Review of Philosophy and Psychology (2024) 15:997–1026
https://doi.org/10.1007/s13164-023-00692-y

Perceptual Similarity: Insights From Crossmodal Correspondences

Nicola Di Stefano[1] · **Charles Spence**[2]

Accepted: 18 July 2023 / Published online: 29 August 2023
© The Author(s) 2023

Abstract

Perceptual similarity is one of the most fiercely debated topics in the philosophy and psychology of perception. The documented history of the issue spans all the way from Plato – who regarded similarity as a key factor for human perceptual experience and cognition – through to contemporary psychologists – who have tried to determine whether, and if so, how similarity relationships can be established between stimuli both within and across the senses. Recent research on cross-sensory associations, otherwise known as crossmodal correspondences – that is, the existence of observable consensual associations, or mappings, between stimuli across different senses – represents an especially interesting field in which to study perceptual similarity. In fact, most accounts of crossmodal association that have been put forward in the literature to date evoke perceptual similarity as a key explanatory factor mediating the underlying association. At the same time, however, these various accounts raise several important theoretical questions concerning the very nature of similarity, with, for example, the sensory, affective, or cognitive underpinnings of similarity judgements remaining unclear. We attempt to shed light on these questions by examining the various accounts of crossmodal associations that have been put forward in the literature. Our suggestion is that perceptual similarity varies from being phenomenologically-based to conceptually-based. In particular, we propose that the nature of the associations underlying similarity judgements – whether these associations are phenomenologically-, structurally-, emotionally-, or conceptually-based – may be represented in a two-dimensional space with associative strength on one axis, and cognitive penetrability on the other.

✉ Nicola Di Stefano
nicola.distefano@istc.cnr.it

[1] Institute of Cognitive Sciences and Technologies, National Research Council of Italy (CNR), Via San Martino Della Battaglia 44, 00185 Rome, Italy

[2] Crossmodal Research Laboratory, University of Oxford, Oxford, UK

1 Introduction

Perceptual similarity has been recognized as a crucial issue in Western philosophy since the time of the Ancient Greeks (Bryan 2012).[1] For Plato, for example, similarity was a key organizational factor of perceptual experience, one that grounded the ability of humans to categorize objects into different classes, that is, perceptual similarity is considered key to concept formation (e.g., see the *Parmenides*, Allen 1997, and Ryle 1939a, b, for authoritative readings; Williams 2002). In the *Republic*, Plato argued that many particular A-things are perceived as one if they are regarded as instantiating a universal A-ness (*Republic*, 596a, Reeve 2004). This idea assumes some sort of similarity between A-things and A-ness. For instance, Greyhounds and Basset hounds are themselves similar because each of them constitutes an instantiation of 'Dog-ness'.[2] From Plato onwards, the topic of similarity has been debated throughout the history of Western thought, among others by Descartes (Gorham 1999), Leibniz (1923, A64 107/P 13), Goodman (1972), and Quine (1960, 1969, 2000; see also Churchland 1976).[3]

The concept of similarity has also been investigated outside philosophy, often being regarded as a crucial prerequisite for human cognition (e.g., Gentner and Medina 1998; Goldstone and Barsalou 1998; Goldstone and Son 2012; Segundo-Ortin and Hutto 2021). Warning about the risks related to overriding philosophical issues in research on similarity,[4] the psychologist Linda Smith (1993) acknowledged that 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). Much of the appeal that has been attributed to

[1] The focus here is restricted to perceptual/sensory similarity (thus, we will not attempt to address the topic of conceptual similarity). This choice is theoretically justified on the assumption, shared by several scholars (e.g., Goldstone and Barsalou 1998; Marks 1978; Smith 1993), that the very concept of similarity must be perceptual in origin. At the same time, however, we acknowledge the existence of issues specifically raised by conceptual similarity, but these will enter our discussion only as and when they are relevant to addressing the question of perceptual similarity (e.g., when we focus on analogical mappings, see §3.4.). The expressions "perceptual/sensory similarity" and "similarity" are used interchangeably throughout the text to enhance the flow and readability.

[2] Such a theoretical claim has been made brilliantly concrete by the contemporary conceptual artist Joseph Kosuth. In one of his works, *One and three chairs* (1965), Kosuth represents a chair in three ways: as a manufactured chair, as a photograph, and as a copy of a dictionary entry for the word "chair." While the three manifestations of the chair account for the "three chairs" mentioned in the title, the "one" might well be the 'Platonic' chair, i.e., the invisible chair thanks to which each of the three visible objects can be recognized as a "chair" (see Wilde 2007).

[3] The issue of similarity, especially when this is conceived across the senses, can be seen as related also to Wittgenstein's reflections on "seeing/hearing as" (Wittgenstein 1953, II, XI; Wollheim 2003), where the philosopher analyzed the ability of humans to shift perceptual attention in order to attribute to the same percept two different meanings. Bistable figures are classic cases of this phenomenon, but Wittgenstein often considers examples from the musical domain, inviting one to think of a melody as an expression of certain feelings (see Scruton 2004; Di Stefano and Oliva 2020, for an embodied reading of the concept).

[4] "One danger in letting the philosophers' worries set the research agenda, however, is that we may forget the psychology of similarity" (Smith 1993, p. 217).

similarity relies on what the same psychologist suggested as essential to similarity, that is, being hybrid perceptual-conceptual in nature: "I propose that 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" (Smith 1993, p. 223). Stressing the "creativity" of perception, Linda Smith seemingly points to a crucial aspect concerning similarity comparisons, suggesting that the perceptual processes underlying sensory similarity aren't completely constrained (e.g., they might admit of individual differences) and thus are, at least to some extent, open (we will deepen this aspect of the discussion in Section 4).

Many psychologists have stressed the phenomenological evidence that similarity is central to many different domains of human experience, from sensory perception through to abstract cognition (e.g., Goldstone and Barsalou 1998; Medin and Ortony 1989; Medin et al. 1990, 1993; Rosch and Mervis 1975; Smith and Medin 1981). Phenomenological similarity is important for the categorization and identification of objects themselves (e.g., think of shape similarity for vision, see Biederman and Ju 1988; Rosch et al. 1976; Tversky and Hemenway 1984). In his foundational paper, Amos Tversky (1977, p. 327) wrote that: "Similarity plays a fundamental role in theories of knowledge and behavior. It serves as an organizing principle by which individuals explain and classify objects, form concepts, and make generalizations. Indeed, the concept of similarity is ubiquitous in psychological theory. It underlies the accounts of stimulus and response generalization in learning, it is used to explain errors in memory and pattern recognition, and it is central to the analysis of connotative meaning." The following year, at the opening of his book *The unity of the senses*, the influential experimental psychologist Lawrence Marks acknowledged the importance of perceptual similarity for the field of sensory psychology: "The theme of this book is similarity among the senses. While writing the book, I became aware (at first only dimly, but gradually more clearly) of the importance and scope of the very concept of similarity itself" (Marks 1978, pp. x-xi).

In the psychological literature, it is commonly accepted that similarity might exist between pairs of stimuli presented within the same sensory modality (e.g., Blank and Mattes 1990; Ekman 1954; Ekman et al. 1964; Shepard 1962, 1974; Tversky 1977). The majority of the studies in the literature on sensory similarity reference vision (e.g., Logothetis and Sheinberg 1996; though see Spence 2022a, for an isolated exception); this bias is not unexpected, given the well-known primacy of vision in Western culture (Classen 1997; Hutmacher 2019; Jenks 2002; see also Levin 1993). At the same time, however, talking about sensory similarity between pairs of stimuli presented in different sensory modalities would appear to be a much more controversial topic (e.g., Helmholtz 1878/1971; Marks 1978). As Marks (1978) pointedly observed: "There is no way that we can gradually modulate a chord played on a piano until it becomes indistinguishable from the fragrance of a rose, just as there is no way the sight of the pen can become the feel of it" (Marks 1978, p. 188). This quotation seems to complete the previous one, thus conveying the following idea: assuming that two stimuli presented to different senses cannot be conceived in terms of mere identity, some degree of similarity needs to be admitted to account for sensory experience. However, the question of how to explain similarity

in the crossmodal domain has attracted, and continues to attract, a diverse range of theoretical responses (Spence 2022a, for a recent review on perceptual similarity in the chemical senses).

As will become clear, the issue of perceptual similarity raises a number of questions, which have both philosophical and psychological implications, such as: what are the origins and properties of perceptual similarity? Is perceptual similarity 'of a single sort' or are there multiple, intrinsically different kinds of perceptual similarity? If there are different kinds of perceptual similarity, what are they and how do they differ? In this paper, we try to address some of these questions by analyzing the notion of similarity with some insight from psychological works on crossmodal associations, a phenomenon which has seemingly received less attention than others in the literature on similarity (e.g., categorical perception, see Goldstone and Hendrickson 2010). The goal is to ascertain whether similarity can be perceived across the senses or rather is inferred as a result of some form of cognitive mediation and, relatedly, whether the similarity relationship can be understood in terms of the sharing of phenomenological or structural properties. To answer these questions, we will restrict our interests to similarity judgements for pairs of stimuli that happen to be presented in different sensory modalities, that is, in the crossmodal domain.

In a nutshell, we propose that what is generally referred to as "perceptual similarity" deals with a wide variety of processes that vary, depending on the context. In some cases, the similarity relationship is established on the basis of the contents of sensory apprehension; in others, the similarity process originates from perception but is finally accomplished at a cognitive level. To consider the effect of exposure and learning on the perception of similarity, we conceive this distinction in terms of a continuum, thus admitting that some similarities that are based on sensory apprehension might be altered by cognitive information as well as vice versa, that is, some cognitively established similarities might become perceptual following repeated exposure. At the same time, however, we also acknowledge that the continuum is not arbitrarily flexible, and we evoke the concept of (cognitive/perceptual) impenetrability to account for those phenomenal(/conceptual) similarities that would appear resistant to any cognitive(/perceptual) information.

2 Perceptual Similarity Across the Senses

The possibility of perceiving similarities across the senses has long been debated by psychologists/psychophysicists. For instance, in the eighteenth century, the eminent psychophysicist, Herman Ludwig von Helmholtz was skeptical in this regard, stating that: "The distinctions among sensations which belong to different modalities, such as the differences among blue, warm, sweet, and high-pitched, are so fundamental as to exclude any possible transition from one modality to another and any relationship of greater or less similarity. For example, one cannot ask whether sweet is more like red or more like blue. Comparisons are possible only within each modality; we can cross over from blue through violet and carmine to scarlet, for example, and we can

say that yellow is more like orange than like blue!" (Helmholtz 1878/1971, p. 77).[5] Others, like Marks, quoted earlier have expressed apparently opposite ideas: "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).[6]

The two quotes would appear to have different implications for sensory processing. While Helmholtz suggests that different sensory information is mutually exclusive, meaning that the same information cannot be picked out from/processed by two sensory systems, Marks believes that the different senses work in close communication and have rather to be considered as a potentially unified, or unifying, processing system. Despite their opposite tone, however, both quotes converge on identifying the field of crossmodal associations – conceived of as the deliberate and consistent matching between sensory stimuli, attributes, or dimensions from different sensory domains that are observed in normal (i.e., non-synaestheses[7]) people (see Marks 2004; Spence 2011 for a review) – as one of the natural domains in which to investigate the nature of perceptual similarity. Marks clearly sees the phenomenon of crossmodal matching as illuminating (cross-sensory) similarity. Commenting on some of his own findings on pitch-brightness associations, Marks writes: "In a cross-modality matching task, for example, virtually all subjects will set higher sound frequencies to match greater visual intensities (Marks 1974, 1978), thereby revealing a universal appreciation of similarity between the dimension of pitch on the one hand and that of brightness on the other" (Marks 1989, p. 58). The link between crossmodal matching and similarity is incidentally suggested by Goldstone and Barsalou (1998), where the authors stress the importance of people's natural tendency to link distinct sensory domains, namely crossmodal matching, in the case of analogical reasoning (cf. Goldstone and Barsalou 1998, p. 253).

Going back to the earliest documented audiovisual associations (Köhler 1929, 1947), for example, the fact that the terms 'baluba' and 'takete' are associated with curved and angular lines, respectively, has been explained in terms of sound symbolism, which might, in turn, be related to the existence of some sort of sensory similarity between the pseudowords and the curvilinearity of the shapes (e.g., Bremner et al. 2013; Margiotoudi and Pulvermüller 2020; Passi and Arun 2022; Sidhu et al. 2021). Or, moving on to research on crossmodal associations between scent and sound, one might attempt to explain the observed associations between blackberry and piano,

[5] At the same time, however, Helmholtz's quotation does not exclude the possibility that some perceptual properties/stimulus dimensions can be perceived through different senses. For example, spatial properties could be perceived through different modalities (shape and size perceived through vision and touch).

[6] Surprisingly, Marks makes no direct reference to Helmholtz's claim.

[7] For the purposes of this paper, synaesthesia can be defined as the rare neurological condition in which an individual experiences conscious sensations in one sensory modality (referred to as concurrents) when an unrelated stimulus (often presented in a different sensory modality, and known as the inducer) is present. Note that the relationship between inducer and concurrent is idiosyncratic in nature, consistent over time within an individual, and automatically elicited (see Deroy and Spence 2013).

musk and brass, or fruit odours and high-pitched notes hypothesizing that the paired stimuli share some sensory characteristics (Crisinel and Spence 2012; Deroy et al. 2013; see also Piesse 1867 and Di Stefano et al. 2022, Spence 2022a, b). At the same time, however, the exact nature of similarity relationships in these associations remains unclear (see also Belkin et al. 1997; Cohen 1934; Hartshorne 1934).

Marks (1987) suggested that similarities can appear in diverse forms in cross-sensory perception, namely informationally, psychophysically, and phenomenologically. 'Informationally' refers to the fact that different sensory systems can provide information about the same quality, i.e., the shape of a cube through sight or touch (e.g., Gibson 1966). Psychophysically refers to the functional similarities in the 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). For the time being, in line with previous researchers, we will refer to these stimulus parameters in terms of amodal qualities and briefly present them in §3.1 (though for a broader discussion of the concept of amodality see Spence and Di Stefano, submitted). Finally, 'phenomenologically' refers to those similarities that are directly perceived between qualities of perceptual experiences in different sensory modalities (Marks 1978; Spence 2022a).

Attempting to explain the wide range of observed associations between pairs of stimuli presented in different sensory modalities, psychologists have elaborated various hypotheses, each assuming, or conveying, a different view of similarity, especially regarding its relationship with sensory perception. For example, the idea of relative positioning is based on the concept of structural, rather than perceptual, similarity.[8] The structure, in this case, is that of the stimulus dimension rather than the stimulus itself. According to this view, perceivers can establish a connection between stimuli that are represented along sensory dimensions that share structural properties, such as their relative position along a prothetic (as compared to metathetic) sensory dimension (Stevens 1957).[9] However, perceivers can experience crossmodal correspondences between prothetic and metathetic dimensions too (e.g., as in the case of the frequently-studied size-pitch correspondence; Gallace and Spence 2006; Parise and Spence 2009). Moreover, crossmodal mappings have frequently been documented in the absence of any obvious phenomenological sense of perceptual similarity (e.g., Crisinel and Spence 2012), thus forcing researchers to search for hypotheses that are not based exclusively on (any sort of comparison between) sensory inputs. One of the most widely accepted accounts (of crossmodal

[8] This modern idea might be dated back to early observations by Aristotle and to his image of the signet ring to explain how analogy works. When a signet ring is pressed into warm wax, the wax receives the form of the ring (De Anima, Bk. II, Ch. 12, see Hicks 1907). In this way, a *structural* similarity is established between the ring and the wax, despite no material properties being shared between them (Identical twins represent rare, and often striking, occurrences of structural similarities amongst humans too.).

[9] Stevens (1957) defines 'prothetic dimensions' as quantitative perceptual continua that have a clear 'more than' and 'less than' end (such as loudness, brightness, and roughness). Metathetic dimensions, by contrast, tend to obey a well-structured perceptual organization without necessarily having a 'more than' or 'less than' end (e.g., pitch, 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; see also Spence and Di Stefano 2022).

correspondences), namely the emotional mediation hypothesis, holds that it is the similarity between the affective meanings associated with sensory stimuli that links the stimuli presented across sensory modalities (see Spence 2020a, b, for a review; Palmer et al. 2013, on colour-music association; Di Stefano et al. 2022 and Spence 2022a, b, on olfaction and literature/music).[10] In the next section, therefore, the various accounts of crossmodal associations that have been put forward in the psychological literature, and that have implications for the idea of perceptual similarity are briefly examined. As will become clear, several philosophical issues are at stake, such as the perceptual vs. cognitive nature of similarity relationships, the putative role of shared perceptual structures/intersensory qualities that might mediate similarity (e.g., isomorphism/amodal quality), and the role of affective factors (e.g., emotional meanings) in establishing similarity relationships.

3 Similarity in the Major Account of Crossmodal Associations/ Correspondences

According to Mellers and Birnbaum (1982), there are two major accounts of cross-modality matching, namely mapping theory and relation theory. The former holds that psychophysical values of stimuli from different continua can be compared directly because they are mapped onto a common scale of sensation. For instance, a cross-modality match in the intensity domain likely occurs when sensations of equal perceived strength are elicited by stimuli on different continua. According to relation theory, relationships (e.g., ratios) between pairs of stimuli from different continua are compared. For example, while it would be difficult to compare pitch and brightness directly, it seems possible to compare the ratio between pitches with the ratio between brightnesses. While mapping theory assumes the existence of a perceptual, crossmodal link, relation theory holds that matching across senses is rooted in the way in which a stimulus dimension is perceived unimodally (with respect to scaling within each sensory modality), with no direct implications for the nature of cross-modality judgements.

Given that most experimental protocols investigating crossmodal perception have had participants match stimuli across different senses (e.g., first presenting a stimulus in one modality and then asking which of the various stimuli from a different sensory modality matches best), findings obtained in this way might be more straightforwardly interpreted in terms of mapping theories, rather than relation theories. Moreover, the latter seemingly work only for what Stevens (1957) defined as 'prothetic dimensions' (see footnote 7). However, in order to account for specific

[10] Additionally, crossmodal matches have also occasionally been conceived in metaphorical terms, via the concept of 'synaesthetic metaphors', namely, linguistic expressions in which a concept from one sensory domain is described in terms of another sensory domain (e.g., 'dark harmonies'). However, here we will not focus on metaphor as it would require much more space and, most notably, because evidence for cross-sensory similarities in early infancy (e.g., Lewkowicz and Turkewitz 1980) seems to point to a possible non-semantic basis (see Di Stefano et al. 2022 and Spence 2022a, b, on odour-based synaesthetic metaphors).

perceptual matches, such as those between odours and sounds, which involve metathetic dimensions (i.e., pitch), researchers have formulated additional hypotheses, such as analogical mapping or affective mediation (Spence 2020a). Given their allegedly universal nature, crossmodal correspondences have been conceived in terms of the compatibility between attributes or dimensions of a stimulus in different sensory modalities (see Spence 2011, for a review).[11] Evidence supporting this phenomenon comes from both speeded classification tasks (Brunel et al. 2015; Evans and Treisman 2010; Marks 1987) and unspeeded psychophysical tasks (Gallace and Spence 2006; Parise and Spence 2009).

In what follows, we briefly present five major accounts of crossmodal similarity, namely, amodal (or suprasensory), relative positioning, emotional mediation, analogical mapping, and statistical learning (see Table 1). It is worth noting that these accounts of crossmodal correspondences should not be considered as mutually exclusive and they may all have some degree of validity, or relevance, in terms of explaining consensual matches.

3.1 Amodal, or Suprasensory, Qualities

Literally meaning "without" modality (see Bahrick 2009), the term 'amodal' is often taken to mean that the same information can be picked-up regardless of the sensory source, or modality, by which that information was acquired (Walker-Andrews 1994; and see Spence and Di Stefano, submitted, for a critical review). Assuming the existence of amodal qualities, thus, 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. Qualitatively distinct and dissimilar sorts of sensation, say of sight and sound, may thus link up because both share a perceptual aspect, for example, they are equally intense. This view suggests that the various sense modalities 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. Among those properties, duration is one of the most thoroughly investigated, with early studies demonstrating that observers' ratings of duration tend to be highly consistent across the senses (e.g., Loeb et al. 1966; Marks 1987). Moreover, magnitude, size (extension), and brightness have often been considered as amodal properties (see Spence and Di Stefano, submitted). In addition to properties, it might be worth mentioning here psychological principles or laws that are assumed to be valid for different sensory domains, such as the Weber's (1978) law of discriminability and Stevens's (1957) law of sensory intensity.

[11] Compatibility effects are the name given by experimental psychologists to a broad range of behavioural phenomena whereby congruent, or compatible, combinations of stimuli give rise to faster and/or more accurate performance (e.g., see Dematte et al. 2007). Note that there is also an extensive literature on stimulus–response compatibility effects (Kornblum et al. 1990).

Table 1 The major accounts of crossmodal associations and their implications for similarity

Account/key idea	Implications for similarity	Perceptual (P) or cognitive (C)	Property-based (PB) or isomorphism-based (IB)	Relevant Sources
Amodal or suprasensory qualities	Sensations in different modalities share a common feature	P	PB	Von Hornbostel (1931) Lewkowicz and Turkewitz (1980) Giannos et al. (2021) Spence and Di Stefano (submitted)
Relative positioning	The values of perceptual attributes in different modalities are ordered equivalently	P & C	IB	Cohen (1934) Mellers and Birnbaum (1982) Marks et al. (1986) Proctor and Cho (2006) Spence (2019)
Emotional mediation	The stimuli evoke or manifest similar (or identical) emotional features or meanings	P & C	PB	Cowles (1935) Schifferstein and Tanudjaja (2004) Palmer et al. (2013) Levinson and Majid (2014) Levitan et al. (2015) Gilbert et al. (2016) Winter (2016a) Guetta and Loui (2017) Spence (2020a, b)
Analogical mapping	A conceptual criterion (e.g., a geometric projection) allows observers to establish an analogical mapping between two domains	C	IB	Ravignani and Sonnweber (2017) Weinberger et al. (2022)
Statistical learning	Crossmodal associations merely reflect the statistical regularities of the environment perceived by observers exposed to the same environment	P	PB	Walker-Andrews (1994) Ernst (2007) Parise and Spence (2013) Parise et al. (2014)

One of the earliest empirical studies of putatively amodal perceptual dimensions was published by Von Hornbostel (1931). He investigated brightness as 'a universal dimension of sensory experience', conducting a study in which three participants matched different sounds (different pitches) and different scents to greyscale values. The transitivity of the resulting matches led von Hornbostel to infer that 'sensory brightness' is common to all of the senses (i.e., 'allen sinnlichen Erscheinungen gemeinsame Eigenschaft', p. 519, 'property common to all sensory phenomena', or 'eine intermodale Eigenschaft', p. 537, 'an intermodal property'), and hence a universal dimension of sensory experience. Given that neither the resemblance between pitch and brightness nor that between loudness and brightness rests on any property common to the objects that generate the experiences (i.e., sound emitting source and luminous objects), calling a high-pitched or a loud sound bright says nothing intrinsic about a connection between the source objects. In some fundamental sense, the similarities between pitch and brightness and between loudness and brightness reside in shared features of sensory processing. With respect to Marks's (1987) above-mentioned distinction, therefore, the amodal account grounds crossmodal associations on psychophysical similarity.

The findings of a recent audiovisual study seemingly suggest that roughness might be considered as an amodal quality. Giannos et al. (2021) hypothesized that non-tonal and highly dissonant harmonic stimuli would have been associated with rough images, while more consonant stimuli would be associated with the images of low visual roughness. To test this hypothesis, the authors harmonized the same melody in seven styles (which varied in the use of consonances and dissonances) and asked their participants to associate the melodies with images of variable roughness. The latter were black and white 2D and 3D images that represented surfaces with different degrees of smoothness/roughness. Showing that participants tend to consistently associate auditory dissonance with visual roughness, the results provide support to the amodal nature of roughness perception, at least if defined in terms of the pick-up by two or more senses (see also Liew et al. 2017, 2018; Ludwig and Simner 2013; Slobodenyuk et al. 2015; and Di Stefano and Spence 2022 for a review).

3.2 Relative Positioning or Structural Alignment

The existence of amodal dimensions ('intermodale Eigenschaft') has been questioned since Von Hornbostel published his findings, and is still being debated (Spence and Di Stefano, submitted). The psychologist Cohen (1934), for instance, argued that crossmodal mappings were actually better conceptualized as relative/relational judgements (cf. Hartshorne 1934, for an extended discussion on this theme). That is, according to Cohen (1934), Von Hornbostel's experimental stimuli were 'analogous' rather than 'identical'. The psychologist explained as follows: "It would not be unreasonable then to suppose that cross-modality comparison should be based (physiologically, if not introspectively) upon relative positions within different 'absolute' scales. According to this view, equation with respect to

brightness of two experiences of different modalities would involve nothing more than the identity of relative positions upon two wholly independent scales." (Cohen 1934, p. 119). A similar view was put forth by Gombrich (1960), who suggested that crossmodal similarity might be better explained in terms of 'structural relationships' within sensory systems rather than the similarity of specific intersensory elements.[12]

According to this view, which might be labelled as 'relative positioning' or 'structural alignment', what is being compared when someone makes a similarity judgement is the relative position of the two sensory features in their own scales, with no implications regarding their actually being similar across the scales/dimensions (cf. Cohen 1934; Marks et al. 1986; Mellers and Birnbaum 1982; Moul 1930; Simpson et al. 1956, p. 100; Stevens 1957). 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 would not necessarily imply that dark colours are in any intuitive way perceptually similar to low pitched tones. Relatedly, certain mappings between different sensory stimuli have been also explained in terms 'polar dimensions' (e.g., Proctor and Cho 2006). The suggestion here is that stimuli might be coded as having positive and negative polarity along several dimensions, and the similarity is established based on the polarity.

Relative positioning thus holds that the concept of similarity makes little sense when referring to stimuli presented in different senses, while it can more properly describe the 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 et al. 1995).

Referring to the different ways in which stimulus dimensions are organized within each sensory modality, relative positioning or structural alignment requires that the organization of the stimulus dimensions be similar across the senses (Mellers and Birnbaum 1982, p. 600). As noted above, this account intuitively works for prothetic dimensions, which can be clearly ordered along a perceptual continuum. Moreover, this account could probably be applied to metathetic dimensions as well, considering that there might be individual differences in which end of the two scales are lined up (e.g., see Marks 1974, on the alignment of loudness/darkness). For example, any associations between a sound, for instance a low key of the piano, and a colour hue, such as dark blue, can be explained only assuming the similarity of the position of the sound along the frequency scale and of dark blue along the hue scale assuming that the two

[12] "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).

dimensions are commensurable (see Spence 2020a and Spence and Di Stefano 2022, for reviews of colour-sound associations).[13]

Marks (1987) talked in terms of 'superimposable' dimensions, where the attributes at stake stand in a one-to-one relation, with no need to hypothesize a third, amodal or suprasensory quality actually picked-up by the various senses.[14] An alternative approach along similar lines could be to investigate the similarities between the psychophysics of different senses. Several studies by von Békésy (e.g., von Békésy 1957) went in this direction, highlighting several similarities between the processing of vibratory stimuli presented to the skin and to the ear, with resulting similarities between sensations of vibratory touch and of hearing. Such robust psychophysical similarities might pave the way for cross-sensory similarity judgements based on similar processing dynamics or mechanisms, with no implication for the similarity of the stimuli in themselves.[15] However, one might still wonder whether the robust psychophysics (e.g., of transitivity) that is obtained when comparing judgements across various pairs of senses (Ellermeier et al. 2021, on the ratio-based crossmodal matching of visual brightness and sound intensity; cf. Luce et al. 2010), reflects anything more than merely the application of ratio-based mappings within qualitatively distinct unimodal dimensions (see Cohen 1934; Root and Ross 1965; Stevens 1957, 1971; Stevens and Guirao 1963).

3.3 Emotional Mediation

The emotional mediation account has been presented as one of the most powerful and general explanations for a wide range of crossmodal correspondences involving different stimuli (see Spence 2020a, b, for reviews). According to this hypothesis, two stimuli are perceived as similar because they are associated with a similar emotional or, more broadly, affective meaning. Similar to relative positioning, emotional mediation does not imply anything about the actual perceived similarity between the pairs of sensory stimuli that are matched, and thus might well explain the existence of crossmodal mappings that have been documented in the absence of any obvious phenomenological sense of perceptual similarity (e.g., Crisinel and Spence 2012).

The conceptual premises of the emotional mediation account can be traced back to the long history of research investigating the feeling value, or affective tone, of sensory stimuli (Spence 2020a). The work of Charles Osgood and colleagues on the semantic differentiation of implicit affective meanings is worth mentioning here (e.g., Osgood 1952; Osgood et al. 1957). According to this approach, it is possible to determine the meaning of any stimulus (simple or complex, semantic, conceptual, or sensory) by asking people to rate their feelings about it on three

[13] Of course, in the case of colours-sounds one could put forward that matchings might be based on the physics of the stimuli themselves (e.g., wavelengths; see Spence and Di Stefano 2022).

[14] One might be tempted to ask whether similarity based on amodal dimension might vary in strength, depending on the salience of the mediating quality.

[15] Conversely, one could point to how the very same physical stimulus, such as a 200 Hz vibration, affects different senses (von Békésy 1957).

semantic differential scales, namely valence (good/bad), activity (active/passive), and potency (strong/weak).

A clear example of emotional mediation in crossmodal correspondence is represented by Palmer et al.'s (2013) study of audiovisual associations, in which participants had to associate musical excerpts to colour patches and rate both stimuli for their emotional valence using eight affective descriptors (happy, sad, angry, calm, strong, weak, lively, and dreary). The results demonstrated that participants consistently associated music excerpts to colours and that the matched stimuli have similar emotional characters (evidence of some role of 'mood' in mediating music-painting associations is provided by the early study of Cowles 1935; see Spence 2020a, b, for reviews).

Considering all of the evidence reviewed by Spence (2020a), it would seem reasonable to conclude that emotion offers a crucial explanatory concept underpinning audiovisual associations (at least when the sensory cues are not paired semantically, e.g., as when we pair the barking sound with the picture of the dog; Chen and Spence 2010).[16] In addition, further evidence suggests that emotion is also relevant for those crossmodal associations that involve olfaction (e.g., see Levitan et al. 2015, for odour-music associations; Schifferstein and Tanudjaja 2004; Spence 2020b, and Gilbert et al. 2016, for colour-taste/smell associations). The findings reported by Winter (2016b) might also indirectly support such a central role for emotion, explaining why it is that taste (gustation), in particular, is a common source domain in most of the crossmodal correspondences, as taste and smell are thought to be more strongly emotionally-valenced than the other senses (see Levinson and Majid 2014; Winter 2016a). Additionally, Guetta and Loui's (2017) findings revealed that crossmodal associations between complex auditory and gustatory stimuli might be mediated by emotional valence. Given all the above, however, one major caveat to the emotional mediation account 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.[17] 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 et al. 2018), with other studies using complex constructs such as peacefulness, transcendence, and tenderness (see Janowski and Chełkowska-Zacharewicz 2019; Juslin 2013; Vuoskoski and Eerola 2011).

3.4 Analogical Mapping

Several psychologists have suggested a cognitive view of similarity in crossmodal matching. Often proposed with the term 'analogous mappings', this idea can be

[16] According to Young (1978), this conclusion could be generalized to comparisons in themselves: "even the simplest act of comparison involves emotional factors" (p. 194).

[17] In their review on aesthetic emotions, Menninghaus et al. (2019) included as prototypical terms for identifying emotions words such as joy, amusement, nostalgia, surprise, being moved, being shattered, fascination, and boredom.

traced back to the medieval notion of "analogy of imitation" (or participation), which was introduced to explain the alleged relation of likeness between an immaterial, and thus unperceivable, entity, namely God, and sensible creatures, namely humans. The similarity between God and human creatures could be grounded only in some sort of conceptual similarity relation (e.g., mediated by an abstract concept, such as goodness or justice, see Ashworth 2008). A secular example of analogous mapping is the Mercator projection, a technique for making a two-dimensional map of a spherical surface. Most world maps that we are used to are the result of a Mercator projection, thus establishing a one-to-one link between each point on the earth and a related point on the projection. A Mercator projection is inherently perspectival, and it necessarily distorts the pattern of lines on the real surface of the globe. Once you choose the projection criterion, no other choices need to be made, everything follows from the criterion. Another famous example of similarity based on analogical mapping is the Rutherford's model of the atom. In this model, the atom is seen as analogous to a solar system, with the nucleus of the atom being associated with the sun and the electrons with the planets on the basis of the shared relation of attraction which causes the revolution of the later around the former.[18]

Analogical mappings can be established based on different factors which vary in the extent to which they have a perceptual rather than just 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, and there would be no problem in notating the same sound duration differently, based on a different criterion.[19] However, other mappings maintain a perceptual basis, though feeble. The shadows series by the contemporary artists Tim Noble and Sue Webster represents an interesting example: These artists pile up waste and old objects and, through the skilful use of light, project shadow figures that represented real things – such as human beings, animals, or everyday objects – on the wall. Observers find it difficult to visually perceive in the raw materials the image that will be projected on the wall. By carefully choosing the angle of the light, the artists establish an analogical mapping between the pile of waste objects and the shadows resembling, for example, a human figure.

The findings of a recent study investigating analogical mapping across sensory modalities supports the existence of a general analogy factor (Weinberger et al. 2022).[20] 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,

[18] It is not by chance that we find relevant examples of analogies in the history of science, as analogical reasoning has been identified as one of the leading strategies of scientists (Dunbar 1995, 2000) and, more extremely, of all intellectual acts (Spearman 1923).

[19] In fact, the notation changed periodically throughout the history of music (e.g., see Treitler 1982).

[20] Neurophysiological findings revealed that analogical reasoning activates specific brain regions, namely frontopolar cortex (e.g., Green et al. 2010). One might thus speculate that relational crossmodal

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). Participants were then asked to judge whether A:B::C:D is true or false. In many cases, the stimuli not only present an explicit structural relation, but involve 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 participants performed well-above-chance in the identification of crossmodal second-order relations, providing robust evidence of analogy across modalities.

The existence of a general analogy factor might be related to the way in which evolutionary forces have shaped human brains to detect sensory regularities in environmental stimuli. Evidence supporting this idea could be provided by studies on non-human primates. For instance, Ravignani and Sonnweber (2017) tested cross-modal isomorphisms – the ability to detect analogies between structural features across domains – in two chimpanzees (*Pan troglodytes*). After an initial unimodal training during which individuals learned to choose structurally 'symmetric' image sequences (two identical geometrical shapes separated by a different shape) presented beside 'edge' sequences (two identical shapes preceded or followed by a different one), individuals were involved in a multimodal task. In the experimental protocol, the presentation of visual stimuli was preceded by the playback of three sounds that could mimic the structure of either symmetric or edge visual stimuli. Individuals more readily chose symmetric sequences when presented with symmetric, rather than edge, sound triplets, thus leading researchers to conclude that the chimpanzees spontaneously detected a crossmodal, namely visual-auditory, isomorphism. Together with the findings that have been reported in infants (Lewkowicz and Turkewitz 1980), this study further indicates that basic cross-modal abstraction capacities transcend linguistic abilities and may involve evolutionary, low-level sensory mechanisms.

3.5 Statistical Associations

The statistical account of crossmodal association is briefly presented, although, as it will become clear, it has few, if any, meaningful implications regarding similarity. The approach moves from the factual evidence that many physical properties of the sensory stimuli are often correlated in nature. Therefore, the sensory system acquires, together with the stimuli, information on their regular binding, thus learning the statistical regularities of the environment and the correlations between multiple sensory cues (Ernst 2007; Glicksohn and Cohen 2013; Parise and Spence 2013). Reflecting the properties of the environment (and in some cases also the laws of physics), most

Footnote 20 (continued)
similarity judgements ought to evoke a similar brain response were the same analogical mapping process to be involved.

statistical correspondences are likely to be universally perceived (at least by those subjects exposed to the same environment). However, and intriguingly, perceivers can also successfully learn novel associations that are not statistically frequent in the environment (e.g., Brunel et al. 2015). As far as similarity is concerned, neither the universally perceived nor the experimentally-induced statistical associations imply that the matched stimuli are, in any perceptually salient way, similar.

4 Accounting for Similarity Across the Senses

We now examine the implications of the reviewed accounts for the concept of similarity across the senses. The diverse accounts are grouped according to Alistair's (2013) general distinction between property-based and isomorphism-based views of similarity. We then progress on to explore some issues related to the hybrid perceptual, affective, and cognitive nature of similarity. Finally, we highlight the key different factors that allegedly contribute to providing a multidimensional definition of (cross-sensory) similarity.

Alistair (2013) distinguished between two possible strategies for conceiving of similarity: property-based and isomorphism-based. The property-based strategy treats two objects as similar when they share phenomenal qualities. The isomorphism-based strategy treats two objects as similar if a mapping between their component parts exists. Apples, blood, and traffic lights might be similar as they share "redness". A Mercator projection of the world can be similar to the world as they can be related by a one-to-one mapping. Based on Alistair's distinction, accounts of crossmodal correspondences in terms of amodal dimensions and emotional mediation are property-based, while structural alignment and analogical mapping are isomorphism-based (see Table 1). In what follows, we go deeper into property-based and isomorphism-based accounts of similarity, evidencing their implications for the perceptual, affective, or cognitive nature of similarity judgements.

In the unisensory domain, the idea of property-based similarity appears to be intuitively and straightforwardly applied. In this context, two perceptual experiences are similar when they present some shared phenomenological aspect (see also Leibniz 1923, A64 107/P 13). For instance, when I experience the blue sky and the ocean blue, the common experience of blueness guarantees that they are similar in some analogous respect (i.e., colour). No mapping is needed to perceive the similarity between the sky and ocean, just the apprehension of a common intramodal property.[21] However, famous critics such as Goodman (1972) have pointed to the fact that any two objects share at least one phenomenal (intramodal or intermodal) quality and thus such view of similarity would simply be a universal relation – 'everything

[21] Interestingly, similar ideas of similarity hold central for the definition of concepts in several theories in cognitive psychology. For example, Rosch and Mervis's (1975) prototype theory considers that an object belongs to a concept based on how similar it is to its prototype. The exemplar theory (see Smith and Medin 1981) rests upon similarity as well, with the difference that one takes into account similarity relations to several exemplars instead of one prototype.

is similar to everything else' – and therefore similarity claims are uninformative (see also Rodriguez-Pereyra 2002).[22]

Conceived of in terms of property sharing, similarity also raises issues in the cross-sensory context. For example, it has been observed that a funeral march (i.e., an auditory stimulus) is associated with a weeping willow (a visual stimulus) as they share a property, that is, they are both perceived as sad (e.g., see Davies 2017). Or, analogously, a happy song might be associated with a sunflower. Similar associations are often explained in terms of emotional expressiveness, noting that both the funeral march (/happy song) and the weeping willow (/sunflower) express the same emotional quality, namely sadness (/happiness), which therefore grounds the emotionally mediated similarity between funeral marches (/happy songs) and weeping willows (/sunflowers).

Two observations should be made here regarding the implications of emotionally mediated correspondences for the concept of property-based similarity. First, the fact that people tend to associate X more to A than to B, for instance, the song "happy birthday" to sunflowers more than to weeping willows, does not necessarily imply that X is phenomenologically *similar* to A. As has been observed elsewhere (see Spence and Levitan 2021; Spence et al. 2015), merely demonstrating a statistically significant crossmodal correspondence between stimuli only implies that it is the best of the options that happened to be available to participants at the time they were asked, and does not imply that the stimuli are perceptually similar in any respect (rather the association may simply reflect learned statistical correlations in the environment, see Section 3.5). If people were presented with the word 'apple' and with the images of an apple and of an apple tree and were asked to match the word with one of the two images, they will likely match the word to the image of the apple though, clearly, the word 'apple' is not in an intuitive way more similar to the apple than to the tree.[23]

Second, emotional meanings can only metaphorically be considered as 'perceived' properties. In fact, according to the emotional mediation account, affective meanings are rather *attributed* or *associated* with perceived stimuli, but are not directly perceived in themselves (as a property-based definition would require). This is in line with the cognitivist's view of musical emotions (e.g., Kivy 1991; Radford 1989) claiming that listeners *recognize* emotional qualities of music owing to cognitive processes. These considerations stress the fact that emotions are either felt or cognitively superimposed on perceptual stimuli, leaving open the question as to whether the sadness of the weeping willow is actually *perceived* (e.g., Davies 2017).

Taken together, the above considerations would appear to suggest that property-based similarity can hardly be invoked to account for the broad range of emotionally mediated correspondences, which are, as said, one the most widely documented

[22] Alternative views of similarity in unisensory contexts have been presented, such as geometrical or transformational models (Goldstone and Son 2012).

[23] However, as noted earlier, Marks (1989) interpreted most of his research on crossmodal matching as illuminating the problem of similarity across the senses.

correspondences (especially audiovisual). We are thus led to examine whether the second view of similarity presented by Alistair (2013), namely isomorphism-based, can more broadly account for the existence of crossmodal correspondences.

According to the isomorphism-based definition, similarity relationships are established by the perceiver based on an implicit or explicit cognitive criterion. Merely recognizing that two objects are perceptually similar has no meaning unless one specifies *how* they are the same (Goodman 1972; Smith 1993). An apple is more similar to a pear than to a carrot. But it is likely more similar to a carrot than to a train. Thus, we could generalize by saying that perceptual similarity always varies with the attributes that are attended to (Nosofsky 1984; Shepard 1964, 1987). For example, cats and lions are similar as they share features such as their shape, retractile claws, locomotive behaviour. But humans might be similar to lions as well, when they show aggressive behaviours as lions do, or to sloths, when they are lazy and sleepy. In the first case, the association is based on the similarity of properties that are actually perceived (property-based). In the latter, it seems that we analogically *infer* similarity between humans and lions or sloths, based on non-perceptual characteristics.

The isomorphism-based similarity account denies that phenomenological similarity can provide a solid basis for establishing similarity relationships across the senses. Even accepting that perceptual or phenomenological similarity exists, isomorphism-based view stresses the fact that similarity relationships based on property sharing are either universal or indeterminate and, in either case, there cannot be a solid link mediating the correspondence or association between stimuli across the senses. Such an objection clearly goes against the explanations of crossmodal correspondences in terms of amodal dimensions (even if such dimensions were to exist, they would not be capable of accounting for all demonstrated occurrences) and emotional mediation, for the reasons highlighted above. One is thus led to assume that cognition must play a role in crossmodal similarity, which naturally raises the question about what triggers the similarity process, when this is conceived of as cognitively mediated.

Armary et al. (2018) suggested that based on the specific context different forms of salience emerge as a possible heuristic in driving the similarity process: sensory, categorical, and operational salience. Grounding on biological, developmental, and evolutionary explanations, sensory salience assumes that the processing and attentional mechanisms of the human perceptual system generate a hierarchy amongst sensory properties, with some appearing more salient than others. For instance, in humans, features such as colour, shape, orientation, symmetry, will likely be more salient than odours. By contrast, for animals with exceptional olfactory abilities, such as certain breeds of dog, odours will be prominent. Categorical salience reflects the hierarchy in the cognitive system and knowledge of the perceiver. Given a concept, several others will follow as more saliently related to the former. For example, despite the different possible uses of a chair, the concept ‘sitting’ will be prominent, while ‘climbing’ less prominent. Finally, operational salience refers to the ongoing salience attributed to one property during the analogical process. It is the most flexible form of salience, influenced by sensory inputs, categorical knowledge, and goals.

Such a multidimensional concept of salience could allow us to explain why certain similarities are perceived more intuitively than others and why it is that we ascribe different weights to different properties of things. Even though ladybirds and tomatoes share the salient perceptual characteristic of red colour, but ladybirds and flies do not, flies may be judged as more similar to ladybirds (both being flying insects) than tomatoes are to ladybirds. Presumably, in this example, the property of being a flying insect overrides having the same colour. Referring to salience can also explain hierarchies in the way crossmodal similarities are established by the perceiver, with the sensory ones being used first, if available, while the most cognitive ones are triggered only when perceptual similarities are unavailable.[24]

Returning to the previous examples, when colours and sounds are presented to a perceiver, s/he might search for a perceptually meaningful mapping criterion (e.g., dark colours to low pitches). By contrast, when one cannot establish an immediate sensory connection, such as between sounds and the notes on the score, or between a pile of rubbish and a female figure, the perceiver might search for cognitively inferred mediating factors instead, such as analogies. Another way to put the same idea is by referring to the role of stimulus/subject in perceptual similarity. In some cases, similarity is based on the phenomenal qualities of the stimulus, and thus the similarity process is exogenous, or stimulus-driven (e.g., dark colours and low pitches); in other cases, the phenomenal qualities of the stimuli do not allow to establish any perceptually meaningful similarity (e.g., sounds and notes on the score), and similarity is rather established by the observer thank to an explicit cognitive or abstract mediation, and is therefore 'endogenous', with respect to the participant.

Interestingly, hierarchies might also take place within conceptually-mediated similarities. The similarity process can, in fact, act as a sorting function that helps to determine the importance of properties in a given context, determining, in turn, which analogy best suits that context. For example, in the analogy mentioned between the atom and the solar system, the spherical shape of the planets and electrons is less relevant than is the causal relation of attraction between the centre and the peripheral elements. Similarly, in the analogy between a pile of rubbish and a human figure, shape similarity overrides other evident differences (e.g., matter, bidimensionality; and see Spiro et al. 1989, on the use of, or need for, multiple analogies when trying to explain/understand complex systems).

Taken together, the above observations suggest a first tentative answer to one of our initial questions, namely, whether similarity comparisons involving sensory qualities from different senses result from perception or are rather cognitively inferred. We suggest that similarity can be perceived across the senses and therefore, at least in certain occurrences is, in our terminology, stimulus-driven. This conclusion stresses that similarity is a quintessential component of human perception

[24] In this respect, it seems possible to trace a distinction between crossmodal associations and synaesthesia, with the former being essentially prompted by contextual factors (e.g., the available matchings), while the latter is independent of the perceptual context (i.e., the inducer always triggers the perception of the concurrent).

(see also Quine 1960, 2000; Churchland 1976, for criticisms of Quine's position). More specifically, we suggest that similarity judgements involving stimuli across the senses have their origins in perception, and perception places important constraints on them (see also Goldstone and Barsalou 1998, for perceptual similarity in general). For example, a tone of 400 Hz and a tone of 402 Hz sound similar, and nothing can be done to alter this sensory similarity. We therefore acknowledge that phenomenological similarity simply exists, and it refers to the phenomenal resemblance between the objects in the source and target and their properties (Keane et al. 1994). However, as shown, a stimulus-driven account of similarity could not explain several occurrences of similarities across the senses, thus suggesting that more cognitive factors might mediate similarity judgements when perceptually driven paths are unavailable or insufficient (e.g., in the case of analogical mappings). Therefore, the previous conclusion has to be completed by stating that similarity processes may, depending on the context, be perceptual or conceptual in nature. When perceptual, the similarity process follows an exogenous, or stimulus-driven, path, starting and ending with sensory apprehension; when conceptual, the similarity process follows an endogenous, or subject-driven, path, starting with perception but being finally accomplished at a cognitive level (see Fig. 1).[25]

Several important observations should now be made in order to achieve the objective of this paper and, thus, to provide a more fine-grained account of perceptual-cognitive dynamics in the case of cross-sensory similarity judgements. First, if it is true that nothing can be done to alter phenomenal similarity, it is also true that this statement is valid under certain circumstances. For instance, it is likely that individual differences play a role in the ability to differentiate between specific classes of stimuli, such as pitches or colours, impacting the ability to perceive similarities among those stimuli. For example, listeners with absolute pitch might perceive two tones at 400 Hz and 402 Hz as sounding different. This would lead us to consider sensory discrimination thresholds, a topic that falls beyond the scope of this paper. However, we would like to point out that discrimination thresholds are likely much more important for similarity comparisons in the same sensory modality than across the senses. Beyond individual differences, universal psychological mechanisms can influence perception, such as the functional unitization/differentiation of perceptual components due to categorization experience (Goldstone 2000). For example, in an early study by Katz (1963), two groups of children were required to associate four similar geometrical shapes with nonsense syllables. In the first group, the association was such that two shapes had the same name while, in the second group, each syllable was associated to a different shape. The results of a subsequent similarity rating task revealed that children from the first group judged the shapes that were named identically as being the same more often than children from the second group. Similar findings were obtained on adults by Livingston et al. (1998), who found that,

[25] Noteworthy here, analogical reasoning might reflect very general properties of cognition, which also extend to non-human primates. For example, using forced choice or same/different protocols, Gillan and colleagues (1981) demonstrated that a chimpanzee was able to solve simple tasks of analogical reasoning. These results demonstrate that basic forms of analogical reasoning do not depend on language.

when participants learned to classify objects (animal body parts or artificial cells) into a single category, these objects were rated as more similar than objects that belonged to different categories, or than objects that were not categorized at all (see also Kurtz 1996).

Second, a caveat should be mentioned here against the radicalization of the dichotomy between perceived vs. cognitively inferred similarities. In fact, in several cases, similarities that were once effortful and inferred based on conceptual means might become perceptual or, at least, less demanding cognitively. Especially when dealing with analogical mapping, perceivers might struggle to see the link between the paired entities at first, while gradually becoming so familiar with it that they do not need to recall the cognitive mediation anymore. Roughly-speaking, this is the process of *perceiving* what was once a *conceptual* similarity. For instance, the student musician explicitly uses (cognitive) rules to decipher the link between sounds and sings on the score. Then, with time, reading music ceases to be effortful and rule based, and becomes perceptual and phenomenologically direct. Importantly, when this occurs, the 'acquired' similarity can be used to trigger new similarities (as often occurs in the case of scientific discoveries, e.g., Gentner 2002).[26] Moreover, as observed by Marks (1987), crossmodal associations might be due to some kind of communality between parallel stages in processing attributes of different sensory modalities; such communality may not be explained in terms of phenomenological properties of the stimulus, but rather in terms of stimulus processing after sensory apprehension.

Third, and relatedly, it should also be noted that cognitive/conceptual information does not necessarily make perceptual stimuli seem any more similar (nor dissimilar). For example, studies in the unisensory context show that certain olfactory molecules called enantiomers, namely odorants which are identical except for their chirality, sometimes smell very different (e.g., Laska and Teubner 1999). Even though the perceiver might be cognitively aware of their similarity, such awareness does not alter the way in which the molecules are actually perceived. In the crossmodal context, the same vibratory stimulus at a frequency beyond 200 Hz is not perceived similarly when experienced as an auditory versus tactile vibration.[27]

To recapitulate, phenomenologically and cognitively mediated similarities are wrongly conceived as mutually exclusive, and rather need to be conceived in terms of continuity and, at least to a certain extent, flexible categories. This implies that phenomenological similarity might be altered or influenced by cognitive information, and that cognitive similarity might become perceptual after exposure or training. At the same time, however, phenomenological similarity would occasionally appear resistant to any cognitive *penetration*, and conversely,

[26] Something similar can occur when an association between X and Y is mediated by an element A which is no more consciously related to X or Y by the perceiver. For example, the association between scarlet and trumpet might be based on a forgotten associative link, namely army soldiers, whose uniform evokes scarlet and whose parades are typically accompanied by trumpet (cf. Harrison 2001, p. 209).

[27] Observers may perceive the similarity at lower frequencies (e.g., Altinsoy and Merchel 2010; Sharma et al. 2022), but discriminatory performance decreases for frequencies beyond 150-180 Hz, thus suggesting that similarity perception might be orthogonal to discriminatory ability.

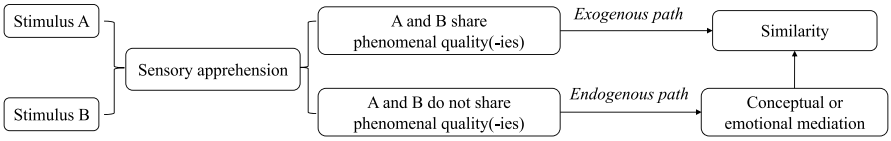

Fig. 1 Two paths leading to the establishment/recognition of crossmodal similarity. When two stimuli in different sensory domains (A and B) are perceived, similarity is directly perceived if they share phenomenal qualities. If not, conceptual, or affective links might trigger similarity comparisons

cognitive information is *impenetrable* by perceptual information. We therefore now want to account for the difference between phenomenological and cognitive similarity using the concept of (cognitive/perceptual) impenetrability.

Perceptual contents might be considered to be cognitively impenetrable if they cannot be influenced or affected by the contents of higher cognitive states or factors such as beliefs, inferences, habits, or knowledge (Pylyshyn 1984; Siegel 2012; Stokes 2013; see also Macpherson 2012); vice versa, perceptual contents are cognitively penetrable when cognitive information or knowledge alters how they are perceived. An illuminating example of the cognitive penetrability of perception comes from studies testing the use of inverted lenses. These lenses turn one's visible world upside down, impacting the way we visually perceive and interact with the world. Typically, the lenses are initially experienced as radically disorienting for the wearer but, after a relatively short learning period of consistently using and acting with the lenses, participants quickly recover to perform perceptual and motor tasks normally (Harris 1965; Kottenhoff 1957; Stratton 1897). This example demonstrates that participants could readily adjust their perception according to their knowledge concerning how the real, non-inverted world is (see also the use of pseudophone to induce sensory incongruity in audition, Spence 2022b).

These observations allow us to refine our cognitive-perceptual polarized model, suggesting that similarity can be rather represented as a continuum along two main dimensions, namely, associative strength and impenetrability (see Fig. 2; see also Deroy and Spence 2013, for a similar bidimensional model to account for crossmodal correspondences and canonical synaesthesia). Associative strength refers to the extent to which the link created by the similarity relationship between two stimuli is binding (i.e., low associative strength implies that the bond is weak, and hence their similarity does not constrain perception/cognition). Impenetrability refers to the extent to which a similarity relationship can be altered by cognitive factors. In the present model, similarity judgements that involve stimuli from different sensory domains can be based on phenomenological, structural, emotional, or conceptual grounds. Although there might not be clear-cut distinctions amongst the different kinds, according to the model, phenomenological matchings are highly impenetrable and perceptual in nature, while conceptual, or analogy-mediated, matchings are penetrable and cognitive in nature. In between these two extreme cases, structural might be considered as less penetrable than emotional, which is likely more cognitive and penetrable.

Before concluding, it might also be worth pointing out that these reflections could not have been made starting from the analysis of what has been often considered – though incorrectly (Deroy and Spence 2013) – a peculiar case of crossmodal matching, namely, synaesthesia (Ramachandran and Hubbard 2001). In fact, while it was possible to highlight a number of reasons why crossmodal associations might be putatively relevant for similarity, the latter concept seems not to be straightforwardly applicable to synaesthesia. With respect to the specific stimuli that are being linked/associated, in fact, synaesthesia seems to be more related to identity, rather than necessarily similarity since, for a synaesthete, the inducer (e.g., the sound of the trumpet) and the concurrent (colour scarlet) are simply part of one and the same, cognitively impenetrable, perceptual experience (that is, the inducer is always co-experienced with the concurrent).

5 Conclusions

Returning to our initial question about the perceptual vs. cognitive nature of similarity across the senses, a tentative answer can now be provided, by observing that crossmodal similarity is, in its allegedly fundamental manifestation, rooted in perception and can therefore be explained in terms of the sharing of phenomenal qualities. However, in other cases, explanations based on phenomenal qualities fail to account for the observed matchings, which can rather be explained in terms of cognitive or affective mediation, according to which, for example, similarity judgements are based on the sharing of similar emotional meanings. While perceptual-based matchings are cognitively impenetrable, conceptually/affectively mediated ones are cognitively penetrable.

At this point, an underlying question regarding the implications of the literature on crossmodal matchings for the notion of (cross-sensory) similarity seemingly remains open. In fact, as briefly observed earlier, according to some commentators (e.g., Spence et al. 2015; Spence and Levitan 2021), the evidence of statistically significant (or consensual) crossmodal correspondence between stimuli does not say anything about crossmodal similarity, being rather explainable in terms of, say, learned statistical correlations in the environment. At the same time, however, influential psychologists such as Marks (e.g., Marks 1989) are strongly convinced that investigating cross-sensory matchings sheds light on the very nature of similarity. From a psychological perspective, the point seemingly remains controversial, as no crossmodal matching protocol can generate results solely on similarity, excluding the possible contribution of mechanisms other than similarity (e.g., statistical learning) in triggering the (eventually discovered) association. In other words, when

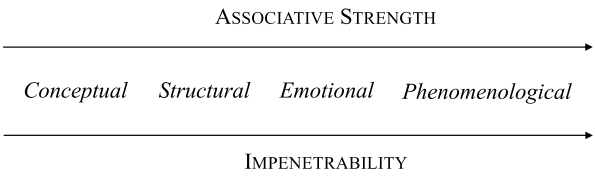

Fig. 2 The different kinds of cross-sensory similarities according to their associative strength and the extent they are impenetrable/penetrable

testing primates or infants, how could we actually know whether it is analogy, emotional mediation or perceptual similarity driving their behavioural responses? Is there some implicit assumption experimenters hold about the kinds of comparisons that can legitimately be based on perceptual similarity? The issue remains controversial from the philosophical side as well, and the controversy might be framed within the never-ending dispute on the nature of perception itself and its objects (universal vs. particular, conceptual vs. non-conceptual, propositional vs. sensory; see Foster 2000).

Several additional questions remain open for future investigation. First, one might be tempted to ask about the implication of our inquiry for similarity in the unisensory domain, for example, by asking whether the emotional mediation account can also work for similarity judgements between colours (e.g., do we judge certain pairs of colours as more similar because they are associated with the same emotion?). Relatedly, it could be worth investigating whether fundamental distinctions between stimuli in one sensory dimension, such as major/minor in sound perception, might have a correlate in a different modality, such as vision or touch. This would open to the investigation of the intriguing possibility of translating between the senses (see Spence and Di Stefano, in press), and the putative role of similarity criteria in mediating the process of sensory translation.

Declarations

Conflict of Interest None.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Funding Open access funding provided by Consiglio Nazionale Delle Ricerche (CNR) within the CRUI-CARE Agreement.

References

Allen, R.E. 1997. *Plato's Parmenides*. New Haven: Yale University Press.

Alistair, M.C.I. 2013. Objective similarity and mental representation. *Australasian Journal of Philosophy* 91 (4): 683–704.

Altinsoy, M.E., and S. Merchel. 2010. Cross-modal frequency matching: Sound and whole-body vibration. In *International workshop on haptic and audio interaction design*, 37–45. Berlin: Springer.

Armary, P., J. Dokic, and E. Sander. 2018. The problem of context for similarity: An insight from analogical cognition. *Philosophies* 3 (4): 39. https://doi.org/10.3390/philosophies3040039.

Ashworth, E.J. 2008. *Les théories de l'analogie du XIIe au XVIe siècle [Theories of analogy from the 12th to the 16th century]*. Paris: Vrin.

Bahrick, L.E. 2009. Amodal perception. In *Encyclopedia of perception*, ed. E.B. Goldstein, 44–46. Thousand Oaks, CA: Sage Publications.
Belkin, K., R. Martin, S.E. Kemp, and A.N. Gilbert. 1997. Auditory pitch as a perceptual analogue to odor quality. *Psychological Science* 8 (4): 340–342.
Biederman, I., and G. Ju. 1988. Surface versus edge-based determinants of visual recognition. *Cognitive Psychology* 20 (1): 38–64.
Blank, D.M., and R.D. Mattes. 1990. Sugar and spice: Similarities and sensory attributes. *Nursing Research* 39 (5): 290–292.
Bremner, A., S. Caparos, J. Davidoff, J. de Fockert, K. Linnell, and C. Spence. 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.
Brunel, L., P.F. Carvalho, and R.L. Goldstone. 2015. It does belong together: Cross-modal correspondences influence cross-modal integration during perceptual learning. *Frontiers in Psychology* 6: 358.
Bryan, J. 2012. *Likeness and likelihood in the Presocratics and Plato*. Cambridge, UK: Cambridge University Press.
Chen, Y.-C., and C. Spence. 2010. When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition* 114 (3): 389–404.
Churchland, P.S. 1976. How Quine perceives perceptual similarity. *Canadian Journal of Philosophy* 6 (2): 251–255.
Classen, C. 1997. Foundations for an anthropology of the senses. *International Social Science Journal* 49 (153): 401–412.
Cohen, N.E. 1934. Equivalence of brightness across modalities. *American Journal of Psychology* 46: 117–119. https://doi.org/10.2307/1416240.
Cowles, J.T. 1935. An experimental study of the pairing of certain auditory and visual stimuli. *Journal of Experimental Psychology* 18 (4): 461–469. https://doi.org/10.1037/h0062202.
Crisinel, A.-S., and C. Spence. 2012. A fruity note: Crossmodal associations between odors and musical notes. *Chemical Senses* 37 (2): 151–158.
Davies, S. 2017. Music matters: Responding to Killin, Ravasio, and Puy. *Debates in Aesthetics* 13 (1): 52–67.
Demattè, M.L., D. Sanabria, and C. Spence. 2007. Olfactory-tactile compatibility effects demonstrated using the Implicit Association Task. *Acta Psychologica* 124: 332–343.
Deroy, O., and C. Spence. 2013. Why we are not all synesthetes (not even weakly so). *Psychonomic Bulletin & Review* 20 (4): 643–664.
Deroy, O., A.-S. Crisinel, and C. Spence. 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.
Di Stefano, N., M. Murari, and C. Spence. 2022. Crossmodal correspondences in art and science: Odours, poetry, and music. In *Olfaction. An interdisciplinary perspective*, eds. N. Di Stefano and M.T. Russo, 155–189. Springer.
Di Stefano, N., and S. Oliva. 2020. Insights into the aesthetic experience through an embodied approach to Wittgenstein's hearing-as. *Reti, Saperi, Linguaggi* 7 (2): 277–292.
Di Stefano, N., and C. Spence. 2022. Roughness perception: A multisensory/crossmodal perspective. *Attention, Perception, & Psychophysics* 84: 2087–2114.
Dunbar, K. 1995. How scientists really reason: Scientific reasoning in real-world laboratories. In *The nature of insight*, eds. R.J. Sternberg and J.E. Davidson, 365–395. Cambridge, MA: MIT Press.
Dunbar, K. 2000. How scientists think in the real world: Implications for science education. *Journal of Applied Developmental Psychology* 21 (1): 49–58.
Ekman, G. 1954. Dimensions of color vision. *Journal of Psychology* 38: 467–474.
Ekman, G., T. Engen, T. Kunnapas, and R. Lindman. 1964. A quantitative principle of qualitative similarity. *Journal of Experimental Psychology* 68: 530–536.
Ellermeier, W., F. Kattner, and A. Raum. 2021. Cross-modal commutativity of magnitude productions of loudness and brightness. *Attention, Perception, & Psychophysics* 83 (7): 2955–2967.
Ernst, M.O. 2007. Learning to integrate arbitrary signals from vision and touch. *Journal of Vision* 7 (5): 7–1-14.
Evans, K.K., and A. Treisman. 2010. Natural cross-modal mappings between visual and auditory features. *Journal of Vision* 10 (1): 6.
Forbus, K.D., D. Gentner, and K. Law. 1995. MAC/FAC: A model of similarity-based retrieval. *Cognitive Science* 19 (2): 141–205.

Foster, J. 2000. *The nature of perception*. Oxford: Clarendon Press.
Gallace, A., and C. Spence. 2006. Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics* 68: 1191–1203.
Gentner, D. 2002. Analogy in scientific discovery: The case of Johannes Kepler. In *Model-based reasoning*, 21–39. Boston, MA: Springer.
Gentner, D., and J. Medina. 1998. Similarity and the development of rules. *Cognition* 65 (2–3): 263–297.
Giannos, K., G. Athanasopoulos, and E. Cambouropoulos. 2021. Cross-modal associations between harmonic dissonance and visual roughness. *Music & Science* 4: 20592043211055484.
Gibson, J.J. 1966. *The senses considered as perceptual systems*. London: George Allen and Unwin Ltd.
Gilbert, A.N., A.J. Fridlund, and L.A. Lucchina. 2016. The color of emotion: A metric for implicit color associations. *Food Quality and Preference* 52: 203–210.
Gillan, D.J., D. Premack, and G. Woodruff. 1981. Reasoning in the chimpanzee: I. Analogical reasoning. *Journal of Experimental Psychology: Animal Behavior Processes* 7 (1): 1–17.
Glicksohn, A., and A. Cohen. 2013. The role of cross-modal associations in statistical learning. *Psychonomic Bulletin & Review* 20: 1161–1169.
Goldstone, R.L. 2000. Unitization during category learning. *Journal of Experimental Psychology: Human Perception and Performance* 26 (1): 86–112.
Goldstone, R.L., and L.W. Barsalou. 1998. Reuniting perception and conception. *Cognition* 65 (2–3): 231–262.
Goldstone, R.L., and A.T. Hendrickson. 2010. Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science* 1 (1): 69–78.
Goldstone, R.L., and J.Y. Son. 2012. Similarity. In *The Oxford handbook of thinking and reasoning*, eds. K.J. Holyoak and R.G. Morrison, 155–176. Oxford, UK: Oxford University Press.
Gombrich, E.H. 1960. *Art and illusion*. London: Phaidon Press.
Goodman, N. 1972. Seven strictures on similarity. In *Problems and projects*, ed. N. Goodman, 437–446. Indianapolis: Bobs-Merril.
Gorham, G. 1999. Causation and similarity in Descartes. In *New essays on the Rationalists*, eds. R.J. Gennaro and C. Huenemann, 296–309. Oxford, UK: Oxford University Press.
Green, A.E., D.J. Kraemer, J.A. Fugelsang, J.R. Gray, and K.N. Dunbar. 2010. Connecting long distance: Semantic distance in analogical reasoning modulates frontopolar cortex activity. *Cerebral Cortex* 20 (1): 70–76.
Guetta, R., and P. Loui. 2017. When music is salty: The crossmodal associations between sound and taste. *PLoS One* 12 (3): e0173366.
Harris, C.S. 1965. Perceptual adaptation to inverted, reversed, and displaced vision. *Psychological Review* 72 (6): 419–444.
Harrison, J. 2001. *Synaesthesia: The strangest thing*. Oxford, UK: Oxford University Press.
Hartshorne, C. 1934. *The philosophy and psychology of sensation*. Chicago, IL: University of Chicago Press.
Hicks, R.D. 1907. *Aristotle's De Anima*. Cambridge, UK: Cambridge University Press.
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., and M. Chełkowska-Zacharewicz. 2019. What do we actually measure as music-induced emotions? *Roczniki Psychologiczne* 22 (4): 373–403.
Jenks, C. 2002. The centrality of the eye in western culture: An introduction. In *Visual culture*, ed. C. Jenks, 1–25. London, UK: Routledge.
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.
Katz, P.A. 1963. Effects of labels on children's perception and discrimination learning. *Journal of Experimental Psychology* 66 (5): 423–428.
Keane, M.T., T. Ledgeway, and S. Duff. 1994. Constraints on analogical mapping: A comparison of three models. *Cognitive Science* 18 (3): 387–438.
Kivy, P. 1991. *Music alone: Philosophical reflections on the purely musical experience*. Ithaca, NY: Cornell University Press.
Köhler, W. 1929. *Gestalt Psychology*. New York: Liveright.
Köhler, W. 1947. *Gestalt psychology: An introduction to new concepts in modern psychology*. New York, NY: Liveright Publication.
Kornblum, S., T. Hasbroucq, and A. Osman. 1990. Dimensional overlap: Cognitive basis for stimulus-response compatibility – a model and taxonomy. *Psychological Review* 97: 253–270.

Kottenhoff, H. 1957. Situational and personal influences on space perception with experimental spectacles: Part one: Prolonged experiments with inverting glasses. *Acta Psychologica* 13: 79–97.
Kurtz, K. J. 1996. Category-based similarity. In G.W. Cottrell ed. *Proceedings of the eighteenth annual conference of the cognitive science society* (290). Hillsdale, NJ: Erlbaum.
Laska, M., and P. Teubner. 1999. Olfactory discrimination ability of human subjects for ten pairs of enantiomers. *Chemical Senses* 24 (2): 161–170.
Leibniz, G.W. 1923. *Sämtliche Schriften und Briefe [Complete Writings and Letters]*. Berlin: Akademie Verlag.
Levin, D.M. 1993. *Modernity and the hegemony of vision*. Berkeley, CA: University of California Press.
Levinson, S.C., and A. Majid. 2014. Differential ineffability and the senses. *Mind & Language* 29 (4): 407–427.
Levitan, C.A., S. Charney, K.B. Schloss, and S.E. Palmer. 2015. The smell of jazz: Crossmodal correspondences between music, odor, and emotion. *CogSci* 1: 1326–1331.
Lewkowicz, D.J., and G. Turkewitz. 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.
Liew, K., P. Lindborg, R. Rodrigues, and S.J. Styles. 2018. Cross-modal perception of noise-in-music: Audiences generate spiky shapes in response to auditory roughness in a novel electroacoustic concert setting. *Frontiers in Psychology* 9: 178.
Liew, K., S. J., Styles, and P. Lindborg. 2017. Dissonance and roughness in cross-modal perception. In *Proceedings of the 6th Conference of the Asia Pacific Society for the Cognitive Sciences of Music*.
Livingston, K.R., J.K. Andrews, and S. Harnad. 1998. Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24 (3): 732–753.
Loeb, M., I. Behar, and J.S. Warm. 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.
Logothetis, N.K., and D.L. Sheinberg. 1996. Visual object recognition. *Annual Review of Neuroscience* 19 (1): 577–621.
Luce, R.D., R. Steingrimsson, and L. Narens. 2010. Are psychophysical scales of intensities the same or different when stimuli vary on other dimensions? Theory with experiments varying loudness and pitch. *Psychological Review* 117 (4): 1247–1258. https://doi.org/10.1037/a0020174.
Ludwig, V.U., and J. Simner. 2013. What colour does that feel? Tactile–visual mapping and the development of cross-modality. *Cortex* 49 (4): 1089–1099.
Macpherson, F. 2012. Cognitive penetration of colour experience: Rethinking the issue in light of an indirect mechanism. *Philosophy and Phenomenological Research* 84 (1): 24–62.
Margiotoudi, K., and F. Pulvermüller. 2020. Action sound–shape congruencies explain sound symbolism. *Scientific Reports* 10 (1): 1–13.
Marks, L.E. 1974. On associations of light and sound: The mediation of brightness, pitch, and loudness. *The American Journal of Psychology* 87: 173–188.
Marks, L.E. 1978. *The unity of the senses: Interrelations among the modalities*. New York, NY: Academic Press.
Marks, L.E. 1987. On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance* 13 (3): 384–394.
Marks, L.E. 1989. For hedgehogs and foxes: Individual differences in the perception of cross-modal similarity. In *Psychophysics in action*, eds. G. Ljunggren and S. Dornic, 55–65. Berlin: Springer Verlag.
Marks, L.E. 2004. Cross-modal interactions in speeded classification. In *Handbook of multisensory processes*, ed. G.A. Calvert, C. Spence, and B.E. Stein, 85–105. Cambridge, MA: MIT Press.
Marks, L.E. 2011. Synesthesia then and now. *Intellectica* 55 (1): 47–80.
Marks, L.E., R. Szczesiul, and P. Ohlott. 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.
Medin, D., and A. Ortony. 1989. Psychological essentialism. In *Similarity and analogical reasoning*, eds. S. Vosniadou and A. Ortony, 179–195. Cambridge, UK: Cambridge University Press.
Medin, D.L., R.L. Goldstone, and D. Gentner. 1990. Similarity involving attributes and relations: Judgments of similarity and difference are not inverses. *Psychological Science* 1 (1): 64–69.

Medin, D.L., R.L. Goldstone, and D. Gentner. 1993. Respects for similarity. *Psychological Review* 100 (2): 254–278.
Mellers, B., and M.H. Birnbaum. 1982. Loci of contextual effects in judgment. *Journal of Experimental Psychology: Human Perception & Performance* 8: 582–601.
Menninghaus, W., V. Wagner, E. Wassiliwizky, I. Schindler, J. Hanich, T. Jacobsen, and S. Koelsch. 2019. What are aesthetic emotions? *Psychological Review* 126 (2): 171–195. https://doi.org/10.1037/rev0000135.
Moul, E.R. 1930. An experimental study of visual and auditory "thickness." *The American Journal of Psychology* 42 (4): 544–560.
Nosofsky, R.M. 1984. Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10 (1): 104–114.
Osgood, C.E. 1952. The nature and measurement of meaning. *Psychological Bulletin* 49: 197–237.
Osgood, C.E., G.J. Suci, and P.H. Tannenbaum. 1957. *The measurement of meaning*. Urbana, IL: University of Illinois Press.
Palmer, S.E., K.B. Schloss, Z. Xu, and L.R. Prado-León. 2013. Music–color associations are mediated by emotion. *Proceedings of the National Academy of Sciences* 110 (22): 8836–8841.
Parise, C.V., and C. Spence. 2009. 'When birds of a feather flock together': Synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS One* 4 (5): e5664.
Parise, C., and C. Spence. 2013. Audiovisual cross-modal correspondences in the general population. In *The Oxford handbook of synesthesia*, eds. J. Simner and E.M. Hubbard, 790–815. Oxford, UK: Oxford University Press.
Parise, C.V., K. Knorre, and M.O. Ernst. 2014. Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences of the USA* 111: 6104–6108.
Passi, A., & Arun, S.P. 2022. The bouba–kiki effect is predicted by sound properties but not speech properties. *Attention, Perception, & Psychophysics*, 1–15. https://doi.org/10.3758/s13414-022-02619-8.
Piesse, G.W.S. 1867. *The art of perfumery and the methods of obtaining the odors of plants: with instructions for the manufacture of perfumes for the handkerchief, scented powders, odorous vinegars, dentifrices, pomatums, cosmetics, perfumed soap, etc., to which is added an appendix on preparing artificial fruit-essences, etc*. Philadelphia: Lindsay & Blakiston.
Proctor, R.W., and Y.S. Cho. 2006. Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin* 132 (3): 416–442.
Pylyshyn, Z.W. 1984. *Computation and cognition*. Cambridge, MA: MIT Press.
Quine, W.V. 1960. *Word and objects*. Cambridge, MA: MIT Press.
Quine, W.V. 1969. Natural kinds. In *Ontological relativity and other essays*, 114–138. New York, NY: Columbia University Press.
Quine, W.V. 2000. Three networks: Similarity, implication, and membership. *The Proceedings of the Twentieth World Congress of Philosophy* 6: 287–291.
Radford, C. 1989. Emotions and music: A reply to the Cognitivists. *The Journal of Aesthetics and Art Criticism* 47 (1): 69–76.
Ramachandran, V.S., and E.M. Hubbard. 2001. Synaesthesia – a window into perception, thought and language. *Journal of Consciousness Studies* 8 (12): 3–34.
Ravignani, A., and R. Sonnweber. 2017. Chimpanzees process structural isomorphisms across sensory modalities. *Cognition* 161: 74–79.
Reeve, C.D.C. 2004. *Plato. The Republic*. Indianapolis: Hackett.
Rodriguez-Pereyra, G. 2002. *Resemblance nominalism: A solution to the problem of universals*. Oxford, UK: Oxford University Press.
Root, R.T., and S. Ross. 1965. Further validation of subjective scales for loudness and brightness by means of cross-modality matching. *The American Journal of Psychology* 78 (2): 285–289.
Rosch, E., and C.B. Mervis. 1975. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* 7: 573–605.
Rosch, E., C.B. Mervis, W.D. Gray, D.M. Johnson, and P. Boyes-Braem. 1976. Basic objects in natural categories. *Cognitive Psychology* 8 (3): 382–439.
Ryle, G. 1939a. Plato's Parmenides (I). *Mind* 48 (190): 129–151.
Ryle, G. 1939b. Plato's Parmenides (II). *Mind* 48 (191): 302–325.
Schifferstein, H.N.J., and I. Tanudjaja. 2004. Visualizing fragrances through colors: The mediating role of emotions. *Perception* 33 (10): 1249–1266. https://doi.org/10.1068/p5132.
Scruton, R. 2004. Wittgenstein and the understanding of music. *The British Journal of Aesthetics* 44 (1): 1–9.

Segundo-Ortin, M., and D.D. Hutto. 2021. Similarity-based cognition: Radical enactivism meets cognitive neuroscience. *Synthese* 198: 5–23.
Sharma, D., K. Ng, I. Birznieks, and R.M. Vickery. 2022. Auditory clicks elicit equivalent temporal frequency perception to tactile pulses: A cross-modal psychophysical study. *Frontiers in Neuroscience* 16: 1006185.
Shepard, R.N. 1962. The analysis of proximities: Multidimensional scaling with an unknown distance function. *II. Psychometrika* 27: 219–246.
Shepard, R.N. 1964. Circularity in judgments of relative pitch. *The Journal of the Acoustical Society of America* 36 (12): 2346–2353.
Shepard, R.N. 1974. Representation of structure in similarity data: Problems and prospects. *Psychometrika* 39: 373–421.
Shepard, R.N. 1987. Toward a universal law of generalization for psychological science. *Science* 237: 1317–1323.
Sidhu, D.M., C. Westbury, G. Hollis, and P.M. Pexman. 2021. Sound symbolism shapes the English language: The maluma/takete effect in English nouns. *Psychonomic Bulletin & Review* 28 (4): 1390–1398.
Siegel, S. 2012. Cognitive penetrability and perceptual justification. *Noûs* 46: 201–222.
Simpson, R.H., M. Quinn, and D.P. Ausubel. 1956. Synesthesia in children: Association of colors with pure tone frequencies. *The Journal of Genetic Psychology* 89 (1): 95–103.
Slobodenyuk, N., Y. Jraissati, A. Kanso, L. Ghanem, and I. Elhajj. 2015. Cross-modal associations between color and haptics. *Attention, Perception, & Psychophysics* 77 (4): 1379–1395.
Smith, E.E., and D.L. Medin. 1981. *Categories and concepts*. Cambridge, MA: Harvard University Press.
Smith, L.B. 1993. The concept of same. *Advances in Child Development and Behavior* 24: 215–252.
Spearman, C. 1923. *The nature of "intelligence" and the principles of cognition*. London, UK: Macmillan.
Spence, C. 2011. Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics* 73: 971–995.
Spence, C. 2019. On the relative nature of (pitch-based) crossmodal correspondences. *Multisensory Research* 32 (3): 235–265.
Spence, C. 2020a. Assessing the role of emotional mediation in explaining crossmodal correspondences involving musical stimuli. *Multisensory Research* 33: 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. https://rdcu.be/b9oDJ.
Spence, C. 2022a. Searching for perceptual similarity within, and between, the (chemical) senses. *i-Perception* 13 (5): 20416695221124150.
Spence, C. 2022b. Proprioceptive art: How should it be defined, and why has it become so popular? *i-Perception* 13(5). https://doi.org/10.1177/20416695221120522.
Spence, C., and N. Di Stefano. (submitted). What, if any, can be considered amodal? *Psychonomic Bulletin & Review.*
Spence, C. and N. Di Stefano. (in press). Sensory translation between audition and vision. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-023-02343-w.
Spence, C., and N. Di Stefano. 2022. Coloured hearing, colour music, colour organs, and the search for perceptually meaningful correspondences between colour and sound. *i-Perception* 13 (3): 20416695221092800.
Spence, C., and C.A. Levitan. 2021. Explaining crossmodal correspondences between colours and tastes. *i-Perception* 12 (3): 20416695211018224.
Spence, C., X. Wan, A. Woods, C. Velasco, J. Deng, J. Youssef, and O. Deroy. 2015. On tasty colours and colourful tastes? Assessing, explaining, and utilizing crossmodalcorrespondences between colours and basic tastes. *Flavour* 4: 1–17.
Spiro, R.J., P.J. Feltovich, R.L. Coulson, and D.A. Anderson. 1989. Multiple analogies for complex concepts: Antidotes for analogy-induced misconception in advanced knowledge acquisition. In *Similarity and analogical reasoning*, eds. S. Vosniadou and A. Ortony, 498–531. Cambridge, UK: Cambridge University Press.
Stevens, S.S. 1957. On the psychophysical law. *Psychological Review* 64: 153–181.
Stevens, S.S. 1961. To honor Fechner and repeal his law: A power function, not a log function, describes the operating characteristic of a sensory system. *Science* 133 (3446): 80–86.

Stevens, S.S. 1971. Issues in psychophysical measurement. *Psychological Review* 78: 426–450. https://doi.org/10.1037/h0031324.

Stevens, S.S., and M. Guirao. 1963. Subjective scaling of length and area and the matching of length to loudness and brightness. *Journal of Experimental Psychology* 66: 177–186. https://doi.org/10.1037/h0044984.

Stokes, D. 2013. Cognitive penetrability of perception. *Philosophy Compass* 8: 646–663.

Stratton, G.M. 1897. Vision without inversion of the retinal image. *Psychological Review* 4 (4): 341–360.

Treitler, L. 1982. The early history of music writing in the west. *Journal of the American Musicological Society* 35 (2): 237–279.

Tversky, A. 1977. Features of similarity. *Psychological Review* 84 (4): 327–352.

Tversky, B., and K. Hemenway. 1984. Objects, parts, and categories. *Journal of Experimental Psychology: General* 113 (2): 169–193.

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: 1059–1069.

Von Békésy, G. 1967. Mach band type lateral inhibition in different sense organs. *The Journal of General Physiology* 50 (3): 519–532.

Von Hornbostel, E.M. 1931. Über Geruchshelligkeit [On smell brightness]. *Pflügers Archiv Für Die Gesamte Physiologie Des Menschen Und Der Tiere* 227: 517–538. https://doi.org/10.1007/BF01755351.

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

Vuoskoski, J.K., and T. Eerola. 2011. Measuring music-induced emotion: A comparison of emotion models, personality biases, and intensity of experiences. *Musicae Scientiae* 15 (2): 159–173.

Walker-Andrews, A. 1994. Taxonomy for intermodal relations. In *The development of intersensory perception: Comparative perspectives*, eds. D.J. Lewkowicz and R. Lickliter, 39–56. Hillsdale, NJ: Lawrence Erlbaum.

Weber, E.H. 1978. *The sense of touch*. San Diego, CA: Academic.

Weinberger, A. B., N. M. Gallagher, G. Colaizzi, N. Liu, N. Parrott, E. Fearon, and A.E. Green. 2022. Analogical mapping across sensory modalities and evidence for a general analogy factor. *Cognition*, 223:105029.

Whiteford, K.L., K.B. Schloss, N.E. Helwig, and S.E. Palmer. 2018. Color, music, and emotion: Bach to the blues. *i-Perception* 9 (6): 1–27.

Wilde, C. 2007. Matter and meaning in the work of art: Joseph Kosuth's "One and Three Chairs." In *Philosophy and conceptual art*, eds. P. Goldie and E. Schellenkens, 119–137. Oxford, UK: Oxford University Press.

Williams, T. 2002. Two aspects of Platonic recollection. *Apeiron* 35 (2): 131–152.

Winter, B. 2016a. Taste and smell words form an affectively loaded part of the English lexicon. *Language, Cognition and Neuroscience* 31 (8): 975–988.

Winter, B. 2016b. The sensory structure of the English lexicon. PhD Dissertation, University of California Merced. https://escholarship.org/uc/item/885849k9.

Wittgenstein, L. 1953. *Philosophical investigations*, 1986. Oxford, UK: Basil Blackwell.

Wollheim, R. 2003. In defense of seeing-in. In *Looking into pictures: An interdisciplinary approach to pictorial space*, eds. H. Hecht, R. Schwartz, and M. Atherton, 3–15. Cambridge, MA: MIT Press.

Young, J.Z. 1978. *Programs of the brain*. New York, UK: Oxford University Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.