

Explaining Multisensory Experience

Matthew Fulkerson

Comment on Kevin Connolly's "Making Sense of Multiple Senses,"
forthcoming in *Consciousness Inside and Out: Phenomenology, Neuroscience, and the
Nature of Experience*. Edited by Richard Brown: Springer.

--Penultimate Version: Do not cite without permission--

Introduction

Our experience of the world involves a number of senses, including (but perhaps not limited to) sight, hearing, touch, taste, and smell. These senses are not isolated from one another. They work together, providing a robust and coherent awareness of our environment. Consider entering a good restaurant: one sees the décor and the other patrons, smells the pleasing odors wafting from the kitchen, hears the pleasant music and sound of conversation, feels the comfort of the seating, and, finally, savors the taste of the food. It seems obvious that, in some sense at least, our perceptual awareness of the restaurant is multisensory. Saying exactly what it is for perceptual awareness to be multisensory is more challenging than it appears, however.

One might suppose, for instance, that there is no single "experience of the restaurant." To say that our awareness of the restaurant is multisensory is just shorthand for saying that it involved many distinct perceptual experiences contributed by different senses. This notion of what it means for an experience to be multisensory is not especially robust or interesting: we have one experience and then another, or perhaps we have several different experiences at the same time.¹ In

¹As we'll see, this way of carving up perception into separate 'experiences' generates some problems in our understanding of multisensory interaction. See Byrne 2009 for a pointed criticism of the philosophical notion of 'experience,'

some cases, this probably is what we mean by multisensory.² But this can't be the whole story. Consider what happens when the food is tasted: at this moment the aroma, taste, feel, and temperature seem to blend into a novel whole. Anyone who has eaten a favorite meal with a bad cold, or when it is at the wrong temperature, or after it's been ground up in a food processor, can attest to the influence of many senses on our experience of food. In these cases, it does not seem as though there are several separate experiences going on at the same time, but rather that there is one, unified experience of the flavor that results from the coordinated operation of more than one sense (Auvray and Spence, 2008). Call these types of multisensory experience *multimodal*. Such experiences are not at all explained by the mere conjunction of distinct sensory experiences.

In addition to multimodal experiences, there are also cases of *crossmodal* experience, where the operations of one sensory modality influence or make a difference in the operations of another. Empirical studies reveal that the senses have strong influences on one another (see, e.g., Calvert and Thesen, 2004; Driver and Spence 2000; Ernst *et al.* 2007). In crossmodal cases too, something more than mere conjunction of distinct perceptual experiences seems required. But how exactly ought we distinguish those cases where the senses are merely co-occurrent from those where they somehow blend into one another, from those that have strong influences on each other?³ And what to make of the many other forms of interaction that are also more than mere conjunctions, but that do not fully blend into single experiences or involve direct influence on another modality?

The recent realization that perceptual modalities are often deeply intertwined might lead some to call for abandoning the very notion of a unisensory experience (see e.g., Shimojo and Shams 2001; also Driver and Spence 2000).

²As when a restaurant critic describes the overall meal as a delightful, 'multisensory' experience.

³I attempt in Fulkerson (2011) to describe and motivate one way of distinguishing these distinct forms of multisensory interaction.

Multisensory experience, on this perspective, requires radically jettisoning our standard conceptions of perceptual experience, at least for a range of paradigm multisensory interactions. For convenience, call those who want to resist such radical moves "sensory conservatives" and the view they defend "sensory conservatism." (A defender of sensory conservatism for vision is Pylyshyn 2006). In this volume, Kevin Connolly defends a limited sensory conservatism for crossmodal (but not multimodal) experiences. Now, to be clear, sensory conservatism does not deny that many of our perceptual experiences are multisensory. It simply takes such experiences to be nothing more than conjunctions of unisensory experiences (or, more accurately, to be a complex formed *somehow* by a combination of nothing but unisensory components). The sensory conservative holds the intuitive and highly plausible view that a perceptual experience is multisensory if it involves more than one sense.

As a sensory moderate, I welcome the parsimony and intuitive appeal of the conservative viewpoint, but believe it ultimately fails to do justice to the complexity and messiness of actual sensory interactions. On the other hand, I don't think we ought to abandon the very concept of a sensory modality as the radicals suggest. Making space in this middle ground is not easy, since, unlike the conservative, we can't simply take for granted that there are perfectly individuated, informationally-encapsulated sensory modalities. For this reason, we cannot account for multisensory interactions as mere conjunctions of such constituents without saying quite a bit more about how we are carving up the individual modalities and experiences. After all, whether there are such constituents is one of the main ongoing theoretical questions.

At any rate, I will set these worries aside for now, and focus my efforts elsewhere. In particular, I distinguish two versions of sensory conservatism, and use these more precise formulations to put some pressure on the conservative position. As we'll see, Connolly denies the first version, and only accepts the second for a limited range of cases. This makes him a rather lukewarm defender of the conservative line. Indeed, there is an inherent tension in both trying to defend the

conservative position because of its relative simplicity and explanatory parsimony, while acknowledging that the view is true only for a limited range of cases.

Clarifying the Target: Sensory Conservatism

First, make the simplifying assumption that there are only 5 distinct sensory modalities: audition, vision, touch, olfaction, and gustation. Second, let each modality have a set of sensible features, so that $\{a_1...a_n\}$ is the set of sensible features available to audition; $\{v_1...v_2\}$ is the set of features available to vision, and so on. For convenience, label these sets of sensible features A, V, T, O, and G. Finally, a perceptual experience E has content E(F), where F is the set $\{f_1...f_n\}$ of sensible features represented by E (for my purposes, this is equivalent to the claim that F is the *content* of E).⁴

Using this terminology, we can say that for a sensory conservative an experience E(F) is multisensory if F contains sensible features from more than one sensory modality. So an experience that represents a blue dot and a C# is multisensory, since it represents features available to two distinct sensory modalities (I discuss this issue in more detail in my 2011). This account is intuitive inasmuch as it assumes that we already have a good grasp of how to individuate sensory modalities and the sensible features that belong to them.

When Connolly suggests that crossmodal cases can be entirely explained by appeal to unimodal features, he is defending perceptual conservatism with respect

⁴Of course, these are gross simplifications. In addition to represented features, perceptual content will also have a spatial distribution and (perhaps many) internal relations (like binding). And there are likely many more than five modalities, often with obscure or entangled contents. I focus in what follows on this simplified account, since I believe it helps clarify the position Connolly defends. But as we'll see, these simplifications can make the conservative viewpoint seem more plausible than it actually is.

to certain crossmodal cases. In order to properly assess this position, we must clarify the issues and present as precise a formulation of the options as possible.⁵

We must distinguish two versions of sensory conservatism. The first is defined by the following thesis:

The Proprietary Content Thesis (PC): No sensory content is shared among the senses.

The paradigm for this kind of view is Fodor's (1981) classic modular view of the senses as (among other things) hard-wired, informationally-encapsulated, domain specific input systems. Still, we can give a more precise formulation. Let P be the collection of modality-specific sensory feature sets (so, e.g., P_V is the set of visual features; P_A is the set of auditory features, etc.). Now, PC is the claim that the members of P are pairwise disjoint: $P_i \cap P_j = \emptyset$. That is, there are no sensory features found in more than one modality, and therefore no sensory content (representing sensory features) is to be found in more than one sensory modality.

PC is one way to resist the idea that perceptual experiences are inherently or radically multisensory. If an experience contains content from more than one sensory modality, then, according to PC, we can always *decompose* this content into its constituent modalities. This accords well with the intuitive idea that a multisensory experience is nothing more than a combination of unisensory components. Note that PC can be violated, and the conservative position undermined, even if we restrict the represented properties to the so-called 'basic' sensibles (if, for example, two or more senses represent such basic features as *number*, *location*, *size*, or *shape*, then PC would be violated). According to PC, even in cases where the senses seem to represent the same sensory features, they do not

⁵What follows is my attempt to give a more precise and careful account of the view Connolly defends. It is hoped that it avoids some of the many difficulties that arise trying to discuss relations between experiences, where many levels of explanation (including qualitative, informational, and functional) interact.

literally *share* content. Instead, each sensory modality represents that feature in its own proprietary format (this story can be told in several different ways).⁶

The second version of sensory conservatism is defined by the following thesis:

The Exhaustive Content Thesis (EC): The content of perceptual experience consists only in the sensible features found in the individual modalities.

More formally, for any perceptual experience $E(F)$, for all $f \in F, f \in P$. Note that the truth of EC is independent of the truth of PC. First, EC can be false when PC is true. This would be the case, for instance, if there were novel features not available in any of the individual modalities, but which occur in (multisensory) perceptual experiences. Call this the possibility of *multimodally emergent* content. Conversely, there could be contents shared among the individual senses (violating PC), yet these modality-specific features may nevertheless exhaust the contents of perception (ensuring the truth of EC). In all likelihood, the most interesting and cohesive versions of sensory conservatism will endorse *both* PC and EC.⁷

So where does Connolly stand on these theses? It seems he believes that both PC and EC are false for some range of perceptual experiences. He agrees with Nudds (2001), O'Callaghan (2008), and others that PC is (perhaps often) violated. Indeed, it's *very* difficult to see how one could defend PC without subscribing to either a extremely naive view of the senses or some version of Fodorian modularism (e.g., Pylyshyn 2006).⁸ But Connolly also thinks EC is false, at least for our experiences of

⁶The main argument of O'Callaghan (2008) seems to be directed at PC. He argues after all that there must be some "shared content" between distinct sensory modalities. Such contents are, in a real sense, inherently multisensory.

⁷I believe these theses are not always clearly distinguished in Connolly's discussion.

⁸To appreciate how naive the view would have to be, consider that even Aristotle's account of the senses (which individuates them according to their unique or 'proper' sensibles) allowed that the senses shared sensory features, which he called 'common' sensibles.

flavor. Flavor experiences, he allows, seem to have emergent content that is not found in P. Since he denies both theses associated with sensory conservatism, it's difficult to see any general theoretical motivation behind Connolly's position.

Instead, Connolly's focus is on particular cases: whereas some (notably for him, Bayne) seem to hold that certain paradigm crossmodal experiences violate EC, Connolly does not believe they do. The paradigm cases he discusses are the McGurk Effect, The Motion-Bounce Illusion, and the Ventriloquism Effect.⁹ It's not clear if these cases are thought to form a natural kind, allowing us to generalize to the falsity of EC for all crossmodal cases, or if his claims are just restricted to these three cases. At any rate, we can ask what hangs on whether these crossmodal cases involve emergent content or not? Connolly seemingly asks this question: *can* we explain crossmodal cases as representing only features in P? He then proceeds to give a consistent account of just such an explanation. Yet it's not clear what hangs on this mere possibility, especially if we allow that the members of P may be shared among multiple sensory modalities. Without maintaining PC, several modalities could share complex sensory features like *cause* or *speech* or even *bounce*. Indeed, such shared contents offer the best explanation of crossmodal experiences, since differences in the shared or overlapping contents would explain the need for such contents to be reconciled. There are many known mechanisms for such reconciliation, usually falling under the rubric of *multisensory integration*. Such mechanisms include sensory suppression, dominance, and facilitation (see Calvert, Spence, and Stein 2004). This is an interesting and rich notion of multisensory interaction, one that would pose many problems for a typical sensory conservative. But Connolly does not really take issue with this possibility. In fact, he seems to think it is right.¹⁰

⁹As he does a fine job describing these effects, I will not redescribe them here.

¹⁰I base this claim on the following (representative) passage: "The idea is that in a crossmodal case, the inputs in two different modalities conflict because they are predicated of a common source or cause (whether it be an individual, object, or event). This conflict requires the reconciliation between the inputs, and what we

His main target is the idea (contra EC) that there are emergent sensible features (e.g., *sound-sources, bounces*) that are represented only when more than one sensory modality is combined. He thinks this idea is false for the paradigm crossmodal illusions (though it may well be true for other multimodal experiences). Connolly's argument, essentially, is that we can explain crossmodal cases by appeal only to unisensory contents. Whenever there is a purported content not in P, he suggests we can find some conjunction of elements in P that would do as well explaining the target experience. There is no need, he suggests, to posit such emergent content. In order to properly assess this claim, we need to discharge some of our simplifications.

If a unisensory experience just is one representing only members of V or T or G, then *what*, if EC is false, is being combined to generate the emergent content? Experiences qua experiences are not at the right level of explanation, and neither is the level of content. But Connolly seems concerned about accounting for "a new kind of property," as though what is being considered is the possibility that two distinct sensory features can be combined in experience and thereby generate a novel, emergent sensible feature. This kind of metaphysical emergence *would* be quite mysterious. But multimodally emergent content is not metaphysically emergent at all. In fact, such emergent content is entirely metaphysically innocent. We are talking about perception as an information processing/representational system, and there is nothing ontologically problematic about a sensory process that combines distinct sensory inputs and produces simplified, coherent representations of more complex features. This confusion is apparent in the way Connolly frames the debate in terms of a causal-constitutive distinction (a distinction at the level of processes)

experience is the product of that reconciliation" (ms 11). Now, it's possible that the senses could predicate features of the same objects and events without sharing content, and perhaps this is Connolly's view (and so maybe he wants to also deny PC). But such a view is only plausible if we assume that sensible features alone enter into perceptual content. But as noted above, this is a gross simplification. Indeed, to even make sense of co-predication, some shared spatial or temporal contents will be required. For this reason, I believe this quote commits Connolly to the falsity of PC.

but then worries about novel properties (at the level of ontology). The causal-constitutive distinction does not make any sense at the level of features or experience, only at some lower level of functional or physiological sensory processing, where there are no worries at all about ontological parsimony or economy. Consider an analogous case of testimony. I hear from one source that my friend Jill is in the next room. I hear from another source that Jill is expecting a child. From this I come to the belief that *a pregnant woman is in the next room*. The content of this belief is not provided by either source alone, nor is it merely a conjunction of the two source reports. But there is nothing ontologically mysterious about where the belief gets its content!¹¹

Setting aside the misplaced metaphysical worries about ontological austerity, it must be that, for some range of typical crossmodal cases, Connolly denies that there are any processes or subsystems that serve to represent content over and above the content available in the individual modalities.¹² So, he is denying that there are any modality-independent "causal detectors" or similar higher-level systems that take as input perceptual information from multiple sources and have as output novel contents. For example, a system taking as input an auditory speech sound and a visual mouth movement and (using temporal and spatial proximity among other factors) generating as output the novel multimodal content *sound-source*.

¹¹The analogy with testimony can be fruitfully extended to cover crossmodal cases: such cases arise when the testimony of multiple senses come into conflict. When one sense/source is more trusted, cases of sensory dominance arise, generating typical crossmodal illusions. Notice that such an account only makes sense if the two sources "share" content.

¹²Actually, Connolly *seems* even to allow such contents that 'outstrip' the sensory modalities, without realizing that this would be a violation of EC. He writes, "If we characterize the individuals, objects, or events in the second way, that is, in modality-independent terms, then we are not positing multimodal content. We are positing amodal content" (13). But amodal content of this kind is clearly emergent content. If this is right, then it's simply not clear what position Connolly intends to defend.

It's not clear that we are given much reason for rejecting such modality-independent systems (indeed, as noted above, Connolly seems to let them in by accident). The main motivation seems to be that it is *possible* that there are no such systems, and as this would be a more elegant or parsimonious account, we ought to accept it. As he writes, "If we reject fused audio-visual content, and appeal instead to audio content and visual content, our account of content is also more economical, since we don't need to posit a new kind of property" (ms 7). But ultimately this is an empirical question (though one desperately in need of conceptual clarification). And there is little reason to think our perceptual systems are wholly subject to constraints of parsimony (if they were, wouldn't we have only a single sensory modality?). And there seems to be strong evidence that there are systems above the level of the individual modalities that both make a difference to and contribute to perceptual contents. Once the door is opened to such systems (for example, for flavor perception, causal awareness, etc.) it's very difficult to use parsimony as a reason for closing the door in other cases. And as noted, there are few general theoretical motivations for the sensory conservative view. Ultimately, I don't see us adjudicating these matters through philosophical argument alone. While much work is needed to clarify our target and to distinguish the various forms of sensory interaction, real progress will be made only when we applying these more precise formulations to the actual data. We need to look more closely at the empirical evidence—the constituent sensory systems and their interactions—in order to tease apart the contributions of the individual modalities from the higher-level systems that integrate and coordinate the input from those modalities.

Conclusion

We find ourselves in the following position: our experience of the world is largely multisensory. We experience the world with all of our senses, and these senses interact at many levels of processing, and in many different ways. The empirical research literature, and philosophy of mind along with it, is beginning to recognize

the importance of these interactions. Instead of focusing on the individual senses in unrealistic isolation (along the lines popularized by Fodor's modular account), philosophers and psychologists are starting to take seriously the idea that the individual senses are deeply intertwined, and that our perceptual experience is inherently multisensory. This move towards a more interactive perspective might make it seem as though there were a clearly defined notion of multisensory interaction, and that it can be easily contrasted with our intuitive notion of unisensory experience. While it may be true, in some narrow sense, that all of our perceptual experiences are multisensory, in reality our perceptual experiences are subserved by a wide range of distinct sensory interactions. Sometimes, for some purposes, we focus on mere conjunctions, other times we focus on the functional interactions between the systems that generate our perceptual experiences, and yet other times we focus on the contents of our experiential states. These different purposes yield distinct and often incompatible notions of multisensory interaction.

The embrace of a multisensory perspective of perceptual experience is a good thing. The idea that the senses are wholly separate and disconnected forms of experience is surely wrong. But we should not suppose on this basis that there are no sensory modalities. The senses are a messy, heterogeneous, complex jumble of distinct interactions at many levels of explanation, but depending on our purposes we can isolate patterns of unity and coherence characteristic of our intuitive notion of the senses. Understanding sensory experience requires that we look more closely at the specific (and pervasive) interactions between the senses, but also that we keep in mind those elements that make the individual senses significant and interesting. That is, we ought to preserve what is important about the individual senses while acknowledging the many varied interactions between them. This is sensory moderation. It is, of course, a difficult task, but these are still early days in our thinking about sensory interaction.

Works Cited

- Auvray, M., & Spence, C. (2008). "The multisensory perception of flavor." *Consciousness and Cognition* 17(3): 1016–1031.
- Byrne, A. (2009). "Experience and Content." *The Philosophical Quarterly* 59(236): 429–451.
- Calvert, G., Spence, C., & Stein, B. E. (2004). *The Handbook of Multisensory Processes*. Cambridge, MA: MIT Press.
- Calvert, G., & Thesen, T. (2004). Multisensory integration: Methodological approaches and emerging principles in the human brain. *Journal of Physiology-Paris*, 98, 191–205.
- Ernst, M. O., Lange, C., & Newell, F. N. (2007). Multisensory recognition of actively explored objects. *Canadian Journal of Experimental Psychology* 61(3): 242–253.
- Fodor, J. (1981). *The Modularity of Mind*. Cambridge, MA: MIT Press.
- Fulkerson, M. (2011). "The Unity of Haptic Touch." *Philosophical Psychology* 24(4): 493-516.
- Milner, A. D., & Goodale, M. A. (1995). *The Visual Brain in Action*. New York: Oxford.
- Nudds, M. (2001). "Experiencing the Production of Sounds." *European Journal of Philosophy* 9(2): 210–229.
- O'Callaghan, C. (2008). "Seeing what You Hear: Cross-Modal Illusions and Perception." *Philosophical Issues* 18(1): 316-338.
- Pylyshyn, Z. (2006). *Seeing and Visualizing: It's not what you think*. Cambridge, MA: MIT Press.
- Shimojo, S., & Shams, L. (2001). "Sensory modalities are not separate modalities: plasticity and interactions." *Current Opinion in Neurobiology* 11(4): 505–509.