


Synthese  
<https://doi.org/10.1007/s11229-020-02862-2>

STRUCTURE OF PERCEPTUAL OBJECTS



# Structure of perceptual objects: introduction to the Synthese topical collection

Alfredo Vernazzani<sup>1</sup>  · Błażej Skrzypulec<sup>2</sup>  · Tobias Schlicht<sup>1</sup>

Received: 1 September 2020 / Accepted: 4 September 2020

© The Author(s) 2020

## 1 Introduction

Much of the current debate in philosophy of perception centers on the nature of our perceptual commerce with the world. Recently, the debate has been fueled by the re-emergence of naïve realist and relationalist perspectives that reject, on the whole, a mediated access to the objects of perception (e.g. Brewer 2011; Campbell 2002; Fish 2009; Martin 1997). Call this the “mainstream” debate in philosophy of perception.

To a large extent, the mainstream debate has almost exclusively focused on vision, and neglected other sense modalities. This “visuocentrism” is often coupled with the tacit assumption that what is true about vision may be simply transferrable to other sense-modalities as well. (For recent exceptions, see Fulkerson 2013; O’Callaghan 2018). Taken together, the mainstream interest in the nature of perception and visuocentrism have led philosophers to neglect an equally important topic of investigation, the structure of perceptual objects, obscuring significant structural differences among the sense-modalities, and the fact that perceptual experiences are frequently multi-modal (e.g. Kubovy and von Valkenburg 2001; O’Callaghan 2012, 2015).

We think that the project of clarifying the structure of perceptual objects is of paramount importance for perception studies in at least three different senses. First, clarifying the nature and structure of visual objects’ structure will shed light not only on the way we experience objects through our sense modality, but also provide a “benchmark” that can be useful for thinking about the differences between visual objects and objects of other sense modalities. In other words, a clarification of the structure of visual objects will be instrumental also to break with the rampant visuocentrism of much of current philosophy of perception. Second, clarifying the structure of perceptual objects, more generally, will shed light on the distinctive manner of composition of different kinds of objects of perception, and how interaction with other sense modal-

---

✉ Błażej Skrzypulec  
[blazej.skrzypulec@gmail.com](mailto:blazej.skrzypulec@gmail.com)

<sup>1</sup> Institut für Philosophie II, Ruhr-Universität Bochum, Universitätsstr. 150, 44780 Bochum, Deutschland

<sup>2</sup> Institute of Philosophy, Jagiellonian University, Grodzka 52, 31-044 Krakow, Poland

ities, cognition, or action may contribute to shaping our perceptual commerce with the world. Third, and more obviously, shedding light on the structure of perceptual objects means clarifying the structure of perceptually subjective appearances.

In the remainder of this Introduction to the Topical Collection “The Structure of Perceptual Objects” we will first (Sect. 2) provide some background about studies on the structure of perceptual objects and then (Sect. 3) present an overview of the papers, clustering them into two groups.

## 2 State of the art

There are different dimensions along which we can talk about the structure of perceptual objects. Here, we will review the core areas which have been explored in this topical collection. They include: the feature-object binding problem, multisensory binding, whether (and which) sense modalities have objects, and finally the relation between structure of perceptual objects and bodily actions. We will briefly review them in this order.

### 2.1 Feature-object binding

The feature-object binding problem is, essentially, the problem of explaining how different visual properties or features come to be attached to the same perceptual individual (Clark 2000; Cohen 2004; Matthen 2005; Treisman 1996). In his now classic contribution, Austen Clark (2000) framed the problem in the following way. Suppose the following holds:

1. S sees something red;
2. S sees something triangular;
3. S sees something both red and triangular (a red triangle).

As Clark pointed out, (3) does not follow from the mere conjunction of (1) and (2). Seeing something red and seeing something triangular is different from seeing a red triangle. The problem is compounded if we introduce a further object, say, a blue square. In some way, the task of the cognitive system is that of correctly sorting out the perceptual attributes to the right things: “red” and “triangular” should go together, whereas “blue” and “square” should belong to a different item.

Clark’s own solution to the riddle, as we know, was that of proposing a theory of sensory individuals (the term is due to Cohen 2004), i.e. perceptual properties are attributed to distinct individuals. On Clark’s own view (2000, pp. 164ff), visual properties are predicated of places. S sees redness and triangularity here, and blueness and square there. In this sense, and to borrow Gareth Evans’ terminology, location in space would provide the “fundamental ground of difference” (Evans 1982, p. 107) that allows the correct featural attribution (Clark 2004, pp. 136–144). This is usually called the feature-placing hypothesis.

The feature placing hypothesis is an attractive and elegant solution to the puzzle of feature-object binding, however, it has faced substantial theoretical and empirical criticism (Cohen 2004; Matthen 2004, 2005; Pylyshyn 2007; Siegel 2002; Vernazzani

forth.). In particular, it is claimed that the feature placing hypothesis cannot account for cases of superimposed object perception (Blaser et al. 2000; Pylyshyn 2007), that is, cases in which we apparently see two objects transparently superimposed like two Gabor patches. Furthermore, it seems at variance with evidence of dynamic feature-object binding, i.e. cases in which perceptual objects move in space, thus making it impossible to attribute multiple features to the same ‘here’ or ‘there’ (Matthen 2005).

## 2.2 Multisensory binding

Multimodal experiences are experiences which happen due to interaction of at least two distinct sensory systems. From a philosophical perspective, the most interesting types of multimodality are those in which multimodal processing has consequences for the phenomenal character of experiences. In fact, it seems that the majority of usual conscious perceptual experiences phenomenally present elements associated with distinct modalities. For instance, we may see an object while touching it or experience that an object has a certain look and makes some sound. The majority of philosophers follows this intuition and it is widely accepted that in fact we often have phenomenal, multimodal experiences (see Briscoe 2016 for a review). However, it should be noted that according to some authors the contemporary empirical knowledge concerning multimodal processing is also consistent with an alternative theory such that what we intuitively consider as a multimodal experience is in fact a series of unimodal mental states (Spence and Bayne 2014).

While it is generally agreed that there are phenomenal, multimodal experiences, it is more controversial what are the ways in virtue of which multimodal phenomenal character is organized. For instance, it may be claimed that in multimodal experiences elements associated with distinct modalities simply co-occur (see Kubovy and Schutz 2010; O’Callaghan 2015), that they are experienced as standing in intermodal spatiotemporal relations (O’Callaghan 2014; Richardson 2014), or that multimodal phenomenal character involves novel elements which cannot be ascribed to any of the usual modalities (Connolly 2014).

A mode of multimodal organization which is particularly important in this context is “multimodal binding” (e.g. Macpherson 2011; O’Callaghan 2015). Experiences involving multimodal binding are those which present that a single entity, like an object or an event, possesses characteristics related to distinct modalities. For instance, when perceiving a person speaking we do not merely experience that there is somebody speaking and somebody making facial movements but rather that both visual and auditory elements belong to the same entity.

In the literature, one may find three major approaches to characterizing experiential structures of objects presented in experiences involving multimodal binding. According to the first, reductive approach multimodal binding is simply a special case of experiencing intermodal relations in which elements related to distinct modalities are experienced as spatiotemporally overlapping (see Briscoe 2017; Deroy et al. 2014). Such an account may be attractive due to its theoretical simplicity and empirical data showing the crucial role of spatiotemporal factors in multimodal binding perception (see Vatakis and Spence 2007). However, according to some authors the

reductive approach do not find a strong support in the phenomenology of multimodal experiences and has difficulties in accommodating data suggesting the presence of multimodal object-files (e.g., Jordan et al. 2010; Zmigrod et al. 2009). In consequence, two types of nonreductive theories of multimodal binding have been developed which characterize binding in terms of instantiation or in terms of parthood.

According to instantiation theory, the structure of multimodal objects should be characterized in terms of a subject who instantiates properties represented by distinct modalities (Macpherson 2011; O'Callaghan 2008). Such an approach can account for both phenomenal unity related to multimodal binding and is consistent with presence of multimodal object-files. However, it may be doubtful whether instantiation theories are able to adequately grasp the experiential nature of elements presented in multimodal experiences. For instance, sound does not seem to be experienced as properties of some multimodal subjects but rather they themselves are subjects instantiating properties such as pitch or loudness. Such problems may be omitted by mereological accounts that interpret multimodal perceptual objects as wholes composed of various unimodal parts as each such part may be a subject instantiating its own properties (O'Callaghan 2014, 2016). However, mereological accounts are not without their own controversies. For instance, in usual perceptual experiences parts of perceptual objects are spatially (like in vision) or temporally (as in audition) separated by qualitative edges. However, in experiences involving multimodal binding the elements ascribed to a common object seem to spatiotemporally coincide.

### 2.3 What modalities have objects?

The fundamental question regarding the investigations about structures of perceptual objects is whether what is perceptually presented should be characterized in terms of objects. The affirmative answer is intuitively true in the case of vision that, in usual circumstances, seems to simultaneously present numerically distinct objects possessing a variety of properties and having mereological structure (see MacCumhaill 2015; Martin 1993; Richardson 2010 for considerations about structural aspects of vision). However, it is less clear whether the same position can be accepted in case of experiences related to other perceptual modalities.

In the contemporary literature, this issue has been debated primarily in the context of auditory and olfactory modalities. First, it is discussed which entities are auditorily or olfactorily presented as objects. In the case of auditory perception, it has been postulated that sounds are represented as objects, that sounds' sources are represented as objects, or that both these types of entities achieve object-like representations (see Cohen 2010; Matthen 2010; Nudds 2014; O'Callaghan 2011). Similarly, in discussions concerning olfaction odors and their sources are the main candidates for perceptual objects (see Batty 2014; Lycan 2000; Mole 2010; Young 2016). Second, it is debated whether the auditorily or olfactorily presented entities are in fact presented in a way which justifies treating them as perceptual objects. For instance, against the idea of olfactory objecthood it has been argued that spatial abilities of human olfaction are too limited (Batty 2010), that olfaction does not obey relevant rules of perceptual constancy (Barwich 2019), or that olfaction cannot present odors as figures differentiated

from ground (Keller 2016). On the other hand, other authors have argued for the opposite thesis by claiming that olfactory abilities for spatial representation may be quite robust (Aasen 2019; Young 2020), that there are constancy phenomena in olfaction (Millar 2019), or that odors are perceptually presented as having part/whole structure (Skrzypulec 2019; Young 2016). Similarly, in considerations regarding audition, it has been debated whether sounds are presented as subjects instantiating properties (O’Callaghan 2008), mereological wholes (Nudds 2014; O’Callaghan 2011), and persisting individuals (Cohen 2010; Davies 2010; Skrzypulec 2020) even if principles of auditory scene organization are significantly different from principles of visual organization.

The above debates, which are far from being concluded, point towards a deeper question regarding the notion of ‘object’ itself. In particular, it is unclear how to characterize the notion of object in order to formulate it in a way that is suitable for analyzing the structure of perceptual objects. For instance, it can be asked whether the characteristics crucial for perceptual objecthood should be those which are proposed in ontological debates (such as being a subject or properties, having mereological structure, or persisting through time) or rather should be described in reference to the abilities of a perceptual modality (like distinguishing figure/from ground, achieving perceptual constancy, or simultaneously tracking several individuals, see Stevenson 2014 for such an approach applied to the chemical senses). Furthermore, there is no consensus whether the notion of perceptual object should be univocal in the sense that there is a single notion of objecthood which can be applied to all modalities or rather being an object means different things depending on the considered sensory system (see Green 2019; O’Callaghan 2016 for attempts in providing a univocal notion). Finally, it is worth pondering whether the concept of perceptual object can be considered as a gradual notion, such that among modalities presenting some entities as objects some of them may present objects in a stronger or a weaker sense.

## 2.4 How bodily actions structure perception

It is generally agreed that at least visual perceptual objects have some sort of spatial structure, but how do they exactly acquire such spatial structure?

According to some philosophers, the structure of perceptual objects is partially due to an appropriate connection between somatosensory and proprioceptive inputs and motor outputs, making our motor actions a fundamental actor in structuring the spatial content of our perceptual experience (Sect. 3.2). Furthermore, according to many philosophers, most often following the Gibsonian tradition (1979), we do not only perceive objects as merely ‘being there’, we also perceive the possible actions they afford, like “being graspable” or affording “sitting on” etc. It may be thus claimed that such action-affording properties of perceptual objects may contribute to building the coherent, structured wholes we perceive.

### 3 Overview of the SI

We can single out two major threads or topics in this topical collection's contributions. The first is a focus on the modes of unification (Sect. 3.1); the second (Sect. 3.2) relates dispositions for actions to the structure of experience.

#### 3.1 First topic: modes of unification

As stated in Sect. 2.2, one of the most important topics regarding the structure of perceptual objects is their mode of unification. Most researchers agree that the structure of perceptual experiences is not merely conjunctive, i.e. we do not perceive a perceptual scene as a set of separate, co-occurring elements such as *redness and greenness and circularity and squareness* (e.g., Clark 2004; Kubovy and Schutz 2010; O'Callaghan 2015; Vernazzani forth.). Instead, the perceived elements are experienced as organized into complex entities, and objects are believed to be the most important of such entities. Such problems are particularly complicated when one considers multimodal experiences in which a unified object is experienced due to cooperation between more than one modality. For instance, it seems likely that in a perceptual state one may see, hear, and touch something and experience this situation not as involving three separate entities but as an experience of a single object combining properties discovered by each of the considered senses. Nevertheless, if distinct modalities organize objects according to different principles, a question arises how these various structures are unified into a single multimodal entity.

*E. J. Green's* paper, "Binding and differentiation in multisensory object perception", addresses the issue of multimodal unification by proposing a typology of ways in which multimodal objects are integrated and providing a review of empirical literature in order to assess which types of unification are likely to occur in case of human perception. The first general type of unification, multimodal binding, is introduced by referring to the psychological notion of object-files, i.e. representations that represent properties of an individual object and allow tracking an object through time (see Kahneman et al. 1992 for a classic study). According to Green, the basic form of multimodal binding consists of two object-files, each associated with distinct modality, referring to the same object. A deeper form of binding involves creating a single, multimodal object-file which contains information about properties detected by distinct modalities. Furthermore, information processed by distinct modalities may be simply gathered in an object-file but, when used in representing the same individual, may also undergo modification. The most profound of such modifications, named 'constitutive binding', happens when in virtue of creation of an object-file a genuinely multimodal property starts to be represented which cannot be represented by any single modality. The second general type of multimodal integrations described by Green, 'multimodal differentiation', is characterized by reference to the notion of figure/ground discrimination (see Craft et al. 2007; Matthen 2010; Millar 2019 for considerations regarding various modalities). In this case, an object is differentiated from the surrounding by functioning of more than one modality. Sometimes this may happen merely because both modalities independently process information about the same fragment of the

environment. However, also more substantial forms of multimodal differentiation are possible, in which two or more modalities actually coordinate each other functioning.

The issue of multimodal unifications is also addressed in the paper “The structure of audio-visual consciousness” by *Błażej Skrzypulec*. The author investigates three modes of unification, property instantiation, part/whole relationship, and grouping, which occur in unimodal experiences and considers whether they can be applied to describe the structure of audio-visual objects. In the perceptual unification by instantiation a form of existential dependence is present, i.e. properties are not experienced as uninstantiated and objects are not experienced as propertyless. Such dependence is not present in case of mereological unification and perceptual grouping as a part of a whole or an element of a group can be experienced separately, without forming any mereological whole or being a member of any group. However, while grouping combines disjoint, similar elements (e.g. Elder and Goldberg 2002), parts of a perceptual whole are spatially connected and do not need to exhibit any significant level of similarity (Palmer and Rock 1994). Skrzypulec argues that there are audio-visual phenomena, such as cross-modal dynamic capture (e.g., Sanabria et al. 2005), that can be plausibly characterized in terms of multimodal grouping. However, other phenomena, such as audio-visual binding (e.g., Bertelson 1999), require a different treatment. The author proposes that the audio-visual binding should be characterized as a situation in which one is experiencing a numerically same entity (event or an object) that instantiates some properties determined by relying on the visual and auditory elements of an experience. In other words, the structure of audio-visual binding is as follows: (a) there are auditory elements instantiating auditory properties; (b) there are visual elements instantiating visual properties; and (c) there is a common entity instantiating properties that are determined by properties of visual and auditory elements.

The problems concerning perceptual unification are not restricted to multimodal contexts. The paper by *Anna Drożdżowicz*, “Bringing back the voice: on the auditory objects of speech perception”, investigates a unimodal, auditory case of unification. The author considers a relation between speech perception and voice perception (see also Di Bona 2017 for an investigation about high-level auditory properties perception). When hearing a speech we are also hearing it as expressed in some voice and so the meaning of what we hear is somehow unified with the auditory properties such as pitch or loudness. However, it is not the case that auditory properties of voice simply determine the perceived speech. For instance, we may hear the same speech uttered in a low-pitched and high-pitched voice. Furthermore, we can recognize speech even if each word is spoken in a different voice. Drożdżowicz proposes that the experienced unity between speech and voice can be characterized by applying the mereological account developed by O’Callaghan and argues that it has advantages over the proposals characterizing unification in terms of instantiation or causality (O’Callaghan 2017). According to the author’s account, speech and voice are experienced as unified in virtue of the fact that speech sound is a part of a voicing event, i.e. an event of using a voice on a particular occasion. Furthermore, Drożdżowicz proposes a Voice Shaping Model aimed to explain how the unity between voice and speech is obtained at the level of psychological processing. It is claimed that in virtue of time-dependent characteristics of the acoustic signal stable voice characteristics can be retrieved that allow for speech recognition.



### 3.2 Second topic: dispositions for action and structure of experience

The present contributions also show a second important thread for thinking about the structure of perceptual experience. Perception does not only present the properties of distal objects, but also what actions are possible given these properties. Thereby, a relevant question emerges: how does this influence the structure of our perceptual experience? Two papers in our collection tackle this issue.

In her contribution on “The Where of Bodily Awareness,” *Alisa Mandrigin* provides a dispositional account of the structure of the spatial content of bodily awareness inspired by Gareth Evans’ (1982) egocentric spatial content. By bodily awareness, Mandrigin refers to «conscious experiences of properties of the body arising from the processing of information generated from the inside». Bodily awareness, as she points out, is not an inchoate and indistinct perceptual muddle, the perceptual uptake of our own bodies presents body parts as located relative to the location of other body parts and as within the bounds of the body as a whole (Bermúdez 2017, p. 126). Mandrigin defends an original Dispositional View according to which bodily awareness has spatial content in virtue of a set of connections established between somatosensory and proprioceptive inputs and motor outputs, thereby showing that the spatial content of bodily awareness constitutively depends on our motor actions.

The Dispositional View has a number of theoretical advantages. First, it can easily explain why bodily awareness experiences facilitate immediate actions (Evans 1982, pp. 155–156), since the spatial content of bodily awareness is connected with having dispositions to perform appropriate bodily actions. Second, it naturally fits an evolutionary approach on perception, on which the preeminent biological function of perception is to enable appropriate bodily movements in relation to the environmental properties (Briscoe 2014).

Yet, the Dispositional View faces challenges coming from experimental studies on alleged dissociations between conscious bodily awareness and action. For instance, patients affected with numbness report that they do not have any proprioceptive or tactile experience in affected limbs. So, when asked to localize tactile stimulation on the affected limb they are at chance. However, when asked to point to the site of stimulation on the affected limb they can do so reliably (Wong 2015). In neurological healthy subjects, dissociations between action and bodily awareness is manifest in cases like the rubber hand illusion (Botvinick and Cohen 1998). In this case, patients may mis-locate their own hand in the direction of the rubber hand, rather than their own biological hand. This is standardly taken to indicate that the subject’s bodily experiences have an illusory spatial content. Finally, work on the Two Visual System Hypothesis seems to provide support against the idea of a constitutive link between visual perception and action, given that action and perception are processed along different neural paths (Clark 1999; cf also Briscoe 2009). This body of work is interpreted by de Vignemont (2018) as providing a challenge against action-based accounts of the spatial content of bodily awareness.

Yet, Mandrigin shows how we can resist the standard explanation of the empirical challenges against the Dispositional View. Mandrigin’s proposal rests on the importance of different reference frames in interpreting the tasks used to assess alleged



dissociations between bodily awareness and actions. As she points out, this is not a direct argument for the Dispositional View, but it shows that a certain challenge against the Dispositional View may actually be sidestepped when empirical evidence is interpreted along the novel explanation she proposes.

In his contribution “Objects, seeing, and object-seeing”, *Mohan Matthen* discusses three questions about what it means to see a material object: what it is to see, or *seeing-as-such*, as it differs, for instance, from episodes of visual recalling; what it means to see something *as an object*; and finally, what is it to see a *particular* object. Much of the philosophical discussion has centered on the first and latter question, leaving out the question of what it means to see something as an object. He points out that while causal theories of perception (Grice 1962) have been useful in providing accounts of the particularity of perception, and in establishing which object we see when we seem to see an object, they do not provide a convincing account of what it means to see an object as an object. He therefore sets out to articulate an account of the first and second question.

In his paper, Matthen articulates his stance as follows. First, he argues that seeing-as-such is a type of veridical experience brought about by external things by looking. Looking, in turn, is something we actively do and of which we can be self-aware. Second, he argues that seeing something as an object requires seeing something as the bearer of features with which we can interact. Finally, taking at face value his account of object perception, he shows how some prominent theories of vision that reduce phenomenology to a causal relation to external things are, ultimately, wrong.

The latter point crucially centers on an understanding of perceivers as active agents in a potentially dangerous environment. Thus, Matthen emphasizes that material object perception—our perception of things like trees, chairs, etc.—is shaped not only by standard features like their spatial location or visible properties, but also by their availability for interaction. When we see a material object, unlike depicted objects, we see it as if it can be grasped or touched. Material objects, Matthen says, «cohere in complicated ways» as we perceive material objects as causally connected wholes.

## 4 Conclusion

In the current debates concerning the structural aspects of sensory experiences one may observe a growing interest in investigating the structure of multimodal objects and the way in which our abilities to act influence the experiential structures. The papers published in this collection address these problems by proposing novel ideas regarding the ways in which contents of unimodal experiences are unified in multimodal ones, the relationships between higher and lower-level properties of perceptual objects, the dependency of spatial content of bodily awareness on motor actions, and the way in which availability for interaction contribute to our perception of objects as material entities. These considerations not only contribute to the current debates, but also can serve as a basis for further investigations regarding the structure of perceptual objects. In particular, it may be asked whether the general ideas concerning structures of multimodal objects can be applied beyond audio-visual and visuo-tactile contexts, to types of multimodal experiences which have attracted less attention in philosophical

debates. Furthermore, while the studies concerning the structures of perceptual objects were mainly focusing on exteroceptive experiences, which present the external entities and not the states of one's body, the works regarding influence of motor actions and bodily awareness on experiential structures open a perspective of investigating the relevance of interoceptive sensations for the structure of perceptual objects.

**Acknowledgements** We wish to thank all reviewers for their work on all papers of the topical collection and Catarina Dutilh Novaes, one of Editors-in-Chief of Synthese for her help in managing the issue. Tobias Schlicht was kindly supported by the Volkswagen Stiftung, grant no. 87 105-1 for the project "Situated Cognition. Perceiving the world and understanding other minds".

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Informed consent and animal welfare** The research did not involve human participants or animals.

**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/>.

## References

- Aasen, S. (2019). Spatial aspects of olfactory experience. *Canadian Journal of Philosophy*, 49(8), 1041–1061.
- Barwich, A.-S. (2019). A critique of olfactory objects. *Frontiers in Psychology*, 10, 1337.
- Batty, C. (2010). A representational account of olfactory experience. *Canadian Journal of Philosophy*, 40, 511–538.
- Batty, C. (2014). Olfactory objects. In S. Biggs, D. Stokes, & M. Matthen (Eds.), *Perception and its modalities* (pp. 222–224). New York: Oxford University Press.
- Bermúdez, J. (2017). Ownership and the space of the body. In F. de Vignemont & A. Alsmith (Eds.), *The subject's matter: self-consciousness and the body* (pp. 117–144). Cambridge, MA: MIT Press.
- Bertelson, P. (1999). Ventriloquism: A case of crossmodal perceptual grouping. In G. Aschersleben, T. Bachmann, and J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events*, Elsevier Science B. V.: Amsterdam, pp. 347–362.
- Blaser, E., Pylyshyn, Z., & Holcombe, A. (2000). Tracking an object through feature space. *Nature*, 408, 196–199.
- Botvinick, M., & Cohen, J. (1998). Rubber hands "feel" touch that eyes see. *Nature*, 391, 756.
- Brewer, B. (2011). *Perception and its objects*. Oxford: Oxford University Press.
- Briscoe, R. (2009). Egocentric spatial representation in action and perception. *Philosophy and Phenomenological Research*, LXXIX(2), 423–460.
- Briscoe, R. (2014). Spatial content and motoric significance. *AVANT*, V(2), 199–217.
- Briscoe, R. (2016). Multisensory processing and perceptual consciousness: Part I. *Philosophy Compass*, 11(2), 121–133.
- Briscoe, R. E. (2017). Multisensory processing and perceptual consciousness: Part II. *Philosophy Compass*, 12(12), e12423.
- Campbell, J. (2002). *Reference and consciousness*. New York: Oxford University Press.

- Clark, A. (1999). Visual awareness and visuomotor action. *Journal of Consciousness Studies*, 6, 1–18.
- Clark, A. (2000). *A theory of sentience*. Oxford: Oxford University Press.
- Clark, A. (2004). Feature-placing and proto-objects. *Philosophical Psychology*, 17(4), 443–469.
- Clark, A. (2011). Cross-modal cuing and selective attention. In F. Macpherson (Ed.), *The senses: Classic and contemporary philosophical perspectives* (pp. 375–396). Oxford: Oxford University Press.
- Cohen, J. (2004). Objects, places, and perception. *Philosophical Psychology*, 17(4), 471–495.
- Cohen, J. (2010). Sounds and temporality. *Oxford Studies in Metaphysics*, 5, 303–320.
- Connolly, K. (2014). Making sense of multiple senses. In R. Brown (Ed.), *Consciousness inside and out: Phenomenology, neuroscience, and the nature of experience* (pp. 351–364). Dordrecht: Springer.
- Craft, E., Schütze, H., Niebur, E., & von der Heydt, R. (2007). A neural model of figure-ground organization. *Journal of Neurophysiology*, 97(6), 4310–4326.
- Davies, S. (2010). Perceiving Melodies and Perceiving Musical Colors. *Review of Philosophy and Psychology*, 1, 19–39.
- de Vignemont, F. (2018). *Mind the body*. Oxford: Oxford University Press.
- Deroy, O., Chen, Y., & Spence, C. (2014). Multisensory constraints on awareness. *Philosophical Transactions of the Royal Society B*, 369, 20130207.
- Di Bona, E. (2017). Towards a rich view of auditory experience. *Philosophical Studies*, 174(11), 2629–2643.
- Elder, J. H., & Goldberg, R. M. (2002). Ecological statistics of gestalt laws for the perceptual organization of contours. *Journal of Vision*, 2(4), 324–353.
- Evans, G. (1982). *The varieties of reference*. New York: Oxford University Press.
- Fish, W. (2009). *Perception, hallucination and illusion*. New York: Oxford University Press.
- Fulkerson, M. (2013). *The first sense: A philosophical study of human touch*. Cambridge, MA: MIT Press.
- Gibson, J. J. (1979/1986). *The ecological approach to visual perception*. New York: Psychology Press.
- Green, E. J. (2019). A theory of perceptual objects. *Philosophy and Phenomenological Research*, 99(3), 663–693.
- Grice, H. P. (1962). Some remarks about the senses. In R. J. Butler (Ed.), *Analytic philosophy first series*. Oxford: Basil Blackwell.
- Jordan, K., Clark, K., & Mitroff, S. (2010). See an object, hear an object file: Object correspondence transcends sensory modality. *Visual Cognition*, 18, 492–503.
- Kahneman, D., Treisman, A. M., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, 24(2), 175–219.
- Keller, A. (2016). *Philosophy of olfactory perception*. New York: Palgrave Macmillan.
- Kubovy, M., & Schutz, M. (2010). Audio-visual objects. *Review of Philosophy and Psychology*, 1, 41–61.
- Kubovy, M., & von Valkenburg, D. (2001). Auditory and visual objects. *Cognition*, 80, 97–126.
- Lycan, W. (2000). The slighting of smell. In S. Rosenfeld (Ed.), *Of minds and molecules* (pp. 273–290). New York: Oxford University Press.
- MacCumhaill, C. (2015). Perceiving immaterial paths. *Philosophy and Phenomenological Research*, 90(3), 687–715.
- Macpherson, F. (2011). Cross-modal experiences. *Proceedings of the Aristotelian Society*, 111, 429–468.
- Martin, M. G. F. (1993). Sense modalities and spatial properties. In N. Eilan, R. McCarthy, & B. Brewer (Eds.), *Spatial representation: Problems in philosophy and psychology* (pp. 206–218). Oxford: Oxford University Press.
- Martin, M. G. F. (1997). The reality of appearances. In M. Sainsbury (Ed.), *Thought and ontology*. Milan: Franco Angeli.
- Matthen, M. (2004). Features, places, and things: Reflections on Austen Clark’s theory of sentience. *Philosophical Psychology*, 17(4), 497–518.
- Matthen, M. (2005). *Seeing, doing, and knowing*. New York: Oxford University Press.
- Matthen, M. (2010). On the diversity of auditory objects. *Review of Philosophy and Psychology*, 1, 63–89.
- Millar, B. (2019). Smelling objects. *Synthese*, 196, 4279–4303.
- Mole, C. (2010). The content of olfactory experience. *Journal of Consciousness Studies*, 17(11–12), 173–179.
- Nudds, M. (2014). Auditory appearances. *Ratio*, 27, 462–482.
- O’Callaghan, C. (2008). Object perception: Vision and audition. *Philosophy Compass*, 3(4), 803–829.
- O’Callaghan, C. (2011). Hearing properties, effects, or parts. *Proceedings of the Aristotelian Society*, 111, 375–405.
- O’Callaghan, C. (2012). Perception and multimodality. In E. Margolis, R. Samuels, & S. Stich (Eds.), *The Oxford handbook of philosophy of cognitive science*. New York: Oxford University Press.

- O'Callaghan, C. (2014). Intermodal binding awareness. In D. J. Bennett & C. Hill (Eds.), *Sensory integration and the unity of consciousness*. Cambridge, MA: MIT Press.
- O'Callaghan, C. (2015). The multisensory character of perception. *Journal of Philosophy*, *112*(10), 551–569.
- O'Callaghan, C. (2016). Objects for multisensory perception. *Philosophical Studies*, *173*, 1269–1289.
- O'Callaghan, C. (2018). *Beyond vision: Philosophical essays*. Oxford: Oxford University Press.
- Palmer, S., & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin & Review*, *1*(1), 29–55.
- Pylyshyn, Z. (2007). *Things and places*. Cambridge, MA: MIT Press.
- Richardson, L. (2010). Seeing empty space. *European Journal of Philosophy*, *18*(2), 227–243.
- Richardson, L. (2014). Non sense-specific perception and the distinction between the senses. *Res Philosophica*, *91*(2), 215–239.
- Sanabria, D., Soto-Faraco, S., Chan, J. S., & Spence, C. (2005). Intramodal perceptual grouping modulates multisensory integration: Evidence from the crossmodal dynamic capture task. *Neuroscience Letters*, *377*, 59–64.
- Siegel, S. (2002). Review of a theory of sentience, by Austen Clark. *Philosophical Review*, *111*(1).
- Skrzypulec, B. (2019). The nonclassical mereology of olfactory experiences. *Synthese*. <https://doi.org/10.1007/s11229-018-02072-x>.
- Skrzypulec, B. (2020). Visual endurance and auditory perdurance. *Erkenntnis*, *85*, 467–488.
- Spence, C., & Bayne, T. (2014). Is consciousness multisensory? In D. Stokes, M. Matthen, & S. Biggs (Eds.), *Perception and its modalities* (pp. 95–132). New York: Oxford University Press.
- Stevenson, R. J. (2014). Object concepts in the chemical senses. *Cognitive Science*, *38*, 1360–1383.
- Treisman, A. (1996). The binding problem. *Current Opinion Neurobiology*, *6*, 171–178.
- Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the 'unity assumption' using audiovisual speech stimuli. *Perception and Psychophysics*, *69*(5), 744–756.
- Wong, H. W. (2015). The body schema as a condition of possibility for action. In Y. Coello & M. H. Fischer (Eds.), *Perceptual and emotional embodiment* (pp. 135–148). London: Routledge.
- Vernazzani, A. (forth.). Do we see facts? *Mind & language*.
- Young, B. D. (2016). Smelling matter. *Philosophical Psychology*, *29*(4), 520–534.
- Young, B. D. (2020). Perceiving smellscapes. *Pacific Philosophical Quarterly*, *101*(2), 203–223.
- Zmigrod, S., Spapé, M., & Hommel, B. (2009). Intermodal event files: integrating features across vision, audition, tactation, and action. *Psychological Research*, *73*, 674–684.

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