

Multisensory Integration Workshop Report

By

Kevin Connolly, Aaron Henry, Zoe Jenkin, and Andrew MacGregor

This report highlights and explores five questions that arose from the workshop on Multisensory Integration at the University of Toronto, on May 9th to 10th, 2014:

- 1. What Is Multisensory Integration?*
- 2. Do Multisensory Percepts Involve Emergent Features?*
- 3. What Can Multisensory Processing Tell Us about Multisensory Awareness?*
- 4. Is Language Processing a Special Kind of Multisensory Integration?*
- 5. What Is the Purpose of Multisensory Integration?*

1. What Is Multisensory Integration?

Imagine that you are sucking on a menthol sweet or candy.¹ A menthol candy has a bitter taste, a minty aroma, and a cool sensation. Take any one of those three away, and it is not the flavor of menthol, since flavor experience requires taste, touch, and smell (see Smith, 2012). As menthol and other flavors show, experiences can be multisensory: single, unified perceptual experiences, which are the result of multisensory integration—an integration of taste, smell, tactile, and other sensations. Other examples of multisensory experiences include one raised by Mohan Matthen: experiences sometimes involve conflict between sense modalities. Some such experiences, as when someone gets spun around, and then asked to walk straight while her vision is moving, may even produce a feeling of sickness. In such cases, there is a sensation that indicates to you something is wrong, and that sensation is not just visual or just proprioceptive. Such an experience is a multisensory experience.

Multisensory experiences are the product of a multisensory integration process. What is multisensory integration? According to Matthew Fulkerson, we should not think of it as a natural kind. That is to say, while multisensory integration exists in many different instances, we should not expect to find necessary and sufficient conditions for it. As theorists, we are

¹ Barry Smith outlined this example at the workshop.

interested in different groupings of the senses at different times—for eating a meal, we will be interested in a different group than in playing a basketball game. Given this, Fulkerson suggested that we should embrace *sensory pluralism*: the idea that there are lots of distinct, equally valid ways of dividing up the senses. On this view, there is no natural way of thinking of the senses as doing just one task. Instead, we ought to seek a better understanding of multisensory integration by outlining different ways of categorizing the senses.

Even if we cannot offer necessary and sufficient conditions for multisensory integration, there are still ways in which we can classify these interactions. In his talk, Casey O’Callaghan classified six different types of multisensory awareness. The first grade is *minimally multisensory awareness*, whereby at a given time, a subject has co-conscious perceptual awareness associated with more than one sensory modality. For instance, a subject might be aware of the fan whirring, while at the same time be aware of the light flickering. This is co-conscious awareness associated with audition and vision. Grade two is *coordinated multisensory awareness*, which is a type of multisensory awareness where stimulation in one modality influences experience in another. For example, in the common case of ventriloquism, seeing the movement of the ventriloquist dummy’s mouth changes your experience of the auditory location of the vocals. Vision influences your experience of auditory location. The third grade is *Intermodal feature binding awareness*, which occurs when you consciously perceive multiple features from more than one sense modality jointly to belong to the same object or event. For example, if you are listening to live jazz and the drummer begins a solo, you might see the cymbal jolt and hear the clang, and be aware that the jolt and the clang are part of the same event. The fourth grade is *awareness of novel feature instances*, whereby one perceives feature instances that are accessible only multimodally. O’Callaghan used the example of baseball umpires, who determine whether the runner is out by watching the runner’s foot strike the base while listening for the sound of

the baseball hitting the fielder's glove. This is an example where they multisensorily perceive a temporal relation, order, or interval, which would be inaccessible unimodally. Grade five is *multisensory awareness of novel feature types*. For example, flavor is an emergent feature of a type that can't be experienced unimodally (as in the menthol example). The sixth and final grade of multisensory awareness is *novel awareness in a sense modality*. Experiences might be associated with only one modality, but at the same time, not be possible without an experience in another modality. An example of this is cross-modal completion, a version of amodal completion that is multimodal. For example, you might hear an event that has visible features that you don't see. If these features affect your experience of its audible aspects, this would be a case of cross-modal completion.

There is a question as to how O'Callaghan's account relates to Fulkerson's. Fulkerson's focus is on perceptual processing, while O'Callaghan's focus is on perceptual awareness (see question three of this report for a more detailed discussion of this issue). On its face, however, O'Callaghan's account of multisensory integration allows us accept Fulkerson's point that multisensory integration is not a natural kind, while still allowing us to have a substantial account of multisensory integration. By providing an account of different grades of multisensory awareness, we can have an informative account of multisensory integration in lieu of providing the necessary and sufficient conditions for multisensory integration.

References:

Smith, Barry C. (2013). "Taste, Philosophical Perspectives." In Pashler, Harold E. (Ed.)

Encyclopedia of the mind. Thousand Oaks, Calif: SAGE Publications, Inc.

2. Do Multisensory Percepts Involve Emergent Features?

At the workshop, Matt Fulkerson, Barry Smith and Casey O'Callaghan all raised the possibility of genuinely emergent features in multimodal awareness, distinctively multisensory percepts that are not reducible to the respective senses' contributions. It was clear from discussion that the notion of emergence here was a strong one, and would not be satisfied by some of the forms of multisensory integration that were proposed. It would not be sufficient for multisensory emergence that, for example, the contributions of multiple senses are represented as coinstantiated (whether through feature-binding and/or association - see question 3 below), or that the inputs of one sense affect how we represent with another, as in the ventriloquist illusion or McGurk effect.

The awareness of emergent features would instead involve the representation of novel feature types not accessible to any one of the contributing senses alone. In asking whether there might be such emergent features, the workshop speakers were questioning whether the content and character of multisensory experience must be reducible to the contributions of the respective senses, or whether the interaction between them might generate some novel, irreducibly multisensory content.

Several speakers cited flavor as an example: here, the percept appears to be something unitary and not reducible to a mere conjunction of the respective contributions of taste and smell. Smith suggested that other examples might include the perception of balance and self-motion where, as described by Jennifer Campos, vision, proprioception, and the vestibular system work in concert. Whether we should allow that there is any genuinely emergent percept here will depend on whether we can identify a novel feature type that cannot be accessed by any of the relevant senses alone. It is not obvious what that would be in this case: for example, one might think that, in its contribution to self-motion perception, proprioception gives us awareness of something--body position--that is in principle accessible to vision, even if less efficiently or accurately. (Smith also offered speech perception as

possibly involving emergence, although any emergence here arises substantially out of diverse features accessible to a *single* mode, i.e. the various auditory objects one might think are accessible in speech perception--sounds, a voice, words, words with meaning, etc.)

Even if we take flavor as the least controversial instance of a novel and emergent feature type, there remains a question about how we might go about demonstrating that it is genuinely emergent. The claim for emergence is based, at least in part, on intuitions about the phenomenal character of flavor experiences and the phenomenal character of taste, smell and touch experiences, each taken in isolation. The intuition is that the mere co-instantiation of the latter is not sufficient for the former. But this relies heavily on our capacity to imaginatively reconstruct, for example, what it's like to experience menthol from what it's like separately to taste bitterness, smell mint and feel coolness on the tongue (see question 1). The reliability of our imagination in this respect is easily questioned.

The challenge here is compounded by the frequent difficulty of subjectively distinguishing the relative contributions of the various senses, sometimes because we are so accustomed to experiencing them together. As Matt Fulkerson pointed out, some senses are especially hard to disentangle phenomenologically (as with taste and smell, or touch and kinesthesia), while some ostensibly unitary senses (vision for example) might actually involve the interaction among several sub-systems that respond to different features of the world.

We might therefore look instead for some empirical evidence, perhaps some measurable differences in subjects' behaviors or powers of discrimination when presented with, say, a flavorful object to various combinations of the relevant senses. In this vein, relying on data from an array of studies on how information from the sensory modalities is fed into higher-level conceptual systems, Auvray and Spence (2008) argue that flavor should be considered a separate perceptual system due to its unique functional interaction with cognition. However,

empirical evidence for functional unity of the system still leaves open whether the experiential properties themselves are emergent, or remain specific to the olfactory and gustatory modalities.

References:

Auvray, M., & Spence, C. (2008). "The multisensory perception of flavour." *Consciousness & Cognition*, 17, 1016-1031.

3. What Can Multisensory Processing Tell Us about Multisensory Awareness?

Casey O'Callaghan emphasized that claims about multisensory processing don't translate directly into claims about multisensory awareness. For example, even if the perceptual processing in distinct sense modalities exhibits a high degree of cross-modal coordination and interaction, it may still be that the subject's perceptual awareness is merely a collection of modality-specific, albeit highly coordinated, experiences. Further, implicit measures of multisensory processing sometimes conflict with measures of multisensory awareness. In a study by Mitroff, Scholl, and Wynn (2005), subjects were shown an ambiguous stimulus of two objects travelling diagonally from opposite corners of a display and which can be perceived either to bounce or to stream through each other when they meet at the centre of the display (the "bouncing/streaming display"). They investigated whether implicit and explicit measures of the resulting percept agreed or disagreed. They found that their implicit measure (provided by measuring object-specific preview benefits) was strongly correlated with the bouncing percept, while their explicit measure (provided by subjects' conscious reports) was strongly correlated with the streaming percept.² Similar dissociations between implicit and explicit measures have been found for intermodal illusions as well (see

² An object-specific preview benefit occurs when information on an object (e.g., a letter) is recognized more quickly because it reappears on the same object than on a different object.

Zmigrod and Hommell 2011). Thus claims about perceptual processing don't translate uncontroversially into claims about perceptual awareness.

Even once we recognize the above points, there remains the possibility that the concepts used at one level of description can affect and inform our understanding of phenomena at another level. Consider the concept of feature binding. As psychologists use the term, "feature binding" refers to the sub-personal mechanism that binds distinct feature representations into the representation of an individual with multiple attributes or parts. Feature binding is "intermodal" when the feature representations being bound belong to distinct sense modalities, as when an auditory feature is bound with a visual feature. In his talk, O'Callaghan also used "feature binding" to characterize a type of perceptual awareness: the awareness of multiple properties as being jointly instantiated by a single object or event. This can be illustrated in terms of a phenomenological contrast between, on the one hand, being aware of a thing's being F and a thing's being G and, on the other hand, being aware of a thing's being both F and G, where only the latter is an instance of feature binding awareness. Moreover, O'Callaghan defended the existence of *intermodal* feature binding awareness—the awareness of features from more than one sense modality as jointly belonging to the same object or event. Although O'Callaghan's case for the existence of intermodal feature binding awareness drew primarily on phenomenological considerations, it also involved an appeal to empirical evidence about intermodal feature binding processing.

In his commentary, Kevin Connolly proposed an account of what O'Callaghan had been calling "feature binding awareness" as resulting not from a feature binding mechanism but from an associative mechanism called "unitization". Unitization consists in the integration of distinct parts of complex stimulus into a single functional unit as a result of perceptual learning, and Connolly argued that this process can occur both intramodally and intermodally. He prefaced his discussion of unitization by suggesting that our conception of

multisensory awareness may be shaped by the view that one takes of the underlying mechanisms—e.g., whether we regard it as resulting from a binding mechanism or an associative mechanism. Partly to mark its associative basis, Connolly referred to the resulting intermodal awareness as “associative awareness” instead of “intermodal feature binding awareness”.

Two main questions arise in light of Connolly’s commentary. First, one can ask whether feature binding and unitization really are competing sub-personal accounts of multisensory integration. In particular, it may turn out that the process of unitization is one type of feature binding. If that is right, then the process of unitization may be one way of generating the sort of states of multisensory awareness whose existence O’Callaghan is concerned to defend, and so there may not be a substantive disagreement between Connolly and O’Callaghan. Supposing, however, that unitization and feature binding are genuinely distinct types of perceptual processing, then a second question arises concerning the implications that each type of processing might have for our understanding of multisensory awareness. Given empirical evidence for both feature binding and unitization, this may give us initial reason to posit and investigate two distinct types of multisensory awareness: feature binding awareness and unitization awareness. There may be a sort of unitization awareness that, like O’Callaghan’s feature binding awareness, is something over and above mere awareness of associated unisensory properties--we may be able to experience them *as* operating together as a functional unit. Further investigation is thus needed to determine whether each type of perceptual process underlies a distinctive type of multisensory experience.

References:

Mitroff, S. R., Scholl, J., and Wynn, K. (2005). “The relationship between object files and conscious perception,” *Cognition* 96: 67-93.

Zmigrod, S. and Hommel, B. (2011). “The relationship between feature binding and consciousness: Evidence from asynchronous multi-modal stimuli,” *Consciousness and Cognition* 20(3): 586-593.

4. Is Language Processing a Special Kind of Multisensory Integration?

Language processing involves multisensory integration. For instance, as Teresa Blankmeyer Burke pointed out, the process of speech reading involves visual sensitivity not just to areas around the speaker’s mouth, where one takes articulation to occur, but also to other regions of the face. Drawing in part from her own experience with speech reading, Burke described how speech reading is impaired not only if the speaker has facial hair (obscuring fine mouth movements) but also if they are wearing sunglasses.

In both research in perceptual psychology and in philosophical discourse, language is often treated as a special case of perception. This is certainly for good reason—language has many features that are distinct from features of perception more generally. For example, our language-processing faculty gives us the ability to understand and generate compositional linguistic structures, which is arguably unique to the domain of language (as opposed to say, olfactory or tactile states, which might well have content but do not display the same sort of recombinable structure). Questions about reference and content may also have different answers when applied to linguistic perceptual representations as opposed to when they are applied to perceptual states more generally. Might the question of how multisensory integration works in the linguistic domain similarly receive a fundamentally different answer than it does in non-linguistic domains?

Casey O’Callaghan offered an argument that language processing is not as special in this respect as it is often made out to be. In his talk, O’Callaghan outlined six grades of multisensory integration that occur in various different types of perceptual processing, none of which are necessarily or even predominantly linguistic. When asked where speech

processing fits into his map of the multisensory landscape, he claimed that it could be thoroughly explained as a combination of the different types of multisensory integration that he had already described. On this view, there would be no need to posit a unique species of multisensory perception for language, either in terms of the processing mechanisms involved, in terms of the types of contents, or in terms of the phenomenal character of such multisensory linguistic perceptual states. If this is the case, the multisensory aspects of speech processing would be by no means uninteresting, but they would be better thought of as among a group of special phenomena (multisensory perceptions) than distinctly special on their own, at least in this respect.

Barry Smith provided some reasons one might think that language processing is a special kind of multisensory integration--different from standard cases involving audition but not speech. First, while in cases of non-linguistic audition, the perceptual object is sounds, it is unclear what the perceptual object of speech would be. Speech perception need not be the perception of sounds, but could be any number of things: sounds with meanings, meanings, a voice, the speaker, someone saying such-and-such, or some further option. Potentially, this makes speech special from cases of non-linguistic audition. Second, the brain treats hearing a human voice do something non-linguistic (such as groaning, laughing, or crying) differently from hearing a human voice speak. Smith cited research showing that immediately upon a subject's identification of electronically produced sounds as speech, there is a correlated transfer from the general auditory cortex to specialized areas in the language centers involved in the processing of speech. This seems to indicate, again, that speech is special, in that it is distinct from other kinds of auditory perception.

In her talk, Janet Werker presented an intermediate position on the question of whether language processing is special or not when compared to other types of multisensory processing. On the one hand, Werker asserted that language processing is special in that

many central features of language mastery, such as categorical distinctions between phonemes, aren't acquired by domain-general learning mechanisms like association, but instead draw upon a proprietary set of representations that prepare the infant for language acquisition. There are, Werker noted, significant constraints on the forms that can serve as a possible signal in a language. For example, a human being cannot learn a language composed wholly of mechanical sounds. Thus the possible forms that can serve as linguistic units are non-arbitrary, even if the meaning that we assign to these units is arbitrary. On the other hand, Werker denied that speech is unique in these respects. Apart from speech, sign language has also been shown to depend on categorical distinctions within the perceptual system to which human infants are sensitive from early in life and which prepare the infant for learning that language. Werker also referred to evidence suggesting that other natural sound signals, such as the sound of water, may qualify as "special" in the relevant sense. Thus she took a similar position to O'Callaghan, although for different reasons, that language processing is not entirely unique, nor is it entirely common. Instead, it's among a group of unique things.

5. What Is the Purpose of Multisensory Integration?

The two core sets of questions about multisensory integration that received the most focus at the workshop revolved around 1) how the perceptual processing involved in multisensory integration works, and 2) the content and character of multisensory experiences. However, a third important type of question focuses instead on the role that multisensory integration plays for an individual. What use does it have, and why might our perceptual systems have evolved so as to integrate information coming from multiple modalities?

One way of approaching this question is to consider what the advantages of multisensory perception might be over unisensory perception. We certainly benefit from

having more than one sensory modality, because it allows us to access more information about our environment—both quantitatively more overall, and more types of information—which can be useful for both navigation and survival. For example, if a predator has the ability to track its prey by sight and by scent, it will be more likely to be successful at hunting even if its sight is obscured by a forest, or if it is dark out.

However, this sort of case does not yet illuminate the advantage that multisensory integration per se gives us over and above the advantage of access to input from multiple modalities, each of which may be processed and experienced in isolation from the others. In her talk, Jennifer Campos highlighted some such unique advantages in the domain of self-motion perception, such as increased accuracy due to flexibility of input weightings for different contexts. In self-motion perception, inputs to the vestibular, proprioceptive, visual, and auditory systems are integrated to represent our own locations and trajectories in space. Ophelia Deroy noted in her commentary on Campos that the input sources that are combined in self-motion perception are particularly interesting, because they involve both interoceptive (vestibular and proprioceptive) and exteroceptive (vision and audition) senses. Campos presented research on integration of such inputs for balance regulation and for estimating distance travelled. When there are multiple inputs available about an individual property, the brain must combine this information using a weighting algorithm, which dictates how much to rely on each particular source. Campos has conducted studies investigating how such weightings are influenced by variations in the input sources and their contents, and has found that subjects generally weight toward the sensory source that provides the most stable and reliable information. This means that often, we are actually responding more to proprioceptive cues than to visual ones, despite the fact that reflection on our own experience might lead us to believe that we are predominantly visual creatures. Deroy made the point that Campos's results indicate that even when vision may be necessary or sufficient for a

given task, it may still not be the dominant modality in play, in terms of the weightings given to sensory cues. The dependence of the relative contribution of any given input source on its reliability indicates that our perceptual systems have evolved to make use of the information that is most likely to accurately represent the world. A core function of multisensory integration is to combine information sources to facilitate the production of such accurate representations across varying contexts.

In a similar vein, Connolly noted that multisensory integration might also facilitate the production of accurate representations by increasing the efficiency of perception. If the perceptual system can generate representations with contents that incorporate properties derived from multiple sensory modalities fused together, this may eliminate the need for certain sequences of reasoning. For example, if upon hearing a clang and seeing a cymbal being hit, our perceptual system can on its own generate a single representation with multisensory contents representing the sound as emanating from the cymbal, this will eliminate the need to consider the auditory and visual percepts independently, and judge that the properties represented in both attach to the same object. In general, perceptual processing proceeds more rapidly than deliberate inference, so multisensory integration may save us time in coping with the environment.

Both of these accounts of the function of multisensory integration appeal to the idea that it allows us to better complete certain crucial tasks (for example, calculating distance estimates, and thereby generating the appropriate motor responses for a situation), due to the incorporation of multiple information sources that increase accuracy and efficiency. It does not, however, posit a distinct *type* of function that multisensory perception serves. On this view, the integration of multisensory inputs merely increases the likelihood that the function (or at least one of the major functions) of unisensory perception (accurately representing the world) will be fulfilled. A more radical answer to the question of the purpose of multisensory

integration might say that it gives rise to truly novel sorts of information, which we could not even in principle access through unisensory processing, and that this novel information plays a crucial functional role. Casey O’Callaghan mentioned some multisensory experiences of this type in his talk, which he labeled as belonging to the “5th grade of multisensory awareness.” Flavor might be an example of one such novel feature type that is constitutively dependent on inputs to olfaction, gustation, and tactition, and on their combination in a particular way. There might be certain cases in which detection of flavor properties (as opposed to detection of smell, taste, and/or touch properties) is distinctively useful in terms of making environmental discriminations that guide behavior, such as determining which foods are beneficial to an animal and would be worthwhile to pursue. It is also plausible that the particular hedonic response that such flavor experiences lead to are not achievable merely through the experience of their constituents of taste, touch, and smell, and these responses might also be useful for overall well-being. While the extent of the usefulness of flavor is an empirical question in evolutionary perceptual psychology, it seems quite plausible that there are at least some cases in which perceptual awareness of novel feature types serves a particular function, related to the evolutionary success of an animal.

Another way of approaching the question of the purpose of multisensory integration is to ask what it would be like for us if we were not able to integrate multiple sources of information. We can glean some insight into this issue by looking at cases of selective impairments. Campos discussed studies that she had conducted on the interaction between vision, audition, and balance in subjects with cochlear implants. She found that deaf children who use cochlear implants had difficulty maintaining their balance when standing on one foot, but that with their implants in, their balance improved. This indicates that auditory cues, in addition to proprioceptive, vestibular, and visual cues, are used in balance, and so the ability to rely on and combine multiple types of perceptual cues is instrumental for navigating

the environment. Campos's research in the area of the purpose of multisensory integration also has clear practical applications—for example, when strategizing ways to help older people who have trouble with balance, we should take into consideration potential deficits in all the senses involved in the integration processes, as well as errors in the way they are combined, as opposed to focusing exclusively on impairments to a single sensory modality.