Consciousness and Cognition 140 (2026) 104019

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/yccog

Review article

On sensory similarities

Charles Spence [a,*], Nicola Di Stefano [b]

[a] Crossmodal Research Laboratory, Department of Experimental Psychology, Oxford University, Oxford, UK
[b] Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy

ARTICLE INFO

Keywords:
Similarity
Senses
Sensory modalities
Analogy
Synaesthesia
Crossmodal correspondences

ABSTRACT

Similarity has been a central concern for millennia, drawing the attention of philosophers, scientists, and artists. The issue of similarity is especially intriguing in the domain of the senses (and of sensory perception), where considerable research effort has been devoted to describing the affinities and divergences between different sensory modalities. In this work, we critically review the multiple different ways in which 'sensory similarity' has been theorized, highlighting how the similarities and analogies that researchers have proposed between the senses tend to constrain the theorization about how the senses operate. Similarities between different pairs of senses have been identified at multiple levels including in terms of stimuli/energy, transduction mechanisms, information-processing, synthetic/analytic analysis, perceptual organization, similarity of structure of the sensory space, and/or of the temporal evolution of different kinds of sensations. Based on the reviewed evidence, we propose an integrated framework, suggesting that sensory similarity can be understood as arising from (i) shared phenomenal qualities, (ii) analogous structural or organizational patterns, or (iii) convergent affective or semantic meanings, each grounded in increasingly cognitive and culturally-mediated processes.

1. Introduction

From the Greek philosophers through to contemporary scientists, the senses have long captivated human inquiry. Since Aristotle's *De Anima*, in which sensory perception was first systematically classified and examined, thinkers have sought to understand how sensory experience contributes to our knowledge of the world (see Everson, 1997; Macpherson, 2011). The nature and function of sensory perception have inspired the enduring philosophical debate between empiricism and innatism, that is, whether the information gathered from the senses is sufficient to construct the knowledge of the world, or whether some form of prior, 'innate' knowledge is required (Meyers, 2014). Contemporary discussions in psychology and cognitive science, opposing nativist and constructivist accounts (e.g., Fodor, 1975; Gibson, 1966), can be seen as a continuation of this classical dispute.

Today, questions about how the senses operate, whether in isolation or together, remain central to theoretical and empirical work in psychology, psychophysics, and cognitive science (Bremner, Lewkowicz, & Spence, 2012; Calvert, Spence, & Stein, 2004; Marks, 1978; Stein, 2012). One key issue concerns whether the senses provide distinct kinds of information about the world or, alternatively, whether some sensory modalities overlap in the information they convey about external properties (see the debate on amodal sensory qualities; Spence & Di Stefano, 2024c, for a review). Relatedly, one may ask whether information delivered by different senses is mutually exclusive or can rather be partially similar. An intriguing question, then, is in what sense the concept of similarity can be applied to sensory contents across the senses (or even sensory systems).

Philosophers, artists, and scientists, have long been fascinated by the search for similarities between the senses (e.g., vision, hearing, touch, taste, and olfaction). It is important to recognize that similarity relationships between the senses can be established in many different ways which might refer to different sensory/cognitive domains (e.g., Marks, 1983, 1987a, b). For instance, Marks (1983) distinguished between informational, psychophysical, and phenomenological similarity. 'Informational similarity' refers to the fact that different sensory systems

* Corresponding author at: Department of Experimental Psychology, University of Oxford, Life & Mind Building, Oxford OX1 3PS, UK.
E-mail address: charles.spence@psy.ox.ac.uk (C. Spence).

https://doi.org/10.1016/j.concog.2026.104019
Received 23 November 2025; Received in revised form 15 January 2026; Accepted 10 February 2026
Available online 23 February 2026

Table 1
Categorization of different types of sensory stimuli: Simple, mid-level, and complex.

Type of correspondence	Visual stimuli	Auditory stimuli	Tactile stimuli	Olfactory stimuli	Gustatory stimuli
Simple (or low-level)	Colour patch, hue, brightness	Pure tone, pitch, loudness	Intensity	Discretely perceived olfactory note, like citrus, or rose	Basic taste, intensity
Mid-level	Spatiotemporal pattern, Colour harmony	Rhythm, musical mode, temporal pattern, leitmotiv	Spatiotemporal pattern	?	Bittersweet or sweet–sour dish
Complex (or high- level)	Painting, abtract animation, film clips,	Music clips	Pattern of tactile stimulation treated as conveying emotion	Complex olfactory stimulus, like fine freagrance or wine	Complex flavour experience such as fine wine, whisky, or coffee

can provide information about the same quality, i.e., the shape of a cube via sight or touch (e.g., Gibson, 1966; Spence & Di Stefano, 2024a). 'Psychophysical similarity' refers to the similar ways in which sensations and perceptions depend on how certain stimulus parameters scale (e.g., intensity, qualitative structure, and distribution in space and time, see Stevens, 1961; von Békésy, 1967). Finally, the most problematic, though clearly most intuitive, case of analogy between the senses is 'phenomenological similarity', that is, the idea that sensory contents may be similar in terms of their phenomenal appearance, or quality (Marks, 1978; 2011; Spence, 2022b; see Di Stefano & Spence, 2024a, for a critical review).[1] Consider, for instance, the numerous attempts that have been made to associate musical pitch (or specific keys) with particular colours (see Spence & Di Stefano, 2022, for a review of such efforts, which date back at least as far as Isaac Newton; Newton, 1704). In some, but by no means all, such attempts have been premised on the notion of crossmodal perceptual similarity. In similar cases, the proposed relationship between 'red' and, say, the key of E major is thought to arise from a perceived similarity between the phenomenal qualities of the two perceptual experiences.

All too frequently (and mistakenly in our view), researchers interested in phenomenal similarity have been attracted by the case of synaesthesia (e.g., Marks, 1978, 1983, 1985, 1987a, b, 1989a, 1995; Marks & Bornstein, 1987; Marks, Hammeal, & Bornstein, 1987; Melara, 1989). However, it is important to make clear at the outset that while synaesthesia has been positioned as a paradigmatic case of associative phenomena linking sensory contents, often (but by no means always) from distinct sensory modalities, thereby bearing on their purported similarity, that the mere association, or co-occurrence of sensations, says nothing about any kind of phenomenal similarity that may, or may not exist, between the contents of experience in the different senses.[2] To be absolutely clear, synaesthesia refers to idiosyncratic, automatic, unidirectional associations between sensory (or, on occasion, conceptual) associations between an inducer and a concurrent. By contrast, crossmodal correspondences refer to shared bidirectional associations between sensory features, attributes, or dimensions across the senses, that are typically shared by the majority of those within a population (see Deroy & Spence, 2013, on this important distinction).

Beyond phenomenal similarity, there are different approaches to similarity (and difference; Ben-Artzi & Marks, 1995) between the senses that consider structural and organizational aspects of sensory perception, rather than by any kind of similarity between the sensory inputs themselves (e.g., Spence & Di Stefano, 2024b, 2025a, b). For instance, similarities across the senses can also be mediated by shared configurations or patterns of sensory stimulation (e.g., Harris, 1950; Julesz & Hirsh, 1972). In such cases, observers may recognize that the patterns of stimuli that are presented in different sensory domains share comparable spatiotemporal dynamics, even though they do not necessarily share any kind of phenomenal qualia. For instance, a musical crescendo/diminuendo and an increase/decrease in brightness, can both be perceived as parallel phenomena because they both unfold in a similar manner over time (Spence & Di Stefano, 2025c). Notably, empirical studies demonstrate that participants can detect far more complex crossmodal pattern configurations beyond simple rising or falling dynamics, such as, for example, symmetry or fragmentation (see Wanke, Ansani, Di Stefano, & Spence, 2025).

As will become clear, crossmodal correspondences represent a relevant phenomenon when it comes to the discussion of sensory similarity. Importantly, however, Spence and Di Stefano (2025a) recently highlighted a distinction between simple, mid-level, and complex crossmodal correspondences (see Table 1). Note that phenomenal similarity has sometimes been postulated as an explanation for the existence of crossmodal correspondences between simple stimuli (such as individual colours, timbres, pitches, etc.; Di Stefano & Spence, 2024a). By contrast, cognitively-mediated analogy is more often mentioned as an explanation in the case of the crossmodal correspondences that have been reported between mid-level stimuli (typically comprising a small number of, possibly dynamic, stimuli in each sensory modality; Spence & Di Stefano, 2025c). Meanwhile, affective, hedonic, or emotional correspondence, or similarity, is typically mentioned in the context of crossmodal correspondences between complex stimuli, such as, for example, pieces of music and paintings (Actis-Grasso, Zavagno, Lega, Zani, Daneyko, & Cattaneo, 2017; Spence, 2020a).[3]

Further analogies between the senses have been put forward in terms of the existence, and number, of basic categories of sensation (or not; Spence & Di Stefano, in press). For instance, the existence of basic colours and tastes might highlight a similarity between the two senses as well as a difference with olfaction, touch, and audition that seemingly lack any commonly agreed upon basic categories of sensation. Furthermore, the distinction between analytic vs. synthetic processing of sensory stimuli might be taken to highlight similarities between sensory modalities based on the similar ways that sensory stimuli/objects are processed and constructed in perception (Breslin, 2001).

Finally, similarities also exist between the senses in terms of similar processing mechanisms (von Békésy, 1957, 1959, 1964a, b), or information processing capacities (Gallace, Ngo, Sulaitis, & Spence, 2012), and/or a similar psychophysics applying across different sensory dimensions, such as loudness and brightness (Marks, 1985; Stevens, 1957,

[1] Note here that sensory metaphors provide another means of aligning sensory impressions (Gardner, 1974; Marks, 1991; Winner, Rosentiel, & Gardner, 1976).

[2] In fact, there is a trope of published commentaries discussing the similarity of the sound of the trumpet and the colour scarlet (e.g., Kandinsky, 1977; Leibniz, 1896; Locke, 1690; Ortmann, 1933; see also Donnell-Kotrozo, 1978). Relevant here, though, the neuroscientist MacDonald Critchley (1979) put forward a purely associative account for this particular association: "the familiar story of trumpet blasts provoking a photism of red, may stem from the fact that such a sound immediately calls up in some persons an imagery of soldiers on parade. Ordinarily they shall be in dress uniform. This evokes a mental picture of scarlet. Should the middle part of this notion eventually become submerged, there will remain a synaesthetic linkage of trumpet-calls with redness." Note that if one accepts such an account then there is absolutely no need to invoke phenomenal similarity.

[3] Though note that crossmodal correspondences operating between simple sensory characteristics of these complex classes of stimuli have, on occasion, also been proposed (Duthie, 2013; Duthie & Duthie, 2015).

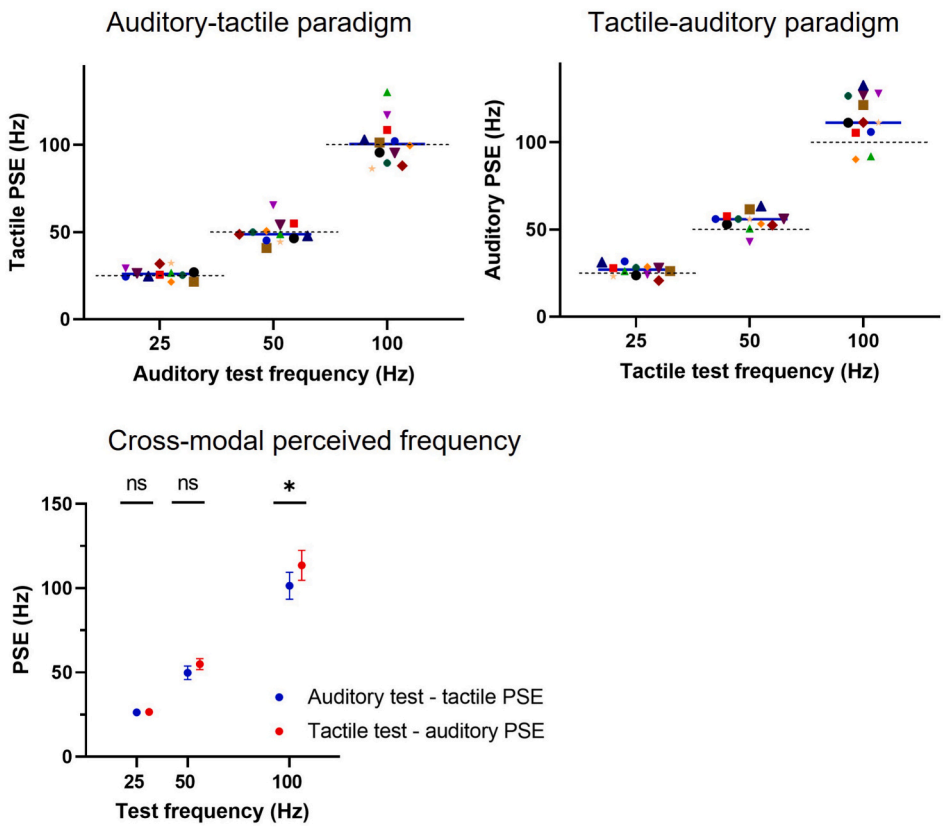

Fig. 1. Cross-modal point of subjective equality (PSE) for regular tactile and auditory stimuli in a study by Sharma et al. (2022) showing that the same frequency of vibration is matched in the auditory-tactile paradigm. (A) Individual subject PSEs obtained for auditory test conditions using tactile comparisons (auditory-tactile paradigm). A tactile PSE represents the frequency of a vibrotactile pulse train that is equally likely to be judged higher or lower than the auditory test stimulus. A solid line indicates a median PSE (n = 12) value for a test frequency, whereas a dashed line represents the expected unimodal PSE (or the physical frequency) of the test stimulus. (B) Individual subject PSEs for tactile test conditions using auditory comparisons (tactile-auditory paradigm), conventions as in panel (A). (C) Cross-modal equivalent mean frequencies for both auditory and tactile stimuli of identical frequency, * represents $p < 0.05$ and error bars indicate 95% CI. [Figure reprinted from Sharma et al. (2022).].

1958, 1961). For instance, one might ask whether the shared vibratory nature of haptic and auditory stimulation affects how observers experience two distinct stimuli that happen to be based on the same physical stimulus, such as a pure tone at 220 Hz that can be both heard and felt. Indeed, the studies of crossmodal matching that have been published to date have largely reported that people subjectively match the same frequency stimulus when tasked with matching vibratory stimuli presented independently to ear and skin (Occelli, Gillmeister, Forster, Spence, & Zampini, 2009; Sharma, Ng, Birznieks, & Vickery, 2022; see Fig. 1), if not necessarily to ear and eye (e.g., Welch, DuttonHurt, & Warren, 1986). Here, therefore, similarity in the underlying physical stimulus between audition and touch is a condition but perhaps not sufficient to guarantee the phenomenal similarity of the sensations themselves (see Deiana, Ansani, Di Pino, Formica, Spence, & Di Stefano, 2025).[4] Or rather, whatever crossmodal matching is observed is not based on physical identity leading to phenomenal similarity but rather operates at the level of structural similarity between the senses (Spence & Di Stefano, 2025c).[5]

1.1. Outline

Having laid out some of the groundwork and challenges in the study of crossmodal similarity, in the sections that follow, we start by discussing phenomenal crossmodal similarity (Section 2). In Section 3, we then go on to examine the case of crossmodal correspondences (Spence, 2011) to assess whether, and how, these perceptual phenomena may shed light on sensory similarity. Thereafter, in Section 4, the literature on structural/organizational similarities and other cases of analogies between the senses is summarized, as well as the important distinction between prothetic, metathetic, and polar sensory dimensions. In Section 5, the putative individual differences, group differences, and differences between different (pairs of) sensory modalities in terms of the concept of phenomenal similarity are briefly discussed. We also question the assumption that there need be a unitary answer to the question of

[4] Here, one might also consider the case of olfactory enantiomers, namely volatile chemicals that have a very similar physical structure (one being the mirror-image of the other), yet which can, on rare occasion, smell completely different, such as orange vs. lemon (limonene), or spearmint vs. caraway (carvone; see Brookes, 2010; Brookes, Horsfield, & Stoneham, 2009; Sell, 2006). Or, on the flip side, metameric tastes that are phenomenologically indistinguishable despite being associated with compounds having very different chemical structures. Consider here only the wide range of sapid compounds that people describe as tasting 'bitter' (Behrens, Gunn, Ramos, Meyerhof, & Wooding, 2013; Lang, Di Pizio, Risso, Drayna, & Behrens, 2023).

[5] Here, it is perhaps worth noting that there is an extensive literature on crossmodal matching. However, it is important to recognize that the fact that people match stimuli across the senses has been taken to indicate amodal, or sometimes merely structural, matching. Therefore, the mere existence of robust crossmodal matching and scaling says nothing, in particular, about crossmodal phenomenal similarity (Marks, Stevens, Bartoshuk, Gent, Rifkin, & Stone, 1988). However, dealing with the literature on crossmodal matching in any detail lies beyond the scope of the present study.

sensory similarity. In Section 6, we briefly consider testable predictions and methodological guidance for future research. Finally, we bring everything together in the Conclusions (Section 7): We highlight the dangers of relying on superficial (rather than structural) similarities between the senses, and suggest the need for multiple analogies between the senses in order to get a better, more representative, picture of the similarities and differences between the senses (cf. Spiro, Feltovich, Coulson, & Anderson, 1989).[6]

This narrative historical review (see Ferrari, 2015; Furley & Goldschmied, 2021, on the strengths of the narrative style review) helps to fill an important gap in the literature, and is particularly timely, given the growing interest in crossmodal correspondences that has exploded over the last 15 years (since the publication of Spence's, 2011, review, which has already been cited well over 2,000 times). A growing number of researchers are asking the question of what explains the multitude of crossmodal correspondences that have been documented over the years. Often, perceptual similarity[7] has been put forward as a putative explanation, without necessarily being theoretically, nor empirically, justified (in part, presumably because of Marks' influential early writings where he often suggested sensory similarity as an explanation for a wide range of crossmodal associations).

2. Phenomenal crossmodal similarity

One of the most intuitive (and experientially prominent) ways in which to conceive of sensory similarity rests on the notion that certain sensory contents are phenomenologically more alike than others, and that such distinctions are rooted in the qualitative nature of sensory experience itself (Marks, 1987a, b; 1989a, b). For instance, when I perceive a red Ferrari and a red apple, the common experience of redness guarantees that the two objects are similar in some analogous respect (i.e., in terms of their colour).[8] Phenomenal similarity has been recognized as key for the categorization and identification of objects and, more generally, for concepts formation, learning and memory (e.g., Marks, 1978; Rosch & Mervis, 1975; Tversky, 1977). Several decades ago, the psychologist Linda Smith (1993) acknowledged that phenomenal similarity is the source of learning, and is crucial even for the simplest of cognitive abstractions: "If we teach a girl to call her collie *dog*, she will call Labradors *dog*, and perhaps goats *dog*, but she will not call a motor scooter *dog*" (Smith, 1993, p. 223, italics in the original).

While phenomenal similarity within a single sensory modality is widely accepted (Helmholtz (1878/1971), difficulties arise when conceiving of similarity based on shared or common qualia across different sensory modalities. Lawrence Marks, one of the most vocal proponents of this notion of phenomenal similarity operating across the senses (e.g., Marks, 1978, 1987a, b; 1989a, b; 2011), had the following to say when, commenting on the "perceptual similarities between and among sensory experiences in different modalities":

> "Much as the color aqua is more similar phenomenologically to cerulean than to pink, the flavour of lime more similar to lemon than to banana, so too are low notes played on a bassoon or an organ more like dark colors such as brown or black than bright colors such as yellow or white, while the higher notes played on clavier or a flute resemble yellow or white more than brown or black." (Marks, 2011, p. 52).

While Marks's description is highly intuitive and vividly conveys the phenomenological associations that have sometimes been reported between the senses, it is important to remember that the examples remain purely descriptive in nature. And while the notion of crossmodal 'phenomenological' similarity may be highly evocative, it does not clarify why such correspondences arise in the first place, nor does it offer any kind of explanatory grounding. Of a completely different tone is the well-known, and much earlier, observation by the German psychophysicist Hermann von Helmholtz (1878/1971, p. 77), who insisted that the distinctions among senses—such as between colour, temperature, taste, and pitch—"are so fundamental as to exclude any possible transition from one modality to another and any relationship of greater or less similarity".[9] Note that, according to Helmholtz, one cannot meaningfully ask whether sweet is more like red or blue; such comparisons, or so he famously asserted, are only possible within the same sense. The contrast between these two views neatly captures the enduring tension between descriptive phenomenology and explanatory constraint in theories of crossmodal similarity.[10] It remains something of an open question as to which of these two opposing views has more support currently.

Doubts about the very notion of meaningful similarity were famously raised by Nelson Goodman (1972), who argued more than half a century ago that any two objects share at least one phenomenal quality, whether intramodal or crossmodal, rendering similarity claims effectively universal and therefore uninformative ("everything is similar to everything else", see also Rodriguez-Pereyra, 2002). Complicating matters further, similarity claims are inherently comparative and relative. A tomato is similar to a Ferrari (both are red) and to a tennis ball (both are roughly spherical). Confronted with these three objects, any judgment of phenomenal similarity immediately raises the question: Similar in what respect? This dependence on a chosen dimension of comparison echoes Goodman's (1972) early critique: Since any two objects share some property, similarity becomes ubiquitous unless constrained by an explicit, and possibly arbitrary criterion.

A related challenge was later formalized by Tversky (1977), whose contrast model shows that similarity is not an intrinsic relation but a function of how features are weighted, selected, and contrasted relative to a task or context. In this light, phenomenal similarity risks losing explanatory force: Rather than revealing a deep structural relation between sensory contents, it may simply track arbitrary or experimenter-imposed dimensions of comparison, thereby rendering similarity judgments highly context-sensitive and potentially indeterminate.

2.1. Phenomenal crossmodal similarity and synaesthesia

Finally, within the domain of phenomenal crossmodal similarity, the

[6] There is an analogy here to the experimental literature on structural vs. surface similarity in the context of analogical problem solving (Gentner, Holyoak, & Kokinov, 2001; Gentner & Markham, 1977; Holyoak & Koh, 1987; Medin & Ortony, 1989; Novick, 1988, 1990).

[7] According to Smith (1993, p. 223): "similarity is a complex and diverse set of processes that in their mutual interactions yield both a system of perceptual comparison that is inherently creative and a unitary concept of same that transcends specific perceptual features." (see also Bianchi & Burro, 2023).

[8] Note that this phenomenological claim does not imply that the internal character of "redness" is the same across different observers, nor that similarity judgments are grounded in identical qualitative experiences. The question of whether phenomenal qualities can be shared, compared, or even properly characterized across subjects has been extensively debated in philosophy (see, e.g., Kripke, 1980; Nagel, 1974; Shoemaker, 1975).

[9] In fact, Helmholtz's criterion has sometimes been used in order to determine whether pairs of sensory stimuli should be associated with the same or different sense (McBurney & Gent, 1979).

[10] At the same time, however, one should perhaps also acknowledge the assumption that there must be a unitary answer to the question of the existence of crossmodal phenomenal similarity. Even disregarding the appeal to synaesthesia (as providing evidence of individual differences in crossmodal perception; Marks, 1989b), one might nevertheless wonder whether the possibility should at least be floated that Marks and Helmholtz could both be correct in terms of their own phenomenal experience? We note that most researchers would appear to assume that the same answer must apply to everyone (though see Churchland, 1998), though we return, briefly, to consider this question, in **Section 5**.

case of synaesthesia is often taken as the most vivid and intuitive example. As Marks (1987a) has observed: "This last category [i.e., phenomenological similarity] is most dramatically illustrated in the phenomena of synesthesia" (p. 384) In fact, despite being a highly idiosyncratic or 'private' condition, synaesthesia involves a systematic and involuntary coupling between distinct sensory contents that are experienced as part of a co-occurring (and hence, in some minimal sense, unified) percept (Ramachandran, Marcus, & Chunharas, 2020). However, as mentioned briefly earlier, the mere existence of an association between an inducer and its associated concurrent (that is, the co-occurrence of independent sensations), has no direct implication for the phenomenal similarity between the contents of experience in the different senses. That is, two sensations are regularly experienced together, thus arguably constituting a case of identity rather than similarity. From a theoretical standpoint, this distinction is crucial, as cases of genuine crossmodal similarity need the recognition of two separate sensory contents whose qualities are compared rather than unwillingly merged or unified (cf. Auvray & Spence, 2008, on this issue as it applies to the case of olfactory and gustatory perception). Understanding this difference clarifies why the inducer-concurrent pair does not count as a paradigmatic case of crossmodal similarity or correspondences, just as the barking of the dog is not phenomenally similar to the sight of a dog though the two undoubtedly commonly co-occur (Chen & Spence, 2010, 2017; Deroy & Spence, 2013).

While crossmodal correspondences have been conflated with synaesthesia ever since researchers first started discussing the two phenomena (Deroy & Spence, 2013; Dimova, 2024; Martino & Marks, 2001), the academic literature on the crossmodal correspondences has increasingly started to diverge from the discussion of synaesthesia (Spence, 2011, 2025). This divergence helps to explain the timeliness of this review given the central role that sensory similarity has played in many researchers' thinking about these phenomena. That said, the distinctive nature of synaesthesia and crossmodal correspondences has not yet become the consensus view. Indeed, a number of researchers still appear to believe that the two phenomena are merely weaker and stronger versions of the same underlying mechanism. In the next section, we examine how crossmodal correspondences might contribute to understanding sensory similarity.

3. Crossmodal correspondences: Another example of perceptual similarity between the senses?

Crossmodal correspondences have been defined as consensual associations between stimuli, attributes or dimensions of experience, either physically presented, or else merely imagined in different modalities, experienced no matter whether one is a synaesthete or not (Spence, 2011). Often (mis)labelled as 'synaesthetic correspondences' or 'synaesthetic metaphors,' crossmodal associations have long been treated as central to discussions of sensory similarity, precisely because they involve a form of sensory matching in which two stimuli or inputs from different sensory modalities are brought into relation with one another (Motoki, Marks, & Velasco, 2023). While this is intuitively compelling, it leaves open the question of the level at which the similarity is established or grasped, since different pairings may rely on quite different mediating factors.

As Marks (1983, p. 2) notes: "As appealing as these cross-sensory similarities may be – in the present example, mainly similarities between color and temperature – they seem not to provide a practicable basis for a sensory substitution system. Despite the fact that the correspondences between color and temperature may be widely acknowledged and appreciated, color-temperature associations seem not to represent any special, intrinsic sensory similarity, but to be cross-modal correlations that are learned – and learned rather slowly at that (Morgan et al., 1975). The first conclusion, then, is that not all phenomenal similarities represent intrinsic correspondences between the senses."

For instance, the fact that people have been shown to match louder sounds with brighter lights (Bahrick, 2010; Walker-Andrews, 1994; Werner, 1934; note that one-month-old babies also appear to treat such crossmodal stimuli similarly; Lewkowicz & Turkewitz, 1980) might be taken to suggest that observers perceive similarity between the positions that stimuli occupy within their respective sensory dimensions (or continua), i.e., loudness and brightness, rather than between the stimuli themselves. Such structural correspondences, or perhaps better said, physiological correspondences (see Spence & Di Stefano, 2024b, on this distinction), often reflect structural alignment or relative positioning. Such an account is particularly useful when it comes to explaining crossmodal correspondences that involve those sensory dimensions that admit of being arranged along a continuum, that is, those that can be experienced as more or less of something (that is, prothetic sensory dimensions).[11] Judgments about such structural mappings will likely be made relative to the other stimuli that happen to have been presented.[12] Note, however, that structural alignment cannot be used to explain crossmodal matches between say, pink and sweetness (given there is no agreed or obvious organization of taste qualities), or occurrences of multisensory integration such as matching the seen and felt shape of an object (e.g., as in the well-known 'Molyneaux problem'; Marks, 1983; see Spence & Di Stefano, 2024a, for a review), or heard and seen speech. In what follows, we take a closer look at the different classes of crossmodal correspondences that have been suggested to date in order to identify potentially different ways in which phenomenological similarity can be established between sensory experiences.

3.1. Similar affective/emotional meaning

One explanation that has been put forward to account for at least a subset of the crossmodal correspondences that have been documented to date is in terms of hedonic, or what are sometimes referred to as emotionally-mediated, or affective, crossmodal correspondences (Spence, 2020a, b). According to the emotional mediation account, two stimuli may elicit a crossmodal correspondence because they happen to be associated with a similar emotional or, more broadly, affective meaning. The findings of a number of recent studies have been taken to support this hypothesis in the audiovisual domain, showing that participants consistently match musical excerpts with complex and realistic drawings and that matched stimuli tend to have a similar emotional (or affective) profile (e.g., Di Stefano, Ansani, Focaroli, Borsella, Formenti, Velardi, Schiavio, & Spence, 2026; Di Stefano, Ansani, Schiavio, &

[11] S. S. Stevens (1957, p. 154) distinguished between prothetic (magnitude) and metathetic (qualitative) perceptual continua. The former describe continua having to do with quantity (i.e., how much), while the latter describe continua where the stimuli are arranged in qualitative terms (i.e., what kind or where). Loudness, size, duration, and rate of change all constitute prothetic dimensions with a clear "more than" end (e.g., loud, bright, big) and another, "less than" end (e.g., quiet, dark, small; Walsh, 2003). By contrast, pitch constitutes a metathetic dimension, since a high-pitched tone is different in kind from a low-pitched tone, without necessarily being meaningfully related in a more than/less than way (shape is also metathetic; see Smith & Sera, 1992). The exact correspondence between these two classes of continua and their underlying neural representations isn't altogether clear (Lewkowicz & Turkewitz, 1980; Stevens, 1971). As has been noted elsewhere, in adults, prothetic dimensions tend to possess a unitary and well-ordered psychophysics, whereas metathetic dimensions do not.

[12] Notice how auditory pitch is *metathetic* in terms of phenomenal experience (i.e., a 500 Hz tone is not experienced as 'more' or 'less' that a 200 Hz tone), the same for hue, while both pitch and hue can be arranged *prothetically* in terms of the wavelength of the underpinning stimulus.

Spence, 2024; see also Palmer, Schloss, Xu, & Prado-León, 2013, on the matching of music with colour patches).

While a similar emotional response does indeed appear to provide the best explanation for a number of crossmodal correspondences, especially in the context of more emotionally meaningful, and thus often complex, stimuli (e.g., see Spence, 2020a, for a review), emotional mediation does not imply anything about the phenomenal similarity between the pairs of sensory stimuli that are so matched. As such, emotional mediation might well explain the existence of crossmodal mappings that have been documented in the absence of any obvious phenomenological sense of perceptual similarity (Crisinel & Spence, 2012).

The affective similarity that appears to connect very diverse stimuli often extends well beyond the class of emotions, at least as narrowly defined. As such, one major caveat to the emotional mediation account of crossmodal correspondences is the fact that, as Spence (2020a) and many others have pointed out, emotions remain a poorly defined concept since different experimental and theoretical contexts adopt very different definitions or paradigmatic examples. For instance, besides referring to primary emotions such as happiness or anger, studies have used much less intuitive emotional descriptors such as 'strong', 'dreary', 'complex', 'spicy', 'whimsical', and 'dissonant' (see Palmer et al., 2013; Whiteford, Schloss, Helwig, & Palmer, 2018), while other studies have chosen to use complex constructs such as peacefulness, transcendence, and tenderness (see Janowski & Chełkowska-Zacharewicz, 2019; Juslin, 2013; Eerola and Vuoskoski, 2011).

This is where the semantic differential technique (SDT) popularized by Osgood and his colleagues may help (see Karwoski, Odbert, & Osgood, 1942; Snider & Osgood, 1969; Osgood et al., 1957).[13] The approach involves people responding on scales anchored by pairs of adjectives (e.g., good-bad; weak-strong; active–passive). Emotional / affective correspondences may then be understood in terms of the stimuli that are so matched having similar semantic differential profiles. Thus, it can be argued that affective matching (as most frequently assessed by the SDT) provides an explanation of the correspondence that people appear to experience between (especially affectively-meaningful) stimuli (see also Wang, Wang, & Spence, 2016). Importantly, however, this kind of affectively-mediated crossmodal correspondence does not have any necessary implications for the phenomenal similarity of the contents of sensory experience. It would now appear that the consensus view is that emotional mediation provides an adequate explanation for some proportion of all the crossmodal correspondences that have been reported to date.

3.2. Similar (same) amodal perceptual qualities

Assuming the existence of amodal perceptual qualities, that is, sensory properties that can be picked-up by different senses (see Spence & Di Stefano, 2024c, for a review), the crossmodal association between A and B, with A and B being stimuli pertaining to two different sensory domains, might be straightforwardly explained by observing that the same sensory quality X is perceived in both stimuli/objects. This view suggests that the various senses might share a few properties of sensation that might be called 'suprasensory', meaning that those categories, or dimensions, of experience are not limited to a single sensory modality, but rather, can be applied to most or to all modalities (see also von Hornbostel, 1950, p. 212). Amongst those properties, duration has been one of the most thoroughly investigated, with early studies demonstrating that subjects' ratings of duration tend to be highly consistent across the senses (e.g., Loeb, Behar, & Warm, 1966; Marks, 1987b; see Di Stefano & Spence, 2025, for a review). Other attributes that have frequently been suggested to be putatively amodal, or intersensory, in the literature include brightness (e.g., Cohen, 1934; Hartshorne, 1934; von Hornbostel, 1950), roughness (Di Stefano & Spence, 2022), stimulus intensity (Bahrick, 2010; Külpe, 1893; Ryan, 1940; Walker-Andrews, 1994; Werner, 1934),[14] and even thickness/volume (Moul, 1930; though see Spence & Di Stefano, 2024c).

That said, even if we accept the existence of amodal sensory qualities, it is important to acknowledge that these qualities, being shared across the senses, would appear to support crossmodal *identity* rather than necessarily crossmodal *similarity*. If what is perceived in vision and audition is literally the same quality (e.g., duration or roughness), then the relationship between the stimuli would be one of sameness, not resemblance. Yet this is not how intermodal relations ordinarily present themselves in experience. We typically say that the temporal profile of stimulus X resembles that of stimulus Y (see Bratzke, 2025), or that the roughness of a surface feels somewhat like the one of a dissonant interval, not that they share the identical property. This suggests that, at least in ordinary experience, crossmodal relations are experienced in a comparative sense.

Moreover, if amodal qualities operated as genuinely modality-independent contents, then the specific sensory modality through which they are accessed would become irrelevant: One could, in principle, experience "roughness-as-such" or "intensity-as-such" without any sensory embedding. But this runs counter to the fact that we never encounter amodal features in isolation; that is, they are always instantiated in particular sensory contents, which help to qualify the percept. Thus, while the notion of amodality could potentially illuminate some structural commonalities between the senses, it cannot fully account for the experienced similarity between stimuli presented in different sensory modalities. At present, there is no consensus as to whether amodel sensory properties actually exist.

3.3. Structural alignment: similar positioning or alignment of component stimuli

The 'relative positioning' or 'structural alignment' account of crossmodal matching is often conceived as a weak, or alternative, explanation compared to amodality. The former account implies that what is being compared when someone makes a similarity judgment is actually the relative position of the two sensory features in their own respective unisensory scales, with no implications regarding their actually being similar (cf. Cohen, 1934; Hartshorne, 1934; Marks, Szczesiul, & Ohlott, 1986; Mellers & Birnbaum, 1982; Moul, 1930; Simpson, Quinn, & Ausubel, 1956; Stevens, 1957). As Marks (1995, pp. 230-231) notes: "both cross-modality matching and cross-modal similarity have strong relativistic components. Indeed, Krantz (1972) argued that cross-modal intensity matches are based on communality of relations in different modalities, not matches of magnitudes per se." According to Mellers and Birnbaum (1982, p. 600): "In cross-modality judgments […] It appears that subjects compare the relative position of a stimulus in its distribution with the relative position of a stimulus of another modality to its distribution." For example, dark colours can be matched to lower keys of the piano, based on the common position they have on their respective unisensory scales, but such a match, no matter how consensual it might be, would not necessarily imply that dark colours are phenomenologically similar to low-pitched tones.

Relative positioning thus holds that the concept of similarity is

[13] This may link to the literature on the 'feeling value' of, for example, lines (Poffenberger & Barrows, 1924; cf. Kimura, Wada, & Noguchi, 2005).

[14] As Marks (1995, p. 210) noted some years ago: "The attribution of a common (and commensurable) dimension of intensity to percepts belonging to different sense modalities implies a kind of similarity among them—a similarity among qualitative dissimilars. Sounds have their defining qualities such as pitch or musical tonality; lights have their colours, tastes their qualities such as saltiness or sourness; feelings have their qualities such as vibration or stickiness of pain. But these qualitative dissimilarities become transparent, in some fashion, through the equivalences of perceived intensity.".

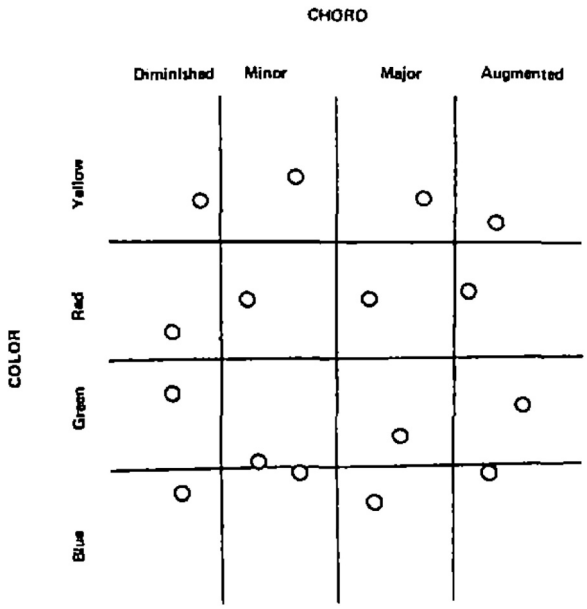

Fig. 2. Two-dimensional scaling solution for the 16 color-chord bimodal stimuli used in Webb's (1976) study. According to Webb (1976, p. 1034): "the centres of the open circles are the points representing the 16 compound stimuli in the best-fitting two-dimensional solution, after rotation to bring the striking regular lattice-like arrangement of the points into a convenient horizontal-vertical orientation." (The straight lines and row and column designations have been added to facilitate identification of the individual points.) [Figure reprinted from Webb (1976).].

unnecessary when referring to stimuli that are presented in different senses, while it can more properly describe the formal relationship that is recognized between the component parts of one stimulus that are mapped onto relative parts of the other stimulus. Therefore, structural similarity refers to the resemblance in the underlying systems of relations between the elements of the sources and the elements of the target. Structural similarity exists if the relations between the objects in the source are similar to the relations between the objects in the target, independently of the similarity between the objects themselves (Forbus, Gentner, & Law, 1994; Markman & Gentner, 2000). Though, it might be noted that one rarely comes across such unidimensional sensory stimuli outside the confines of the psychophysics laboratory, thus perhaps questioning the ecological validity of this kind of research.

A study by Webb (1976) nevertheless provides a fine example of structural alignment operating in the context of multisensory stimuli. In the protocol, participants had to rate the experienced similarities among 16 compound visual-auditory stimuli. Specifically, each of four colours (red, yellow, green, and blue) was combined with each of 4 musical chords (major, minor, diminished, and augmented). The participants in this particular study were tasked with rating the similarity of each pair of colour-chord impressions on a scale from 10 to 1, indicating maximum vs. minimum similarity, respectively.[15] The results suggested that the colour-chord combinations produced psychologically integrated impressions which varied along orthogonal dimensions of darkness (versus lightness) and spread (versus compactness) of chords (see Fig. 2).

Marks (1989a) also conducted a small study of crossmodality similarity scaling. Eight participants were presented with pairs of stimuli comprising nine pure tones differing in frequency and intensity and five white lights differing in luminance. Intramodal and crossmodal pairs of stimuli were presented on each trial. The participant's task was to rate the degree of similarity between the two stimuli on a graphic scale, a line anchored by the labels 'extremely dissimilar' and 'extremely similar'. The results for all possible pairings were then submitted to multidimensional scaling (MDS). Intriguingly, the similarity of brightness judgments did not need to be separated by the modality of the stimuli involved (see also Wicker, 1968).

It is worth noting how the conclusion that stimulus dimensional structure has no implications for crossmodal perceptual similarity would appear to contradict the claims made by the eminent experimental psychologist Lawrence Marks in his influential book, *The unity of the senses*, when he asserts that: "Sensory correspondence is not a domain of inquiry restricted to scientists, a matter solely for experimental scrutiny and empirically based theory. The plain fact is that sensory analogies do exist; they are important to the ways that we sense, perceive, and cognize; they are significant properties of the bodies and minds of people" (Marks, 1978, p. 7). In contrast, according to the view that has been espoused here, while 'sensory analogies' may bias (and/or be based on) the kinds of sensory connections that people find it easy to establish cognitively, we do not believe that they directly determine the nature, or strength, of the crossmodal correspondences that are based on perceived similarity. However, while apparently directly contradicting Marks' (1978) view, our own conclusion would appear to be much more in tune with the suggestion coming from the eminent art historian Sir Ernst Gombrich that: "The problem of synesthetic equivalences will cease to look embarrassingly arbitrary and subjective if … we fix our attention not on likeness of elements but on structural relationships within a scale or matrix" (Gombrich, 1960, p. 314; see also Di Stefano & Spence, 2024b). Note that this comment was made in the context of attempts to use synaesthetic inducer-concurrent mappings to crossmodally link pitch to colour in the case of 'colour music'.

3.4. Analogical mappings between the senses

Analogical mappings can be established based on different factors which vary in the extent to which they have a perceptual, rather than merely a conceptual, basis. For example, music notation establishes a conventional mapping between sound duration and the shape of notes on the musical score. However, there is no perceptual similarity between the duration of, for instance, quarter and eight notes and the way these are represented in the score. As such, there would be no problem in notating the same sound duration differently, based on a different criterion.

The findings of a study by Weinberger and colleagues investigating analogical mapping across the senses supports the existence of a general analogy factor (Weinberger, Gallagher, Colaizzi, Liu, Parrott, Fearon, Shaikh, & Green, 2022). In particular, the authors tested relations between information presented in different modalities (e.g., words, sounds, lines). The participants were presented with two pairs of stimuli (A:B and C:D) and had to make binary true–false judgements about whether the relation conveyed in the first pair of stimuli (A:B) was analogous to the relation conveyed in the second pair (C:D). Stimuli included unimodal pairs, such lines-to-lines, words-to-words, as well as crossmodal stimuli, such as lines-to-sounds, lines-to-words, words-to-sounds. For example, a sequence of vertical and horizontal short segments (A) and a solid and longer segment (B) were paired with the words 'indecision' (C) and 'certainty' (D). The participants were then asked to judge whether A:B::C:D is true or false. In many cases, the stimuli not only presented an explicit structural relation, but involved mappings based on quantitatively specified ratios (e.g., the differences in pitch between a series of rising and falling tones is quantitatively comparable to the differences in distance between a series of lines on the screen). Weinberger et al.'s results demonstrated that their participants performed well-above-chance in the identification of crossmodal second-order relations, thus providing robust evidence of the use/existence of analogy across the senses.

[15] Note that the participants were also allowed to mark an X if they were unable to form a judgment, though apparently few of the participants marked one or more Xs.

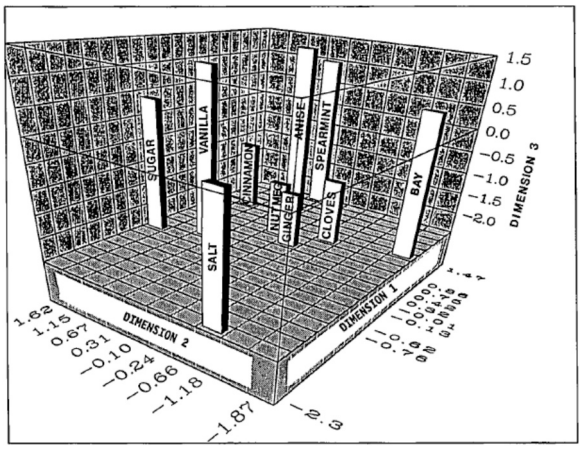

Fig. 3. Multidimensional scaling (MDS) configuration of spices achieved in a study by Blank and Mattes (1990). Dimension 1 was interpreted by participants as the intensity of odour, e.g., weak (salt) to strong (spearmint). Dimension 2 was interpreted as the compatibility with sweetness (e.g., more compatible (sugar) to less compatible (bay)). Dimension-3 was interpreted as the degree of bitter taste (e.g., more bitter (ginger) to less bitter (anise)). These three dimensions accounted for 69% of variance in the data. Notice here that sugar (a tastant) is rated as more similar to vanilla (an odour or flavour) than it is to another white crystalline tastant, namely salt. This may be one of the only examples of crossmodal phenomenological similarity (i.e., rating a tastants as more similar to an odorant than to another tastant). Of course, salt and sugar would have been expected to lie very close to each other if participants were to have judged the visual similarity of these pairs of stimuli instead.

These findings have clear implications for how sensory similarity is understood. In particular, if analogical mappings can be successfully constructed between the senses, even in the absence of shared perceptual qualities, then similarity judgments need not be grounded solely in phenomenal properties, but may instead rely on cognitively-imposed structures, such as ratios, relational patterns, or conceptual associations. In this sense, what appears as "sensory similarity" may sometimes reflect relational alignment rather than necessarily any kind of crossmodal perceptual resemblance. More generally, such results suggest that the experience of similarity can be mediated, scaffolded, or even generated by high-level interpretive criteria, thus blurring the line between perceptual and conceptual bases of crossmodal correspondence. This obviously immediately complicates any attempt to treat cross-sensory similarity as a purely perceptual phenomenon.

3.5. *Statistical associations between the senses*

The statistical account of crossmodal association is briefly presented, although, as will become clear, we believe that it has few, if any, meaningful implications regarding phenomenal similarity between the senses. The approach moves from the fact that many physical properties of sensory stimuli are often correlated, co-occurring in nature. The observer acquires, together with the stimuli, information concerning their regular co-occurrence, thus learning the statistical regularities of the environment and the correlations between multiple sensory cues (Shepard, 1994). Reflecting the properties of the environment (and in some cases also the laws of physics), most statistical correspondences are likely to be universally perceived (at least by those individuals who have been exposed to the same environment). For instance, consider only the pitch-size correspondence, which reflects a universal physical law (namely that larger objects tend to make lower-pitched noises when struck or sounded). That being said, it is important to note that multiple other factors also influence the pitch made by an object (e.g., consider only the phenomenon of 'dishonest signalling' in the animal kingdom), not to mention the material properties of the stimulus (Evans & Treisman, 2010; Gallace & Spence, 2006; Spence & Zampini, 2006). Once again, though, as far as similarity is concerned, statistical associations do not imply that the matched stimuli are, in any perceptually-salient sense, phenomenologically similar.

According to Spence (2022b), perhaps the only examples of crossmodal perceptual similarity resulting from statistical association may occur in the chemical senses.[16] In particular, those combinations of olfactory and gustatory stimuli that frequently co-occur together in foods. So, for example, consider only the fact that the smell of vanilla taking on the sweetness of sugar because the stimuli are so often combined in foods (Blank & Mattes, 1990; Spence, 2022b; Stevenson & Boakes, 2004). Note also that according to this account, there could well be individual differences in crossmodal perceptual similarity, depending on the particular food culture an individual grew up in (cf. Blank & Mattes, 1990). As such, perceptual (or phenomenological) similarity would only appear to exist between a subset of olfactory and gustatory stimuli (see Fig. 3). As noted by Spence (2024), this makes the consequences for phenomenal similarity of the statistical association between, say, odour and taste, importantly different from what might be a similar association between colour and taste. At the same time, similarity is restricted to gustatory quality, rather than tastes taking on the olfactory qualities with which it regularly co-occurs. While the fact that odours take on the qualities of the tastants with which they are commonly paired is now generally accepted, the most appropriate explanation for such effects is still the subject of debate (see Auvray & Spence, 2008 Stevenson & Boakes, 2004; Stevenson & Tomiczek, 2007).

The case of the chemical senses raises the question of how temporal factors (e.g., simultaneity and co-occurrence) shape the experience of phenomenal similarity. For instance, vanilla and sweetness, or citrus aromas and acidity, are often judged as "similar" or mutually fitting (Spence, 2022b, 2024; Stevenson & Boakes, 2004; Stevenson & Tomiczek, 2007). Yet these pairings may simply reflect statistical learning, thus making it unclear whether any felt similarity is being reinforced or rather induced by experience rather than grounded in intrinsic phenomenal qualities. This highlights the importance of being able to distinguish crossmodal congruency[17]—arising from learned associations or ecological regularities—from crossmodal similarity, implying a directly experienced affinity (cf. Stevenson, Rich, & Russell, 2012).[18]

4. Crossmodal similarity based on the structure and dimensions of the senses

We now turn to a different way of conceiving similarity relations, that is, those grounded not in the phenomenal contents of sensory experience, but in the structures, organizational principles, or processing mechanisms characteristic of different sensory modalities. Instead of asking whether the contents of two senses resemble one another (e.g., whether a Ferrari is red like a tomato, or a tone is "bright" like a colour), the approach considers whether two modalities exhibit parallel forms of information processing, analogous representational architectures, or else shared functional parts. For instance, it has long been suggested that

[16] It this context, it is interesting to note that Helmholtz (1878/1971) doesn't explicitly mention the phenomenal similarity, if any, between olfactory and gustatory sensations. As such, it is difficult to know, for certain, whether he thought that perceptual similarity judgments were equally difficult/impossible between this pair of distinct senses. However, it can perhaps legitimately be presumed that had that not been the case, he might have mentioned this exemption explicitly.

[17] Congruency can be defined as entities "going well together" (Maille & Fleck 2011) and may, in theory, result from perceptual, conceptual, and/or emotional similarities.

[18] It has been suggested that there may be a very limited range of innate crossmodal correspondences perhaps including intensity (Lewkowicz & Turkewitz, 1980) while the majority of the rest can be learnt via statistical co-occurrence (Spence, 2011; see also Deroy, Crisinel, & Spence, 2013; Shepard, 1981, 1987, 1994).

certain senses operate more "analytically" (e.g., audition, decomposing complex sounds into frequency components), whereas others function more "synthetically" (e.g., olfaction, delivering holistic odour percepts without discriminable parts). Similarly, both vision and touch encode spatial structure via topographic maps (Hsiao, 1998), whereas audition and olfaction primarily rely on temporal (or non-spatial) coding dimensions. Additionally, it must be considered whether similarity relationships might emerge across the senses due to similar background/foreground dynamics of the stimuli involved (e.g., sounds and shapes). According to such a view, sensory similarity can potentially arise from a number of structural/organizational factors, such as isomorphisms in information processing, shared computational principles, or comparable representational geometries, even when the sensory experiences in themselves are not phenomenologically similar.

4.1. Analogies between the senses

Over the years, a number of researchers have considered the parallels between the senses of hearing and vision, in particular, concerning pattern perception (Aksentijević, Elliott, & Barber, 2001; Julesz & Hirsh, 1972). This brings up the similarities in many of the Gestalt grouping principles that operate across the two spatial senses (e.g., see Spence, 2015, for a review). That said, the dimension on which grouping operates has been debated, with some arguing that space in vision maps onto time in audition. Indeed, a few decades ago, an informative debate emerged between two prominent North American psychologists concerning the most appropriate way in which to think about the cross-modal mapping between vision and audition (Handel, 1988a, b; Kubovy, 1988). Kubovy questioned the analogy that 'space is to vision' as 'time is to hearing'.[19] Note, though, that in this case, the analogy is operating at the level of the dimension of experience, rather than at the level of individual sensory stimuli (nor at the level of the analogy between relations between the stimuli presented in each modality; Wells, 1980).

Over the years, Kubovy and his colleagues have continued to investigate the parallels between auditory and visual sensory/perceptual information processing dimensions (e.g., Kubovy, 1981; Kubovy & Van Valkenburg, 2001; Kubovy & Yu, 2012). Difficulties emerge, however, when one attempts to clarify the nature of the correspondence, analogy, or alignment that may be implied in the space:time::vision:audition relationship. One may be inclined to reformulate the question in terms of the underlying basis on which such a relationship rests, possibly referring to the notion of alignment resulting from structural alignment, or the matching of the underlying sensory dimensions (this, according to Spence & Di Stefano, 2024c).

At the same time, however, there are also grounds for considering auditory pitch as analogous to visual space (e.g., when thinking about the retinotopic organization in early vision, and the tonotopic representation in early audition; e.g., Spence & Driver, 1994). As von Békésy (1964b, p. 811) noted: "The tongue is a sense organ with a large surface area. Since this is also true of the skin, the retina, and the organ of Corti, some analogies between these apparently different sense organs are to be expected." (see also Hsiao, 1998; Keidel, 1974; White, 1974).[20] For instance, while noting that similar 'what' vs. 'where' neurophysiological pathways are present in both the visual and tactile modalities, Hsiao (1998, p. 158) is also cautious in his conclusions when stating that: "Whether the similarities in touch and vision go beyond the peripheral level is speculative." Hsiao fails to offer an unequivocal answer to this question, instead suggesting that: "Although the two systems appear to be quite similar in many respects, there are many differences that indicate that visual processing is significantly more difficult and requires more neural machinery than touch." (Hsiao, 1998, p. 159).

Analogies can range widely, from phenomenal/perceptual to conceptual, and even to what might be called conventional correspondences (see Di Stefano & Spence, 2024a). For example, we may agree that vertical arrows pointing upward conventionally indicate forward movement along the horizontal axis, as is the case of traffic signs.[21] Similarly, the criteria that are used to establish the space:time::vision:audition analogy can vary considerably. These may include: (1) Analogies based on shared variation patterns across sensory modalities—for instance, both pitch and spatial height vary from low to high (i.e., they can be considered as polar dimensions though note the earlier point about pitch being metathetic in experience but physically prothetic; Smith & Sera, 1992),[22] despite involving entirely different physical dimensions (frequency vs. elevation)[23]; (2) Structural/statistical relationships internal to each modality, such as the fact that 'larger objects make louder sounds' (Gallace & Spence, 2006), without implying any direct equivalence between size and loudness; or (3) Metaphorical mappings grounded in affective or semantic associations—such as linking harsh timbres with spiky visual forms—where the analogy is mediated by emotional or symbolic interpretation rather than direct perceptual similarity (cf. Marks, 1987a, b; Murari, Rodà, Canazza, De Poli, & Da Pos, 2015; Parise & Spence, 2012). Note also that Boroditsky (2000) has argued for metaphoric structuring being involved in people's understanding of time in terms of spatial metaphors. A number of researchers have also studied the development of metaphoric understanding (Wagner, Winner, Cicchetti, & Gardner, 1981).

If distinct sensory modalities can represent the same external information, this raises the question of whether such correspondences are learned through experience or instead reflect innate representational constraints. Early developmental studies might be taken to support the latter view. For instance, Meltzoff and Borton's (1979) oral–visual matching experiments with one-month-old infants were interpreted as evidence that neonates possess an innate capacity to represent shape in a modality-independent format. According to such a view, spatial form would be available to both touch and vision from the outset, allowing crossmodal matching without prior associative learning. However, subsequent work has cast doubt on this strong innatist conclusion. Both replication failures and methodological critiques suggest that what is transferred across modalities may not be abstract shape *per se*, but rather lower-level features such as texture, hardness, or even numerosity (see Spence & Di Stefano, 2025c). In the same line, the study by De HeviaI, Izard, Coubart, Spelke, and Streri (2014) demonstrated that 0- to 3-day old neonates reacted to a simultaneous increase (or decrease) in spatial extent and in duration or numerical quantity, thus providing evidence that representations of space, time, and number are systematically interrelated prior to the acquisition of language and before extensive experience with the environmental correlations between these

[19] While intriguing, it is worth noting that this particular analogy fails to specify which specific dimension of space should map onto which specific temporal parameter in audition. For instance, does auditory duration correspond to visual distance? Should auditory rhythm be mapped onto spatial repetition, orientation, or motion trajectories? Without specifying a clear correspondence between specific parameters, the analogy risks remaining more rhetorical/qualitative than mechanistic/explanatory (Di Stefano & Spence, 2024).

[20] Though, on the very same point, Marks (1983, p. 9) observed that: "the skin is not a retina. Nor is it a cochlea. And the cochlea is not a retina.".

[21] 'Conventionally' here means that, in a different context, the very same vertical arrow could just as easily be interpreted as indicating upward movement along a spatial axis, without any additional effort on the part of the observer.

[22] Though see Dolscheid and Casasanto (2015) for the suggestion that spatial congruity effects may reveal metaphorical thinking, rather than necessarily polarity correspondences.

[23] At the same time, spatial attention has been shown to be directed vertically as a result of the crossmodal correspondences operating at the level of the relative pitch of auditory stimuli (Spence, 2011).

dimensions.

4.2. Primary sensations and the distinction between analytic vs. synthetic senses

Besides focusing on specific processing mechanisms of sensory perception, researchers have often reflected on the general structure of the senses, that is, asking whether global features emerge that characterize one sensory modality as compared to the others. For instance, the general distinction between spatial and non-spatial, or chemical, senses provide an example of how sensory modalities differ in terms of their ability to represent, or deal with, spatial or chemosensory information. Needless to say, classifications of this sort are relevant to the issue of similarity, allegedly constraining its applicability across the senses.

The notion of basic, or primary, sensory qualities is perhaps worth discussing here in that it refers to the way in which sensory contents are built in perception, namely, that some basic or atomic elements give rise to more complex contents. The existence of basic sensory qualities has been applied by many commentators to taste (gustation), and to the dimension of colour (vision; see Spence & Di Stefano, in press). Though it is by no means always clear whether researchers have quite the same thing in mind when using the term 'basic' with respect to different senses (cf. Byrne & Hilbert, 2008; Twomey, Roberts, Brainard, & Plotkin, 2021). As such, the mere existence of primary (or basic) sensations has implications for the way in which the senses process information, that is, in either an analytic or synthetic manner. While audition and vision are exemplarily considered analytic and synthetic senses, respectively, much blurred or debated is the nature of taste and olfaction. For instance, Breslin (2001) highlights how taste is in some ways a synthetic sense, like vision, while at the same time, sharing something of the status of an analytic sense, like audition. That is, similarity relations can simultaneously pull in different directions.

McBurney and Gent (1979) draw an analogy between gustation and the skin senses instead,[24] due to the suggestion that separate channels of tactile stimulation (just like discrete taste qualities) can perhaps be considered as different sub-modalities of touch. They go on to suggest that the analytic/synthetic question is anyway irrelevant to the discussion of whether there are basic tastes. Relevant to this discussion, Erickson and Covey (1980) conducted an intriguing study of taste primaries in which separate groups of participants were presented with either single or double combinations of colours, pitches, or tastants. The participants in the study were simply asked to indicate whether the stimuli were 'singular' or 'more-than-one'. In line with the suggestion that vision is a synthetic sense, the participants rated both single hues and combinations of two hues as more 'unified' than 'multiple'. In contrast, single tones were rated as unified, whereas pairs of tones were rated as 'more than one', thus supporting the analytic nature of audition. Intriguingly, when it came to gustatory stimuli, participants' responses were neither purely synthetic (like vision), nor purely analytic (like audition), but somewhere in-between, depending on the basic taste/taste mixture that was being evaluated. Erickson and Covey (1980, p. 531) concluded that: "The most conspicuous aspects of the gustatory data are that single stimuli, including the "primary" stimuli, are not consistently judged as singular, and that mixtures are not consistently perceived as being more than one. Instead, a more continuous distribution of ratings from nearly "singular" to nearly "more than one" was found." These results lead Erickson and Covey to conclude that neither the analytic or synthetic viewpoints, are appropriate analogies in the context of thinking about tastes.

4.3. Prothetic, metathetic, and polar sensory dimensions

Stevens (1957) highlighted an important distinction between 'metathetic' and 'prothetic' stimulus dimensions. Prothetic dimensions have a clear 'more than' and 'less than' end. Stevens refers to these as quantitative perceptual continua (i.e., having to do with 'how much') with ratio properties. Among prothetic dimensions, loudness, brightness, lightness, heaviness, duration, roughness, and area can all be mentioned. In contrast, metathetic dimensions obey a well-structured organization without necessarily having a 'more than' or 'less than' end. Stevens classified these dimensions as 'what kind' or 'where (position)'. For instance, both auditory pitch and hue are (perceptually) metathetic dimensions, as are visual position, inclination, and proportion, according to Stevens. Smith and Sera (1992) subsequently added shape to the list. Interestingly, however, pitch and hue can also both be represented as circular dimensions. (In fact, according to Pridmore (1992), they may be the only perceptual dimensions that can be so represented).

Polar dimensions should also be mentioned here. According to Clark (1970), polar dimensions are fundamental to children's development of relational concepts. Dimensions such as big vs. little, high vs. low, but presumably also left vs. right, sweet vs. sour/bitter are all polar. Notice here how a polar organization can be applied to prothetic dimensions (big vs. little; spatial high vs. low), to metathetic dimensions (such as left vs. right; high vs. low pitch), and to pairs of stimuli that cannot be organized as part of either prothetic or metathetic dimensions, such as, for example, between pairs of basic tastes or olfactory stimuli (Proctor & Cho, 2006). Polar dimensions can presumably be aligned with other polar dimensions although which poles people choose to align may, on rare occasions, turn out to lead to divergent mappings (Marks, 1974). Intriguingly, certain of the basic tastes (having no commonly agreed organizational principle) have nevertheless been shown to correspond to the polar dimension of elevation (e.g., high vs. low; see Velasco, Adams, Petit, & Spence, 2019).

These distinctions are likely relevant for theories of sensory similarity. Notice that prothetic and metathetic continua afford different types of similarity relations: prothetic dimensions tend to support graded, metric-like similarity judgments (e.g., two sounds are similar in loudness to the extent that their magnitudes differ only slightly), whereas metathetic dimensions often yield similarity structures based on categorical or positional relations rather than magnitude. Polar dimensions allow similarity to be grounded in opposition rather than continuous distance. Because different sensory domains instantiate these types of dimensional structures to different degrees, recognizing whether a stimulus feature is prothetic, metathetic, circular, or polar helps explain why certain crossmodal matches (or within-modality similarities) may appear natural, while others can feel arbitrary or strained instead. These dimensional properties therefore shape the geometry of perceptual spaces and constrain the kinds of similarity relations that can emerge within and across the senses (see Di Stefano et al., under review, for a comprehensive study on the multisensory assessment of musical intervals).

4.4. Temporal and spatial patterns

Crossmodal similarity may also emerge from shared temporal and spatial patterns of stimuli across modalities. For instance, descriptions of gustatory experience, such as those reported in studies of basic tastes in solution, frequently rely on temporal metaphors (sharp onsets, lingering finishes, or evolving trajectories) that parallel the way in which people describe auditory or visual events (Obrist, Comber, Subramanian, Piqueras-Fiszman, Velasco, & Spence, 2014). Similarly, the classic

[24] McBurney and Gent (1979, p. 165) end their article by concluding that: "If one must make analogies among the senses, and one must, then we argue that the proper senses with which one should compare taste are the skin senses. After all, the sense of taste shares the same receptor surface, nerves, neurons, and central projections as do the other skin senses. When one makes these analogies, one sees that the notion of four basic qualities or submodalities of taste has a great deal of empirical evidence to recommend it and avoids the pitfalls into which the analogy with vision has led us.".

Table 2
Summary of different kinds of similarities between the senses.

Sensory similarity			
	Phenomenal	Structural	Affective/semantic
Claim	Sensory contents are similar	The way in which sensory contents are organized or structured is similar	The affective or semantic meaning attributed to sensory contents is similar
Example(s)	Tomatoes and Ferraris are both red Trumpet and french horn have similar timbre	Intermittent lights and noise Increasing brightness and a rising glissando	A siren call and a warning signal Debussy's *Prélude à l'après-midi d'un faune* and Monet's *Nymphéas*
Sensory/ cognitive	Sensory	Sensory	Sensory and cognitive
Roots	Rooted into the physiology and physical properties of the stimuli	Rooted in the processing of dynamic stimuli	Relies on the affective or cognitive elaboration of the stimuli
Observations	Not (or not significantly) influenced by culture, individual experience, or language (except perhaps in the special case of odours taking on commonly-paired taste qualities)	Might be influenced by individual/ subjective factors	May be influenced by individual experience, cultural background, and language

bouba–kiki effect may be grounded in similarities in how sensory events unfold over time: abrupt, discontinuous changes in shape align with the sharpness of certain acoustic transitions, whereas smoother contours correspond to gradual auditory transformations (Bremner, Caparos, Davidoff, de Fockert, Linnell, & Spence, 2013). These examples suggest that the structural dynamics of a stimulus—its (experienced) temporal profile, rate of change, or spatial contour—can serve as a crossmodal bridge for similarity judgments (see also Wanke et al., 2025).

A related source of evidence here comes from the language that people spontaneously choose to use to describe their sensory experiences. Notice how terms such as harmony, top notes, middle, base, and even chords are regularly applied not only to sounds but also to smells, reflecting a putative similarity in the organizational structure of auditory and olfactory phenomenal experiences (Spence, Di Stefano, Reinoso-Carvalho, Mesz, & Zacharakis, 2026). Such linguistic parallels might be taken to reveal that people often conceptualize sensations through shared structural and/or relational frameworks (Cazeaux, 2002). These convergences in temporal form, spatial patterning, and descriptive vocabulary collectively highlight a level of sensory similarity that is neither strictly phenomenal nor purely conceptual, but rooted in the morphodynamic evolution of perceptual experience.

4.5. Interim summary

To summarize, sensory similarity can be categorized into three different categories: *Phenomenal similarity* concerns those cases in which sensory contents themselves are alike, as when tomatoes and a Ferrari are both red or when a trumpet and a French horn share a similar timbre. These judgments are primarily sensory in nature and are rooted in the physiology of perception and the physical properties of the stimuli (though the story may be different in the unique case where olfactory stimuli take on the taste qualities with which they are commonly paired; Spence, 2022b). *Structural similarity*, by contrast, arises when the organization or dynamics of sensory contents align, as in the parallel between intermittent lights and noise, or between increasing brightness and a rising glissando. Such similarities depend on how the perceptual system processes dynamic patterns, and although still largely sensory, they may vary across individuals. Finally, *affective or semantic similarity* involves those cases in which different stimuli are experienced as alike because they convey comparable emotional, expressive, or conceptual meanings. For instance, the resemblance between a siren call and a warning signal, or the often-perceived affinity between Debussy's music (e.g., *Prélude à l'après-midi d'un faune*) and Monet's impressionist paintings (e.g., *Nymphéas*). Such similarity judgments rely on a combination of sensory input and cognitive, stylistic, cultural, and linguistic mediation, and are therefore the most variable across individuals and contexts (see Table 2 for a summary.

Here, it might be useful to consider briefly how the tripartite division outlined here links to Marks' (1983) earlier distinction between phenomenological, information processing, psychophysics, and neurophysiological similarity. We have chosen not to focus on Marks' 'informational similarity', which seems, at least in our view, circular. That is, Marks suggests that informational similarity is when different sensory systems provide information about the same sensory quality, which to our way of thinking is actually one of the open questions that need to be answered; namely, whether and how different sensory systems can provide information about the very same property. Marks' 'phenomenological similarity' maps on to our phenomenal similarity, but here, by using the term 'phenomenal' we stress that such similarity is established at early stages of sensory processing, while the term 'phenomenological' seems to refer to the configuration and organization stages of sensory contents into unitary and structured percepts. Finally, Marks' notion of 'psychophysical similarity' refers quite closely to our notion of structural similarity, which nevertheless sounds more neutral, or general.

A final complication should be mentioned here, and refers to those rare cases in which similarity of the physical stimuli themselves is high, yet perceptual similarity completely fails. A well-known olfactory example occurs in the case of olfactory enantiomers, molecules that are mirror-image isomers of each other. Enantiomers share identical physico-chemical properties in almost every respect (e.g., molecular weight, bond structure, energy spectra), differing only in terms of their spatial chirality. Despite their similarity, however, humans sometimes perceive them as having entirely distinct odours. A classic demonstration involves the enantiomers of carvone, one of which smells like spearmint, the other like caraway, even though the molecules are chemically identical except for handedness (Brookes et al., 2009), Such phenomena demonstrate that even when physical structure is nearly indistinguishable, perceptual systems may map these stimuli into very different qualitative categories. Such cases can be taken to challenge any view that physical similarity is either sufficient, or necessary, for perceptual similarity, thus making an important point for crossmodal similarity clearer. If physical similarity within a single modality does not guarantee phenomenal similarity, then it will be even less likely to predict crossmodal correspondences in any straightforward way.

5. Individual differences, group differences, and modality-based differences in sensory similarity

Ultimately, it is perhaps important to return, albeit briefly, to the question, raised earlier in this narrative theoretical review, concerning the possibility of individual differences in the perception of crossmodal similarity (e.g., Marks, 1989b; Rader & Tellegen, 1987). While parsimony might speak to the desirability of a unitary solution, as yet, we have not come across any evidence that unequivocally argues for either a unitary or multifarious answer to this question. As such, this is undoubtedly an important question for future research in the area of

Table 3
A set of diagnostic prompts to help distinguish phenomenal, structural, and affective/semantic similarity claims in practice. Note that most empirical phenomena likely involve more than one level, the diagnostic prompts are therefore intended to identify the primary explanatory level of a given claim regarding similarity.

Similarity type	Diagnostic question	Hallmarks	Comments
Phenomenal	Is the claim about similarity in what-it-is-like to experience the stimuli?	Reports of felt resemblance; intramodal analogies extended crossmodally; appeals to sensory qualities (qualia)	Intuitively clear in unisensory context, more problematic across the senses
Structural / organizational	Is the claim about shared relations, dynamics, or spatial/temporal structure rather than sensory feel?	No evident similarity between qualia; Relative position on a scale; similarity between spatio/temporal patterns	Might be mistakenly treated as phenomenal similarity (e.g., low pitch is similar to dark light)
Affective / semantic	Is the similarity mediated by shared emotion, meaning, metaphor, or concept?	Convergence in affective ratings or semantic differentials	Leave the question open as to why the stimuli elicited similar affective / semantic meaning

sensory similarity.[25] In the absence of robust data concerning such individual differences, it is difficult, as yet, to speculate on what theoretical implications, were they to be observed, might have (Churchland (1998). Nevertheless, however, one might consider whether those individuals who exhibit absolute perceptual judgments in some domain, such as, for example, those who have perfect pitch (see Di Stefano & Spence, 2024b, for a review) might exhibit distinct pattern of structural/organizational crossmodal similarity judgments as compared to others where the relative, rather than the absolute, pitch of sounds appears to be far more perceptually relevant. It is also worth noting how cross-cultural research has sometimes delivered a different answer, e.g., in terms of the crossmodal mapping of bitterness to angularity (see Bremner et al., 2013). While synaesthesia can be discounted (as speaking to the issue), there are various other groups of individuals who either fail to experience certain crossmodal correspondences, or else exhibit different ones (see Spence, 2022a). There have also been occasional reports of different polarities of crossmodal correspondence mapping amongst populations (e.g., in the correspondence between visual lightness and auditory pitch; see Marks, 1974).

The one area where there is a clear opportunity for individual differences in sensory similarity is in the case of those odorants that take on the associated taste properties. Notice how it is here, and only here, that cultural differences in cuisine will likely give rise to sometimes different patterns of similarity between food aromas and basic tastes (cf. Blank & Mattes, 1990).

6. Testable predictions and methodological guidance for future research

If similarity reflects genuinely phenomenal overlap between sensory contents, then crossmodal matching should be robust to changes in stimulus orientation, scale, modality-specific exploratory constraints, and context (or range effects; see Cohen, 1934). By contrast, if similarity is structural or relational, then performance should depend on the preservation of higher-order relations (e.g., relative position within a sensory dimension or spatiotemporal pattern) and should be affected by disruption of these relations, even when stimuli features are the same (see Table 3 for a set of diagnostic prompts to help distinguish the different kinds of similarity prompts). In the future, experimental designs that separately manipulate dimensions often conflated in previous work might help to shed light on the relational/structural nature of similarity. For example, studies of crossmodal "shape" transfer should independently manipulate shape, size, texture, and numerosity, rather than treating shape as a unitary property.

7. Conclusions

In conclusion, multiple similarities between the senses have been suggested by researchers over the decades (e.g., Marks, 1983). Most frequently mentioned is the phenomenal, but this is also most problematic, especially in the crossmodal domain, given the widespread appeal to synaesthesia (Deroy & Spence, 2013; Martino & Marks, 2001). However, as has been highlighted in this review, multiple other parallels between different combinations of senses have also been proposed. While researchers could, presumably, have chosen to compare any pair of senses, it is noticeable how the majority of commentators have chosen to focus on the comparison of, and/or analogies between, vision with audition (e.g., Geldard, 1970; Harris, 1950; Julesz & Hirsh, 1972; Stevens, 1958). Partly, it could be imagined, this might be the result of visual dominance (Hutmacher, 2019), and also the similar circular/polar structure of the underlying perceptual continua in the case of auditory pitch and visual hue (Smith & Sera, 1992; Spence & Di Stefano, 2024b).[26] Others, like Von Békésy have been drawn to the parallels between audition and touch, given that, within a narrow frequency band, the same physical stimulus can potentially be both heard and felt.

At the same time, however, as highlighted by Di Stefano and Spence (2024a), similarity is a complex notion, regardless of the type of stimuli (individual sensory stimuli, entire sensory dimensions or modalities, or anything else for that matter) that one happens to be comparing, especially given the multiple sensory, perceptual, and cognitive dimensions along which similarity often operates (e.g., Attneave, 1950; Landahl, 1945; Matthen, 2005; Medin, Goldstone, & Gentner, 1990, 1993; Melara, Marks, & Lesko, 1992; Nosofsky, 1992; Shepard, 1974; Tversky, 1977; Yearsley, Pothos, Barque-Duran, Trueblood, & Hampton, 2022).

It is important to recognize the theoretical work that specific crossmodal analogies are intended to do in terms of problem-solving and reasoning (e.g., Gentner, 1989; Gentner & Holyoak, 1997), including influencing the way in which we think about the operation of the senses (von Békésy, 1959, 1964a).[27] What is more, it should also be stressed that no one analogy is likely to capture the operation of a complex system, and thus multiple analogies have been proposed as the most appropriate route when considering any kind of complex concept (i.e., to help avoid analogy-induced misconception in the acquisition of advanced knowledge; Gentner & Gentner, 1983; Spiro et al., 1989). The same recommendation would appear to be most appropriate when it comes to thinking about sensory similarities too.

It is interesting to at least ask the question of what is known concerning the cognitive neuroscience underpinnings of crossmodal perceptual similarity judgments. Unfortunately, there is little concrete evidence as yet relevant to this theme. That said, one might legitimately

[25] This problem becomes all the more challenging when it comes to assessing questions of sensory similarity in other species (see Spence & Deroy, 2012).

[26] Also relevant here may be the preponderance of audiovisual forms of crossmodal synaesthesia (e.g., consider only the extensive scientific literature and artistic interest in 'colour music'; Karwoski & Odbert, 1938; Lorusso & Porro, 2010; MacDougal, 1898), as compared to other combinations of crossmodal inducer and concurrent (e.g., Day, 2005), which has, in turn, led to a whole movement in the arts towards color music (e.g., Zilczer, 1987).

[27] As von Békésy (1964a, p. 369) put it: "The existence of an analogy between hearing and smelling may help to delineate some electrophysiological correlates of the directional sensation, which is so well developed in both senses." This following von Békésy's (1959, p. 10) earlier suggestion that: "…the similarity between hearing and the skin sensations can be of real value.".

expect different neural substrates depending on the type of similarity under consideration (shared phenomenal qualities, analogous structural or organizational patterns, or convergent affective or semantic meanings, each grounded in increasingly cognitive and culturally-mediated processes), as outlined here, and perhaps also depending on the combination of modalities involved (i.e., presumably audiovisual similarity judgments are likely to recruit very different neural substrates than gustatory-olfactory decisions).

CRediT authorship contribution statement

Charles Spence: Writing – review & editing, Writing – original draft, Conceptualization. **Nicola Di Stefano:** Writing – review & editing, Writing – original draft, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

Actis-Grasso, R., Zavagno, D., Lega, C., Zani, A., Daneyko, O., & Cattaneo, Z. (2017). Can music be figurative? Exploring the possibility of crossmodal similarities between music and visual arts. *Psihologija, 50*(3), 285–306. https://doi.org/10.2298/PSI1703285A

Aksentijević, A., Elliott, M. A., & Barber, P. J. (2001). Dynamics of perceptual grouping: Similarities in the organization of visual and auditory groups. *Visual Cognition, 8*(3–5), 349–358. https://doi.org/10.1080/13506280143000043

Attneave, F. (1950). Dimensions of similarity. *American Journal of Psychology, 63*(4), 516–556. https://doi.org/10.2307/1418869

Auvray, M., & Spence, C. (2008). The multisensory perception of flavor. *Consciousness and Cognition, 17*(3), 1016–1031. https://doi.org/10.1016/j.concog.2007.06.005

Bahrick, L. E. (2010). Amodal perception. In E. B. Goldstein (Ed.), *Encylopedia of perception* (pp. 44–46). Thousand Oaks, CA: Sage Publications.

Behrens, M., Gunn, H., Ramos, P., Meyerhof, W., & Wooding, S. (2013). Genetic, functional, and phenotypic diversity in TAS2R38-mediated bitter taste perception. *Chemical Senses, 38*(6), 475–484. https://doi.org/10.1093/chemse/bjt016

Ben-Artzi, E., & Marks, L. E. (1995). Visual-auditory interaction in speeded classification: Role of stimulus difference. *Perception & Psychophysics, 57*(8), 1151–1162. https://doi.org/10.3758/BF03208371

Bianchi, I., & Burro, R. (2023). The perception of similarity, difference and opposition. *Journal of Intelligence, 11*(9), 172. https://doi.org/10.3390/jintelligence11090172

Blank, D. M., & Mattes, R. D. (1990). Sugar and spice: Similarities and sensory attributes. *Nursing Research, 39*(5), 290–293. PMID: 2399134.

Bratzke, D. (2025). Do "auditory" and "visual" time really feel the same? Effects of stimulus modality on duration and passage–of–time judgements. *Attention, Perception, & Psychophysics, 87*, 2241–2249. https://doi.org/10.3758/s13414-025-03131-5

Bremner, A., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K., & 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. https://doi.org/10.1016/j.cognition.2012.09.007

Bremner, A., Lewkowicz, D., & Spence, C. (Eds.). (2012). *Multisensory development*. Oxford, UK: Oxford University Press.

Breslin, P. A. S. (2001). Human gustation and flavour. *Flavour and Fragrance Journal, 16*(6), 439–456. https://doi.org/10.1002/ffj.1054

Brookes, J. C. (2010). Science is perception: What can our sense of smell tell us about ourselves and the world around us? *Philosophical Transactions of the Royal Society A. Mathematical, Physical and Engineering Sciences, 368*(1924), 3491–3502. https://doi.org/10.1098/rsta.2010.0117

Brookes, J. C., Horsfield, A. P., & Stoneham, A. M. (2009). Odour character differences for enantiomers correlate with molecular flexibility. *Journal of the Royal Society Interface, 6*(30), 75–86. https://doi.org/10.1098/rsif.2008.0165

Byrne, A., & Hilbert, D. R. (2008). Basic sensible qualities and the structure of appearance. *Philosophical Issues, 18*(1), 385–405. https://doi.org/10.1111/j.1533-6077.2008.00153.x

Calvert, G. A., Spence, C., & Stein, B. E. (Eds.). (2004). *The handbook of multisensory processes*. Cambridge, MA: MIT Press.

Cazeaux, C. (2002). Metaphor and the categorization of the senses. *Metaphor and Symbol, 17*(1), 3–26. https://doi.org/10.1207/S15327868MS1701_2

Chen, Y.-C., & Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition, 114*(3), 389–404. https://doi.org/10.1016/j.cognition.2009.10.012

Chen, Y.-C., & Spence, C. (2017). Assessing the role of the 'unity assumption' on multisensory integration: A review. *Frontiers in Psychology, 8*, 445. https://doi.org/10.3389/fpsyg.2017.00445

Churchland, P. M. (1998). Conceptual similarity across sensory and neural diversity: The Fodor/Lepore challenge answered. *The Journal of Philosophy, 95*(1), 5–32. https://doi.org/10.2307/2564566

Clark, H. H. (1970). The primitive nature of children's relational concepts. In J. R. Haveys (Ed.), *Cognition and the development of language* (pp. 269–278). New York, NY: Wiley.

Cohen, N. E. (1934). Equivalence of brightness across modalities. *American Journal of Psychology, 46*(1), 117–119. https://doi.org/10.2307/1416240

Crisinel, A.-S., & Spence, C. (2012). The impact of pleasantness ratings on crossmodal associations between food samples and musical notes. *Food Quality and Preference, 24*, 136–140. https://doi.org/10.1016/j.foodqual.2011.10.007

Critchley, M. (1979). *The divine banquet of the brain and other essays*. New York, NY: Raven Press.

Day, S. (2005). Some demographic and socio-cultural aspects of synesthesia. In L. C. Robertson, & N. Sagiv (Eds.), *Synesthesia: Perspectives from cognitive neuroscience* (pp. 11–33). New York, NY: Oxford University Press.

de Hevia, M. D., Izard, V., Coubart, A., Spelke, E. S., & Streri, A. (2014). Representations of space, time, and number in neonates. *Proceedings of the National Academy of Sciences of the USA, 111*(13), 4809–4813. https://doi.org/10.1073/pnas.1323628111

Deiana, D., Ansani, A., Di Pino, G., Formica, D., Spence, C., & Di Stefano, N. (2025). Auditory and tactile perception of musical intervals: A pilot study. *IEEE Medical Measurements & Applications (MeMeA)*, (pp. 1-6). Chania, Greece. DOI: 10.1109/MeMeA65319.2025.11068071.

Deroy, O., Crisinel, A.-S., & Spence, C. (2013). Crossmodal correspondences between odors and contingent features: Odors, musical notes, and geometrical shapes. *Psychonomic Bulletin & Review, 20*, 878–896. https://doi.org/10.3758/s13423-013-0397-0

Deroy, O., & Spence, C. (2013). Why we are not all synesthetes (not even weakly so). *Psychonomic Bulletin & Review, 20*, 643–664. https://doi.org/10.3758/s13423-013-0387-2

Dimova, P. (2024). *At the crossroads of the senses: The synaesthetic metaphor across the arts in European Modernism*. Pennsylvania State University Press.

Di Stefano, N., Ansani, A., Schiavio, A., & Spence, C. (2024). Prokofiev was (almost) right: A cross-cultural exploration of auditory-conceptual associations in Peter and the Wolf. *Psychonomic Bulletin & Review, 31*, 1735–1744. https://doi.org/10.3758/s13423-023-02435-7

Di Stefano, N., Ansani, A., Focaroli, V., Borsella, R., Formenti, G., Velardi, A., Schiavio, A., & Spence, C. (2026). Auditory-conceptual associations in Peter and the Wolf and Carnival of the animals: Evidence from 6-9 year-old children. *Psychonomic Bulletin & Review, 33*, 28. https://doi.org/10.3758/s13423-025-02804-4

Di Stefano, N., & Spence, C. (2022). Roughness: A multisensory/crossmodal perspective. *Attention, Perception & Psychophysics, 84*, 2087–2114. https://doi.org/10.3758/s13414-022-02550-y

Di Stefano, N., & Spence, C. (2024a). Perceptual similarity: Insights from the crossmodal correspondences. *Review of Philosophy & Psychology, 15*(3), 997–1026. https://doi.org/10.1007/s13164-023-00692-y

Di Stefano, N., & Spence, C. (2024b). Should absolute pitch be considered as a unique example of context-free sensory judgments in humans? *Cognition, 249*, Article 105805. https://doi.org/10.1016/j.cognition.2024.105805

Di Stefano, N., & Spence, C. (2025). Perceiving temporal structure within and between the senses: A crossmodal/multisensory perspective. *Attention, Perception, & Psychophysics, 87*, 1811–1838. https://doi.org/10.3758/s13414-025-03045-2

Di Stefano, N., Ansani, A., Schiavio, A., Saarikallio, S., Toiviainen, P., Brattico, E., & Spence C. (under review). Consonance shapes the multisensory and emotional mappings of musical intervals across Western and Eastern cultures. *Annals of the New York Academy of Sciences*.

Dolscheid, S., & Casasanto, D. (2015). Spatial congruity effects reveal metaphorical thinking, not polarity correspondence. *Frontiers in Psychology, 6*, 1836. https://doi.org/10.3389/fpsyg.2015.01836

Donnell-Kotrozo, C. (1978). Intersensory perception of music: Color me trombone. *Music Educators Journal*, **65(December)**, 32-37. ERIC Number: EJ204081.

Duthie, A. C. (2013). *Do music and art influence one another? Measuring cross-modal similarities in music and art*. Iowa State University.

Duthie, C., & Duthie, B. (2015). Do music and art influence one another? Measuring cross-modal similarities in music and art. *Polymath: An Interdisciplinary Arts and Sciences Journal, 5*(1). https://bradduthie.github.io/Duthie%26Duthie2015.pdf.

Eerola, T., & Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music, 39*(1), 18–49. https://doi.org/10.1177/0305735610362821

Erickson, R. P., & Covey, E. (1980). On the singularity of taste sensations: What is a taste primary? *Physiology & Behavior, 25*(4), 527–533. https://doi.org/10.1016/0031-9384(80)90117-1

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision, 10*(1), 6. https://doi.org/10.1167/10.1.6

Everson, S. (1997). *Aristotle on perception*. Oxford, UK: Oxford University Press.

Ferrari, R. (2015). Writing narrative style literature reviews. *Medical Writing, 24*(4), 230–235. https://doi.org/10.1179/2047480615Z.000000000329

Fodor, J. (1975). *The language of thought*. Harvard University Press.

Forbus, K. D., Gentner, D., & Law, K. (1994). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science, 19*, 141–205. https://doi.org/10.1207/s15516709cog1902_1

Furley, P., & Goldschmied, N. (2021). Systematic vs. narrative reviews in sport and exercise psychology: Is either approach superior to the other? *Frontiers in Psychology, 12*, Article 685082. https://doi.org/10.3389/fpsyg.2021.685082

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics, 68*(7), 1191–1203. https://doi.org/10.3758/BF03193720

Gallace, A., Ngo, M. K., Sulaitis, J., & Spence, C. (2012). Multisensory presence in virtual reality: Possibilities & limitations. In G. Ghinea, F. Andres, & S. Gulliver (Eds.), *Multiple sensorial media advances and applications: New developments in MulSeMedia* (pp. 1–38). Hershey, PA: IGI Global.

Gardner, H. (1974). Metaphors and modalities: How children project polar adjectives onto diverse domains. *Child Development, 45*(1), 84–91. https://doi.org/10.2307/1127753

Geldard, F. A. (1970). Vision, audition and beyond. In W. D. Neff (Ed.), *Contributions to sensory physiology* (Vol. 4, pp. 1–17). New York, NY: Academic Press. https://doi.org/10.1016/b978-0-12-151804-2.50007-3.

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou, & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 199–241). Cambridge, UK: Cambridge University Press.

Gentner, D., & Gentner, D. R. (1983). Flowing waters or teeming crowds: Mental models of electricity. In D. Gentner, & A. L. Stevens (Eds.), *Mental models* (pp. 99–120). Hillsdale, NJ: Lawrence Erlbaum.

Gentner, D., & Holyoak, K. J. (1997). Reasoning and learning by analogy. *American Psychologist, 52*(1), 32–34. https://doi.org/10.1037/0003-066X.52.1.32

Gentner, D., & Markham, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52*(1), 45–56. https://doi.org/10.1037/0003-066X.52.1.45

Gentner, D., Holyoak, K. J., & Kokinov, B. N. (2001). *The analogical mind: Perspectives from cognitive science*. Cambridge, MA: MIT Press.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston, MA: Houghton-Mifflin.

Gombrich, E. H. (1960). *Art and illusion*. London, UK: Phaidon Press.

Goodman, N. (1972). Seven strictures on similarity. In N. Goodman (Ed.), *Problems and projects* (pp. 437–446). Indianapolis: Bobs-Merril.

Handel, S. (1988a). Space is to time as vision is to audition: Seductive but misleading. *Journal of Experimental Psychology: Human Perception and Performance, 14*(2), 315–317. https://doi.org/10.1037/0096-1523.14.2.315

Handel, S. (1988b). No one analogy is sufficient: Rejoinder to Kubovy. *Journal of Experimental Psychology: Human Perception and Performance, 14*(2), 321. https://doi.org/10.1037/h0092865

Harris, J. D. (1950). *Some relations between vision and audition*. Springfield, Illinois: Charles C. Thomas.

Hartshorne, C. (1934). *The philosophy and psychology of sensation*. Chicago, IL: University of Chicago Press.

Helmholtz, H.v. (1878/1971). *Treatise on physiological optics (Vol. II)*. New York, NY: Dover Publications.

Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition, 15*(4), 332–340. https://doi.org/10.3758/BF03197035

Hsiao, S. S. (1998). Similarities between touch and vision. *Advances in Psychology*, **127**, 131-165. (J. W. Morley, Ed., *Neural Aspects of Tactile Sensation*. Elsevier Science B. V.) DOI: 10.1016/S0166-4115(98)80066-6.

Hutmacher, F. (2019). Why is there so much more research on vision than on any other sensory modality? *Frontiers in Psychology, 10*, 2246. https://doi.org/10.3389/fpsyg.2019.02246

Janowski, M., & Chełkowska-Zacharewicz, M. (2019). What do we actually measure as music-induced emotions? *Roczniki Psychologiczne, 22*(4), 373–403. https://doi.org/10.18290/rpsych.2019.22.4-5

Julesz, B., & Hirsh, I. J. (1972). Visual and auditory perception - an essay of comparison. In E. E. David, Jr., & P. B. Denes (Eds.), *Human communication: A unified view* (pp. 283–340). New York, NY: McGraw-Hill.

Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: Towards a unified theory of musical emotions. *Physics of Life Reviews, 10*(3), 235–266. https://doi.org/10.1016/j.plrev.2013.05.008

Kandinsky, W. (1977). *Concerning the spiritual in art, especially in painting* (trans. M. T. H. Sadler). New York, NY: Dover Publications.

Karwoski, T. F., & Odbert, H. S. (1938). Color-music. *Psychological Monographs, 50*(2), i-60. https://doi.org/10.1037/h0093458

Karwoski, T. F., Odbert, H. S., & Osgood, C. E. (1942). Studies in synesthetic thinking. II. The rôle of form in visual responses to music. *Journal of General Psychology, 26*, 199–222. https://doi.org/10.1080/00221309.1942.10545166

Keidel, W. D. (1974). The cochlear model in skin stimulation. In F. A. Geldard (Ed.), *Cutaneous communication systems and devices* (pp. 27–32). The Psychonomic Society Inc.

Kimura, A., Wada, Y., & Noguchi, K. (2005). The effect of the similarity of affective meanings on harmony between shape and color. *Bulletin of Japanese Society for the Science of Design, 52*, 1–8.

Krantz, D. H. (1972). A theory of magnitude estimation and cross-modality matching. *Journal of Mathematical Psychology, 9*(2), 168–199. https://doi.org/10.1016/0022-2496(72)90025-9

Kripke, S. A. (1980). *Naming and necessity*. Oxford, UK: Blackwell.

Kubovy, M. (1981). Concurrent-pitch segregation and the theory of indispensable attributes. In M. Kubovy, & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 55–98). Hillsdale, NJ: Erlbaum.

Kubovy, M. (1988). Should we resist the seductiveness of the space:Time::Vision: Audition analogy? *Journal of Experimental Psychology: Human Perception and Performance, 14*(2), 318–320. https://doi.org/10.1037/0096-1523.14.2.318

Kubovy, M., & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition, 80*(1–2), 97–126. https://doi.org/10.1016/s0010-0277(00)00155-4

Kubovy, M., & Yu, M. (2012). Multistability, cross-modal binding and the additivity of conjoint grouping principles. *Philosophical Transactions of the Royal Society B, 367*(1591), 954–964. https://doi.org/10.1098/rstb.2011.0365

Külpe, O. (1893). *Grundriss der Psychologie*. Leipzig: Engelmann.

Landahl, H. D. (1945). Neural mechanisms for the concepts of difference and similarity. *Bulletin of Mathematical Biophysics, 7*, 83–88. https://doi.org/10.1007/BF02478456

Lang, T., Di Pizio, A., Risso, D., Drayna, D., & Behrens, M. (2023). Activation profile of TAS2R2, the 26th human bitter taste receptor. *Molecular Nutrition & Food Research, 67*(11), Article e2200775. https://doi.org/10.1002/mnfr.202200775

Leibniz, G. W. v. (1896). *On solidity. New essays concerning human understanding* (Originally published, 1704). Macmillan.

Lewkowicz, D. J., & Turkewitz, G. (1980). Cross-modal equivalence in early infancy: Auditory-visual intensity matching. *Developmental Psychology, 16*(6), 597–607. https://doi.org/10.1037/0012-1649.16.6.597

Locke, J. (1690). *An essay concerning human understanding, book III*. Paris, France: Basset.

Loeb, M., Behar, I., & Warm, J. S. (1966). Cross-modal correlations of the perceived durations of auditory and visual stimuli. *Psychonomic Science, 6*(2), 87–88. https://doi.org/10.3758/BF03327970

Lorusso, L., & Porro, A. (2010). Coloured-hearing synaesthesia in nineteenth-century Italy. In F. Clifford Rose (Ed.), *Neurology of music* (pp. 239–256). London, UK: Imperial College Press.

MacDougal, R. (1898). Music imagery; A confession of experience. *Psychological Review, 5*(5), 463–476. https://doi.org/10.1037/h0075017

Macpherson, F. (Ed.). (2011). *The senses: Classic and contemporary philosophical perspectives*. Oxford, UK: Oxford University Press.

Maille, V., & Fleck, N. (2011). Perceived congruence and incongruence: Toward a clarification of the concept, its formation and measure. *Recherche et Applications En Marketing (English Edition), 26*(2), 77–113. https://doi.org/10.1177/205157071102600204

Markman, A. B., & Gentner, D. (2000). Structure mapping in the comparison process. *The American Journal of Psychology, 113*(4), 501–538. https://doi.org/10.2307/1423470

Marks, L. E. (1974). On associations of light and sound: The mediation of brightness, pitch, and loudness. *American Journal of Psychology, 87*(1–2), 173–188. PMID: 4451203.

Marks, L. (1978). *The unity of the senses: Interrelations among the modalities*. New York, NY: Academic Press.

Marks, L. E. (1983). Similarities and differences among the senses. *International Journal of Neuroscience, 19*(1–4), 1–12. https://doi.org/10.3109/00207458309148640

Marks, L. E. (1985). Multidimensional scaling in the assessment of cross-modal similarity. In E. E. Roskam (Ed.), *Measurement and personality assessment* (pp. 165–177). Amsterdam, NL: North-Holland.

Marks, L. E. (1987a). On cross-modal similarity: Auditory-visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance, 13*(3), 384–394. https://doi.org/10.1037/0096-1523.13.3.384

Marks, L. E. (1987b). On cross-modal similarity: Perceiving temporal patterns by hearing, touch, and vision. *Perception & Psychophysics, 42*, 250–256. https://doi.org/10.3758/BF03203076

Marks, L. E. (1989a). On cross-modal similarity: The perceptual structure of pitch, loudness, and brightness. *Journal of Experimental Psychology: Human Perception and Performance, 15*(3), 586–602. https://doi.org/10.1037//0096-1523.15.3.586

Marks, L. E. (1989b). For hedgehogs and foxes: Individual differences in the perception of cross-modal similarity. In G. Ljunggren, & S. Dornic (Eds.), *Psychophysics in action* (pp. 55–65). Berlin: Springer Verlag.

Marks, L. E. (1991). Metaphor and the unity of the senses. In H. T. Lawless, & B. P. Klein (Eds.), *Sensory science theory and applications in foods* (pp. 185–205). New York, NY: Marcel Dekker.

Marks, L. E. (1995). Intermodal similarity and cross-modality matching: Coding perceptual dimensions. In R. D. Luce, D. D. Hoffman, M. D'Zmura, G. Iverson, & A. K. Romney (Eds.), *Geometric representations of perceptual phenomena: Papers in honor of Tarow Indow on his 70th birthday* (pp. 207–233). Hove, UK: Lawrence Erlbaum.

Marks, L. E. (2011). Synesthesia, then and now. In V. Rosenthal (Ed.), *Synesthesia and Intermodality. Intellectica*, **55(1)**, 47-80. DOI: 10.3406/intel.2011.1161.

Marks, L. E., & Bornstein, M. H. (1987). Sensory similarities: Classes characteristics, and cognitive consequences. In R. E. Haskell (Ed.), *Symbolic structures: The psychology of metaphoric transformation* (pp. 49–65). Norwood, NJ: Ablex.

Marks, L. E., Hammeal, R. J., & Bornstein, M. H. (1987). Perceiving similarity and comprehending metaphor. *Monographs of the Society for Research in Child Development, 52*(215), 1–102. https://doi.org/10.2307/1166084

Marks, L. E., Stevens, J. C., Bartoshuk, L. M., Gent, J. F., Rifkin, B., & Stone, V. K. (1988). Magnitude-matching: The measurement of taste and smell. *Chemical Senses, 13*(1), 63–87. https://doi.org/10.1093/chemse/13.1.63

Marks, L. E., Szczesiul, R., & Ohlott, P. (1986). On the cross-modal perception of intensity. *Journal of Experimental Psychology: Human Perception and Performance, 12*(4), 517–534. https://doi.org/10.1037/0096-1523.12.4.517

Martino, G., & Marks, L. E. (2001). Synesthesia: Strong and weak. *Current Directions in Psychological Science, 10*(2), 61–65. https://doi.org/10.1111/1467-8721.00116

Matthen, M. (2005). The sources of sensory similarity. *Seeing, doing, and knowing: A philosophical theory of sense perception* (Oxford, 2005; online edn, Oxford Academic, 20 Apr. 2005), DOI: 10.1093/0199268509.003.0006, accessed 9 Nov. 2025.

McBurney, D. H., & Gent, J. F. (1979). On the nature of taste qualities. *Psychological Bulletin, 86*(1), 151–167. https://doi.org/10.1037/0033-2909.86.1.151
Medin, D. L., Goldstone, R. L., & Gentner, D. (1990). Similarity involving attributes and relations: Judgments of similarity and difference are not inverses. *Psychological Science, 1*(1), 64–69. https://doi.org/10.1111/j.1467-9280.1990.tb00069.x
Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100*(2), 254–278. https://doi.org/10.1037/0033-295X.100.2.254
Medin, D., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou, & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). Cambridge, UK: Cambridge University Press.
Melara, R. D. (1989). Similarity relations among synesthetic stimuli and their attributes. *Journal of Experimental Psychology: Human Perception & Performance, 15*(2), 212–231. https://doi.org/10.1037//0096-1523.15.2.212
Melara, R. D., Marks, L. E., & Lesko, K. E. (1992). Optional processes in similarity judgments. *Perception & Psychophysics, 51*(2), 123–133. https://doi.org/10.3758/BF03212237
Mellers, B., & Birnbaum, M. H. (1982). Loci of contextual effects in judgment. *Journal of Experimental Psychology: Human Perception & Performance, 8*(4), 582–601. https://doi.org/10.1037/0096-1523.8.4.582
Meltzoff, A. N., & Borton, R. W. (1979). Intermodal matching by human neonates. *Nature, 282*(5737), 403–404. https://doi.org/10.1038/282403a0
Meyers, R. G. (2014). *Understanding empiricism*. Routledge.
Morgan, G. A., Goodson, F. E., & Jones, T. (1975). Age differences in the associations between felt temperatures and color choices. *American Journal of Psychology, 88*(1), 125–130. https://doi.org/10.2307/1421671
Motoki, K., Marks, L. E., & Velasco, C. (2023). Reflections on cross-modal correspondences: Current understanding and issues for future research. *Multisensory Research, 37*(1), 1–23. https://doi.org/10.1163/22134808-bja10114
Moul, E. R. (1930). An experimental study of visual and auditory 'thickness'. *The American Journal of Psychology, 42*(4), 544–560. https://doi.org/10.2307/1414876
Murari, M., Rodà, A., Canazza, S., De Poli, G., & Da Pos, O. (2015). Is Vivaldi smooth and takete? Nonverbal sensory scales for describing music qualities. *Journal of New Music Research, 44*, 359–372. https://doi.org/10.1177/2041669519852643
Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review, 83*(4), 435–450. https://doi.org/10.2307/2183914
Newton, I. (1704). Opticks. [Available online at https://archive.org/details/opticksortreatis00newt].
Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology, 43*, 25–53. https://doi.org/10.1146/annurev.ps.43.020192.000325
Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*(3), 510–520. https://doi.org/10.1037/0278-7393.14.3.510
Novick, L. R. (1990). Representational transfer in problem solving. *Psychological Science, 1*(2), 128–132. https://doi.org/10.1111/j.1467-9280.1990.tb00081.x
Obrist, M., Comber, R., Subramanian, S., Piqueras-Fiszman, B., Velasco, C., & Spence, C. (2014). *Temporal, affective, and embodied characteristics of taste experiences, 14*, 2853–2862. https://doi.org/10.1145/2556288.2557007521
Occelli, V., Gillmeister, H., Forster, B., Spence, C., & Zampini, M. (2009). Audiotactile frequency matching. In M. E. Altinsoy, U. Jekosch, & S. Brewster (Eds.), *Proceedings of the 4th International Haptic and Auditory Interaction Design Workshop,* **2**, 44-46.
Ortmann, O. (1933). Theories of synesthesia in the light of a case of colored hearing. *Human Biology, 5*(2), 155–211. https://www.jstor.org/stable/i40070648.
Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana, IL: University of Illinois Press.
Palmer, S. E., Schloss, K. B., Xu, Z., & Prado-León, L. R. (2013). Music-color associations are mediated by emotion. *Proceedings of the National Academy of Sciences of the USA, 110*(22), 8836–8841. https://doi.org/10.1073/pnas.1212562110
Parise, C. V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. *Experimental Brain Research, 220*, 319–333. https://doi.org/10.1007/s00221-012-3140-6
Poffenberger, A. T., & Barrows, B. E. (1924). The feeling value of lines. *Journal of Applied Psychology, 8*(2), 187–205. https://doi.org/10.1037/h0073513
Pridmore, R. W. (1992). Music and color: Relations in the psychophysical perspective. *Color Research & Application, 17*(1), 57–61. https://doi.org/10.1002/col.5080170110
Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin, 132*(3), 416–442. https://doi.org/10.1037/0033-2909.132.3.416
Rader, C. M., & Tellegen, A. (1987). An investigation of synesthesia. *Journal of Personality and Social Psychology, 52*(5), 981–987. https://doi.org/10.1037/0022-3514.52.5.981
Ramachandran, V. S., Marcus, Z., & Chunharas, C. (2020). Bouba-Kiki: Cross-domain resonance and the origins of synesthesia, metaphor, and words in the human mind. In K. Sathian & V. S. Ramachandran (Eds.), *Multisensory perception* (pp. 3-40). Elsevier. DOI: 10.1016/B978-0-12-812492-5.00001-2.
Rodriguez-Pereyra, G. (2002). *Resemblance nominalism: A solution to the problem of universals*. Oxford, UK: Oxford University Press.
Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*(4), 573–605. https://doi.org/10.1016/0010-0285(75)90024-9
Ryan, T. A. (1940). Interrelations of sensory systems in perception. *Psychological Bulletin, 37*(9), 659–698. https://doi.org/10.1037/h0060252
Sell, C. S. (2006). On the unpredictability of odor. *Angewandte Chemie International Edition, 45*, 6254–6261. https://doi.org/10.1002/anie.200600782
Sharma, D., Ng, K. K. W., Birznieks, I., & Vickery, R. M. (2022). Auditory clicks elicit equivalent temporal frequency perception to tactile pulses: A cross-modal psychophysical study. *Frontiers in Neuroscience, 16*, Article 1006185. https://doi.org/10.3389/fnins.2022.1006185
Shepard, R. N. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrika, 39*(4), 373–421. https://doi.org/10.1007/BF02291665
Shepard, R. N. (1981). Psychophysical complementarity. In M. Kubovy, & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 279–341). Hillsdale, NJ: Erlbaum.
Shepard, R. N. (1987). Evolution of a mesh between principles of the mind and regularities of the world. In J. Dupré (Ed.), *The latest on the best: Essays on evolution and optimality* (pp. 251–275). Cambridge, MA: MIT Press.
Shepard, R. N. (1994). Perceptual-cognitive universals as reflections of the world. *Psychonomic Bulletin & Review, 1*(1), 2–28. https://doi.org/10.3758/bf03200759
Shoemaker, S. (1975). Phenomenal similarity. *Crítica: Revista Hispanoamericana de Filosofía, 7*(20), 3–37. http://www.jstor.org/stable/40104007.
Simpson, R. H., Quinn, M., & Ausubel, D. P. (1956). Synesthesia in children: Association of colors with pure tone frequencies. *The Journal of Genetic Psychology, 89*(1), 95–103. https://doi.org/10.1080/00221325.1956.10532990
Smith, L. B. (1993). The concept of same. *Advances in Child Development and Behavior, 24*, 215–252. https://doi.org/10.1016/S0065-2407(08)60304-3
Smith, L. B., & Sera, M. D. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology, 24*(1), 99–142. https://doi.org/10.1016/0010-0285(92)90004-L
Snider, J. G., & Osgood, C. E. (1969). *Semantic differential technique: A sourcebook*. Chicago, IL: Aldine Publishing.
Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics, 73*(4), 971–995. https://doi.org/10.3758/s13414-010-0073-7
Spence, C. (2015). Cross-modal perceptual organization. In J. Wagemans (Ed.), *The Oxford handbook of perceptual organization* (pp. 649–664). Oxford, UK: Oxford University Press.
Spence, C. (2020a). Assessing the role of emotional mediation in explaining crossmodal correspondences involving musical stimuli. *Multisensory Research, 33*(1), 1–29. https://doi.org/10.1163/22134808-20191469
Spence, C. (2020b). Olfactory-colour crossmodal correspondences in art, science, & design. *Cognitive Research: Principles & Implications (CRPI), 5*, 52. https://doi.org/10.1186/s41235-020-00246-1
Spence, C. (2022a). Exploring group differences in the crossmodal correspondences. *Multisensory Research, 35*(6), 495–536. https://doi.org/10.1163/22134808-bja10079
Spence, C. (2022b). Searching for perceptual similarity within, and between, the (chemical) senses. *i-Perception, 13*(5), 1–15. https://doi.org/10.1177/20416695221124154
Spence, C. (2024). Multisensory sweetness enhancement: Comparing olfaction and vision. In L. Stafford (Ed.), *Smell, taste, eat: The role of the chemical senses in eating behaviour* (pp. 17-28). Palgrave Macmillan. DOI: 10.1007/978-3-031-41375-9_2.
Spence, C. (2025). Polina Dimova (2024). At the Crossroads of the Senses: The Synaesthetic Metaphor Across the Arts in European Modernism. The Pennsylvania State University Press. *Perception*, **55(1)**, 90-98. DOI: 10.1177/03010066251371942.
Spence, C., & Deroy, O. (2012). Crossmodal correspondences: Innate or learned? *i-Perception, 3*(5), 316–318. https://doi.org/10.1068/i0526ic
Spence, C., & Di Stefano, N. (2022). Coloured hearing, colour music, colour organs, and the search for perceptually meaningful correspondences between colour and pitch. *i-Perception, 13*(3), 1–42. https://doi.org/10.1177/20416695221092802
Spence, C., & Di Stefano, N. (2024a). Old and new versions of the Molyneux Question: A review of experimental answers. *Philosophy and the Mind Sciences, 5*. https://doi.org/10.33735/phimisci.2024.11337
Spence, C., & Di Stefano, N. (2024b). Sensory translation between audition and vision. *Psychonomic Bulletin & Review, 31*, 599–626. https://doi.org/10.3758/s13423-023-02343-w
Spence, C., & Di Stefano, N. (2024c). What, if anything, can be considered an amodal sensory dimension? *Psychonomic Bulletin & Review, 31*, 1915–1933. https://doi.org/10.3758/s13423-023-02447-3
Spence, C., & Di Stefano, N. (2025a). Augmenting art crossmodally: Possibilities and pitfalls. *Frontiers in Psychology, 16*, Article 1605110. https://doi.org/10.3389/fpsyg.2025.1605110
Spence, C., & Di Stefano, N. (2025b). Gestalt perceptual grouping and crossmodal art. In W. Coppola (Ed.), *Handbook of Gestalt-Theoretic Psychology of Art* (pp. 202-230). Routledge. DOI: 10.4324/9781032694467-11.
Spence, C., & Di Stefano, N. (2025c). Mid-level audiovisual crossmodal correspondences: A narrative review. *Multisensory Research*. https://doi.org/10.1163/22134808-bja10175
Spence, C., & Di Stefano, N. (in press). Basic tastes – A philosophical consideration. Philosophy and the Mind Sciences.
Spence, C., Di Stefano, N., Reinoso-Carvalho, F., Mesz, B., & Zacharakis, A. (2026). 'Play the fragrance': Designing musical soundscapes to match fragrances based on olfactory-auditory crossmodal correspondences. *i-Perception, 17*(1), 1–33. https://doi.org/10.1177/20416695251409265
Spence, C., & Driver, J. (1994). Covert spatial orienting in audition: Exogenous and endogenous mechanisms. *Journal of Experimental Psychology: Human Perception and Performance, 20*(3), 555–574. https://doi.org/10.1037/0096-1523.20.3.555
Spence, C., & Zampini, M. (2006). Auditory contributions to multisensory product perception. *Acta Acustica united with Acustica, 92*, 1009–1025.
Spiro, R. J., Feltovich, P. J., Coulson, R. L., & Anderson, D. A. (1989). Multiple analogies for complex concepts: Antidotes for analogy-induced misconception in advanced knowledge acquisition. In S. Vosniadou, & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 498–531). Cambridge, UK: Cambridge University Press.

Stein, B. E. (Ed.-in-chief), Spence, C. (Section Ed.). (2012). *The new handbook of multisensory processing.* Cambridge, MA: MIT Press.
Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64*(3), 153–181. https://doi.org/10.1037/h0046162
Stevens, S. S. (1958). Some similarities between hearing and seeing. *The Laryngoscope, 68* (3), 508–527. https://doi.org/10.1002/lary.5540680332
Stevens, S. S. (1961). The psychophysics of sensory function. In W. A. Rosenblith (Ed.), *Sensory communication* (pp. 1–33). Cambridge, MA: MIT Press.
Stevens, S. S. (1971). Issues in psychophysical measurement. *Psychological Review, 78*(5), 426–450. https://doi.org/10.1037/h0031324
Stevenson, R. J., & Boakes, R. A. (2004). Sweet and sour smells: Learned synaesthesia between the senses of taste and smell. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processing* (pp. 69–83). Cambridge, MA: MIT Press.
Stevenson, R. J., Rich, A., & Russell, A. (2012). The nature and origin of cross-modal associations to odours. *Perception, 41*(5), 606–619. https://doi.org/10.1068/p7223
Stevenson, R. J., & Tomiczek, C. (2007). Olfactory-induced synesthesias: A review and model. *Psychological Bulletin, 133*(2), 294–309. https://doi.org/10.1037/0033-2909.133.2.294
Tversky, A. (1977). Features of similarity. *Psychological Review, 84*(4), 327–352. https://doi.org/10.1037/0033-295X.84.4.327
Twomey, C. R., Roberts, G., Brainard, D. H., & Plotkin, J. B. (2021). What we talk about when we talk about colors. *Proceedings of the National Academy of Sciences, 118*(39), Article e2109237118. https://doi.org/10.1073/pnas.2109237118
Velasco, C., Adams, C., Petit, O., & Spence, C. (2019). On the localization of tastes and tasty products in 2D space. *Food Quality and Preference, 71*, 438–446. https://doi.org/10.1016/j.foodqual.2018.08.018
von Békésy, G. (1957). Neural volleys and the similarity between some sensations produced by tones and by skin vibrations. *Journal of the Acoustical Society of America, 29*(10), 1059–1069. https://doi.org/10.1121/1.1908698
von Békésy, G. (1959). Similarities between hearing and skin sensations. *Psychological Review, 66*(1), 1–22. https://doi.org/10.1037/h0046967
von Békésy, G. (1964a). Olfactory analogue to directional hearing. *Journal of Applied Physiology, 19*, 369–373. https://doi.org/10.1152/jappl.1964.19.3.369
von Békésy, G. (1964b). Rhythmical variations accompanying gustatory stimulation observed by means of localization phenomena. *Journal of General Physiology, 47*(5), 809–825. https://doi.org/10.1085/jgp.47.5.809
von Hornbostel, E. M. (1950). The unity of the senses. In W. D. Ellis (Ed.), *A source book of Gestalt psychology* (pp. 210-216). London, UK: Routledge and Kegan Paul. A translation of von Hornbostel, E. M. (1927). Die einheit der sinne. The unity of the senses. *Psyche*, **7**, 83-89.
von Békésy, G. (1967). *Sensory inhibition.* Princeton, NJ: Princeton University Press.
Wagner, S., Winner, E., Cicchetti, D., & Gardner, H. (1981). "Metaphorical" mapping in human infants. *Child Development, 52*(2), 728–731. https://doi.org/10.2307/1129200
Walker-Andrews, A. (1994). Taxonomy for intermodal relations. In D. J. Lewkowicz, & R. Lickliter (Eds.), *The development of intersensory perception: Comparative perspectives* (pp. 39–56). Hillsdale, NJ: Lawrence Erlbaum.
Walsh, V. (2003). A theory of magnitude: Common cortical metrices of time, space and quality. *Trends in Cognitive Sciences, 7*(11), 483–488. https://doi.org/10.1016/j.tics.2003.09.002
Wang, Q., Wang, S., & Spence, C. (2016). "Turn up the taste": Assessing the role of taste intensity and emotion in mediating crossmodal correspondences between basic tastes and pitch. *Chemical Senses, 41*(4), 345–356. https://doi.org/10.1093/chemse/bjw007
Wanke, R., Ansani, A., Di Stefano, N., & Spence, C. (2025). Exploring auditory morphodynamics: Audio-visual associations in sound-based music. *i-Perception, 16* (4). https://doi.org/10.1177/20416695251338718
Webb, N. (1976). Impressions in the dark: A multidimensional scaling approach to multimodal experience. *Perceptual and Motor Skills, 42*(3 suppl), 1031–1036. https://doi.org/10.2466/pms.1976.42.3c.1031
Weinberger, A. B., Gallagher, N. M., Colaizzi, G., Liu, N., Parrott, N., Fearon, E., Shaikh, N., & Green, A. E. (2022). Analogical mapping across sensory modalities and evidence for a general analogy factor. *Cognition, 223*, Article 105029. https://doi.org/10.1016/j.cognition.2022.105029
Welch, R. B., DuttonHurt, L. D., & Warren, D. H. (1986). Contributions of audition and vision to temporal rate perception. *Perception & Psychophysics, 39*(4), 294–300. https://doi.org/10.3758/BF03204939
Wells, A. (1980). Music and visual color: A proposed correlation. *Leonardo, 13*(2), 101–107. https://doi.org/10.2307/1577978
Werner, H. (1934). L'Unité des sens [the unity of the senses]. *Journal de Psychologie, Normale et Pathologique, 31*, 190–205.
White, B. W. (1974). What other senses can tell us about cutaneous communication. In F. A. Geldard (Ed.), *Cutaneous communication systems and devices* (pp. 15–19). The Psychonomic Society Inc.
Whiteford, K. L., Schloss, K. B., Helwig, N. E., & Palmer, S. E. (2018). Color, music, and emotion: Bach to the blues. *i-Perception, 9*(6), 1–27. https://doi.org/10.1177/2041669518808535
Wicker, F. W. (1968). Mapping the intersensory regions of perceptual space. *American Journal of Psychology, 81*(2), 178–188. PMID: 5747961.
Winner, E., Rosentiel, A. K., & Gardner, H. (1976). The development of metaphoric understanding. *Developmental Psychology, 12*(4), 289–297. https://doi.org/10.1037/0012-1649.12.4.289
Yearsley, J. M., Pothos, E. M., Barque-Duran, A., Trueblood, J. S., & Hampton, J. A. (2022). Context effects in similarity judgments. *Journal of Experimental Psychology: General, 151*(3), 711–717. https://doi.org/10.1037/xge0001097
Zilczer, J. (1987). "Color music": Synaesthesia and nineteenth-century sources for abstract art. *Artibus et Historiae, 8*, 101–126. https://doi.org/10.2307/1483303