Di Stefano, N. (submitted). What, if anything, can be considered amodal? *Psychon. Bull. Rev.*
- Spence, C. and Levitan, C. A. (2022). Exploring the links between colours and tastes/flavours, *J. Percept. Imaging* **4**, 000408. DOI:10.2352/J.Percept.Imaging.2022.5.000408.
- Spence, C. and Sathian, K. (2020). Audiovisual crossmodal correspondences: behavioural consequences and neural underpinnings, in: *Multisensory Perception: from Laboratory to Clinic*, K. Sathian and V. S. Ramachandran (Eds), pp. 239–258. Elsevier, San Diego, CA, USA. DOI:10.1016/B978-0-12-812492-5.00011-5.
- Spence, C., Deroy, O. and Bremner, A. (2013). Questioning the utility of the concept of amodality: towards a revised framework for understanding crossmodal relations, *Multisens. Res.* **26**, 57. DOI:10.1163/22134808-000S0036.
- Spence, C., Wan, X., Woods, A., Velasco, C., Deng, J., Youssef, J. and Deroy, O. (2015). On tasty colours and colourful tastes? Assessing, explaining, and utilizing crossmodal correspondences between colours and basic tastes, *Flavour* **4**, 23. DOI:10.1186/s13411-015-0033-1.
- Stevens, S. S. (1957). On the psychophysical law, *Psychol. Rev.* **64**, 153–181. DOI:10.1037/h0046162.
- Stevenson, R. J. and Boakes, R. A. (2004). Sweet and sour smells: learned synaesthesia between the senses of taste and smell, in: *The Handbook of Multisensory Processing*, G. A. Calvert, C. Spence and B. E. Stein (Eds), pp. 69–83. MIT Press, Cambridge, MA, USA.
- Stevenson, R. J., Boakes, R. A. and Prescott, J. (1998). Changes in odor sweetness resulting from implicit learning of a simultaneous odor-sweetness association: an example of learned synesthesia, *Learn. Motiv.* **29**, 113–132. DOI:10.1006/lmot.1998.0996.
- Stevenson, R. J., Rich, A. and Russell, A. (2012). The nature and origin of cross-modal associations to odours, *Perception* **41**, 606–619. DOI:10.1068/p7223.
- Stewart, C. R., Sanchez, S. S., Grenesko, E. L., Brown, C. M., Chen, C. P., Keehn, B., Velasquez, F., Lincoln, A. J. and Müller, R. A. (2016). Sensory symptoms and processing of nonverbal auditory and visual stimuli in children with autism spectrum disorder, *J. Autism Dev. Disord.* **46**, 1590–1601. DOI:10.1007/s10803-015-2367-z.
- Styles, S. J. and Gawne, L. (2017). When does maluma/takete fail? Two key failures and a meta-analysis suggest that phonology and phonotactics matter, *i-Perception* **8**, 2041669517724807. DOI:10.1177/2041669517724807.
- Tong, J., Li, L., Bruns, P. and Röder, B. (2020). Crossmodal associations modulate multisensory spatial integration, *Atten. Percept. Psychophys.* **82**, 3490–3506. DOI:10.3758/s13414-020-02083-2.
- Vaught, G. M., Pittman, M. D. and Roodin, P. A. (1975). Haptic-visual form identification in children aged 4 through 13, *Percept. Mot. Skills* **40**, 305–306. DOI:10.2466/pms.1975.40.1.305.
- Velasco, C., Woods, A. T., Deroy, O. and Spence, C. (2015). Hedonic mediation of the cross-modal correspondence between taste and shape, *Food Qual. Pref.* **41**, 151–158. DOI:10.1016/j.foodqual.2014.11.010.
- Velasco, C., Michel, C., Youssef, J., Gamez, X., Cheok, A. D. and Spence, C. (2016). Colour–taste correspondences: designing food experiences to meet expectations or to surprise, *Int. J. Food Design* **1**, 83–102. DOI:10.1386/ijfd.1.2.83_1.

- Wagner, K. and Dobkins, K. R. (2011). Synaesthetic associations decrease during infancy, *Psychol. Sci.* **22**, 1067–1072. DOI:10.1177/0956797611416250.
- Wagner, S., Winner, E., Cicchetti, D. and Gardner, H. (1981). “Metaphorical” mapping in human infants, *Child Dev.* **52**, 728–731. DOI:10.2307/1129200.
- Walker-Andrews, A. (1994). Taxonomy for intermodal relations, in: *The Development of Inter-sensory Perception: Comparative Perspectives*, D. J. Lewkowicz and R. Lickliter (Eds), pp. 39–56. Lawrence Erlbaum, Hillsdale, NJ, USA.
- Walker, L. and Walker, P. (2016). Cross-sensory mapping of feature values in the size–brightness correspondence can be more relative than absolute, *J. Exp. Psychol.: Hum. Percept. Perform.* **42**, 138–150. DOI:10.1037/xhp0000128.
- Walker, P. (2012). Cross-sensory correspondences and cross talk between dimensions of connotative meaning: visual angularity is hard, high-pitched, and bright, *Atten. Percept. Psychophys.* **74**, 1792–1809. DOI:10.3758/s13414-012-0341-9.
- Walker, P. (2016). Cross-sensory correspondences: a theoretical framework and their relevance to music, *Psychomusicology* **26**, 103–116. DOI:10.1037/pmu0000130.
- Walker, P. and Smith, S. (1984). Stroop interference based on the synaesthetic qualities of auditory pitch, *Perception* **13**, 75–81. DOI:10.1068/p130075.
- Walker, P. and Walker, L. (2012). Size–brightness correspondence: crosstalk and congruity among dimensions of connotative meaning, *Atten. Percept. Psychophys.* **74**, 1226–1240. DOI:10.3758/s13414-012-0297-9.
- Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A. and Johnson, S. P. (2010). Preverbal infants’ sensitivity to synaesthetic cross-modality correspondences, *Psychol. Sci.* **21**, 21–25. DOI:10.1177/0956797609354734.
- Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A. and Johnson, S. P. (2014). Preverbal infants are sensitive to cross-sensory correspondences: much ado about the null results of Lewkowicz and Minar (2014), *Psychol. Sci.* **25**, 835–836. DOI:10.1177/0956797613520170.
- Walker, R. (1987). The effects of culture, environment, age, and musical training on choices of visual metaphors for sound, *Percept. Psychophys.* **42**, 491–502. DOI:10.3758/BF03209757.
- Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity, *Trends Cogn. Sci.* **7**, 483–488. DOI:10.1016/j.tics.2003.09.002.
- Wan, X., Woods, A. T., van den Bosch, J. J. F., McKenzie, K. J., Velasco, C. and Spence, C. (2014). Cross-cultural differences in crossmodal correspondences between tastes and visual features, *Front. Psychol.* **5**, 1365. DOI:10.3389/fpsyg.2014.01365.
- Wang, Q.(J.) and Spence, C. (2017). Assessing the role of emotional associations in mediating crossmodal correspondences between classical music and red wine, *Beverages* **3**, 1. DOI:10.3390/beverages3010001.
- Wang, Q.(J.), Wang, S. and Spence, C. (2016). “Turn up the taste”: assessing the role of taste intensity and emotion in mediating crossmodal correspondences between basic tastes and pitch, *Chem. Senses* **41**, 345–356. DOI:10.1093/chemse/bjw007.
- Ward, J., Huckstep, B. and Tsakanikos, E. (2006). Sound-colour synaesthesia: to what extent does it use cross-modal mechanisms common to us all?, *Cortex* **42**, 264–280. DOI:10.1016/S0010-9452(08)70352-6.

- Weinberger, A. B., Gallagher, N. M., Colaizzi, G., Liu, N., Parrott, N., Fearon, E., Shaikh, N. and Green, A. E. (2022). Analogical mapping across sensory modalities and evidence for a general analogy factor, *Cognition* **223**, 105029. DOI:10.1016/j.cognition.2022.105029.
- Whittingham, K. M., McDonald, J. S. and Clifford, C. W. G. (2014). Synesthetes show normal sound-induced flash fission and fusion illusions, *Vision Res.* **105**, 1–9. DOI:10.1016/j.visres.2014.08.010.
- Wicker, F. W. (1968). Mapping the intersensory regions of perceptual space, *Am. J. Psychol.* **81**, 178–188. DOI:10.2307/1421262.
- Wicker, F. W. and Holahan, C. K. (1978). Analogy training and synesthetic phenomena, *J. Gen. Psychol.* **98**, 113–122. DOI:10.1080/00221309.1978.9920862.
- Woods, A. T., Spence, C., Butcher, N. and Deroy, O. (2013). Fast lemons and sour boulders: testing crossmodal correspondences using an Internet-based testing methodology, *i-Perception* **4**, 365–369. DOI:10.1068/i0586.
- Yates, M. J., Loetscher, T. and Nicholls, M. E. R. (2012). A generalized magnitude system for space, time, and quantity? A cautionary note, *J. Vis.* **12**, 9. DOI:10.1167/12.7.9.
- Zangenehpour, S. and Zatorre, R. J. (2010). Cross-modal recruitment of primary visual cortex following brief exposure to bimodal audiovisual stimuli, *Neuropsychologia* **48**, 591–600. DOI:10.1016/j.neuropsychologia.2009.10.022.