BRILL

Multisensory Research 29 (2016) 29–48

MULTISENSORY RESEARCH

brill.com/msr

Crossmodal Correspondences: Four Challenges

Ophelia Deroy [1,*] **and Charles Spence** [2]

[1] Centre for the Study of the Senses, Institute of Philosophy, School of Advanced Study, University of London, Senate House, Malet Street, WC1E 7HU London, UK

[2] Crossmodal Research Laboratory, Department of Experimental Psychology, Oxford University, UK

Received 6 December 2014; accepted 6 March 2015

Abstract

The renewed interest that has emerged around the topic of crossmodal correspondences in recent years has demonstrated that crossmodal matchings and mappings exist between the majority of sensory dimensions, and across all combinations of sensory modalities. This renewed interest also offers a rapidly-growing list of ways in which correspondences affect — or interact with — metaphorical understanding, feelings of 'knowing', behavioral tasks, learning, mental imagery, and perceptual experiences. Here we highlight why, more generally, crossmodal correspondences matter to theories of multisensory interactions.

Keywords

Crossmodal correspondences, sound symbolism, metaphors, multisensory interactions, unity assumption

1. Introduction

For three decades or so, multisensory research has made a great deal of progress, and yet, until the last few years, has really hardly touched on the topic of crossmodal correspondences. Looking back over the field, discussion of the correspondences is notable by its absence from most seminal reviews of the field. For example, neither Welch and Warren's (1986) influential review of human psychophysics, nor Stein and Meredith's classic (1993) book on the neurophysiological underpinnings of multisensory integration mention the correspondences at all. Discussion of the correspondences is also largely absent from Calvert *et al.*'s (2004) *Handbook of Multisensory Processing*, except

* To whom correspondence should be addressed. E-mail: ophelia.deroy@sas.ac.uk

DOI:10.1163/22134808-00002488

for a solitary chapter by Lawrence Marks. Why this topic has effectively been ignored isn't altogether clear. Perhaps at the time these texts were written, the researchers were motivated more by the rapid advances in animal neurophysiology into thinking first and foremost about the importance of space and time as central factors governing multisensory integration: Generally-speaking, it would appear that our brains have been shaped to combine sensory cues coming from different senses when they occur close together in time or from similar locations in space (though see Spence, 2013), and the investigation of the rules governing this binding has dominated the field of multisensory research.

For many years now, the majority of investigations on the topic of the interactions between the senses has tended to focus on trying to understand the spatial and temporal factors modulating cases where our senses contribute to the perception of the same object or property (e.g., see Alais and Burr, 2004; Calvert, Spence and Stein, 2004; Ernst and Bülthoff, 2004; Roach *et al.*, 2006; and see the chapters in Spence and Driver, 2004; Stein, 2012, for reviews). The contribution of prior knowledge concerning which cues go together has mostly focused on the 'unity assumption' (Welch and Warren, 1986) and looked at how integration is modulated by the semantic knowledge that steaming kettles emit whistling sounds, or that dog-like shapes go with barking sounds (e.g., Chen and Spence, 2011; Doehrmann and Naumer, 2008). The matchings encompassed as crossmodal correspondences, and presented in the present issue, should strike one as being of a much broader kind: although some corresponding properties can belong to a single object, like in the case of size and pitch, they can equally concern two distinct objects. In the most famous case of sound-shape correspondences, for instance, people agree that a rounded cloud-like shape is better named 'Maluma' than 'Takete' although the name and the shape are presented as distinct, and although people don't think that cloud-like shapes are regularly experienced in nature emitting a 'Maluma'-like sound (see Fig. 1a, b). Although lemons don't move any faster than any other type of fruit or vegetable, people generally feel inclined to say that lemons are fast: following on Peter Walker's initial anecdotal confirmation that design students would consider lemons to be fast rather than slow, we recently demonstrated that lemons were twice as likely to be considered fast than slow, and that the same feeling extends across several different cultural groups (Woods *et al.*, 2013).

As these crossmodal matchings do not find an immediate explanation in terms of object-based associations, such as those that hold between visual and tactile shape, the smell of lemon and the colour yellow, or a barking sound and the shape or image of a dog, it is easy to understand that the main question that they raise is one of causal origin (see Deroy and Spence, 2013a for a

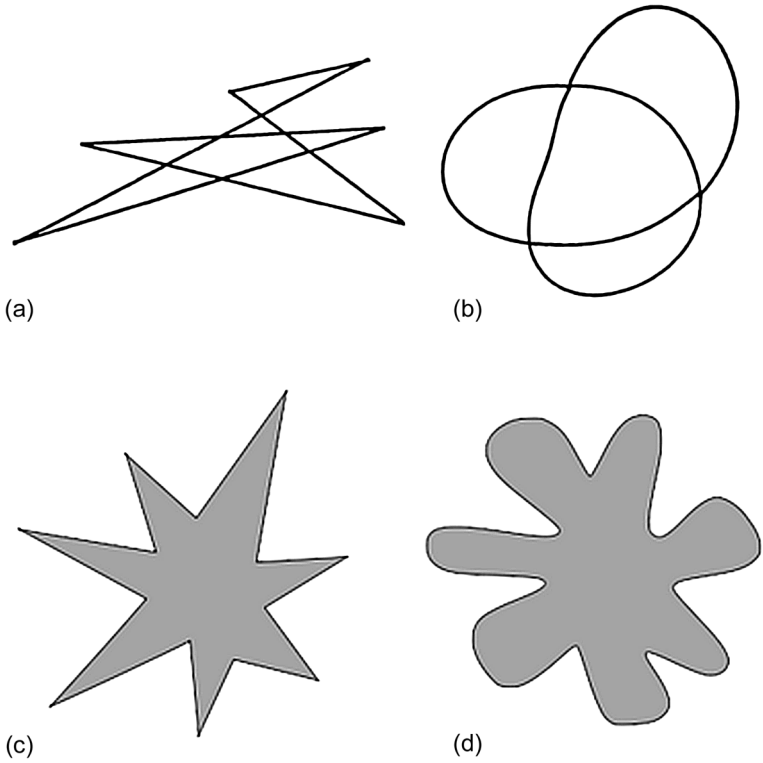

Figure 1. Köhler (1929, 1947) asked participants which of the two shapes (a) and (b) should be called 'Maluma' and which should be called 'Takete'. He observed that most people would instinctively call the spiky shape Takete and the rounded shape Maluma; Ramachandran and Hubbard (2001) had participants choose which of the two shapes (c) and (d) should be called 'Bouba' and which 'Kiki'. 90% of the Western participants apparently agreed that 'Bouba' would be the rounded shape and 'Kiki' the spiky one. Bremner *et al.* (2013) found that illiterate participants, the Himba of Kaokoland in Northern Namibia, converged on the same matching.

review). However, as originally pointed out by Marks (1978), and developed more recently by Spence (2011), it is likely that different correspondences have different origins, or perhaps mixed origins. A prior challenge, in this respect, might be to individuate crossmodal correspondences (Section 2) — a question which will help to solve the second challenge of distinguishing or relating them to other forms of mappings and associations (Section 3). These two questions open onto two others, regarding the pressure on the distinction between perceptual spaces (Section 4) and the need to consider multisensory interactions across different objects, or between objects and sensory context (Section 5).

2. Are All Crossmodal Correspondences of One Kind?

Growing out of the pioneering work on audio-visual correspondences by Lawrence Marks and Robert Melara (e.g., Marks, 1974; Melara and Marks, 1990; see Marks, 2004, for a review), crossmodal matches (or correspondences) have been documented between most sensory dimensions, and across all combinations of sensory modalities: people consistently and reliably pair certain musical timbres with particular smells (Crisinel and Spence, 2012; Deroy *et al.*, 2013), shapes (Adeli *et al.*, 2014), and colours (Haverkamp, 2014); they consider changes in pitch to match difference in height (Doldscheid *et al.*, 2014) and pitch to correspond to visual brightness (Ludwig *et al.*, 2011; Melara, 1989) and size (Evans and Treisman, 2010; Gallace and Spence, 2006; see also Eitan, 2013a, b, for reviews). Visual shapes are matched to tactile hardness, brightness, and pitch (Walker, 2012), to pieces of music (Athanasopoulos and Moran, 2013; Eitan, 2013c), tastes (Velasco *et al.*, 2014), smells (Hanson-Vaux *et al.*, 2013; Seo *et al.*, 2010) and flavours (Deroy and Valentin, 2011), while colours are matched to temperatures (e.g., Ho *et al.*, 2014), music (e.g., Chen, 2013; Palmer *et al.*, 2013), odours (Guerdoux *et al.*, 2014), angles (Albertazzi *et al.*, 2015), shapes (Chen *et al.*, 2015), tastes (Tomasik-Krótki and Strojny, 2008; Wan *et al.*, 2014) and letters (Rouw *et al.*, 2014; Simner *et al.*, 2005). While going from touch and taste to vision can be more frequent in language than going from vision to touch or taste (see Yu, 2003), *all* matchings seem to be equivalently easy to understand by participants (Werning *et al.*, 2006) and bi-directionality seems the rule, rather than the exception, in crossmodal correspondences (Deroy and Spence, 2013b, submitted).

What is also surprising is that matchings often remain consistent from one domain to the next. The matching that holds between shapes and names, for instance, remains consistent with the shapes and names matched to odours, tastes, and flavours (see Spence and Ngo, 2012, for a review): fruity smells and sweeter flavours are both independently matched with the name ‘Maluma’ and with rounded shapes. This said, these matches do not necessarily follow the rules of logic: for instance, lower pitch is considered larger than higher pitch, yet descending pitches shrink, instead of getting larger (Eitan, 2013b; Eitan *et al.*, 2014).

Instead of an anecdotal or isolated phenomenon, crossmodal matchings thus appear to be pervasive, and more or less inter-related. A long list of other sensory connections documented in the older research literature, or in artists’ notebooks, is currently awaiting more robust testing by means of modern psychophysical techniques, such as the weight and texture of colours or the ‘brightness’ of smells (see Alexander and Shansky, 1976; Cohen, 1934; Monroe, 1926; von Hornbostel, 1931). With the recent growth of robust internet-

based testing techniques (e.g., Miller, 2012), a number of these under-explored phenomena can also be tested across different cultures, and at a wide scale (e.g., Wan *et al.*, 2014; Woods *et al.*, 2013). By testing crossmodal correspondences beyond those participants who share the same Western, Educated, Industrial, Rich, and Democratic background (W.E.I.R.D, see Henrich *et al.*, 2010), and quite often come from similar linguistic backgrounds, it becomes possible to see whether these matches have been picked-up through manufactured objects or symbolic associations. Importantly, many crossmodal correspondences are shared across cultures, and some would even appear to be universal (Athanasopoulos and Moran, 2013; Bremner *et al.*, 2013; Dolscheid *et al.*, 2014; Shayan *et al.*, 2011; Woods *et al.*, 2013). What is universal, at least, is the fundamental tendency to match dimensions and features across distinct sensory modalities, or to map one dimension of experience onto another. It is an important characteristic of the human brain and mind: one that might also exist in a few primates and animals, though perhaps only in a more rudimentary form (see Faragó *et al.*, 2010; Ludwig *et al.*, 2011; Morton, 1994).

When studying this general tendency though, it is important to know whether it should be broken down into more specific mechanisms and categories. A first way to draw a distinction is in terms of causal origins: some correspondences are likely to be grounded in the statistical regularities present in the environment (for instance, pitch-size, or pitch-elevation, see Parise *et al.*, 2014); others might originate in structural similarities (e.g., in the coding of sensory magnitudes, such as brightness–loudness, see Stevens and Marks, 1965) or even be semantically (Martino and Marks, 1999; Walker, 2012) or hedonically mediated (Palmer *et al.*, 2013; Velasco *et al.*, 2015) (Table 1). A broader distinction, then, might be between those cases where dimensions are directly associated, and cases where they are associated because of their relation to a third, mediate element. In turn, this distinction might illuminate another difference, between cases of absolute matchings and cases of contextual mappings. In certain correspondences, a definite feature or value on a dimension (e.g., red, lemon) is matched to a definite feature (angularity) or

Table 1.
Distinctions to be taken into account in distinguishing between different kinds of correspondences

Elements	Relation	Origin
Dimension–dimension	Absolute matching	Statistical
Dimension–feature	Relative mapping	Structural
Feature–feature		Semantic mediation
		Hedonic mediation

polarity (heavy, fast); in others, the same value (e.g., pitch) will be mapped onto a different polarity or value in a contextual or relative way (e.g., different degrees of brightness, different relative height). Besides the causal explanation of the correspondence, and the relation itself, correspondences also differ regarding the elements they relate — which can be either psychological dimensions (e.g., brightness, pitch) or sensory features (shape, colour).

3. Are Crossmodal Correspondences Distinct from Other Associations or Mappings?

At first sight, the tendency to match dimensions across domains of experience can seem only to be a manifestation of the well-documented capacity that we all have to create and understand metaphors. It is more than thirty years now since Lakoff and Johnson (1980) first proposed that these ways of crossing sensory domains were instances of metaphors, but of a more fundamental kind than those present in language. In their influential book, *Metaphors We Live by*, their argument was that certain mappings were inherent to the making of our cognitive, but also embodied and emotional life. Their key example was the fact that 'up' is positive and 'down' negative in almost all cultures, and that this mapping crops in all sorts of practices, from linguistic expressions to gestures, and from dance to music (see Eitan *et al.*, in press; and Maes and Leman, 2013, for recent examples and discussion). Positive words are shown to shift spatial attention upward, while negative words shift it downward (Meier and Robinson, 2004). This mapping is so fundamental that it seems impossible to start considering an upward movement as negative. In a similar way, it seems impossible to hear a sound that is increasing in pitch as descending. An increase in pitch is heard as going up, not down. Young children's first dance moves seem to follow this rule (Kohn and Eitan, 2009) and more thorough tests have confirmed that infants as young as seven months of age associate rising pitch and rising visual movement (Jeschonek *et al.*, 2013; see also Dolscheid *et al.*, 2014, for cross-cultural evidence).

It is undeniable that certain cross-sensory mappings appear to be deeply embedded in our language and practices, but labeling them as metaphor might not be sufficient to distinguish them from other cultural associations — such as between white and purity, for instance (Wheatley, 1973) — or from those mappings holding between sensory and abstract domains, or sensory and affective domains. Lawrence Marks has, for decades, been trying to better understand the articulation between metaphors and sensory mappings, to show that the two are not just manifestations of the same underlying correspondence. Marks became interested, for instance, in the way in which children and adults understand the same cross-sensory metaphors. He discovered that children and adults would disagree on the meaning of cross-sensory metaphors,

with children finding that a quiet sunshine was less bright than a loud moonshine, whereas adults tended to consider the reverse mapping more appropriate (Marks, 1978, 1982, 1996). Applying auditory adjectives such as 'loud' or 'quiet' to more or less bright visual objects, such as moonshine and sunshine, is not done in the same way by adults and children. By contrast, adults and children agree on a perceptual version of the matching, and converge on the degree of brightness that corresponds to a specific sound. The capacity for cross-sensory mappings can be influenced by, but still operates independently of, the capacity to understand metaphor.

The brute and intuitive character of these matchings, we would argue, is what singles them out from metaphors and other semantic correspondences: while most metaphors and semantic correspondences can be justified by the person who hears them, crossmodal correspondences seem to afford no further explanation — at the explicit level — than the fact that they just 'feel right'. This criterion might not be sufficient to draw a clear-cut distinction between all metaphors and crossmodal correspondences, as some metaphors might also feel very intuitive, but seems at least to point to an important characteristic of correspondences. When many cognitive neuroscientists, anthropologists, and philosophers are trying, at the moment, to understand the nature of moral and linguistic intuitions, it feels especially timely to investigate the nature of these sensory intuitions. Making sense of this intuitive aspect might be one of the challenges that lie ahead for researchers interested in the difference between correspondences and sensory metaphors (e.g., Forceville, 2006).

4. Do Crossmodal Correspondences Suggest a 'Unity of the Senses'?

Historically, the existence of crossmodal correspondences has been seen as a challenge to the core notion, passed from common sense into philosophy and psychology, that the experiences conveyed by each sense have no immediate connection to one another, or don't welcome any transition. This notion was captured by Helmholtz's forceful early claim 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). The better understanding of the physiological distinction between sensory processing certainly encouraged the claim, expressed here, that their outputs, and noticeably the conscious experiences resulting from the stimulation of different senses, belong to totally separate and unreconciled sensory

domains. The most famous proponent of this view, Bishop Berkeley noted that we should not confuse our judgment that the various experiences refer to the same thing with the fact that experiences remain distinct and are merely associated in our mind: "Sitting in my Study I hear a Coach drive along the street; I look through the Casement and see it; I walk out and enter into it; thus, common Speech would incline one to think, I heard, saw, and touch'd the same thing, to wit, the Coach. It is nevertheless certain, the Ideas intromitted by each Sense are widely different, and distinct from each other; but having been observed constantly to go together, they are spoken of as one and the same thing" (Berkeley, 1709/1995, pp. 51–52). Although many would choose not to be as radical as Berkeley, the vast majority of thinkers would nevertheless still recognize that seeing a cube and touching a cube are different conscious experiences, with no immediate connection (Held *et al.*, 2011). As we get repeatedly exposed to the conjunction of certain visual and tactile experiences (e.g., Connolly, 2014), we might learn to do a fast transition between the two 'ideas' or experiences, but they remain subjectively distinct.

An idea closely related to Helmholtz's and Berkeley's claims, and expressed by the philosopher Austin Clark, is that it is possible to define a sensory modality through sets of experiences that 'match' with one another: "Facts about matching can individuate modalities. Sensations in a given modality are connected by the matching relation. From any sensation in the given modality, it is possible to reach any other by a sufficiently long series of matching steps. Distinct modalities are not so connected. One can get from red to green by a long series of intermediaries, each matching its neighbors; but no such route links red to C-sharp." (Clark, 1993, pp. 140–141). The point made by Clark is directly challenged by the existence of matching relations between colours and sounds (Langlois *et al.*, 2013): one can get from a colour to another colour, but also from a colour to sound, or to textural properties through what seems also to deserve to be called a matching relation. It is possible to argue that a different operation or kind of 'matching' is at stake when we match a shade of grey to a darker shade of grey, or when we match a dark patch to a low-pitched sound, but it is not immediately clear in which sense the two operations might differ.

Shall we then consider that correspondences reveal a kind of 'unity of the senses' (Marks, 1978)? The suggestion, eloquently defended by mystics such as Swedenborg (see Garrett, 1984, or Wilkinson, 1996) and poets such as Baudelaire (1857), was that the intuitive relations captured in the correspondences was accessing a real objective unity: the idea being that the world is not carved according to our senses, but constitutes "a deep and tenebrous unity, Vast as the dark of night and as the light of day" where "Perfumes, sounds, and colors correspond." (Baudelaire, 1857). Several researchers have followed the same idea: Von Hornsbostel famously stressed that "What is essential in the

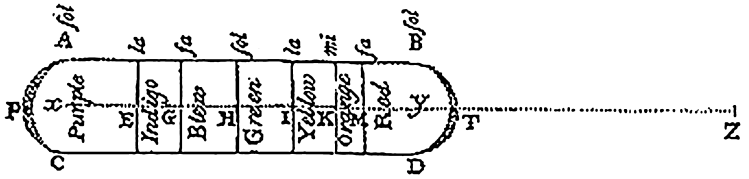

Figure 2. Musical divisions of the prism, proposed by Newton (1675). The seven colours red, orange, yellow, green, blue, indigo and violet, fill the seven intervals between the eight notes, starting from the highest note on the right. Deep violet, on the left, is the most refracted light, and red on the right is the least refracted light (from Newton, 1675/1757).

sensuous-perceptible is not that which separates the senses from one another, but that which unites them; unites them among themselves; unites them with the entire (even with the non-sensuous) experience in ourselves; and with all the external world that there is to be experienced." (Von Hornbostel, 1927, 1950, p. 214), while the Gibsons and their followers thought that our senses were picking up invariant information out there (Gibson, 1966).

The claim of an objective unity behind our matching intuitions is not specific to psychologists: Newton, after all, believed that our intuitions about which colours and sounds went best together was a cue to an objective 'parallel' between lightwaves and soundwaves (Pesic, 2006; see also Duck, 1988, for discussion). Newton's idea originated in the observation that both the vibrations of the light rays and the air vibrations excite sensory nerves according to their wavelength: harmonies of colors, Newton thought, would then depend on the ratio of vibrations propagated through the optical nerve, in the same way as the harmonies of sounds are thought to come from the ratio between air vibrations (see Fig. 2).

These sort of parallelisms have been abandoned by contemporary physicists, but this makes Newton's and von Hornbostel's points all the more challenging: unless there is a real connection out there between light and sound, why is it that we find some colour-to-sound mappings more intuitive than others? Why do we have a 'feeling of knowing' which answer to give when asked whether the sound of a trombone is dark or bright (Haverkamp, 2014) and why does this puzzling association modulate perception? For one thing, we might need to widen our approach to learning, and investigate natural scenes statistics more thoroughly: if it does not seem true that objects and animals which emit high pitch sounds are more often experienced as being high in space, a more general analysis of what our ears receive might find that this is overall the case (Parise *et al.*, 2014). Correspondences might not be internalized object by object, but track general regularities in experience and the environment. We might also need to revise our understanding of how subjective dimensions of experience — such as pitch, elevation, magnitude or brightness — might relate to one another.

With new studies revealing that sensory processing is not as neatly localized in primary sensory areas for each sense as was once thought (Ghazanfar and Schroeder, 2006), and that the sensory processing of stimuli is multisensory from the earliest levels (e.g., Henschke *et al.*, 2014; Liang *et al.*, 2013, for recent studies), the idea of a neat distinction between sensory channels, all having distinct inputs and well-separated outputs, might need to be revised. Finding how early the neural signature of crossmodal correspondences can be detected, in this sense, might play a crucial role in determining whether or not sensory correspondences constitute one of the manifestations of this early cross-sensory interactions (Kovic *et al.*, 2010; Ludwig *et al.*, 2011; Revill *et al.*, 2014; Sadaghiani *et al.*, 2009).

5. Shall We Consider Interactions Across Objects, Rather than Integration Within a Single Object?

Crossmodal correspondences have been shown to bear on the way stimuli from multiple sensory modalities are processed in a variety of tasks: crossmodally congruent shapes make our responses to words and non-words faster (Westbury, 2005), high pitched sounds played in the background speed up our responses to bright stimuli (Ludwig *et al.*, 2011; Spence, 2011), crossmodal correspondences can also influence our performance on visual search tasks (Klapetek *et al.*, 2012; Orchard-Mills *et al.*, 2013a, b) or on shape estimation tasks (Sweeny *et al.*, 2012).

Importantly, these effects do not seem to vary with the spatial congruency between the two cues — something which might also hold more generally for other multisensory interactions (Spence, 2013). Although many of these effects occur when the two corresponding stimuli are presented at the same time, they do not seem to be confined to a strict temporal window either. Crossmodal correspondences have, for instance, been shown to operate in priming and spatial cuing tasks, when one of the two stimuli is deliberately placed before the other (e.g. Chiou and Rich, 2012; Fernández-Prieto *et al.*, 2012; Ho *et al.*, 2014; see Spence and Deroy, 2013a for a review). Again, though, this is not specific to crossmodal correspondences, and might hold true of other semantic correspondences as well — although differences might exist with semantic influences perhaps taking longer than crossmodal ones to show an effect (Chen and Spence, 2011). What seems more distinctive of crossmodal correspondences is the fact that they operate independently of the unity assumption and hold across objects which remain perceived as distinct: the round shapes are not perceived as belonging to the visual words, nor the high pitched sounds to the bright patches we are interested in classifying, and yet they can still influence our responses to these objects. It can be argued, therefore, that crossmodal correspondences present us with a new kind of interaction between the

senses — one that does not require that the senses that are involved to target the same object or property, and can hold even when two distinct objects are perceived as separated in both space and time. This capacity to hold between distinct objects, as it turns out, might explain why correspondences between sounds and visual shapes or other properties come to be so important in the acquisition of language, by helping infants to establish and remember the relation between two objects, i.e., audible words and visual referents (see Asano *et al.*, 2015; Imai and Kita, 2014, for recent reviews).

Crossmodal correspondences are crucial in understanding how our minds and brains respond to natural scenes, that is when a plurality of objects are presented in different modalities, and not just when two properties can be referred to the same object thanks to the 'unity assumption' (see Welch and Warren, 1986; see also Vatakis *et al.*, 2008). When two stimuli are presented in different sensory modalities, the decision as to whether to bind or segregate them depends partly on whether or not our mind — or brain — assumes that the stimuli have a single source or belong to the same object. What falls under this assumption has varied over the years, but it is presumed to depend on the number of physical properties (e.g., shape, size, motion) that are redundantly represented in the stimulus situation (Welch and Warren, 1980; see Spence, 2007) as well as other learned associations between non-redundant information available to each sensory modality (i.e., semantic congruency, see Doehrmann and Naumer, 2008; Laurienti *et al.*, 2004; see Deroy, 2014, for discussion) — both of which depend on the sensitivity to spatio-temporal congruencies and statistical regularities (Altieri *et al.*, 2013).

What correspondences show here is that not all the factors which have an influence on the processing of jointly presented stimuli fall under the category of the 'unity assumption': although they are shown to bear on our propensity to integrate sensory cues and refer them to a single source (e.g. Brunel *et al.*, in press; Parise and Spence, 2009) crossmodal correspondences are not targeting properties that necessarily belong to a single object, and, as detailed in Section 2, not all are obviously internalized because of spatio-temporal congruencies or statistical regularities that are present in the environment. We would argue that, if correct, this discovery is fundamental to perception, but also to the contextual shaping of memory, learning, expectations, mental imagery, the orientation of attention, evaluation, and decision-making. Although these different processes might seem to operate on specific objects, their working might well be shaped by the kind of correspondence or congruence which hold between the target object and the rest of the environment. Our senses are almost always stimulated by multiple sources of information, which can be more or less distinct and independent in terms of their qualities, space, and time: the main challenge is to understand how the senses interact when mul-

tiple objects are at stake, not just when senses collaborate in referring to the same object.

With research in the field of multisensory interactions being strongly driven and influenced by attention-capturing illusions such as the ventriloquist effect, the McGurk effect, the rubber-hand illusion, or the two-flash or bouncing balls sound-induced visual illusions are key examples here (see Alais and Burr, 2004; Botvinick and Cohen, 1998; McGurk and MacDonald, 1976; Sekuler *et al.*, 1997; Shams *et al.*, 2000), it is important to stress that crossmodal correspondences also lead to illusory percepts, albeit apparently more loosely tied to the unity assumptions and spatial or temporal congruence: congruent speech sounds have been shown to affect the perceived size of ovals (Sweeny *et al.*, 2012); rising or descending melodies have been shown to induce a visual after effect (Hedger *et al.*, 2013) and to alter the perception of visual motion (Takeshima and Gyoba, 2013); congruent visual gestures bias the perception of music (Connell *et al.*, 2013; Schutz and Lipscomb, 2007) and colour affects reactions to thermal stimuli (Ho *et al.*, 2014). More surprisingly perhaps, music and soundscapes that have been based on the correspondences can affect the perceived bitterness of a foodstuff (Crisinel *et al.*, 2012).

The effects on consciousness of crossmodal correspondences are certainly not confined to these illusory percepts, and their role has been suggested to extend to mental imagery in unstimulated streams when stimulation is reduced to just one modality (e.g., Spence and Deroy, 2013b), as well as to attention. And here, we would argue that crossmodal correspondences have all the reasons to be also relevant to new fields of multisensory research.

6. Conclusions

In our sense, several important theoretical questions need to be addressed before we can understand the role(s) that crossmodal correspondences might play in our overall cognitive economy. Some of these questions are specific to the correspondences. Others connect correspondences to other phenomena and stress why correspondences offer a general entry into issues about the origins of metaphorical thinking, the distinction between the senses or the existence of multisensory interactions across objects — not to mention the origins of reference and language (Imai and Kita, 2014). These difficult questions, which we have organized into four core challenges, need to be addressed collaboratively, by cognitive neuroscientists, linguists, psycholinguists, anthropologists and philosophers. Is the relation of correspondence a direct one, or mediated by concepts or emotions (e.g., Palmer *et al.*, 2013)? How do they relate to other kinds of mappings and correspondences? Do crossmodal correspondences reveal a kind of unity between the senses, less salient perhaps than their distinction? Instead of helping us to bind information into a single ob-

ject, aren't they showing that the information coming from different sources can still be used to facilitate the processing of the relation between objects, across sensory modalities? The ultimate goal is to explain our cross-sensory intuitions and tendencies to match or translate experiences which apparently belong to very distinct sensory repertoires — something we consider to be forming the general category of 'crossmodal correspondences'.

We see no reason to posit a mysterious unity of the senses leading to a better grasp on the deeper unity of the world, as suggested by Baudelaire or Von Hornbostel. But there is also no reason to think of crossmodal correspondences as nothing more than merely arbitrary connections, governed by language and concepts. What the older conceptions of the unity of the senses remind us of is that the distinction between the senses can be seen as a sort of punishment, or fault, that made us miss out on more general environmental connections and relations: a claim that most certainly contrasts with the alternative claim of modern biologists and psychologists who have stressed that having different senses increases the likelihood that one will perceive things more veridically (e.g., Ernst and Bülthoff, 2004; Stein and Meredith, 1993). Keeping the best of these two views, we suggest that sensory evolution should be seen as an on-going trade-off between increasing specialization on the one hand, and the need to maintain some kind of unity on the other: the more senses a creature has, the harder it is going to be to solve the crossmodal binding problem (Spence, 2010; Treisman, 1996), but also to maintain a sense of all objects fitting in the same scene, in relation with one another. The importance of crossmodal correspondences is, then, perhaps best understood as one of the means by which the nervous system has dealt with this trade-off.

References

Adeli, M., Rouat, J. and Molotchnikoff, S. (2014). Audiovisual correspondence between musical timbre and visual shapes, *Front. Hum. Neurosci.* **8**, 352.

Alais, D. and Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration, *Curr. Biol.* **14**, 257–262.

Albertazzi, L., Malfatti, M., Canal, L. and Micciolo, R. (2015). The hue of angles — was Kandinsky right? *Art Percept.* **3**, 81–92.

Alexander, K. R. and Shansky, M. S. (1976). Influence of hue, value, and chroma on the perceived heaviness of colors, *Percept. Psychophys.* **19**, 72–74.

Altieri, N., Stevenson, R. A., Wallace, M. T. and Wenger, M. J. (2013). Learning to associate auditory and visual stimuli: behavioral and neural mechanisms, *Brain Topogr*. DOI:10.1007/s10548-013-0333-7.

Asano, M., Imai, M., Kita, S., Kitajo, K., Okada, H. and Thierry, G. (2015). Sound symbolism scaffolds language development in preverbal infants, *Cortex* **63**, 196–205.

Athanasopoulos, G. and Moran, N. (2013). Cross-cultural representations of musical shape, *Empir. Musicol. Rev.* **8**, 185–199.

Baudelaire, C. (1857). *Les Fleurs du mal [Flowers of Evil]*. Poulet-Malassis, Paris, France.

Berkeley, G. (1709/1975). An essay toward a new theory of vision, in: *Philosophical Works*, M. R. Ayers (Ed.), pp. 1–59. Dent, London, UK.

Botvinick, M. and Cohen, J. (1998). Rubber hands 'feel' touch that eyes see, *Nature* **391**(6669), 756.

Bremner, A., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. and Spence, C. (2013). Bouba and Kiki in Namibia? A remote culture makes similar shape–sound matches, but different shape–taste matches to Westerners, *Cognition* **126**, 165–172.

Brunel, L., Carvalho, P. and Goldstone, R. L. (in press). It does belong together: cross-modal correspondences influence cross-modal integration during perceptual learning, *Front. Psychol.* **6**, 358. DOI:10.3389/fpsyg.2015.00358.

Calvert, G., Spence, C. and Stein, B. E. (Eds) (2004). *The Handbook of Multisensory Processing*. MIT Press, Cambridge, MA, USA.

Chen, L. (2013). Synaesthetic correspondence between auditory clips and colors: an empirical study, in: *Intelligent Science and Intelligent Data Engineering, Second Sino-Foreign-Interchange Workshop, IScIDE 2011, Xi'an, China, October 23–25, 2011, Revised Selected Papers*, pp. 507–513. Springer, Berlin, Germany.

Chen, Y.-C. and Spence, C. (2011). Crossmodal semantic priming by naturalistic sounds and spoken words enhances visual sensitivity, *J. Exp. Psychol. Hum. Percept. Perform.* **37**, 1554–1568.

Chen, N., Tanaka, K. and Watanabe, K. (2015). Color–shape associations revealed with implicit association tests, *PLoS ONE* **10**, e0116954. DOI:10.1371/journal.pone.0116954.

Chiou, R. and Rich, A. N. (2012). Cross-modality correspondence between pitch and spatial location modulates attentional orienting, *Perception* **41**, 339–353.

Clark, A. (1993). *Sensory Qualities*. Oxford University Press, Oxford, UK.

Cohen, N. E. (1934). Equivalence of brightness across modalities, *Am. J. Psychol.* **46**, 117–119.

Connell, L., Cai, Z. G. and Holler, J. (2013). Do you see what I'm singing? Visuospatial movement biases pitch perception, *Brain Cogn.* **81**, 124–130.

Connolly, K. (2014). Multisensory perception as an associative learning process, *Front. Psychol.* **5**, 1095. DOI:10.3389/fpsyg.2014.01095.

Crisinel, A.-S. and Spence, C. (2012). A fruity note: crossmodal associations between odors and musical notes, *Chem. Sens.* **37**, 151–158.

Crisinel, A.-S., Cosser, S., King, S., Jones, R., Petrie, J. and Spence, C. (2012). A bittersweet symphony: systematically modulating the taste of food by changing the sonic properties of the soundtrack playing in the background, *Food Qual. Pref.* **24**, 201–204.

Deroy, O. (2014). The unity assumption and the many unities of consciousness, in: *Sensory Integration and the Unity of Consciousness*, D. Bennett and C. Hill (Eds), pp. 105–124. MIT Press, Cambridge, MA, USA.

Deroy, O. and Spence, C. (2013a). Are we all born synaesthetic? Examining the neonatal synaesthesia hypothesis, *Neurosci. Biobehav. Rev.* **37**, 1240–1253.

Deroy, O. and Spence, C. (2013b). Why we are not all synesthetes (not even weakly so), *Psychon. Bull. Rev.* **20**, 643–664.

Deroy, O. and Spence, C. (submitted). The role of vision in crossmodal matchings.

Deroy, O. and Valentin, D. (2011). Tasting shapes: investigating the sensory basis of cross-modal correspondences, *Chemosens. Percept.* **4**, 80–90.

Deroy, O., Crisinel, A.-S. and Spence, C. (2013). Crossmodal correspondences between odors and contingent features: odors, musical notes, and geometrical shapes, *Psychon. Bull. Rev.* **20**, 878–896.

Doehrmann, O. and Naumer, M. J. (2008). Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration, *Brain Res.* **1242**, 136–150.

Dolscheid, S., Hunnius, S., Casasanto, D. and Majid, A. (2014). Prelinguistic infants are sensitive to space-pitch associations found across cultures, *Psychol. Sci.* **25**, 1256–1261.

Duck, M. J. (1988). Newton and Goethe on colour: physical and physiological considerations, *Ann. Sci.* **45**, 507–519.

Eitan, Z. (2013a). How pitch and loudness shape musical space and motion: new findings and persisting questions, in: *The Psychology of Music in Multimedia*, S. L. Tan, A. Cohen, S. Lipscomb and R. Kendall (Eds), pp. 161–187. Oxford University Press, Oxford, UK.

Eitan, Z. (2013b). Which is more? Pitch height, parametric scales, and the intricacies of cross-domain magnitude relationships, in: *Musical I: Essays in Honor of Eugene Narmour*, L. Brenstein and A. Rozin (Eds), pp. 131–148. Pendragon Press, Hillsdale, NY, USA.

Eitan, Z. (2013c). Musical objects, cross-domain correspondences, and cultural choice: commentary on "Cross-cultural representations of musical shape" by George Athanasopoulos and Nikki Moran, *Empir. Musicol. Rev.* **8**, 204–207.

Eitan, Z., Schupak, A., Gotler, A. and Marks, L. E. (2014). Lower pitch is larger, yet falling pitches shrink: interaction of pitch change and size change in speeded discrimination, *Exp. Psychol.* **61**, 273–284.

Eitan, Z., Timmers, R. and Adler, M. (in press). The mapping of musical shapes to multiple dimensions, in: *Music and Shape*, D. Leech-Wilkinson and H. Prior (Eds). Oxford University Press, Oxford, UK.

Ernst, M. O. and Bülthoff, H. H. (2004). Merging the senses into a robust percept, *Trends Cogn. Sci.* **8**, 162–169.

Evans, K. K. and Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features, *J. Vis.* **10**(6), 1–12.

Faragó, T., Pongrácz, P., Miklósi, Á., Huber, L., Virányi, Z. and Range, F. (2010). Dogs' expectation about signalers' body size by virtue of their growls, *PLoS ONE* **5**, e15175. DOI:10.1371/journal.pone.0015175.

Fernández-Prieto, I., Vera-Constán, F., García-Morera, J. and Navarra, J. (2012). Spatial recoding of sound: pitch-varying auditory cues modulate up/down visual spatial attention, *See. Perceiv.* **25**, 150–151.

Forceville, C. (2006). Non-verbal and multimodal metaphor in a cognitivist framework: agendas for research, in: *Cognitive Linguistics: Current Applications and Future Perspectives*, G. Kristiansen, M. Achard, R. Dirven and F. Ruiz de Mendoza Ibàñez (Eds), pp. 379–402. Mouton de Gruyter, New York, NY, USA.

Gallace, A. and Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size, *Percept. Psychophys.* **68**, 1191–1203.

Garrett, C. (1984). Swedenborg and the mystical enlightenment in late eighteenth-century England, *J. Hist. Ideas* **45**, 67–81.

Ghazanfar, A. A. and Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends Cogn. Sci.* **10**, 278–285.

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Houghton Mifflin, Boston, MA, USA.

Guerdoux, E., Trouillet, R. and Brouillet, D. (2014). Olfactory–visual congruence effects stable across ages: yellow is warmer when it is pleasantly lemony, *Atten. Percept. Psychophys.* **76**, 1280–1286.

Hanson-Vaux, G., Crisinel, A.-S. and Spence, C. (2013). Smelling shapes: crossmodal correspondences between odors and shapes, *Chem. Sens.* **38**, 161–166.

Haverkamp, M. (2014). *Synesthetic Design: Handbook for a Multi-Sensory Approach*. Birkhäuser Verlag, Basel, Switzerland.

Hedger, S. C., Nusbaum, H. C., Lescop, O., Wallisch, P. and Hoeckner, B. (2013). Music can elicit a visual motion aftereffect, *Atten. Percept. Psychophys.* **75**, 1039–1047.

Held, R., Ostrovsky, Y., de Gelder, B., Gandhi, T., Ganesh, S., Mathur, U. and Sinha, P. (2011). The newly sighted fail to match seen with felt, *Nat. Neurosci.* **14**, 551–553.

Helmholtz, H. (1878/1971). The facts of perception, in: *Selected Writings of Hermann Helmholtz*. Wesleyan University Press, Middletown, CT, USA.

Henrich, J., Heine, S. J. and Norenzayan, A. (2010). The weirdest people in the world? *Behav. Brain Sci.* **33**, 61–135.

Henschke, J. U., Noesselt, T., Scheich, H. and Budinger, E. (2014). Possible anatomical pathways for short-latency multisensory integration processes in primary sensory cortices, *Brain Struct. Funct.* **219**, 1–23.

Ho, H. N., Van Doorn, G. H., Kawabe, T., Watanabe, J. and Spence, C. (2014). Colour–temperature correspondences: when reactions to thermal stimuli are influenced by colour, *PLoS ONE* **9**, e91854. DOI:10.1371/journal.pone.0091854.

Imai, M. and Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution, *Phil. Trans. R. Soc. B Biol. Sci.* **369**(1651), 20130298.

Jeschonek, S., Pauen, S. and Babocsai, L. (2013). Cross-modal mapping of visual and acoustic displays in infants: the effect of dynamic and static components, *Eur. J. Dev. Psychol.* **10**, 337–358.

Johnson, M. and Lakoff, G. (1980). *Metaphors We Live by*. University of Chicago Press, Chicago, IL, USA.

Klapetek, A., Ngo, M. K. and Spence, C. (2012). Do crossmodal correspondences enhance the facilitatory effect of auditory cues on visual search? *Atten. Percept. Psychophys.* **74**, 1154–1167.

Köhler, W. (1929). *Gestalt Psychology*. Liveright, New York, NY, USA.

Köhler, W. (1947). *Gestalt Psychology: An Introduction to New Concepts in Modern Psychology*. Liveright, New York, NY, USA.

Kohn, D. and Eitan, Z. (2009). Musical parameters and children's movement responses, in: *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music*, J. Louhivuori, T. Eerola, S. Saarikallio, T. Himberg and P. S. Eerola (Eds), pp. 233–241. ESCOM 2009, Jyväskylä, Finland.

Koriat, A. (1975). Phonetic symbolism and feeling of knowing, *Mem. Cogn.* **3**, 545–548.

Koriat, A. (2008). Subjective confidence in one's answers: the consensuality principle, *J. Exp. Psychol. Learn. Mem. Cogn.* **34**, 945–959.

Kovic, V., Plunkett, K. and Westermann, G. (2010). The shape of words in the brain, *Cognition* **114**, 19–28.

Langlois, T., Schloss, K. and Palmer, S. (2013). Music–color associations to simple melodies in synesthetes and non-synesthetes, *J. Vis.* **13**, 1325.

Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H. and Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance, *Exp. Brain Res.* **158**, 405–414.

Liang, M., Mouraux, A., Hu, L. and Iannetti, G. D. (2013). Primary sensory cortices contain distinguishable spatial patterns of activity for each sense, *Nat. Commun.* **4**, 1979.

Ludwig, V. U., Adachi, I. and Matzuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (*Pan troglodytes*) and humans, *Proc. Natl Acad. Sci. USA* **108**, 20661–20665.

Maes, P. J. and Leman, M. (2013). The influence of body movements on children's perception of music with an ambiguous expressive character, *PLoS ONE* **8**, e54682. DOI:10.1371/journal.pone.0054682.

Marks, L. E. (1974). On associations of light and sound: the mediation of brightness, pitch, and loudness, *Am. J. Psychol.* **87**, 173–188.

Marks, L. E. (1978). *The Unity of the Senses: Interrelations Among the Modalities*. Academic Press, New York, NY, USA.

Marks, L. E. (1982). Bright sneezes and dark coughs, loud sunlight and soft moonlight, *J. Exp. Psychol. Hum. Percept. Perform.* **8**, 177–193.

Marks, L. E. (1996). On perceptual metaphors, *Metaphor Symb.* **11**, 39–66.

Marks, L. (2004). Crossmodal interactions in speeded classifications, in: *The Handbook of Multisensory Processes*, G. Calvert, C. Spence and B. E. Stein (Eds), pp. 85–105. MIT Press, Cambridge, MA, USA.

Martino, G. and Marks, L. E. (1999). Perceptual and linguistic interactions in speeded classification: tests of the semantic coding hypothesis, *Perception* **28**, 903–924.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices, *Nature* **264**(5588), 746–748.

Meier, B. P. and Robinson, M. D. (2004). Why the sunny side is up: associations between affect and vertical position, *Psychol. Sci.* **15**, 243–247.

Melara, R. D. (1989). Dimensional interaction between color and pitch, *J. Exp. Psychol. Hum. Percept. Perform.* **15**, 69–79.

Melara, R. D. and Marks, L. E. (1990). Processes underlying dimensional interactions: correspondences between linguistic and nonlinguistic dimensions, *Mem. Cogn.* **18**, 477–495.

Miller, G. (2012). The smartphone psychology manifesto, *Perspect. Psychol. Sci.* **7**, 221–237.

Monroe, M. (1926). The apparent weight of color and correlated phenomena, *Am. J. Psychol.* **36**, 192–206.

Morton, E. (1994). Sound symbolism and its role in non-human vertebrate communication, in: *Sound Symbolism*, L. Hinton, J. Nichols and J. Ohala (Eds), pp. 348–365. Cambridge University Press, Cambridge, UK.

Newton, I. (1675/1757). Hypothesis explaining the properties of light, in: *History of the Royal Society*, Vol. III, T. Birch (Ed.), pp. 247–305. A. Millar, London, UK.

Orchard-Mills, E., Alais, D. and Van der Burg, E. (2013a). Amplitude-modulated auditory stimuli influence selection of visual spatial frequencies, *J. Vis.* **13**, 6. DOI:10.1167/13.3.6.

Orchard-Mills, E., Alais, D. and Van der Burg, E. (2013b). Cross-modal associations between vision, touch and audition influence visual search through top-down attention, not bottom-up capture, *Atten. Percept. Psychophys.* **75**, 1892–1905.

Palmer, S. E., Schloss, K. B., Xu, Z. and Prado-León, L. R. (2013). Music–color associations are mediated by emotion, *Proc. Natl Acad. Sci. USA* **110**, 8836–8841.

Parise, C. V. and Spence, C. (2009). When birds of a feather flock together: synesthetic correspondences modulate audiovisual integration in non-synesthetes, *PLoS ONE* **4**, e5664. DOI:10.1371/journal.pone.0005664.

Parise, C. V., Knorre, K. and Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing, *Proc. Natl Acad. Sci. USA* **111**, 6104–6108.

Pesic, P. (2006). Isaac Newton and the mystery of the major sixth: a transcription of his manuscript 'Of Musick' with commentary, *Interdiscipl. Sci. Rev.* **31**, 291–306.

Ramachandran, V. S. and Hubbard, E. M. (2001). Synaesthesia — a window into perception, thought and language, *J. Conscious. Stud.* **8**, 3–34.

Revill, K. P., Namy, L. L., DeFife, L. C. and Nygaard, L. C. (2014). Cross-linguistic sound symbolism and crossmodal correspondence: evidence from fMRI and DTI, *Brain Lang.* **128**, 18–24.

Roach, N. W., Heron, J. and McGraw, P. V. (2006). Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration, *Proc. R. Soc. B* **273**, 2159–2168.

Rouw, R., Case, L., Gosavi, R. and Ramachandran, V. (2014). Color associations for days and letters across different languages, *Front. Psychol.* **5**, 369. DOI:10.3389/fpsyg.2014.00369.

Sadaghiani, S., Maier, J. X. and Noppeney, U. (2009). Natural, metaphoric, and linguistic auditory direction signals have distinct influences on visual motion processing, *J. Neurosci.* **29**, 6490–6499.

Schutz, M. and Lipscomb, S. (2007). Hearing gestures, seeing music: vision influences perceived tone duration, *Perception* **36**, 888–892.

Sekuler, R., Sekuler, A. B. and Lau, R. (1997). Sound alters visual motion perception, *Nature* **385**(6614), 308.

Seo, H.-S., Arshamian, A., Schemmer, K., Scheer, I., Sander, T., Ritter, G. and Hummel, T. (2010). Cross-modal integration between odors and abstract symbols, *Neurosci. Lett.* **478**, 175–178.

Shams, L., Kamitani, Y. and Shimojo, S. (2000). Illusions. What you see is what you hear, *Nature* **408**(6814), 788.

Shayan, S., Ozturk, O. and Sicoli, M. A. (2011). The thickness of pitch: crossmodal metaphors in Farsi, Turkish, and Zapotec, *Sens. Soc.* **6**, 96–105.

Simner, J., Ward, J., Lanz, M., Jansari, A., Noonan, K., Glover, L. and Oakley, D. A. (2005). Non-random associations of graphemes to colours in synaesthetic and non-synaesthetic populations, *Cogn. Neuropsychol.* **22**, 1069–1085.

Spence, C. (2007). Audiovisual multisensory integration, *Acoust. Sci. Technol.* **28**, 61–70.

Spence, C. (2010). Multisensory integration: solving the crossmodal binding problem. Comment on "Crossmodal influences on visual perception" by Shams & Kim, *Physics of Life Rev.* **7**, 285–286.

Spence, C. (2011). Crossmodal correspondences: a tutorial review, *Atten. Percept. Psychophys.* **73**, 971–995.

Spence, C. (2013). Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule, *Ann. NY Acad. Sci.* **1296**, 31–49.

Spence, C. and Deroy, O. (2013a). How automatic are crossmodal correspondences? *Conscious. Cogn.* **22**, 245–260.

Spence, C. and Deroy, O. (2013b). Crossmodal mental imagery, in: *Multisensory Imagery*, S. Lacey and R. Lawson (Eds), pp. 157–183. Springer, New York, NY, USA.

Spence, C. and Driver, J. (Eds) (2004). *Crossmodal Space and Crossmodal Attention*. Oxford University Press, Oxford, UK.

Spence, C. and Ngo, M. (2012). Assessing the shape symbolism of the taste, flavour, and texture of foods and beverages, *Flavour* **1**, 12.

Stein, B. E. (Ed.) (2012). *The New Handbook of Multisensory Processing*. MIT Press, Cambridge, MA, USA.

Stein, B. E. and Meredith, M. A. (1993). *The Merging of the Senses*. MIT Press, Cambridge, MA, USA.

Stevens, J. C. and Marks, L. E. (1965). Cross-modality matching of brightness and loudness, *Proc. Natl Acad. Sci. USA* **54**, 407–411.

Sweeny, T. D., Guzman-Martinez, E., Ortega, L., Grabowecky, M. and Suzuki, S. (2012). Sounds exaggerate visual shape, *Cognition* **124**, 194–200.

Takeshima, Y. and Gyoba, J. (2013). Changing pitch of sounds alters perceived visual motion trajectory, *Multisens. Res.* **26**, 317–332.

Tomasik-Krótki, J. and Strojny, J. (2008). Scaling of sensory impressions, *J. Sens. Stud.* **23**, 251–266.

Treisman, A. (1996). The binding problem, *Curr. Opin. Neurobiol.* **6**, 171–178.

Vatakis, A., Ghazanfar, A. and Spence, C. (2008). Facilitation of multisensory integration by the 'unity assumption': is speech special? *J. Vis.* **8**, 1–11.

Velasco, C., Salgado-Montejo, A., Marmolejo-Ramos, F. and Spence, C. (2014). Predictive packaging design: tasting shapes, typographies, names, and sounds, *Food Qual. Pref.* **34**, 88–95.

Velasco, C., Woods, A. T., Deroy, O. and Spence, C. (2015). Hedonic mediation of the crossmodal correspondence between taste and shape, *Food Qual. Pref.* **41**, 151–158.

Von Hornbostel, E. M. (1931). Über Geruchshelligkeit [On smell brightness], *Pflügers Arch. gesamte Physiol. Menschen Tiere* **227**, 517–538.

Von Hornbostel, E. M. (1950). The unity of the senses, in: *A Source Book of Gestalt Psychology*, W. D. Ellis (Ed.), pp. 210–216. Routledge and Kegan Paul, London, UK.

Walker, P. (2012). Cross-sensory correspondences and cross talk between dimensions of connotative meaning: visual angularity is hard, high-pitched, and bright, *Atten. Percept. Psychophys.* **74**, 1792–1809.

Wan, X., Woods, A. T., Van den Bosch, J., Mckenzie, K. J., Velasco, C. and Spence, C. (2014). Cross-cultural differences in crossmodal correspondences between tastes and visual features, *Front. Psychol. Cogn.* **5**, 1365.

Welch, R. B. and Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy, *Psychol. Bull.* **88**, 638–667.

Welch, R. B. and Warren, D. H. (1986). Intersensory interactions, in: *Handbook of Perception and Performance. Vol. 1. Sensory Processes and Perception*, K. R. Boff, L. Kaufman and J. P. Thomas (Eds), pp. 25-1–25-36. Wiley, New York, NY, USA.

Werning, M., Fleischhauer, J. and Beseoglu, H. (2006). The cognitive accessibility of synaesthetic metaphor, in: *Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society*, pp. 2365–2370. Québec City, QC, Canada.

Westbury, C. (2005). Implicit sound symbolism in lexical access: evidence from an interference task, *Brain Lang.* **93**, 10–19.

Wheatley, J. (1973). Putting colour into marketing, *Marketing* **67**, 24–29.

Woods, A. T., Spence, C., Butcher, N. and Deroy, O. (2013). Fast lemons and sour boulders: testing the semantic hypothesis of crossmodal correspondences using an internet-based testing methodology, *i-Perception* **4**, 365–369.

Yu, N. (2003). Synesthetic metaphor: a cognitive perspective, *J. Lit. Semantics* **32**, 19–34.