Accounting for Action:
Challenging the Traditional View of Multimodal Perceptual Objects

Aleksandra Mroczko-Wąsowicz (aleksandra.mroczko-wasowicz@uw.edu.pl)
University of Warsaw, Faculty of Philosophy, Warsaw, Poland

Spencer Ivy (s.ivy@uw.edu.pl)
University of Warsaw, Faculty of Philosophy, Warsaw, Poland

Michał Bachanek (mj.bachanek@uw.edu.pl)
University of Warsaw, Faculty of Philosophy, Warsaw, Poland

Aleksandra Cząstkiewicz (aleksandra.czastkiewicz@mail.umcs.pl)
Maria Curie-Skłodowska University, Institute of Philosophy, Lublin, Poland

Abstract

In this paper, we argue that action is involved in the creation and representation of perceptual objects. We introduce leading philosophical theories regarding the structure of perceptual objects in modality-independent and multisensory settings. These accounts omit action as a causal factor that can facilitate feature binding and serve as a structural component of perceptual objects. We argue that action does play this causal role due to the connections between the brain's motor system and perceptual processing as evidenced by neurophysiological and behavioral studies. These data include research on view-independent representations, peripersonal space, and event file coding. We conclude that to omit the influence of the motor system on the structure of perceptual objects is to have an incomplete account of object perception. Motor action is often required to drive the integration of sensory features into corresponding perceptual objects.

Keywords: Action; Multisensory Perception; Object Perception; Motor System; Feature Binding; Event Files

I. Introduction

Perception informs us about the features of objects in our environment and, notably, represents potential actions on distal objects (Vernazzani, Skrzypulec & Schlicht, 2021; Mroczko-Wąsowicz & Grush, 2023). Accordingly, in this paper, we explore the following questions: Why is action so closely entwined with perception; and, how does this relationship influence the structure of perceptual objects? We answer these questions by surveying a wide range of empirical evidence regarding the relationship between motor control and perceptual processing. We argue that there is a significant involvement of action in the creation and constitution of perceptual objects. This is because a basic function of perception is enabling appropriate movements with respect to environmental objects. In Section II we introduce leading philosophical theories of perceptual object ontology that focus on the structure of perceptual objects in modality-independent and multisensory settings. Section III emphasizes a significant omission in the current state of the art in specifying conditions for feature binding and perceptual object formation. Namely, the motoric aspect of perception remains undiscussed. Subsequent to this, we present evidence supporting our view that there is a strong connection between the brain's motor system and perceptual processing based on neurophysiological studies. Finally, section IV surveys empirical research on the essential role of action in multisensory integration including studies on view-independent representations, peripersonal space, and event coding. We conclude that to omit the influence of the motor system on the structure of perceptual objects is to have an incomplete account of object perception. Furthermore, we suggest how future research may focus on reforming our understanding of object perception with the influence of action in mind.

II. Leading philosophical accounts of perceptual objects

Current philosophical theories of perceptual objects do not take into account the integral role of the motor system in the process of perceptual object formation. To demonstrate this, we discuss three leading views of the structure of perceptual objects: Casey O'Callaghan's *structured mereologically complex individuals*, Jonathan Cohen's *mereological co-constituency*, and EJ Green's *structural unity schema*. Each of these views is a comprehensive account of perceptual objects' ontology. More specifically, these accounts focus on the rules that link multimodal sensory properties or parts of perceptual objects (such as Gestalt or body principles) into coherent wholes that we perceive.

According to Casey O'Callaghan (2016), perceptual objects can be described as "structured mereologically complex individuals". They are individuals in a sense that they possess certain perceptual features, such as color, or pitch. They are mereologically complex because they are composed of parts, which can be either perceived individually or as parts of a bigger whole. Some of these parts may be accessible only through one sense modality. The structure of perceptual objects depends on specific spatial or temporal relations between their parts for a particular sense modality. Such objects are used by the perceptual system to constrain perceptual attention, to allow for recognition of individual perceptual items and reidentifying them over time

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society*.

as well as to provide items that can be used for demonstrative thought.

Along the same vein, Jonathan Cohen (2023) proposes “the mereological co-constituency view” to explain the structure of multimodal perceptual objects and the process of multisensory feature binding. Within this framework, Cohen argues that the structure of multimodal perceptual objects is contingent on two other accounts. The first of them is the “convergence account” according to which the attribution of sensory features to a single entity leads to the creation of a “mereological complex.” In other words, perceptual objects are complex entities that remain in part-whole relations with their aspects (as described above based on O’Callaghan, 2016). The second account constitutive of Cohen's view is the “association account” according to which unisensory features are represented jointly when they are associated with one another. This means that if two features co-occur in the same spatiotemporal conditions, they are more easily bound together insofar as they are “predicted” by our perceptual system to be one entity/event (Fulkerson, 2011).

Cohen blends the foregoing accounts by discussing two kinds of “sensory individuals” as perceptual objects that obey the laws of association and convergence. Sensory individuals may be bearers of unimodal features and also multimodal complex entities. The relationship between complex multimodal entities and their unimodal features takes the form of part-whole co-constituency. This means that multimodal objects are complexes constituted by unimodal objects/individuals that are directly accessible to our perceptual system. Cohen posits that we do not attribute features convergently to these multimodal mereological complexes (as it is within O’Callaghan’s view according to Cohen, 2023). Instead, modality-specific individuals instantiating unisensory features (as in the association account) are co-constituents of such multisensory complexes. For example, given the case of perceiving a car crash, we have an event with visible and audible parts/aspects (which we can call a mereological complex). Such aspects are treated as exemplifications of unimodal features (e.g., color in case of vision) by relevant unimodal individuals (e.g., material object that we see) forming coordinated representations within the same complex (hierarchically structured) event. Consequently, the individual auditory and visual aspects of the crash are bound together as co-constituents of the complex multimodal event as a whole, yet may also be perceived as individuals in their own right.

Finally, E.J. Green’s (2019) account “the structural unity schema” claims that perceptual objects are sets of features that are causally connected. Since perception parses the environment and singles out entities by picking up certain perceptible regularities, perceptual objects are created by grouping perceptual features that share common, noticeable regularities. For example, a group of musical notes may be perceptually grouped into a singular chord when they stand in harmonic relation to each other. These regularities have to be stable over time, which is the case when they are causally sustained. The reason why these stable perceptible regularities are grouped into perceptual objects is that this is the most resource-saving way of processing perceptual inputs. It is simply more efficient to group and process causally stable perceptual patterns together instead of processing them separately. These groups can then be decomposed into simpler perceptual objects (e.g., when a bird disembarks from its flock), as well as be combined into more complex objects (e.g., when two flocks group into a single murmuration) when necessary. As perceptible regularities may be presented through different sensory modalities, this account may be applied both to modality-specific and multimodal perceptual objects.

These accounts conditionalize the ontology of perceptual objects by attempting to form unified theories of objects applicable across sense modalities. We acknowledge the importance of these structural approaches. We think that the conditions they propose are consistent with the notion that the motor system plays an integral role in shaping perceptual objecthood. It is noteworthy that the above-mentioned accounts share the capacity to encompass both object-like individuals and event-like individuals as combined within mereologically complex perceptual objects. However, for all of their virtues, the foregoing philosophical accounts do not directly address the motor system's causal role in creating perceptual objects. In what follows, we survey three distinct empirical perspectives that substantiate the contribution of action and the motor system to perceptual objecthood. As we explore the relationships between the motor and perceptual systems, we illuminate how action is a crucial ontological element missing from the current state of the art in the philosophy concerning multisensory perceptual objects.

III. Accounting for Action

Perception does not exist in isolation from action. We perceive objects in certain environments wherein we directly act and perceive often not for the sake of perceiving itself, but in order to *do* something with the information that we process. While there are many theories that demonstrate the relationships between the motor and perceptual systems (Proffitt, 2008; Creem-Regehr and Kunz, 2010; Clark, 2015), none seem to address how this interaction directly affects the structure of individual perceptual objects themselves. Consequently, we aim to address this gap by demonstrating how the structure of perceptual objects, understood as object-like and event-like sensory individuals, often requires action-oriented components, and how those components change the constitution of their corresponding objects as mereological complexes.

As mentioned above, one striking feature of the foregoing theories of perceptual objects is that the accounts conspicuously fail to address action and the motor system as constitutive elements of perceptual objects’ structural composition. Explanations for this omission could take several forms. It might be thought that action and perception are of two different ontological and neurophysiological kinds. However, there is a large body of new data that puts the foregoing claim into doubt. Recent EEG and fMRI studies have shown strong correlations between the brain's motor system and perceptual processing where the neural components of action are levied to aid in perception (Binder et al., 2004; Zekveld et al., 2006; Wu et al., 2014; Michaelis et al., 2021; Schmalbrock & Frings, 2022). Although it is still poorly understood whether different brain systems are simultaneously involved in the processing of perception-action integration and to what extent the integration modulates activities in these systems, recent work suggests a close link between learning and the perception-action integration, emphasizing that binding between stimulus features and response may be key for event file coding (Eggert et al., 2021; Hommel 2019, 2004). Further, an example of the entanglement of perception and action can be seen in the fact that there exists a shared neural pathway devoted to both the perceptual and motor system.

Perceptual processing across sensory modalities is typically divided into two pathways: ventral and dorsal (*for audition see:* Rauschecker & Tian, 2000, Rauschecker, 2018; *for olfaction see*: Frasnelli et al., 2012; *for vision see*: Ferretti, 2018, 2019; *for touch:* Reed, Klatzky & Halgren, 2005). The functional dissociation between ventral object processing and dorsal spatial processing reflects a division of labor in the brain that acts as a general, modality-independent organizational principle. The dorsal pathway is specifically devoted to guiding action in space (Gallivan & Goodale, 2018; Freud, Behrmann & Snow, 2020). It processes simple features such as 2D orientation, or figural depth in case of vision by transferring information relevant for motor activity (like graspability) in the parietal cortex. This, in turn, affects the localization of objects in space and the actionable identification of objects that can be manipulated. In contrast, the ventral pathway is commonly characterized as responsible for object identification and recognition in non-action-oriented perception. The characteristic functions of the ventral pathway are to: (1) process a more complex analysis of sensory scenes in the temporal cortex; (2) to guide the work of the dorsal stream by attaching meaning to perceptual object representations in order to optimize interaction (e.g. grasping an object) (Almeida, Mahon & Caramazza, 2010; *see also* Ferretti, 2018 *for a review*). In this conceptualization, the dorsal and ventral pathways exhibit the entanglement of perception with action: where perception is processed in the brain, the motor system also does its work. Additionally, the ventral pathway contributes to action guidance, although indirectly.

Nevertheless, it might be supposed that although action and perception go hand in hand, the former does not play a necessary role for the latter. To the contrary, in what follows, we argue that action contributes significantly to the creation and structural composition of perceptual objects. In the next section of this paper, we survey a series of behavioral data regarding the impact that motor control has on perceptual processing. Empirical research on (1) multimodal view-independent object representations, (2) action-influenced multisensory integration within peripersonal space, (3) event coding in multisensory action orientation and audio-visual speech perception, altogether suggest that the motor and perceptual systems are closely intertwined and influence each other. Thus, as we demonstrate through the series of empirical perspectives, perception often does not occur in a vacuum without associated motor activity. Consequently, we conclude that the motor system may play an ontologically constitutive role in the construction of perceptual objects and that contemporary accounts regarding the structural composition of perceptual objects ought to take this fact into account.

IV. Empirical evidence for cooperation between motor system and perception

In this section of the paper, we survey three separate sets of empirical studies on the connections between the motor and perceptual systems. Each of these sets of data aim at demonstrating from a different perspective how the motor system plays an integral and structural role in generating perceptual objects. The first set addresses "view independence", a form of perspective-taking that often requires action in order to successfully aid in perceptual object recognition. The second set of empirical work regards "peripersonal space" which is the area surrounding one's body. The final set delves into a popular psychological theory called the Theory of Event Coding (TEC). Based on this evidence, we conclude that action does, in fact, play an important role in structuring perceptual objects.

i. View independent representations: Action enables multimodality

A perceptual object representation is "view-independent" when we are aware of how we might perceive the object from different angles. This awareness enables us to recognize perceptual objects without being dependent on any single perspective of the object. View independence is crucial for perceptual objects derived from any modalities whose structure is mostly dependent on their spatial properties, such as vision or touch. What is more, empirical research suggests that view-independence is a link that allows for creating a representation shared between vision and touch. Once a perceptual object becomes view-independent in one modality, it can be used for recognition in another one (*see* Lacey & Sathian, 2023 *for a review*).

In a study by Lacey et al. (2009) participants were presented with objects that at first could only be recognized in one modality based on a specific viewpoint. However, when subjects familiarized themselves with the objects in this modality, they learned to recognize the objects regardless of their orientation. Once the familiarity stage had been achieved, the same objects could be recognized in a view-independent way in the other modality as well. This effect occurs symmetrically, both for visual-haptic and haptic-visual crossmodal recognition (*ibid.*). What is more, object-perception in both vision and touch seems to have the same neural substrate. fMRI studies show activation in lateral occipital complex (LOC) both in vision and tactual object perception and the lesion of this area results both in visual and haptic agnosia (Lacey & Sathian, 2014).]

One component that is crucial for developing this kind of view-independent perceptual object representation is action. In everyday contexts, in order to perceive certain objects from different perspectives we change our positioning towards them. We move around objects and manipulate them by changing their position to perceive them more easily from different angles. One may object to this argument by stating that sometimes certain perceptual objects change their positioning regardless of our actions, and this can be enough for creating a view-independent representation. This counterargument works against the claim that action is necessary for creating view-independent object representations, but it does not undermine the claim that action facilitates creating such representations in many (if not most) cases. Even so, the necessity of action for creating such object representations may be defended based on research concerning sensory substitution studies.

Now famous research by Bach-y-Rita (1984) presented a case in which congenitally blind people used a device to translate camera inputs into a tactile stimulation on their skin. After some training, participants started to experience objects presented to a camera not as a pattern of tactile stimuli, but as vision-like objects, based on which they could make judgements about the objects' shape and distance. What is crucial is that these vision-like objects started to emerge only when the participants had control over the camera by moving it and controlling its zoom or focus. Accordingly, an action-involving exploration of view independence was necessary for creating accurate perceptual object representations. The idea that active exploration is needed for perceiving a sensory substitution device's input as an external object has also been corroborated by more recent findings in this area (Kristjánsson et al., 2016; Hartcher-O'Brien & Auvray, 2014). These findings suggest that action does not only facilitate creating perceptual object representations that are view-independent, but seems to be necessary for creating them. Taking into account that perceptual objects become accessible to different sense modalities, once objects achieve view-independence, we can claim that, in certain instances, action facilitates the combination of unimodal perceptual objects into a single, complex multimodal object. In that sense, action seems crucial for creating multimodal perceptual object representations.

ii. Peripersonal space: Motor interactions facilitate multisensory integration

The relation between motor performance and peripersonal space is relevant for investigating how action facilitates multisensory integration and the formation of multimodal object representations. In particular, it has been studied how the brain integrates multisensory information and motor processing to build multisensory representations of actions and of the space nearby (Brozzoli, Ehrsson & Farnèr, 2014; Di Pellegrino & Làdavas, 2015). Peripersonal space (PPS) is the space directly surrounding the agent's body (typically, up to about 50cm). It is represented by multisensory neurons from premotor and parietal areas which integrate tactile stimuli from the agent's body with visual or auditory stimuli presented within a limited distance from them (de Vignemont et al., 2021).

The evidence demonstrates that PPS may expand to include the external objects that one is acting upon or the space within which these actions are performed. The literature highlights the impact of action performance on the emergence of multisensory integration within PPS. For instance, Noel et al. (2015) demonstrated that hand movements and walking have been shown to reshape PPS and influence multisensory integration processes in the space surrounding the torso. Researchers reported that walking extends the effects of audio-tactile integration of stimuli perceived in the direction of movement. When participants were standing still, sounds boosted tactile processing within 65-100 cm from the participants. However, when participants were walking, peripersonal space was extended with boosted tactile processing at ~1.66 m (Noel et al., 2015). These results show that peripersonal space constitutes a dynamic sensory-motor interface between the moving subject and their environment. Action planning and motor execution may trigger a dynamic reorganization of the peripersonal space around a distal object when the object becomes the target of subjects' movement (Brozzoli et al., 2010; Berger, Neumann & Gail, 2019; Patané et al., 2018).

PPS, as a perception-to-action interface, provides a spatial framework for avoiding or approaching body-objects interactions (Brozzoli et al., 2011). Evidence shows that fronto-parietal regions respond to bodily actions in peripersonal space related to both protective avoidance behaviors and goal-directed approaching (Graziano & Cooke, 2006; Clery et al., 2015; Rizzolatti et al., 1997). This has led to the proposal of a dual model of PPS suggesting that the two functions of PPS - bodily protection and goal-directed action - require distinct sensory and motor processes that obey different principles (de Vignemont & Iannetti, 2015).

Overall, the foregoing research demonstrates that motor actions (both planning and executing them) within PPS lead to highly integrated multisensory representations of that space. These representations include external objects as well as motor responses toward those objects. This provides an additional source of evidence supporting the close relationship between action and perception. As we demonstrate in the next section, this relationship offers good reason to think of perceptual objects as mereologically complex, event-like individuals. To this end, we show how action-oriented perceptual objects are mereologically complex since they integrate different sense modalities, and event-like since they also include the representations of motoric information.

iii. Event coding: Action is constitutive of multimodal perception

According to the Theory of Event Coding (TEC), there is no distinction between the motor system and the perceptual system and the two function as parts of an identical cognitive operation to interact with the world through perceptual input. As Hommel, (2019) defines the theory, “perceptions and action goals are coded in the same way through distributed feature codes which refer to the distal features of the represented event.” In other words, the cognitive operations involved with perception are also the same operations that are used to produce action control. As a result, in TEC the typical ‘object file’ concept of a perceptual object is extended into an ‘event file’ to include action-oriented information.

Within an event file, behavioral and environmental data regarding how to interact with the environment are encoded to perceptual stimuli. The ‘feature codes’ refer to both the representational and neural underpinnings of event files (Wendiggensen, Prochnow & Pscherer, 2023). Evidence from fMRI and EEG studies supports the hypothesis that areas of the brain associated with motor control are simultaneously operative when processing related perceptual stimuli (*see* Hommel, 2022 *for a review of this work*). Moreover, in some behavioral studies reported below, event files have been created and reinforced under lab conditions to elicit consistent and measurable effects which form the backbone of TEC’s core predictions (Hommel, 2004, 2009; Zmigrod, Spapé & Hommel, 2009; Janczyk et al., 2023).

The way that perceptual features are linked to an action control is a well-studied phenomenon (*see* Janczyk et al., 2023 *for a comprehensive overview and replication of the last 20 years of TEC research*). Perceptual stimuli (S) are encoded to behavioral responses (R) such that they combine together as a Stimulus-Response pair (S-R). The S and R components of an S-R pair can then reciprocally trigger one another when associated perceptual stimuli or environmental conditions are present (Spence, 2018). Importantly, it is not merely that perceptual stimuli provoke behavioral responses. The relationship between contents of the S-R pair is symmetric: “[…] hearing often benefits from orienting one’s body or head towards stimulus sources and tactile perception would be virtually impossible without systematically moving one’s effectors across the to-be-perceived surface. This means that perception is just as well the consequence of action than it is its cause.” (Hommel, 2009).

Accordingly, some of the most prevalently measured effects of S-R event files over the last three decades are: (1) response-effect compatibility (Kunde, 2001) which shows that similarity between presented stimuli facilitates faster response time for encoded S-R actions; (2) response-effect learning (Elsner & Hommel, 2001) which demonstrates how stimuli and their consequent effects are learned as bidirectional connections which facilitate goal-directed action; and (3) action-induced blindness (Müsseler & Hommel, 1997) which demonstrates how S-R pairs can worsen/limit the perception of stimuli not associated with encoded actions that are already underway.

For example, in unimodal settings, Wamain et al. (2019) explored whether perceiving a change in visual features as a direct outcome of one’s voluntary motor actions would impact the resulting integration of those features into accurate perceptual object representations used to inform judgment. The participants were required to assess the temporal order of alterations in both the color and position of a visual stimulus. This is because it was found that subjects may allocate more attentional resources to stimuli that they control. Such action-influenced sensory binding happens because subjects expect results from their own motor action which, in turn, helps them to integrate visual features and process changes in a more effective manner (Corveleyn, López-Moliner & Coello, 2015).

Research on response-effect compatibility connected to S-R responses indicates that multisensory cues significantly impact speed and accuracy with perceived feature combinations stored in temporary event files for efficient, later retrieval (Zmigrod et al., 2009). However, partial replication of feature combinations, such as encountering a purple square after processing a yellow square, may negatively influence performance due to spontaneous feature-integration conflicts. This phenomenon was observed also in multimodal (visual-auditory and audio-tactual) conditions, causing a delay in response when encountering partially incompatible features. Similar patterns of biased performance were noted for repeated stimuli and repeated responses. The study demonstrated that integrating stimulus features with actionable responses into event-files or S-R pairs affects both the timing and accuracy of responses.

Accordingly, as we have suggested so far, the organization and representation of multisensory perceptual objects is often intimately linked to the brain’s motor system. Further evidence bears out this conclusion with regard to audio-visual perception in speech. Specifically, the suppression of the μ/beta (8-30 Hz) frequency in left-lateralized dorsal streams of the brain has been shown through EEG to indicate motor activity in cases of decision making (Binder et al., 2004; Zekveld et al., 2006; Callan et al., 2010; D’Ausilio et al., 2012), anticipation (Balser et al., 2014; Denis et al., 2017), and significantly, in multisensory speech perception (Michaelis et al., 2021). In short, μ/beta suppression is strongly correlated with motor activity and this suppression was exactly what Michaelis et al. (2021) found in their research on audio-visual speech perception.

The researchers used EEG to measure the μ/beta frequency in subjects who engaged in a study which manipulated their ability to perceive different types of audible cues. The experiment included four different conditions containing unique types of audible stimuli. Subjects would hear either a regular English word (Auditory Condition), a syllabic phoneme related to the English language (Syllabic Condition), some environmental sounds including animal, traffic, office, and other generic, non-semantic noises (Environmental Condition), or words in conjunction with a visual representation of a speaker saying the word (Audiovisual Condition). Presented stimuli were made more or less difficult by their volume paired to the volume of background ‘pink’ noise (think TV static, but more soothing to the human ear) and a pre-stimulus cue. After the presentation window of the actual cue ended, subjects were asked to report which selection from a list of words/sounds was what they had heard.

As the researchers expected, subjects exhibited a strongly left-lateralized suppression of μ/beta in case of correct responses to the Syllabic and Audiovisual Conditions (no significant suppression was found for the Environmental Condition). Further, the magnitude of this effect depended upon the presence of visual stimuli indicating a significant relationship between multisensory speech perception and motor activity. Thus, the presence of visual stimuli increased the presence of motor activity. Interestingly, there was very little μ/beta suppression in easy Auditory Condition trials. This is likely because unimodal speech cues are largely processed through the ventral stream, as opposed to the motor-related dorsal stream (Milner & Goodale, 2006; McIntosh et al., 2009). However, to the surprise of the researchers, as the difficulty increased in Auditory Condition trials, suppression of μ/beta followed. In other words, incorrect perception of difficult auditory-only speech cues was significantly correlated to an increase in motor activity.

We may draw two important conclusions from these foregoing data. First, as Michaelis et al. (2021) remark, "these findings suggest that the motor system is flexibly engaged to aid perception depending on the nature of the speech stimuli." In other words, motor activity aids in the perception and processing of multisensory speech objects. The second conclusion to be drawn from these data regards the more surprising result that increased motor activity was found to correlate with incorrect perception of auditory-only trials. If the motor system flexibly engages during multisensory speech perception, then it seems to have flexed itself in these more difficult trials as well: but why? After all, as mentioned above, there was little μ/beta suppression in the easy trials of the Auditory Condition. Yet, when speech perception is difficult and unimodal (auditory only), these findings suggest that the motor system engaged itself to help overcome the difficulties. The foregoing discussion of S-R pairs and TEC helps to yield an interesting answer to this question.

The motor system, which was most strongly activated by audiovisual (thus multisensory) speech perception, engaged in auditory-only speech perception when the difficult auditory cues were of very low quality. One way to explain this effect is to recall that the motor system is intimately connected to the processes that produce multisensory objects of audiovisual speech. In other words, there is an encoded integration of motor information in multisensory speech perception in an S-R system. When given multisensory cues, sensorimotor interactions fire up in the brain to process the cues. However, when auditory data is incomplete or 'noisy' (as in phoneme-only and challenging trials), the incomplete stimulus is processed through the same sensory-motor network because a part of the S-R pathway is satisfied by the incomplete stimulus. Consequently, integration is attempted because it's the best that the perceptual system can do with incomplete, non-environmental auditory cues - send them through the closest pre-programmed S-R pathway and hope for the best. This suggests that the motor system not only plays an important role in the construction of perceptual speech objects when perceptual information is multimodal (as evidenced by the Audiovisual Condition), but also when speech information is not clearly unimodal (as evidenced by difficult trials of the Auditory Condition).

To sum up the foregoing discussion, the empirically informed philosophical views on the structure of perceptual objects (*see* Section I) all claim that perceptual objects are created by combining different perceptual features within one complex object-representation. TEC expands upon this claim to include motor information. Mental representations are formed when we are both perceiving and acting; they include not only perceptual features but may also possess action-oriented features. Because of this, an accurate account of perceptual objects should include the fact that they can have distinctly action-oriented aspects. Even if we reject TEC's notion that action representations and perceptual representations are one and the same thing the evidence still suggests that they are often strongly interconnected.

V. Conclusion

Our exploration of the relationship between motor control and perceptual processing reveals evidence that supports the notion that action plays a pivotal role in the formation of perceptual objects. The conclusion to be drawn from the empirical findings we have discussed above can be summarized as follows: Motor action is often required to facilitate the integration of perceptual features into corresponding perceptual objects. This is the case both for unisensory and multisensory perceptual objects. Because of this motoric context in which these object representations are created, perceptual objects either also include features that are action-oriented, or are closely intertwined with action-oriented representations. This makes the content of perceptual objects broader than conventionally suggested in the relevant philosophical literature. By bridging gaps in the current empirical and philosophical understanding of object perception, we have highlighted a crucial link between the motor system and the structure of perceptual objects. Future research should focus on investigating the impact of these insights regarding the connections between the motor and perceptual systems on the ontology of perceptual objects.

Acknowledgements

This work was supported by the National Science Centre, Poland (AMW's grant 2019/35/B/HS1/04386).

References

Almeida, J., Mahon, B. Z., & Caramazza, A. (2010). The Role of the Dorsal Visual Processing Stream in Tool Identification. *Psychological Science*, *21(6)*, 772-778.

Bach-y-Rita, P. (1984). The relationship between motor processes and cognition in tactile vision substitution. In W. Prinz, A.F. Sanders (Eds.), *Cognition and motor processes* (pp. 149-160). Berlin, Heidelberg: Springer.

Balser, N., Lorey, B., Pilgramm, S., Naumann, T., Kindermann, S., Stark, R., Zentgraf, K., Williams, A.M., Munzert, J. (2014). The influence of expertise on brain activation of the action observation network during anticipation of tennis and volleyball serves. *Frontiers in human neuroscience*.;*8*:568.

Berger, M., Neumann, P., & Gail, A. (2019). Peri-hand space expands beyond reach in the context of walk-and-reach movements. *Scientific Reports*, 9(1), 3013.

Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature neuroscience*, *7*(3), 295-301.

Brozzoli, C., Cardinali, L., Pavani, F., & Farnè, A. (2010). Action-specific remapping of peripersonal space. *Neuropsychologia, 48(3),* 796-802.

Brozzoli, C., Ehrsson, H. H., & Farnè, A. (2014). Multisensory representation of the space near the hand: from perception to action and interindividual interactions. *The Neuroscientist*, *20*(2), 122-135.

Brozzoli, C., Makin, T. R., Cardinali, L., Holmes, N. P., & Farnè, A. (2011). Peripersonal space: a multisensory interface for body-object interactions. In Murray, M.M., Wallace, M.T., (Eds). *The Neural Bases of Multisensory Processes*. Taylor & Francis: London.

Callan, D., Callan, A., Gamez, M., Sato, M. A., & Kawato, M. (2010). Premotor cortex mediates perceptual performance. *Neuroimage*, *51*(2), 844-858.

Clark, A. (2015). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*, Oxford University Press, UK.

Cléry, J., Guipponi, O., Odouard, S., Wardak, C., & Hamed, S. B. (2015). Impact prediction by looming visual stimuli enhances tactile detection. *Journal of Neuroscience*, *35*(10), 4179-4189.

Cohen, J. (2023). Multimodal binding as mereological co-constituency. In A. Mroczko-Wąsowicz & R. Grush (Eds.), *Sensory Individuals: Unimodal and Multimodal Perspectives*, Oxford, UK: Oxford University Press.

Corveleyn, X., López-Moliner, J., & Coello, Y. (2015). Sensorimotor adaptation modifies action effects on sensory binding. *Attention, Perception, & Psychophysics*, *77*(2), 626–637.

Creem-Regehr S.H., Kunz, B.R. (2010). Perception and action. *Wiley Interdiscip Rev Cogn Sci.*, 1(6), 800–810.

D'Ausilio, A., Bufalari, I., Salmas, P., & Fadiga, L. (2012). The role of the motor system in discriminating normal and degraded speech sounds. *Cortex*, *48*(7), 882-887.

Denis, D., Rowe, R., Williams, A. M., & Milne, E. (2017). The role of cortical sensorimotor oscillations in action anticipation. *NeuroImage, 146*, 1102-1114.

Di Pellegrino, G., & Làdavas, E. (2015). Peripersonal space in the brain. *Neuropsychologia*, *66*, 126-133.

de Vignemont, F., & Iannetti, G. D. (2015). How many peripersonal spaces? *Neuropsychologia*, 70, 327-334.

de Vignemont, F., Serino, A., Wong, HY, Farnè, A. (2021). Peripersonal space: A special way of representing space. In F. de Vignemont, A. Serino, HY Wong, A. Farnè (Eds.), *The World at Our Fingertips: A Multidisciplinary Exploration of Peripersonal Space*, Oxford, UK: Oxford University Press.

Eggert E, Bluschke A, Takacs A, Kleimaker M, Münchau A, Roessner V, Mückschel M, Beste C. (2021). Perception-action integration is modulated by the catecholaminergic system depending on learning experience. *International Journal of Neuropsychopharmacology, 24*(7), 592-600.

Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology: Human Perception and Performance, 27*(1), 229–240.

Ferretti, G. (2018). Two visual systems in Molyneux subjects. *Phenomenology and the Cognitive Sciences, 17,* 643–679.

Ferretti, G. (2019). The Neural Dynamics of Seeing-In. *Erkenntnis 84*(6):1285-1324.

Frasnelli, J., Lundström, J. N., Schöpf, V., Negoias, S., Hummel, T., & Lepore, F. (2012). Dual processing streams in chemosensory perception. *Frontiers in human neuroscience*, *6*, 288.

Freud, E., Behrmann, M., & Snow, J. C. (2020). What Does Dorsal Cortex Contribute to Perception?. *Open mind : discoveries in cognitive science*, *4*, 40–56.

Fulkerson, M. (2011). The unity of haptic touch. *Philosophical Psychology*, *24*(4), 493–516.

Gallivan, J. P., & Goodale, M. A. (2018). The dorsal "action" pathway. *Handbook of clinical neurology*, 449–466.

Graziano, M. S., & Cooke, D. F. (2006). Parieto-frontal interactions, personal space, and defensive behavior. *Neuropsychologia*, *44*(13), 2621-2635.

Green, E. J. (2019). A theory of perceptual objects. *Philosophy and Phenomenological Research*, *99*(3), 663-693.

Hartcher-O'Brien, J., & Auvray, M. (2014). The process of distal attribution illuminated through studies of sensory substitution. *Multisensory Research*, *27*(5-6), 421-441.

Hommel, B. (2004). Event files: feature binding in and across perception and action. *Trends in cognitive sciences, 8*(11), 494–500.

Hommel B. (2009). Action control according to TEC (theory of event coding). *Psychological research*, *73*(4), 512–526.

Hommel, B. (2019). Theory of event coding (TEC) V2.0: representing and controlling perception and action. *Attention, Perception, & Psychophysics*, *81*:2139–2154.

Hommel, B. (2022). The control of event-file management. *Journal of Cognition*, *5*(1):1.

Janczyk, M., Giesen, C.G., Moeller, B., Dignath, D., & Pfister, R. (2023). Perception and action as viewed from the Theory of Event Coding: a multi-lab replication and effect size estimation of common experimental designs. *Psychological Research 87*, 1012–1042.

Kristjánsson, Á., Moldoveanu, A., Jóhannesson, Ó. I., Balan, O., Spagnol, S., Valgeirsdóttir, V. V., & Unnthorsson, R. (2016). Designing sensory-substitution devices: Principles, pitfalls and potential 1. *Restorative neurology and neuroscience*, *34*(5), 769-787.

Kunde, W. (2001). Response-effect compatibility in manual choice reaction tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(2), 387–394.

Lacey, S., Pappas, M., Kreps, A., Lee, K., & Sathian, K. (2009). Perceptual learning of view-independence in visuo-haptic object representations. *Experimental brain research, 198,* 329-337.

Lacey, S., & Sathian, K. (2014). Visuo-haptic multisensory object recognition, categorization, and representation. *Frontiers in Psychology*, *5,* 730.

Lacey, S., & Sathian, K. (2023). Visuo-haptic object processing in the multisensory brain. In A. Mroczko-Wąsowicz and R. Grush (Eds.) *Sensory Individuals: Unimodal and Multimodal Perspectives*, Oxford, UK: Oxford University Press.

McIntosh A.M., Moorhead T.W., McKirdy J., Hall J., Sussmann J.E., Stanfield A.C., Harris J.M., Johnstone E.C., Lawrie S.M. (2009). Prefrontal gyral folding and its cognitive correlates in bipolar disorder and schizophrenia. *Acta Psychiatrica Scandinavica*. 119(3):192-8.

Michaelis, K., Miyakoshi, M., Norato, G., Medvedev, A. V., & Turkeltaub, P. E. (2021). Motor engagement relates to accurate perception of phonemes and audiovisual words, but not auditory words. *Communications biology*, *4*(1), 108.

Milner, A. D., & Goodale, M. A. (2006). *The visual brain in action* (2nd ed.). Oxford University Press.

Mroczko-Wąsowicz, A., and Grush, R. (2023). Introduction: Sensory individuals: Contemporary perspectives on modality-specific and multimodal objecthood. In A. Mroczko-Wąsowicz & R. Grush (Eds.), *Sensory Individuals: Unimodal and Multimodal Perspectives*, Oxford, UK: Oxford University Press.

Müsseler, J., & Hommel, B. (1997). Blindness to response-compatible stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 23*(3), 861–872.

Noel, J. P., Grivaz, P., Marmaroli, P., Lissek, H., Blanke, O., & Serino, A. (2015). Full body action remapping of peripersonal space: the case of walking. *Neuropsychologia*, *70*, 375–384.

O'Callaghan, C. (2016). Objects for multisensory perception. *Philosophical Studies, 173*(5), 1269-1289.

Patané, I., Cardinali, L., Salemme, R., Pavani, F., Farnè, A., Brozzoli, C. (2018). Action Planning Modulates Peripersonal Space. *Journal of Cognitive Neuroscience*, *31*, 1141–1154.

Proffitt, D. R. (2008). An action-specific approach to spatial perception. In R. L. Klatzky, B. MacWhinney, & M. Behrman (Eds.), *Embodiment, ego-space, and action* (pp. 179–202). Psychology Press. New York.

Rauschecker J. P. (2018). Where, When, and How: Are they all sensorimotor? Towards a unified view of the dorsal pathway in vision and audition. *Cortex, 98,* 262–268.

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proceedings of the National Academy of Sciences 97*(22), 11800–11806.

Reed, C. L., Klatzky, R. L., & Halgren, E. (2005). What vs. where in touch: an fMRI study. *NeuroImage, 25*(3), 718–726.

Rizzolatti, G., Fadiga, L., Fogassi, L., & Gallese, V. (1997). The space around us. *Science*, *277*(5323), 190-191.

Schmalbrock, P., & Frings, C. (2022). A mighty tool not only in perception: Figure-ground mechanisms control binding and retrieval alike. *Attention, Perception, & Psychophysics, 84*(7), 2255–2270.

Spence, C. (2023). Multisensory Perception. *In Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, J.T. Wixted (Ed.).

Vernazzani, A., Skrzypulec, B., & Schlicht, T. (2021). Structure of perceptual objects: introduction to the Synthese topical collection. *Synthese*, *199*(1-2), 1819-1830.

Wamain, Y., Corveleyn, X., Ott, L., & Coello, Y. (2019). Does the motor system contribute to the perception of changes in objects visual attributes? The neural dynamics of sensory binding by action. *Neuropsychologia*, 132, 107121.

Wendiggensen, P., Prochnow, A., Pscherer, C. et al. (2023). Interplay between alpha and theta band activity enables management of perception-action representations for goal-directed behavior. *Communications Biology, 6*(1), 494.

Wu, Z. M., Chen, M. L., Wu, X. H., & Li, L. (2014). Interaction between auditory and motor systems in speech perception. *Neuroscience Bulletin, 30,* 490-496.

Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., & Schoonhoven, R. (2006). Top–down and bottom–up processes in speech comprehension. *Neuroimage, 32(4),* 1826-1836.

Zmigrod, S., Spapé, M., & Hommel, B. (2009). Intermodal event files: integrating features across vision, audition, taction, and action. *Psychological research, 73*(5), 674–684.