

*Introduction*

## Multistability in perception: binding sensory modalities, an overview

Jean-Luc Schwartz<sup>1</sup>, Nicolas Grimault<sup>2</sup>, Jean-Michel Hupé<sup>3</sup>,  
Brian C. J. Moore<sup>4,\*</sup> and Daniel Pressnitzer<sup>5,6</sup>

<sup>1</sup>*Gipsa-lab, UMR 5216 CNRS, Grenoble INP, Université Joseph Fourier, Université Stendhal, Grenoble, France*

<sup>2</sup>*Cognition Auditive et Psychoacoustique, Centre de Recherche en Neurosciences de Lyon, UMR 5292, Université Lyon 1, Lyon, France*

<sup>3</sup>*Centre de Recherche Cerveau et Cognition, Université de Toulouse and Centre National de la Recherche Scientifique, 31300 Toulouse, France*

<sup>4</sup>*Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK*

<sup>5</sup>*Laboratoire de Psychologie de la Perception, Université Paris Descartes and Centre National de la Recherche Scientifique, 75006 Paris, France*

<sup>6</sup>*Département d'Etudes Cognitives, Ecole Nationale Supérieure, Paris, France*

This special issue presents research concerning multistable perception in different sensory modalities. Multistability occurs when a single physical stimulus produces alternations between different subjective percepts. Multistability was first described for vision, where it occurs, for example, when different stimuli are presented to the two eyes or for certain ambiguous figures. It has since been described for other sensory modalities, including audition, touch and olfaction. The key features of multistability are: (i) stimuli have more than one plausible perceptual organization; (ii) these organizations are not compatible with each other. We argue here that most if not all cases of multistability are based on competition in selecting and binding stimulus information. Binding refers to the process whereby the different attributes of objects in the environment, as represented in the sensory array, are bound together within our perceptual systems, to provide a coherent interpretation of the world around us. We argue that multistability can be used as a method for studying binding processes within and across sensory modalities. We emphasize this theme while presenting an outline of the papers in this issue. We end with some thoughts about open directions and avenues for further research.

**Keywords:** multistability; multisensory; binding; perceptual organization

### 1. INTRODUCTION

Multistability occurs when a single physical stimulus produces alternations between different subjective percepts. For more than two centuries, it has been a major conceptual and experimental tool for investigating perceptual awareness in vision. This special issue of the *Philosophical Transactions of the Royal Society B* presents recent advances in the study of multistability not only for vision but also for audition and speech, with a combination of psychophysical, physiological and modelling approaches.

This introduction is not intended as a review of multistability, as many excellent reviews are already available [1–7] and more reviews are available in the present issue [8–10]. Rather, the next section presents the motivation for extending the study of visual

multistability to other modalities. The third section describes how the papers presented in the issue contribute to this relatively recent field of research. Finally, some open questions arising from the current state of the field are listed.

Multistability provides a window into the mind, since it gives a natural and unique dissociation between objective properties of the stimulus and subjective sensations: the stimulus properties are constant, whereas sensations change in a dynamic fashion. The study of multistability in several perceptual modalities has the potential to provide a powerful framework for understanding how the different attributes of objects in the environment are bound together, within our perceptual systems, to provide a coherent interpretation of the world around us. This process is known as binding, and it occurs both within and across sensory modalities. As demonstrated by the rich collection of papers in this issue, the expected benefits of the approach are broad, from fundamental theories of the psychology of perception to their underlying neural mechanisms, and from

\* Author for correspondence (bcjm@cam.ac.uk).

One contribution of 10 to a Theme Issue 'Multistability in perception: binding sensory modalities'.

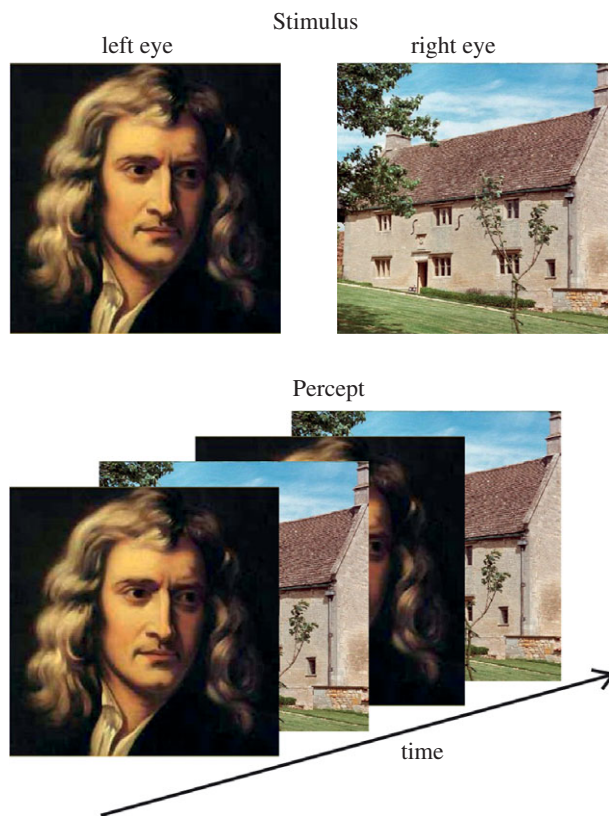


Figure 1. Illustration of binocular rivalry. Different images are presented to the left and right eyes ('Stimulus'). The subject experiences switches from perception of one image (face) to the other (house) ('Percept'). Note that 'mixed percepts' (composed of parts of both images) are also experienced ('piecemeal rivalry'). The phenomenology of binocular rivalry can be experienced with monocular rivalry (see demonstration under the Wikipedia entry).

computational neuroscience to neuro-genetics and the role of spontaneous brain activity in perceptual decision-making.

## 2. MULTISTABILITY FOR DIFFERENT SENSORY MODALITIES

### (a) *Extending the study of multistability and binding from vision to other senses*

Historically, multistability was considered as a visual phenomenon. O'Shea [11] ascribes the first published report of visual multistability to Dutour [12]. This report describes what is now termed 'binocular rivalry' (figure 1). Dutour observed that, when presenting a disc of blue taffeta to one eye and a disc of yellow taffeta to the other eye, he did not see a mixture of the blue and yellow colours. Rather, he was 'unable to detect even the least tint of green'. His conscious experience alternated between blue and yellow. The percept seemed to be dominated by the signal from one of the two eyes at any one time and the eye that was dominant alternated in apparently random fashion. This illustrates the basic characteristic of multistable perception: a static physical stimulus may induce the subjective experience of a percept that is stable over short times, but changes from time to time.

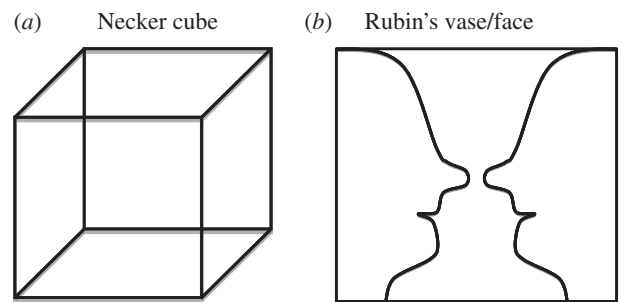


Figure 2. Illustration of ambiguous images. The figure in (a) may be perceived as a cube with either the lower left face or the upper-right face in front. The figure in (b) may be perceived as either a vase or two faces in silhouette. The subject experiences switches from perception of one interpretation to perception of the other.

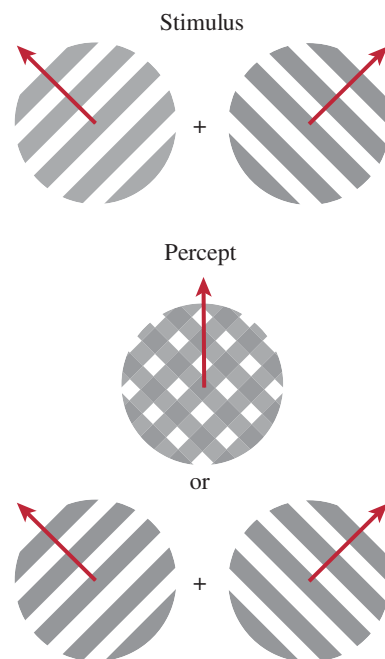


Figure 3. Illustration of moving plaids. Two series of oblique lines (gratings) with orthogonal directions of movements are superimposed ('Stimulus'). The subject may perceive the image ('Percept') as two gratings moving in opposite directions, or as a single cross-hatched object moving upwards (indicated by the arrow). The percept alternates between the two interpretations.

Multistability was also described for ambiguous figures involving depth interpretation (like the two-dimensional outline of a cube, first described by Necker [13]; figure 2a), figure/ground organization (like the Rubin's vase; figure 2b) or motion perception (as in ambiguous motion displays; figure 3). Multistability in binocular rivalry involves perceptual competition between two images, while the multistable perception of ambiguous figures involves competition between interpretations of a single image. Accordingly, the two phenomena have been studied independently during the past two centuries. According to Leopold [14], Walker [15] was the first to suggest that 'a parallel may exist between binocular rivalry and the perceptual reversal of ambiguous figures'. Such a parallel was popularized by Leopold & Logothetis [2].

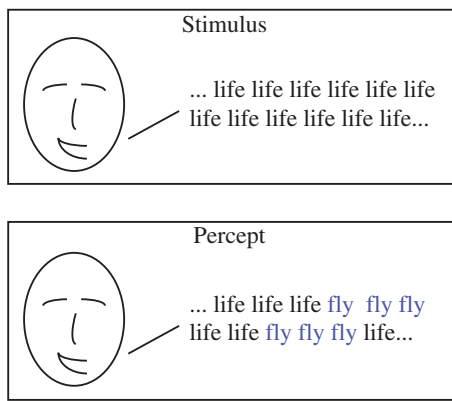


Figure 4. Illustration of the verbal transformation effect. A word is presented repeatedly ('Stimulus', here 'life life life'). After some time, the percept may change ('Percept'), reflecting a different perceptual organization of the sound segments (e.g. 'fly fly fly . . .'), and then may alternate between the two organizations (or other organizations may occur).

Indeed, these apparently disparate stimuli all have some crucial features in common: (i) they have more than one plausible perceptual organization; and (ii) these organizations are not compatible with each other. The perception of such stimuli also shares many similarities: (i) only one interpretation at a time is experienced by observers (and not an 'average' interpretation); (ii) 'flips' in perceived organization occur with prolonged viewing; and (iii) the statistical properties of the multistable alternations are similar across different types of stimuli; they show similar distributions of dominance phases (which percept is dominant), and the distributions are unimodal and asymmetric [2].

Until recently, studies of other perceptual modalities did not capitalize on the large body of accumulated knowledge on visual multistability. However, ambiguous stimuli that gave rise to perceptual alternations had been described for other modalities. For example, when a word or short phrase is presented repeatedly (e.g. 'life life life'), the words or phonemes that are perceived change over time (the 'verbal transformation effect' [16]); for this example, 'fly fly fly' might be perceived after a while (figure 4). Also, when a rapid sequence of tones with different frequencies is presented, the tones may be perceived as coming from a single source, as if played by one instrument (called coherence or fusion), or as multiple sources, as if played by more than one instrument (called stream segregation or fission), and the percept may 'flip' between the two (figure 5, [17]; see Moore & Gockel [8] for a review). It had been noted that the verbal transformation effect provided 'an auditory analogue of the visual reversible figure' and that auditory stream segregation presented a 'striking parallel' with visual apparent motion ([18], p. 21; see also [19]), but the theoretical and experimental tools used to investigate visual multistability were not applied to those stimuli until recently.

Evidence for multistability has been presented for modalities other than vision and audition. Carter *et al.* [20] extended the dynamic dot displays previously used to study visual multistability to touch (figure 6). Zhou & Chen [21] extended binocular

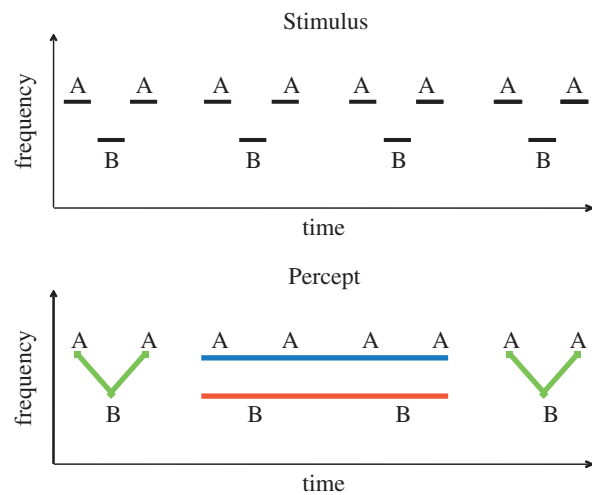


Figure 5. Schematic spectrogram of stimuli used to study auditory streaming. A succession of tones with two different frequencies, A and B, is presented ('Stimulus'). The subject may perceive either a single stream with a 'gallop' rhythm (ABA-ABA-ABA . . ., illustrated by the green lines connecting A and B in 'Percept') or as two regular streams (A-A-A and B-B-B, illustrated by the blue line connecting the A tones and the red line connecting the B tones). The percept can alternate between the two interpretations.

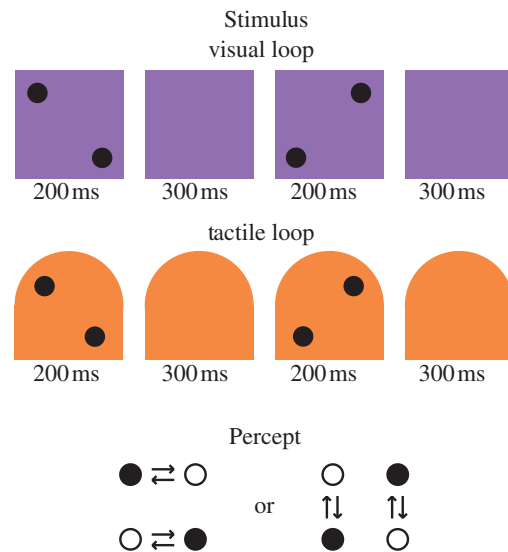


Figure 6. Illustration of motion quartets (visual and tactile). When a subject is presented with two successive visual images ('Visual loop') with two black dots moving from one configuration (on one diagonal) to another (on the inverse diagonal), the subject may perceive either a horizontal or a vertical displacement of the two black dots, and switch from one percept to the other ('Percept'). The same bistability illusion may be obtained with tactile stimuli ('Tactile loop'), using motion touch zones on the thumb.

rivalry to olfaction and reported alternating olfactory percepts when different odorants were presented to the two nostrils (binaral rivalry; figure 7), as well as when presented to the same nostril (mononaral rivalry, supposedly analogous to monocular rivalry). Illusory motion reversals were also reported in proprioception; biceps vibration induces illusory forearm extension, and it was proposed that this phenomenon could be

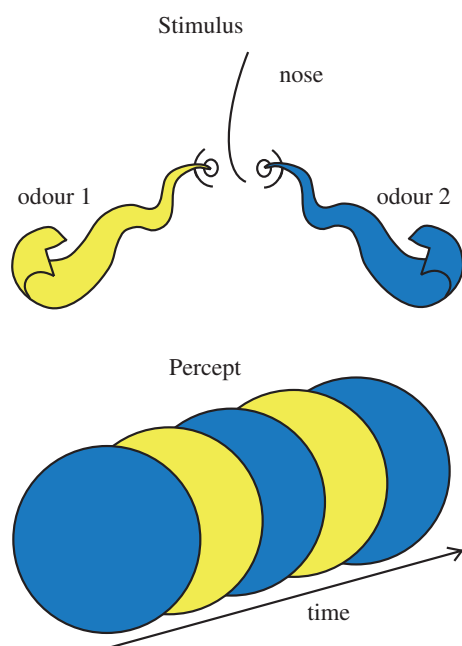


Figure 7. Illustration of binaral rivalry (olfactory). When a subject is presented with two different odours ('Stimulus'), one in each nostril, perception may switch from one odour to the other ('Percept').

an instance of multistability [22]. To our knowledge, no instance of multistability has been reported for taste, but the paradigms developed for olfaction could probably be adapted to taste. In the motor domain, experiments on bimanual rhythmic coordination patterns in response to visual input revealed the presence of a few stable or preferred coordination patterns, suggesting that multistability is also a property of the organization of motor commands [23,24], an argument developed by Kelso [25].

The main rationale for this special issue is that detailed comparisons of the phenomenology of multistability across modalities are now emerging, mostly between vision and audition. Auditory streaming was studied independently by several groups for this purpose [26–28]. For all studies, it was found that the distributions of the random durations of switches in perceptual organization were very similar to those observed for visual multistability. In fact, when measured using the same observers, the dynamics of auditory and visual switching revealed almost identical patterns [26]. Interestingly, bistability for streaming seemed to be the rule rather than the exception, as it could be observed over a surprisingly broad range of stimulus parameters [27,28]. Similar dynamic properties have been observed for the verbal transformation effect [29,30]. Auditory multistability has also been reported with very different stimuli, using rhythmic cues [31].

Leopold & Logothetis [2] proposed that similar mechanisms underlie binocular rivalry and ambiguous figures (a proposal that is still debated; see, for example, Kleinschmidt *et al.* [32]). We propose an extension of this idea and suggest that some common principles might be at work in perceptual organization for different sensory modalities. 'Common principles' could be understood in two ways. Leopold & Logothetis [2]

proposed that perceptual decision-making in multistable perception is triggered by some central, supramodal mechanism. An alternative model is based on the idea that there is more distributed competition [6]. When considering auditory bistability, Pressnitzer & Hupé [26] suggested that functionally similar mechanisms were implemented independently across sensory modalities (see also Hupé *et al.* [33]). According to this view, the specific mechanisms and implementations are likely to differ from one modality to the other, depending on the nature of the physical information and the structure of the sensory inputs. But whatever the modality, the perceptual system must organize the sensory data into a coherent interpretation of the outside world that can be used to guide behaviour. Importantly, when there is more than one plausible interpretation of the sensory evidence, the same phenomenology is observed for all modalities: multistable perception arises, or, in other words, a kind of 'stable instability' seems to be the rule, as Zeki [34] put it for vision.

Extending the study of multistability to sensory modalities other than vision is of interest for at least three reasons. Firstly, it provides a method for studying the neural bases of perceptual organization in those modalities (for hearing, see [30,35–40]). Secondly, the intrinsic characteristics of each sensory modality may extend the scope of the original visual multistability paradigm in important ways. For instance, in audition, the stimuli are by nature time-varying. Competition between perceptual organizations is thus not limited to space or motion direction, but must also involve the time dimension [41]. The interactions between perceptual and motor processes are also quite different between vision (eye movements, [42,43]), touch [20] and speech (e.g. perceptuo-motor theory of speech, [44]). The influence of motor processes can also be more easily controlled in audition, as eye movements are less likely to produce confounding effects. Thirdly, and perhaps most importantly, the extension from vision to other modalities strengthens the hypothesis that multistability is a general property of perceptual systems. Therefore, current research using multistability to probe cognitive processes such as attention, decision-making and consciousness in the visual modality gains further relevance for other modalities.

### (b) *Binding stimulus information within modalities*

Among the great variety of stimuli that evoke multistable perception, it is striking that most are based on competition in selecting and binding stimulus information. An obvious case is provided by binocular rivalry [45]. In this case, multistability involves a perceptual selection between subsets of information from two incompatible sources, one for each eye. The resulting perceived image is sometimes made out of local patches from each of the two eyes, a phenomenon known as piecemeal rivalry. However, most of the time, binding occurs within the image from one eye, and the predominant percept is based on the entire image from that eye. The role of binding is most dramatically illustrated by 'interocular grouping' rivalry [46,47] for which scrambled images are presented

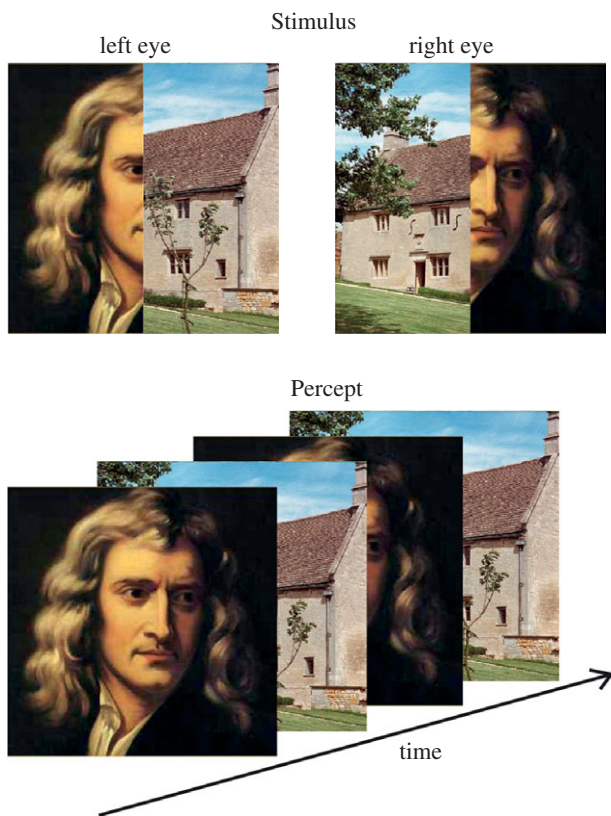


Figure 8. Stimuli producing visual interocular grouping. Different parts of two images are presented to the left and the right eyes ('Stimulus'). The subject perceptually reconstructs the original images and experiences switches from perception of one figure to perception of the other ('Percept').

to each eye (figure 8). Patches of images belonging to a face and a house are presented to one eye, while the complementary patches are presented to the other eye. If multistable competition were purely eye-based, observers would experience alternations of scrambled images. This is not what occurs: instead, binding of elements that form a coherent image occurs, and the result of the multistable competition is usually the percept of either a face or a house. Thus, information from the two eyes is selected and combined to form meaningful objects.

The case of ambiguous images (figure 2) also involves selection and binding. In classical examples such as Rubin's vase–face image, selection involves deciding which parts of the image are assigned to the foreground and which are assigned to the background; the components making up the foreground should be bound together and segregated from components of the background. In other cases, such as the Necker cube, binding may occur within the foreground to determine which segments form the front face of the cube. Dynamic displays also involve binding, both in space and time. The 'moving plaid' stimulus (figure 3) is perceived as one object moving in one direction or two superimposed objects moving in opposite directions, depending on whether the grating components of the plaid are bound or segmented. These represent two very different bindings of components within and across the moving images.

Binding also plays an important role in multistability for audition and speech. For verbal transformations, sound segments are bound in different orders or with different segment boundaries to generate new percepts (figure 4). In the case of auditory streaming, successive sounds are either bound into one stream (heard as if coming from one source), or bound into two streams (heard as if coming from two sources) (figure 5).

We argue here that it is revealing that multistability always involves perceptual binding, especially if one considers that the most common situation leading to perceptual ambiguity is not based on binding and selection. Consider 'boundary' stimuli, which have features close to a boundary between two perceptual categories along a perceptual continuum. For example, in vision, one boundary stimulus is a colour between blue and green. Such a stimulus would appear to possess the correct properties for being ambiguous: one is not certain whether the colour is green or blue. However, to the best of our knowledge, there are no experimental data showing that multistability can occur for such stimuli. Thus, boundary stimuli that do not involve ambiguous binding do not seem to produce multistability. This issue has seldom been considered (but see the discussion of the possible epistemological distinction between multistability and ambiguity in Egré [48]) and remains to be addressed experimentally.

Overall, the view that emerges from studies of different modalities is that multistability reflects processes of competition between different perceptual organizations of the same scene, where the binding of sensory information is always involved.

### (c) *Binding across perceptual modalities*

Since multistability involves binding in various modalities, what kind of multistable phenomena might emerge when more than one modality is simultaneously involved? This raises the question of the level at which multisensory interactions happen, relative to binding within modalities, and more generally, leads to the possibility that multistability could be used as a tool for studying multisensory perceptual organization.

Let us begin by considering the possibility that multistable effects in one modality can be modified by stimuli in another modality. It has been shown that binocular rivalry can be influenced by sounds congruent with one or the other image, but only when the visual stimulus is consciously perceived, not when it is suppressed from awareness [49,50]. In the same vein, Munhall *et al.* [51] have shown a McGurk effect (the identity of a speech sound being influenced by visual information from the face of the talker) with moving lips on an ambiguous face/vase stimulus only when it was perceived as a face. These results suggest that audio-visual integration in these situations happens only *after* binding is resolved within each modality. Moreover, stimuli in one modality may influence the bistable perception of ambiguous stimuli in a second modality, but only when subjects pay attention to the stimulus in the first modality [52]. Hupé *et al.* [33] presented multistable stimuli in the auditory and visual modalities, using auditory stimuli that led to streaming and visual stimuli that led to ambiguous

motion. They reported large cross-modal influences, whose magnitude depended on audio-visual congruence; perceptual switches in one modality could modulate switches in the other modality. However, the timing of the modulation was quite sluggish, suggesting that it was mediated by contextual processes and hence that perceptual organization initially occurred separately in each modality. However, Takahashi & Watanabe [53] showed that changes in auditory stimuli that the subjects were not aware of had weak (and delayed) effects on the dynamics of visual apparent motion. Also, the (supposedly implicit) semantic content of auditory stimuli influenced the balance of percepts in binocular rivalry by a few per cent [54]. This leaves open the possibility that some cross-modal interactions happen before the completion of perceptual organization within each modality. Klink *et al.* [9] further discuss cross-modal effects as a form of ‘contextual information’ used to disambiguate the sensory input.

These (partly conflicting) results raise the question of how multisensory perceptual organization occurs. The issue here is to know at what level perceptual objects are best defined: are perceptual objects constructed independently for each modality before interactions occur at a relatively high level of processing, or can a common object representation be formed for different modalities to mediate interactions at an early stage of processing? This question is well exemplified by the contrast between two positions: Kubovy & Van Valkenburg [55] argue that auditory objects differ from visual objects because of the intrinsic structure of auditory and visual processing, whereas the classical assumption in speech perception is that speech objects are multisensory and hence that there is a common representational format for auditory, visual and motor speech at some level of processing [56,57].

Assuming that multistability is a result of the competition between perceptual organizations, the existence of audio-visual objects would be indicated by simultaneous switches of auditory and visual organizations under conditions of multisensory multistability. The finding of Hupé *et al.* [33] that audio-visual capture for apparent motion did not lead to simultaneous switches in the two modalities led them to argue against the existence of a specific ‘audio-visual apparent motion’ object (see also Kubovy & Yu [58]). However, synchronous lights and sounds are not necessarily perceived as a unified perceptual object, while the speech percept routinely depends on combining information from the auditory and visual modalities [59]. This is what led Sato *et al.* [60] to claim that ‘multistable speech perception is indeed a multisensory effect’. Their set of experiments involved for the first time audio-visual verbal transformations, that is, multistability in speech with multisensory inputs. They showed that the visual input modifies the perceptual stability of the auditory input and that switches applied to the visual input could largely drive the audio-visual percept by inducing rather synchronous switches in perception. Altogether, their results demonstrate the capacity of visual information to control the multistable perception of speech in its phonetic content and temporal course.

Hence, the two modalities seem to be bound together in the multistability phenomenon in this case. This suggests that the multistability paradigm may provide an effective tool for determining if ‘multisensory’ objects exist for speech [61].

### 3. CONTRIBUTION OF THE PAPERS FROM THIS ISSUE

This issue covers most of the facets of multistability and binding that we have just described, in audition and vision. For all of the contributions, a useful distinction to keep in mind is between *what* competes and *how* competition takes place [33]. *What* competes is the content of sensory experience, the components of the stimulus that have to be bound into perceptual objects, corresponding to the ‘neural events associated with the representation of a given perceptual state’ [62]. *How* competition occurs depends on the neural processes ‘that are responsible for switches between alternative perceptual states’ [62]. The contents of perceptual experience, *what* competes, are obviously different for visual and auditory multistability. The question highlighted in this introduction is whether the mechanisms of switching share some principles and/or neural processes in vision and audition. The question is thus related to *how* competition takes place. Phrased differently, the question is ‘what determines the change in perceptual organization after the observer has been perceiving the stimulus in a particular way’ [63].

While the *what* and *how* questions are independent in principle, empirical evidence does not always provide a basis for distinguishing them unequivocally. For example, the effect of intention on the dynamics of bistable perception may be interpreted as revealing the mechanisms of switching, or as ‘simply’ affecting the content of one or the other representation. The same can be said for the effects of attention, adaptation or even ‘noise’. It is therefore paramount to know precisely what factors influence competing percepts in audition and vision before being able to address the question of the switching mechanisms.

The two papers following this introduction mostly focus on the question of *what* competes in individual sensory modalities. One deals with auditory streaming, which is a topic of several contributions in this issue. As we have seen, auditory streaming has recently been a key paradigm for building bridges between studies of visual and auditory multistability. Moore & Gockel [8] provide an up-to-date overview of the streaming paradigm. Importantly, they provide a comprehensive survey of the many kinds of acoustic cues and other experimental parameters (like attention, time and sudden changes in the stimulus sequence) that can affect auditory binding in the streaming paradigm. In the next contribution, Klink *et al.* [9] review the many factors influencing multistability, mostly for vision but also including cross-modal influences. They consider multistability as an optimal paradigm for studying how perceptual systems can produce context-driven inference and decisions. They consider four major kinds of context (temporal, spatial, multisensory, and associated with the subject’s internal state).

The next three contributions explore the *what* question by comparing multistability for audition and vision. Hupé & Pressnitzer [64] present experimental data comparing the initial phase of perceptual organization for visual plaids and auditory streaming. This phase exhibits a peculiar pattern: it is longer than later phases and is biased towards one object for both modalities. In vision, they show that it is the tristable nature of plaid perception that produces the longer percept, whereas the evidence is less clear cut for audition. In fact, tristability for auditory streaming remains to be shown. The bias towards integration is discussed in terms of local versus global organization cues. Kubovy & Yu [58] re-examine the similarities and differences between the requirements of perceptual scene analysis in audition and vision. They suggest that cross-modal causality is a necessary prerequisite for efficient cross-modal binding, but they express doubts that such cross-modal binding could result in multisensory multistability, because of the intrinsic difference in nature between perceptual objects in different modalities (here, audition and vision). However, they consider the hypothesis that speech could be an exception to this general view. Directly related to this conjecture, Basirat *et al.* [61] demonstrate how the verbal transformation effect can be a valuable tool for studying the perceptual organization of speech, returning to old but key questions in speech perception, such as the role of perceptuo-motor interactions and the nature of the representation units. They claim that since the objects of speech perception are intrinsically multisensory, they may lead to multisensory multistability.

The last four contributions examine the switching mechanisms (*how* competition takes place) from the perspective of its neural bases in audition [65] and vision [32], from a theoretical perspective in audition [10], or in terms of computational processes and dynamic systems, whatever the sensory modality [25].

Kashino & Kondo [65] compare the neural bases of switching for two multistability paradigms in audition, auditory streaming and the verbal transformation effect. Functional MRI data acquired using similar paradigms for both phenomena allow them to make a direct comparison, and reveal the role of motor-based processes in multistability for both non-speech and speech sounds, in addition to the involvement of sensory regions dedicated to audition (auditory cortex and thalamus). Moreover, activity in the motor structures is shown to be correlated with individual switching rates. This variability may be a result of genetically determined differences in the catecholaminergic system. Kleinschmidt *et al.* [32] review the variations of neural activity that have been observed in relation to visual multistability, mostly with ambiguous figures. Importantly, they use the association between neural fluctuations and switches in perceptual states to decide 'where does brain activity reflect perceptual dominance' (our *what* question) 'and where does brain activity reflect perceptual alternations' (our *how* question), and they discuss to what extent studies of the relative timing of neural and perceptual events and studies using transcranial magnetic stimulation can resolve this issue. Like Kashino & Kondo, Kleinschmidt *et al.*

discuss individual differences in brain state fluctuations, and relate them to neuroanatomical substrates, suggesting that there could be a genetic basis for differences between subjects in how they behave in multistability paradigms, focusing on a causal role of parietal regions in perceptual inference.

Winkler *et al.* [10] present a theoretical framework for auditory streaming based on the idea of 'predictive coding'. In this approach, the goal of perceptual organization is to find regularities in the incoming sensory information, in order to predict the pattern of future sounds. Interestingly, the competition in their framework is not between sensory representations, but rather between abstract rules that bind successive sounds together. They also suggest a new computational approach for understanding the competition between those rules, related to the *how* question.

Finally, Kelso [25] describes how multistable perception can be considered as part of a wide range of multistable phenomena in living systems. He relates these to adaptability and the ability to dynamically define self-organizing functional grouping of individual elements to optimize specific behaviours. In this sense, he switches from the *how* to a possible *why* question, in which multistability appears as one component of a global process that allows a creative organism to adapt and invent solutions to deal with a highly complex environment.

#### 4. OPEN QUESTIONS AND DIRECTIONS

##### (a) *Extending the range of multistable phenomena in various modalities*

We have seen how the study of multistability has been extended from vision to other modalities. However, the range of multistable phenomena in these modalities is still rather limited compared with the rich set available to visual scientists. This range will probably be extended in the future. For instance, until now, auditory multistability has involved stimuli that unfold over time, such as rapid sound sequences, but there has been no published report of an effect of 'binaural rivalry' comparable to binocular rivalry. Deutsch [66] discovered a kind of 'interaural grouping' illusion, but it is unclear whether this produces multistable perception. The difference between vision and hearing could occur because totally different images in the two eyes are highly unlikely and thus incompatible, whereas sounds often differ somewhat at the two ears owing to head-shadow effects, and the sounds at the two ears can be very different when the sound sources are very close to the ears. This usually results in the perception of multiple sound sources at different positions in space, rather than rivalry between perceived sources. A possible auditory analogue of interocular grouping rivalry could occur under conditions where perception of a sound depends on combining information across the two ears, for example, when part of a speech sound (e.g. the first and second formants) is presented to one ear and the remainder (e.g. the third formant) to the other ear [67]; see also the further experiments on 'duplex perception' conducted by Liberman *et al.* [68].

Analogues of the verbal transformation effect might also be found for non-speech sounds or modalities other than hearing. An aspect of verbal transformations is that they involve different ways of sorting sounds into segments (the 'segmentation' problem, which is crucial for speech perception). Similar transformations might occur for non-speech sounds or visual or even tactile stimuli, provided that there are multiple possible ways of segmentation, each of which gives rise to a plausible perceptual interpretation of the input. We could also ask whether multistability in touch extends beyond motion perception, by looking for a touch analogue of the auditory streaming paradigm.

**(b) *Is conflicting binding required for multistability?***

As discussed earlier, boundary stimuli can have more than one interpretation, but such stimuli do not seem to trigger multistable perception, perhaps because they do not involve conflicting binding cues. This conjecture requires further experimental testing. It seems likely that a given boundary stimulus, such as an image containing a colour between green and blue or a synthetic speech sound between 'ba' and 'da', could lead to responses that vary over time. But this may reflect the ambiguous nature of what is perceived rather than reflecting a flip from one percept to another. This could be assessed by collecting confidence judgements of the subject and not only categorical decisions. Another assessment method would be to use reaction times: ambiguity in categorical judgements leads to increased response latency [69] while bistable perception does not [70]. Comparison of response latencies for stimuli involving conflicting binding cues (e.g. binocular rivalry) and those not involving such cues (e.g. a colour between green and blue) could reveal whether or not the latter involve bistability.

Such methodologies could be used in other modalities, for example olfaction. The results of Zhou & Chen [21] on mononasal rivalry suggested that it is not possible to experience two different odours at the same time. The generality of this finding should be tested for a large variety of pairs of odours. Also, it needs to be determined whether the responses reflect a categorical judgement (see also Gottfried [71]). Further evidence for multistability in mononasal olfaction would clarify the issue of what constitutes a 'perceptual object' in olfaction. Following the general framework set up in this introduction, the rivalry between two odours may imply that each odour represents a different perceptual object. Similar methods could be applied to assess the phenomenological level of perceptual organization in touch, proprioception and perhaps even taste.

**(c) *Subjectivity, individual differences and multistability***

Multistability opens a window on the subjective experience of the perceiver, by using stimuli that are physically stable but lead to a rich and a diverse phenomenology. The study of multistability can therefore play a crucial role in understanding the characteristics of the construction of perceptual awareness. It has been known

for a long time that individuals differ markedly in the rate of the alternation between alternative perceptual organizations, but not in the general distribution of stability periods. The extent to which these individual differences are consistent across stimuli is still a matter of debate.

Intra-subject consistency has been observed within a modality ([72] for vision; and, in this issue, within an auditory non-speech and a speech task [65]). Results across modalities are contradictory, with negative results for audition and vision [26], and for touch and vision [20], but significant correlations across all bistable paradigms, in vision and audition [73]. However, the development of measures and adequate paradigms for assessing inter-task and inter-modality correlations is far from trivial. The switching rate between alternative percepts depends on the stimuli used for a given task, with more switches when the two interpretations are equally likely, which leads to equal dominance of the percepts [74]. Hence, the stimulus parameters should be carefully matched across tasks and modalities and calibrated for each subject, in order to achieve equal dominance, which has not always been done and is, in any case, difficult to do. A number of contextual parameters (such as the general level of attention and arousal) and possible artefacts (e.g. the role of eye movements) are likely to introduce variability and biases into the results.

In conclusion, the study of multistability, extending from vision to audition and speech in the present issue, and potentially to many other sensory modalities, provides a window for examining many factors that are associated with or influence perceptual binding, both within and across sensory modalities. These include: attention, decision-making and consciousness; the cognitive state of the perceiving brain; mental disorders, pharmacology and genetics; and even creativity and culture. A striking example of the latter is a type of French slang called 'verlan', in which words are created by reversing the order of the syllables. The word *verlan* itself is a *verlan* word: the French for 'reverse' is 'l'envers', and pronouncing 'l'envers' with the syllables in reverse order results in 'verlan'. The word 'verlan' and more generally most slang constructions in *verlan*, may have arisen through repetition of 'l'envers' via the verbal transformation effect. Artists such as Salvador Dali have used ambiguous figures in their work, for example, 'Slave market with the disappearing bust of Voltaire', and composers such as Bach have exploited auditory streaming to create the impression of two melodic lines coming from an instrument such as the flute, which produces only one note at a time (see the cover of this issue).

In summary, multistability in vision and audition has long fascinated researchers, artists, composers, and philosophers. The study of multistability is being extended to most sensory and motor modalities and to cross-modal perception. This fascinating landscape is explored in the wide-ranging papers in this volume.

This work was supported by the 'Agence Nationale de la Recherche' (ANR-08-BLAN-0167-01, project Multistap). The work of BCJM was supported by the MRC (UK). We thank Agnès Léger for creating the figures used in this article.



## REFERENCES

- 1 Blake, R. 1989 A neural theory of binocular rivalry. *Psychol. Rev.* **96**, 145–167. (doi:10.1037/0033-295X.96.1.145)
- 2 Leopold, D. A. & Logothetis, N. K. 1999 Multistable phenomena: changing views in perception. *Trends Cogn. Sci.* **3**, 254–264. (doi:10.1016/S1364-6613(99)01332-7)
- 3 Blake, R. & Logothetis, N. K. 2002 Visual competition. *Nat. Rev. Neurosci.* **3**, 13–21. (doi:10.1038/nrn701)
- 4 Long, G. M. & Toppino, T. C. 2004 Enduring interest in perceptual ambiguity: alternating views of reversible figures. *Psychol. Bull.* **130**, 748–768. (doi:10.1037/0033-2909.130.5.748)
- 5 Kim, C. Y. & Blake, R. 2005 Psychophysical magic: rendering the visible ‘invisible’. *Trends Cogn. Sci.* **9**, 381–388. (doi:10.1016/j.tics.2005.06.012)
- 6 Tong, F., Meng, M. & Blake, R. 2006 Neural bases of binocular rivalry. *Trends Cogn. Sci.* **10**, 502–511. (doi:10.1016/j.tics.2006.09.003)
- 7 Sterzer, P., Kleinschmidt, A. & Rees, G. 2009 The neural bases of multistable perception. *Trends Cogn. Sci.* **13**, 310–318. (doi:10.1016/j.tics.2009.04.006)
- 8 Moore, B. C. J. & Gockel, H. E. 2012 Properties of auditory stream formation. *Phil. Trans. R. Soc. B* **367**, 919–931. (doi:10.1098/rstb.2011.0355)
- 9 Klink, P. C., van Wezel, R. J. A. & van Ee, R. 2012 United we sense, divided we fall: context-driven perception of ambiguous visual stimuli. *Phil. Trans. R. Soc. B* **367**, 932–941. (doi:10.1098/rstb.2011.0358)
- 10 Winkler, I., Denham, S., Mill, R., Böhm, T. M. & Bendixen, A. 2012 Multistability in auditory stream segregation: a predictive coding view. *Phil. Trans. R. Soc. B* **367**, 1001–1012. (doi:10.1098/rstb.2011.0359)
- 11 O’Shea, R. P. 1999 Translation of Dutour (1760). See [http://psy.otago.ac.nz/r\\_oshea/WebPubs/dutour60.html](http://psy.otago.ac.nz/r_oshea/WebPubs/dutour60.html) (retrieved 28 November 2011).
- 12 Dutour, E.-F. 1760 Discussion d’une question d’optique [Discussion on a question of optics]. *l’Acad. Sci. Mém. Math. Phys. Présentés Divers Savants* **3**, 514–530.
- 13 Necker, L. A. 1832 Observations on some remarkable optical phenomena seen in Switzerland; and on an optical phenomenon which occurs on viewing a figure of a crystal or geometrical solid. *Lond. Edinb. Phil. Mag. J. Sci.* **1**, 329–337.
- 14 Leopold, D. A. 1997 Brain mechanisms of visual awareness. Using perceptual ambiguity to investigate the neural basis of image segmentation and grouping. PhD thesis, Baylor College of Medicine, Texas, USA.
- 15 Walker, P. 1975 Stochastic properties of binocular rivalry alternations. *Percept. Psychophysol.* **18**, 467–473. (doi:10.3758/BF03204122)
- 16 Warren, R. M. & Gregory, R. L. 1958 An auditory analogue of the visual reversible figure. *Am. J. Psychol.* **71**, 612–613. (doi:10.2307/1420267)
- 17 Miller, G. A. & Heise, G. A. 1950 The trill threshold. *J. Acoust. Soc. Am.* **22**, 637–638. (doi:10.1121/1.1906663)
- 18 Bregman, A. S. 1990 *Auditory scene analysis: the perceptual organization of sound*. Cambridge, MA: Bradford Books, MIT Press.
- 19 Anstis, S. & Saida, S. 1985 Adaptation to auditory streaming of frequency-modulated tones. *J. Exp. Psychol. Human Percept. Perf.* **11**, 257–271. (doi:10.1037/0096-1523.11.3.257)
- 20 Carter, O., Konkle, T., Wang, Q., Hayward, V. & Moore, C. 2008 Tactile rivalry demonstrated with an ambiguous apparent-motion quartet. *Curr. Biol.* **18**, 1050–1054. (doi:10.1016/j.cub.2008.06.027)
- 21 Zhou, W. & Chen, D. 2009 Binaral rivalry between the nostrils and in the cortex. *Curr. Biol.* **19**, 1561–1565. (doi:10.1016/j.cub.2009.07.052)
- 22 Holcombe, A. O. & Seizova-Cajic, T. 2008 Illusory motion reversals from unambiguous motion with visual, proprioceptive, and tactile stimuli. *Vis. Res.* **48**, 1743–1757. (doi:10.1016/j.visres.2008.05.019)
- 23 Zanone, P. G. & Kelso, J. A. S. 1992 The evolution of behavioral attractors with learning: nonequilibrium phase transitions. *J. Exp. Psychol. Hum. Percept. Perf.* **18**, 403–421. (doi:10.1037/0096-1523.18.2.403)
- 24 Kelso, J. A. S. & Zanone, P. G. 2002 Coordination dynamics of learning and transfer across different effector systems. *J. Exp. Psychol. Hum. Percept. Perf.* **28**, 776–797. (doi:10.1037/0096-1523.28.4.776)
- 25 Kelso, J. A. S. 2012 Multistability and metastability: understanding dynamic coordination in the brain. *Phil. Trans. R. Soc. B* **367**, 906–918. (doi:10.1098/rstb.2011.0351)
- 26 Pressnitzer, D. & Hupé, J. M. 2006 Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Curr. Biol.* **16**, 1351–1357. (doi:10.1016/j.cub.2006.05.054)
- 27 Denham, S. L. & Winkler, I. 2006 The role of predictive models in the formation of auditory streams. *J. Physiol. Paris* **100**, 154–170. (doi:10.1016/j.jphysparis.2006.09.012)
- 28 Kashino, M., Okada, M., Mizutani, S., Davis, P. & Kondo, H. M. 2007 The dynamics of auditory streaming: psychophysics, neuroimaging, and modeling. In *Hearing—from sensory processing to perception* (eds B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Upperkamp & J. Verhey), pp. 275–283. Berlin, Germany: Springer.
- 29 Sato, M., Schwartz, J. L., Abry, C., Cathiard, M. A. & Loevenbruck, H. 2006 Multistable syllables as enacted percepts: a source of an asymmetric bias in the verbal transformation effect. *Percept. Psychophysol.* **68**, 458–474. (doi:10.3758/BF03193690)
- 30 Kondo, H. M. & Kashino, M. 2007 Neural mechanisms of auditory awareness underlying verbal transformations. *NeuroImage* **36**, 123–130. (doi:10.1016/j.neuroimage.2007.02.024)
- 31 Repp, B. H. 2007 Hearing a melody in different ways: multistability of metrical interpretation, reflected in rate limits of sensorimotor synchronization. *Cognition* **102**, 434–454. (doi:10.1016/j.cognition.2006.02.003)
- 32 Kleinschmidt, A., Sterzer, P. & Rees, G. 2012 Variability of perceptual multistability: from brain state to individual trait. *Phil. Trans. R. Soc. B* **367**, 988–1000. (doi:10.1098/rstb.2011.0367)
- 33 Hupé, J. M., Joffo, L.-M. & Pressnitzer, D. 2008 Bistability for audio-visual stimuli: perceptual decision is modality specific. *J. Vis.* **8**(7), 1. (doi:10.1167/8.7.1)
- 34 Zeki, S. 2004 The neurology of ambiguity. *Conscious. Cogn.* **13**, 173–196. (doi:10.1016/j.concog.2003.10.003)
- 35 Cusack, R. 2005 The intraparietal sulcus and perceptual organization. *J. Cogn. Neurosci.* **17**, 641–651. (doi:10.1162/0898929053467541)
- 36 Micheyl, C., Tian, B., Carlyon, R. P. & Rauschecker, J. P. 2005 Perceptual organization of sound sequences in the auditory cortex of awake macaques. *Neuron* **48**, 139–148. (doi:10.1016/j.neuron.2005.08.039)
- 37 Gutschalk, A., Micheyl, C., Melcher, J. R., Rupp, A., Scherg, M. & Oxenham, A. J. 2005 Neuromagnetic correlates of streaming in human auditory cortex. *J. Neurosci.* **25**, 5382–5388. (doi:10.1523/JNEUROSCI.0347-05.2005)
- 38 Basirat, A., Sato, M., Schwartz, J. L., Kahane, P. & Lachaux, J. P. 2008 Parieto-frontal oscillatory synchronization during the perceptual emergence and stabilization of speech forms. *NeuroImage* **42**, 404–413. (doi:10.1016/j.neuroimage.2008.03.063)

- 39 Pressnitzer, D., Sayles, M., Micheyl, C. & Winter, I. M. 2008 Perceptual organization of sound begins in the auditory periphery. *Curr. Biol.* **18**, 1124–1128. (doi:10.1016/j.cub.2008.06.053)
- 40 Snyder, J. S., Alain, C. & Picton, T. W. 2006 Effects of attention on neuroelectric correlates of auditory stream segregation. *J. Cogn. Neurosci.* **18**, 1–13. (doi:10.1162/089892906775250021)
- 41 Winkler, I., Denham, S. L. & Nelken, I. 2009 Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends Cogn. Sci.* **13**, 532–540. (doi:10.1016/j.tics.2009.09.003)
- 42 Ellis, S. R. & Stark, L. 1978 Eye movements during the viewing of Necker cubes. *Perception* **7**, 575–581. (doi:10.1068/p070575)
- 43 Sabrin, H. W. & Kertesz, A. E. 1980 Microsaccadic eye movements and binocular rivalry. *Percept. Psychophysol.* **28**, 150–154. (doi:10.3758/BF03204341)
- 44 Schwartz, J. L., Basirat, A., Ménard, L. & Sato, M. 2010 The perception-for-action-control theory (PACT): a perceptuo-motor theory of speech perception. *J. Neurolinguist.* **1**–19. (doi:10.1016/j.jneuroling.2009.12.004)
- 45 Blake, R. & O'Shea, R. P. 2009 Binocular rivalry. In *Encyclopedia of neuroscience*, vol. 2 (ed. L. Squire), pp. 179–187. Oxford, UK: Academic Press.
- 46 Kovacs, I., Papatomas, T. V., Yang, M. & Feher, A. 1996 When the brain changes its mind: interocular grouping during binocular rivalry. *Proc. Natl Acad. Sci. USA* **93**, 15 508–15 511.
- 47 Logothetis, N. K. 1998 Single units and conscious vision. *Phil. Trans. R. Soc. Lond. B* **353**, 1801–1818. (doi:10.1098/rstb.1998.0333)
- 48 Egré, P. 2009 Soritical series and Fisher series. In *Reduction between the mind and the brain* (eds H. Leitgeb & A. Hieke), pp. 91–115. Heusenstamm, Germany: Ontos-Verlag.
- 49 Kang, M. S. & Blake, R. 2005 Perceptual synergy between seeing and hearing revealed during binocular rivalry. *Psychologija* **32**, 7–15.
- 50 Conrad, V., Bartels, A., Kleiner, M. & Noppeney, U. 2010 Audiovisual interactions in binocular rivalry. *J. Vis.* **10**(10), 27. (doi:10.1167/10.10.27)
- 51 Munhall, K., ten Hove, M., Brammer, M. & Pare, M. 2009 Audiovisual integration of speech in a bistable illusion. *Curr. Biol.* **19**, 735–739. (doi:10.1016/j.cub.2009.03.019)
- 52 van Ee, R., van Boxtel, J. J. A., Parker, A. L. & Alais, A. 2009 Multisensory congruency as a mechanism for attentional control over perceptual selection. *J. Neurosci.* **29**, 11 641–11 649. (doi:10.1523/JNEUROSCI.0873-09.2009)
- 53 Takahashi, K. & Watanabe, K. 2010 Implicit auditory modulation on the temporal characteristics of perceptual alternation in visual competition. *J. Vis.* **10**(4), 11. (doi:10.1167/10.4.11)
- 54 Chen, Y. C., Yeh, S. L. & Spence, C. 2011 Crossmodal constraints on human perceptual awareness: auditory semantic modulation of binocular rivalry. *Front Psychol.* **2**, 212. (doi:10.3389/fpsyg.2011.00212)
- 55 Kubovy, M. & Van Valkenburg, D. 2001 Auditory and visual objects. *Cognition* **80**, 97–126. (doi:10.1016/S0010-0277(00)00155-4)
- 56 Summerfield, Q. 1987 Some preliminaries to a comprehensive account of audio-visual speech perception. In *Hearing by eye: the psychology of lip-reading* (eds B. Dodd & R. Campbell), pp. 3–51. London, UK: Lawrence Erlbaum Associates.
- 57 Schwartz, J. L., Robert-Ribes, J. & Escudier, P. 1998 Ten years after Summerfield ... a taxonomy of models for audiovisual fusion in speech perception. In *Hearing by Eye, II. Perspectives and directions in research on audiovisual aspects of language processing* (eds R. Campbell, B. Dodd & D. Burnham), pp. 85–108. Hove, UK: Psychology Press.
- 58 Kubovy, M. & Yu, M. 2012 Multistability, cross-modal binding and the additivity of conjoined grouping principles. *Phil. Trans. R. Soc. B* **367**, 954–964. (doi:10.1098/rstb.2011.0365)
- 59 Campbell, R. 2008 The processing of audio-visual speech: empirical and neural bases. *Phil. Trans. R. Soc. B* **363**, 1001–1010. (doi:10.1098/rstb.2007.2155)
- 60 Sato, M., Basirat, A. & Schwartz, J. L. 2007 Visual contribution to the multistable perception of speech. *Percept. Psychophysol.* **69**, 1360–1372. (doi:10.3758/BF03192952)
- 61 Basirat, A., Schwartz, J.-L. & Sato, M. 2012 Perceptuo-motor interactions in the perceptual organization of speech: evidence from the verbal transformation effect. *Phil. Trans. R. Soc. B* **367**, 965–976. (doi:10.1098/rstb.2011.0374)
- 62 Kang, M. S. & Blake, R. 2010 What causes alternations in dominance during binocular rivalry? *Atten. Percept. Psychophysol.* **72**, 179–186. (doi:10.3758/APP.72.1.179)
- 63 Rock, I. 1975 *An introduction to perception*. New York, NY: Macmillan.
- 64 Hupé, J.-M. & Pressnitzer, D. 2012 The initial phase of auditory and visual scene analysis. *Phil. Trans. R. Soc. B* **367**, 942–953. (doi:10.1098/rstb.2011.0368)
- 65 Kashino, M. & Kondo, H. M. 2012 Functional brain networks underlying perceptual switching: auditory streaming and verbal transformations. *Phil. Trans. R. Soc. B* **367**, 977–987. (doi:10.1098/rstb.2011.0370)
- 66 Deutsch, D. 1974 An auditory illusion. *Nature* **251**, 307–309. (doi:10.1038/251307a0)
- 67 Rand, T. C. 1974 Dichotic release from masking for speech. *J. Acoust. Soc. Am.* **55**, 678–680. (doi:10.1121/1.1914584)
- 68 Liberman, A. M., Isenberg, D. & Rakerd, B. 1981 Duplex perception of cues for stop consonants: evidence for a phonetic mode. *Percept. Psychophysol.* **30**, 133–143. (doi:10.3758/BF03204471)
- 69 Ratcliff, R. & Rouder, J. 1998 Modeling response times for two-choice decisions. *Psychol. Sci.* **9**, 347–356. (doi:10.1111/1467-9280.00067)
- 70 Takei, S. & Nishida, S. 2010 Perceptual ambiguity of bistable visual stimuli causes no or little increase in perceptual latency. *J. Vis.* **10**(4), 23. (doi:10.1167/10.4.23)
- 71 Gottfried, J. A. 2009 Olfaction: when nostrils compete. *Curr. Biol.* **19**, R862–R864. (doi:10.1016/j.cub.2009.08.030)
- 72 Carter, O. & Pettigrew, J. 2003 A common oscillator for perceptual rivalries? *Perception* **32**, 295–305. (doi:10.1068/p3472)
- 73 Kondo, H. M., Kitagawa, N., Kitamura, M., Nomura, M. & Kashino, M. In press. Separability and commonality of auditory and visual bistable perception. *Cereb. Cortex* (doi:10.1093/cercor/BHR266)
- 74 Moreno-Bote, R., Shapiro, A., Rinzal, J. & Rubin, N. 2010 Alternation rate in perceptual bistability is maximal at and symmetric around equi-dominance. *J. Vis.* **10**(11), 1. (doi:10.1167/10.11.1)