

De Gelder, B., & Vroomen, J. (1994). A new place for modality in a modular mind. *Cahiers de Psychologie Cognitive*, 13, 84-91.

## A New Place for Modality in a Modular Mind

<sup>1</sup> *Beatrice de Gelder* and <sup>2</sup> *Jean Vroomen*

<sup>1</sup> *Cognitive Psychology and Psychophysiology unit, Tilburg University, The Netherlands, and Laboratoire de Psychologie Experimentale, Université Libre de Bruxelles, Belgium*

<sup>2</sup> *Cognitive Psychology and Psychophysiology unit, Tilburg University, The Netherlands*

The tension between modality-specific sensory processes and abstract concepts is an old one. The notion that one stratum of the mind has a modular organisation is rather recent. Not much attention has been paid so far to the way sensorial modality and mental modularity might combine. For example, Massaro (1987) argued that facts which transcend modality-specificity of speech present arguments against the modularity of mind. We have disagreed with this in the past (de Gelder and Vroomen, 1989), and in the present comment we pursue our analysis of the modality versus modularity debate, maintaining that both are orthogonal issues while there is, at the same time, room for modality-specificity within the realm of modular processes. The occasion for these remarks is the paper by Radeau (this volume). Materially, the paper is drawn from a series of studies over more than twenty years with Paul Bertelson, presenting beautiful experiments on audio-visual spatial interaction. Radeau tries to build a bridge between that research on sensory integration processes and the more recent notion of mental modules. The central claim of our comments is that issues of modality must be distinguished from issues of modularity. The reasons for keeping to this distinction are the same as the reasons for not assimilating the research on spatial integration with the research on the multi-modality of speech input. It follows that data from studies on audio-visual speech do

NOT lend support to the notion of a special module for audio-visual non-speech interactions.

### *1. Modality versus modularity*

Modality and modularity are complex and ambiguous notions. This ambiguity derives from the fact that the two pop up at various stages in psychological models of knowledge and in philosophical analyses of its foundations. Yet, the two notions can be used unambiguously when the nature of the explanation they figure in is made explicit. In what follows, we will rely on the one specific meaning of modality and modularity that is the essential one for the issues at stake.

Philosophers have traditionally been interested in modality-specificity, worrying whether the senses must be considered as anchors of knowledge or as sources of its contamination. Berkeley is a notorious example, with his claim that our knowledge begins and ends with experiences locked into sensorial modality-specificity. But in the context of present day empirical psychological theories, such a view on sensory specific knowledge is hard to place. The notion of modality-specificity that is of most immediate concern in psychological research relates to sense-specific coding of information, e.g., vision provides visual information, hearing provides auditory information, etc. Note that we are not talking here about sensory objects, nor about sensory concepts, but only about sensory-specific featural information existing at a level of processing in which objects do not yet have existence. The differences between the modalities thus relate to differences in the physical properties they inform us about. In due course, information from the various modalities gets integrated and we achieve objective knowledge and entertain concepts, etc. Depending on one's philosophical view and psychological theory, one may hold the view that modality-specificity percolates upward the information processing system all the way leading to a concept of e.g., a visual circle which is different from that of a haptic circle.

The alternative view is that modality-specificity begins and ends with the senses, and that what comes next is a matter of abstract concepts and propositional knowledge including knowledge about sensory objects. The research Radeau reminds us of was, we believe, very much in the tradition of contrasting sensory and conceptual

processes, perceptions and cognitions in the accepted sense of these terms. For example, experiments were set up to examine conceptual influence on sensorial integration where subjects' conceptual knowledge of what loudspeakers were for was the critical variable. If, just if, the notion of a mental module has a birthright, it hangs on the conceptual consistency and the empirical plausibility of a modular level of processing that is neither captured by sensory analysis nor by conceptual labour.

What, if anything, has changed with the arrival of modularity? Even if we push aside the epistemological issues and the way the modularity issue gets invested with epistemological nobility and its role in the great debate about theory-neutral observation (Fodor, 1984; de Gelder, in press), modularity is still a very rich notion. There is little doubt that on the conceptual side, the leading notion motivating Fodor's psychological theory is the Chomskyan notion of a module. On Chomsky's account (see Chomsky, 1986), the notion of a module is central to the analysis of a speaker's knowledge of language. The language module and its submodules are the grammatical theory and its separate subparts. The claim of psychological reality which combines with this notion of knowledge of language suggests that these analytical parts correspond to separate subtasks a language learner and a language processing subject is capable of. Whatever the differences between the explanatory project of linguistic theory and that of psychological theory in the business of figuring out how the mind goes about its daily job, this origin of the concept of a mental module cannot be ignored in the discussion since it is responsible for the one and only distinctive property of modularity, i.e., *domain specificity*.

Matters of modality of the senses are thus very different from issues of the modularity of mind. The distinctive characteristic of a module is its domain-specificity. Claims about modality relate to questions on the mode of information input and correspond to ways of carving information input up into sensorial regions. In contrast, claims about modularity carve information up into types of knowledge or more specifically and less ambiguously, into semantic domains. The contrast is thus no longer one between data-driven and concept-driven processes. Instead, one must now face up to a three layered picture with sensory states, modular states, and belief states. The specific claim that is new with this notion of modularity is that there is of a level of

processing that is intermediate between sensations and full blown concepts. Such a level might be called a level of shallow objects, shallow because not yet integrated in the network of real world knowledge. The example from language is helpful.

Sensorially there are sounds, centrally there are meanings or concepts, while at the intermediate level there are linguistic objects or linguistic representations.

Modularity is thus not a claim to be taken lightly or to be given a relative interpretation, as if there were such a thing as being 'relatively domain-specific'. Of course, Fodor has somewhat inadvertently advertised modules as species one recognises by crossing off traits on a checklist. It is clear that matters of speed, impenetrability or pre-wiredness blur the picture of what a true module looks like, since they are found across the board off all information processes. The mechanism of audio-visual spatial integration appears to score on most of these traits, but that does NOT make it a module. For example, it is a trivial fact that sensory processing is not under conceptual or doxastic influence of the kind usually assimilated with a subjects' cogitations. But that does not make it a module in the only sense that matters, i.e., a device that is operational in a specific semantic environment. The study of modular processing is that of the processing abilities of a system qua functional architecture, in its biological sense. The latter is likely to exhibit some degree of species specificity as the example from language does bring out. In contrast, the study of modality of input concerns physical properties of the stimulus input. Principles, Gestalt ones like grouping, common fate, and others or stimulus properties like signal intensity are general and found in vision just like in audition.

## *II. Relations between audio/visual pairing and audio/visual speech*

At first sight, it may look as if the pairing mechanism which underlies audio-visual pairing of non-speech does fulfil the criteria of a Fodorian module in a similar way as the module for audio-visual speech appears to do. Indeed, both score high on Fodor's checklist. But that is not the procedure that can settle the debate. What is critical is that the analysis of the audio-visual pairing mechanism requires a description of the object in its domain. And this, we argue, is not the case. Instead, audio-visual pairing relies on general Gestalt principles of common fate and proximity which

transcend and precede (or follow) domain specificity. They thus apply to both speech and non-speech input, but they are by themselves not part of a module. Gestalt principles may apply at a pre-modular or post-modular stage, but whatever the outcome, it seems clear that there is some kind of hierarchic organization such that Gestalt principles operate independent of modular processes. One can imagine that Gestalt principles operate at a pre-modular level such that they group perceptual primitives (uni-modal and multi-modal) into a unitary event. On the basis of such primitive grouping, the domain of the event might be inferred. For instance, formants might be grouped together because of similar onset/offset times (Darwin and Sutherland, 1984), and then referred to the speech module for speech processing.

Such a hierarchic perceptual organization has several consequences worth considering. The first is that rejections in the perceptual grouping of the primitives by the Gestalt principles percolate up to the domain-specific modular processes. For instance, it seems clear that auditory and visual speech will not be integrated if they are too asynchronous, and, although there are to our knowledge no studies which have addressed this issue, it might also be the case that McGurk-like fusions will not occur if the locations of the auditory and visual information are too disparate. The refusal of the perceptual system to integrate these cross-modal inputs is caused at a non-modular stage at which Gestalt principles operate. They should thus also be found with other stimuli and modalities, like those for instance mentioned by Radeau. Indeed, it was found that bongo's or light flashes cause adaptation and that recalibration is decreased if the spatial separation is increased. The Gestalt principles thus seem to operate across modalities and stimulus-domains in a non-specific way.

There are reasons of a different kind besides violations of Gestalt principles which prevent crossmodal interactions. One is that intra-modular resistance against cross-modal interactions will occur whenever one of the inputs does not match the domain of the module. For instance, Summerfield (1979) observed that auditory speech is not integrated with a Lissajou ring which corresponds to the centers and corners of the lips. Presumably, audio-visual integration of speech only occurs if the input of both modalities is processed as speech, and a Lissajou figure does not fulfil this criterion. It thus seems that Gestalt-like and module-like criteria for integration are

different: synchronicity and spatial adjacency are triggers for the Gestalt principles, but they are not sufficient to guarantee module-like audio-visual speech integration.

A further distinction between Gestalt principles and intra-modular interactions is that the former have their own signature, namely adaptation and recalibration. When two cues of the same perceptual parameter (e.g., depth, location etc) arrive at different values, adaptation, and later recalibration takes place. This situation is of course well demonstrated by the work of Radeau and Bertelson. However, adaptation and recalibration phenomena do not take place in the phonetic module. Thus, in the McGurck-situation, conflicting phonetic cues from audition and vision are presented, and the outcome is a more or less optimal solution which fits both information sources (Massaro, 1987; Vroomen, 1992). There is, however, in the McGurk-situation no adaptation to the strange combination of auditory and visual cues, and even more important, so far nobody has reported any evidence for recalibration processes. It thus looks that conflicting cues in the phonetic module are solved differently if compared with conflicting cues in the spatial or temporal domain: the former do not lead to recalibration, the latter do. This strongly suggests that audio-visual speech and audio-visual pairing are two sides of two different coins: audio-visual speech is specific for the speech module, audio-visual pairing applies to audio-visual speech and to non-speech in a non-domain-specific way.

### *III. Modalities within modules*

Modularity has challenged the traditional contrast between sensory modalities and abstract concepts. We now turn to another aspect of this challenge: the existence of modality-specificity within the module. The critical distinction needed to make room for modality in the modular mind is the one between pre-modular modality aspects versus post-modular modality aspects. Let us turn to aspects of cross-modality in relation to modularity and our analysis of McGurk-like conflicts already mentioned. One might argue that such interactions, whether integration, as in the spatial domain, or conflicts as in the speech domain, overrides our principled distinction between modalities or modules.

It follows from the previous remarks that it is potentially misleading to compare

the McGurck-illusion with the case of audio-visual pairing, since the former is most clearly a case of a domain specific conflict and since it shows that there still is modality after modularity. Evidently, there is only an integration problem because there is successful modular processing, and there is only conflict once the module has operated successfully, that is, once linguistic information has been extracted from the two sensory modalities. The conflict between the input from the two modalities would not occur if in each of the modalities the speech module had not detected linguistic information. Likewise it should be very clear that conflict and conflict resolution of the kind found in fusions and blends is a process under modular control and not a sensorial interaction of the kind so well illustrated by the Bertelson/Radeau research. Nor is it a conceptual integration issue, or a conflict whose origin or resolution is under conceptual control.

Empirical illustrations for the independence of modality and modularity come from cases of modular impairments in the absence of sensorial disorders as observed for example by Campbell (1993) and ourselves. We observed impairments in speech processing and in memory for speech input in developmental phonological dyslexics. This fact is intriguing because it concerns a modular impairment, i.e., in the speech-processing domain. If poor readers suffer from phonological processing impairments, there is no reason to limit these to problems in the auditory modality. As a matter of fact, we have consistently found that in these subjects phonological impairments are also found when speech is lipread (de Gelder and Vroomen, 1988; de Gelder and Vroomen, in press, Vroomen, 1992). Modular impairments thus ignore sensory modalities but they do not jump domains. Similarly, a face processing deficit does not lead to a visual speech perception deficit (Campbell, this volume; de Gelder, Vroomen and van der Heide, 1991; de Gelder, Gepner, and de Schonen, submitted). Empirical evidence for modality-specificity within the speech module are for example found in studies on short-term memory of heard and lipread speech (de Gelder & Vroomen, 1992; de Gelder & Vroomen, in press).

In conclusion, there was room for modality effects before modularity appeared. There is still just as much room for them when modularity is brought into the picture. Actually, there is a new place for modality effects within the modular mind. Rather than

assimilating the old modality effects to the new modularity notion, a new notion of post-modular modality effects might be needed. Moreover, modality effects from pre-modular and post-modular processes are likely to be very different, computationally as well as neuropsychologically.

## Acknowledgements

Research reported in the present paper was partly supported by a grant from the Belgian Ministere de l'Education de la Communaute Francaise ('Action de recherche concertee' (Language processing in different modalities: comparative approaches)).

## References

- Chomsky, N. (1986). *Knowledge of language*. Cambridge, Mass: MIT Press
- Campbell, R. (this volume).
- Darwin, C. J., & Sutherland, N. S. (1984). Grouping frequency components of vowels: when is a harmonic not a harmonic? *The Quarterly Journal of Experimental Psychology*, **36A**, 193-208.
- de Gelder, B. (in press). Modularity and logical cognitivism. In A. Clarck (Ed.), *Folk psychology and common sense*. Oxford: Oxford University Press.
- de Gelder, B., & Vroomen, J. (1988). *Bimodal speech perception in young dyslexics*. Paper presented at the 6th Australian Language and Speech Conference, Sydney.
- de Gelder, B., & Vroomen, J. (1989). Models in the mind, modules on the lips. *Behavioral and Brain Sciences*, **124**, 762-763.
- de Gelder, B., & Vroomen, J. (1992). Abstract versus modality-specific memory representations. *Memory and Cognition*, **20**, 533-538.
- de Gelder, B. & Vroomen, J. (in press). Memory for consonants versus vowels in heard and lipread speech. *Journal of Memory and Language*.
- de Gelder, B., Vroomen, J., & van der Heide, L. (1991). Face recognition and lip-reading in autism. *European Journal of Cognitive Psychology*, **3**, 69-86.
- Fodor, J. (1984). Observation reconsidered. *Philosophy of Science*, **51**, 23-43.
- Massaro, D. W. (1987). *Speech perception by ear and eye: a paradigm for psychological inquiry*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Radeau, M. (this volume).
- Summerfield, Q. (1979). Use of visual information for phonetic perception. *Phonetica*, **36**, 314-331.
- Vroomen, J. (1992). *Hearing voices and seeing lips: investigations in the psychology of lipreading*. Doctoral dissertation, Tilburg University.