Bistability for audiovisual stimuli: Perceptual decision is modality specific

Jean-Michel Hupé   CerCo, Université de Toulouse & UMR5549, CNRS, Toulouse, France

Lu-Ming Joffo   LPP, CNRS-Université Paris Descartes & DEC, ENS, Paris, France

Daniel Pressnitzer   LPP, CNRS-Université Paris Descartes & DEC, ENS, Paris, France

Ambiguous stimuli can produce spontaneous perceptual alternations in the mind of the observer, even though the stimulus itself remains the same. Common features in the temporal dynamics of bistability have been observed for various types of stimuli, both visual and auditory. This raises the question of whether bistable perception results from stereotyped, local competition between stimulus-specific representations or whether it is triggered by some central, supramodal mechanism. We tested the distributed versus centralized hypothesis by asking observers to simultaneously monitor their bistable perception of ambiguous auditory and visual stimuli. Strong interactions between auditory and visual perceptual switches would indicate a central decision mechanism. We used streaming stimuli in the auditory modality and either plaids or apparent motion stimuli in the visual modality. The use of two different sensory modalities allowed the distinction of contextual interactions due to the similarity between stimuli from interactions linked to perceptual decision itself. The long-term dynamics of bistable perception were identical in unimodal and bimodal presentations for all types of stimuli. Surprisingly, even strong short-term cross-modal interactions, when present, did not alter these dynamics. We conclude that bistability can co-occur independently in different sensory modalities. This observation supports models of distributed competition for perceptual decision and awareness.

Introduction

Competition for awareness between alternative interpretations of complex scenes is a key issue in perceptual organization of sensory information. Observers experience spontaneous transitions between percepts when presented with ambiguous stimuli, a phenomenon known as perceptual bistability. Major advances have been recently made concerning our understanding of the neural basis of bistability, with correlates found at several levels of processing (Long & Toppino, 2004; Tong, Meng, & Blake, 2006). To make sense of the various results, however, it is useful to distinguish "*what*" competes during bistability from "*how*" it competes (or to distinguish ambiguity from reversibility, to use the terminology of Long & Toppino, 2004). The "what" question addresses the nature of the representations of conscious percepts. The "how" question aims at uncovering the causal mechanisms responsible for perceptual switches. It is important to note that these conceptual distinctions overlap with, but are not equivalent to, other classical distinctions found in the literature. For instance, the

distinction of "low-level" vs. "high-level" processes can refer to the neural representations that correlate with perceptual reports (what) but also to the locus of origin of the perceptual switches (how). Similarly, "bottom-up" and "top-down" can characterize neural processes that follow perceptual switching or, alternately, that cause a switch. In the following experiments, we investigate the "how" question, namely, the causal mechanisms of perceptual decision during bistable perception.

A fundamental unresolved issue concerning the "how" of bistable perception contrasts distributed vs. central origins of perceptual switching. Does a switch originate because of competition distributed throughout many levels of processing, in a stimulus-specific manner, or is it the result of a supramodal mechanism that weights sensory signals, possibly coming from different sensory modalities, in order to achieve perceptual decision (Figure 1)?

Explanations in terms of distributed mechanisms posit that perceptual decisions are the consequence of local competition between networks of neurons coding contradictory interpretations. Computational models can indeed reproduce the characteristic dynamics of perceptual bistability with relatively simple adaptation mechanisms or
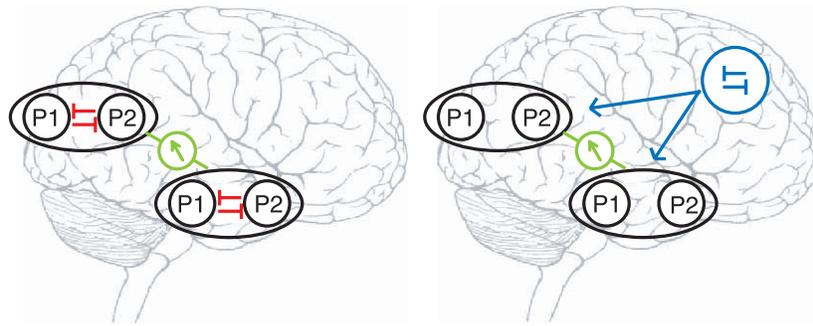
Figure 1. Testing distributed vs. centralized hypotheses of bistability with an audiovisual paradigm. Competing percepts (P1 and P2) are thought to be coded within the auditory and visual pathways, potentially at various cortical and subcortical processing stages. If competition mechanisms are distributed, switching should co-occur independently across modalities (left). In contrast, if a supramodal switching mechanism is involved in both tasks, it should always cause some interaction in the switching statistics (right). In both models, contextual cross-modal effects are expected to occur but they should covary with cross-modal congruence and be independent of switching mechanisms (green arrow).

noise and reciprocally connected neural populations (Kelso, 1995; Laing & Chow, 2002; Moreno-Bote, Rinzel, & Rubin, 2007; Noest, van Ee, Nijs & van Wezel, 2007; Shpiro, Curtu, Rinzel, & Rubin, 2007; Wilson, 2003). On the other hand, central explanations postulate brain structures distinct from the sensory cortices, for example, in the frontal cortex, to resolve ambiguities. Such structures would be necessary to initiate percept changes (Sterzer & Kleinschmidt, 2007). They would do so by sending signals to the sensory cortices (Cosmelli et al., 2004). A single, supramodal, network exploring the perceptual scene and thus generating bistability (Carter & Pettigrew, 2003; Leopold & Logothetis, 1999) could easily explain why experienced subjects can voluntarily modulate some aspects of the dynamics of bistable switches (Flugel, 1913; Von Helmholtz, 1925). It could also explain the striking similarities in general dynamics of perceptual alternations for many if not all ambiguous stimuli in vision (Leopold & Logothetis, 1999) as well as in audition (Pressnitzer & Hupé, 2006).

Growing behavioral and neurophysiological evidence shows that, at least for binocular rivalry, bistability involves both low- and high-level perceptual representations (Blake & Logothetis, 2002; Tong et al., 2006). Low-level representations are at the level of, for instance, eye-based retinotopic cues, whereas high-level representations are at the level of a whole perceptual object such as a face. The evidence that competing representations are found at multiple levels is compatible with both the hypothesis that competition is mediated by distributed mechanisms and with the hypothesis that it is initiated by a single central mechanism and feedback connections. Clearly, the issue yet to be resolved is how competition takes place. A decision between the two possible frameworks has profound implications, as in one case perceptual awareness results from monitoring sensory representations, whereas in the other it emerges from distributed processing.

A powerful technique to distinguish between these hypotheses is to present several bistable stimuli simultaneously. Here we used a double bistable task in vision and audition. A supramodal structure should be equally involved in the switching of perceptual decisions for auditory and visual bistability and therefore generate some degree of either entrainment or interference in the dynamics of a double, bimodal task compared with simple, unimodal tasks. The major advantage of testing bistability in two modalities, as opposed to within vision alone (Alais & Blake, 1999; Flugel, 1913; Grossmann & Dobbins, 2006; Long & Toppino, 2004), is that it is easy to control for contextual effects, that is, facilitation or inhibition between percepts, which are unrelated to bistability. With audiovisual presentation, cross-modal influences between stimuli are expected but they can be manipulated explicitly by varying the similarity between the perceptual contents of auditory and visual stimuli. Cross-modal coherence should modulate cooperative processes unrelated to bistability but should not affect the mandatory involvement of a supramodal switching structure.

We now describe two experiments where bistable stimuli were presented visionally, auditorily, and audiovisually. In Experiment 1, stimuli are unrelated, whereas in Experiment 2 they display a strong level of audiovisual congruence.

Methods

Stimuli and procedure

We performed two main experiments. In Experiment 1, we used sequences of pure tones of alternating frequencies as auditory stimuli (Bregman, 1990; Van Noorden, 1975) and visual plaids as visual stimuli (Hupé & Rubin, 2003;

Wallach, 1935; Wuerger, Shapley, & Rubin, 1996). Both stimuli are bistable when presented unimodally, with similar dynamics (Pressnitzer & Hupé, 2006). They are subjectively unrelated however, except that for both there is a *grouped* interpretation (one auditory stream or one plaid) and a *split* interpretation (two auditory streams or two gratings). Stimuli were presented unimodally or bimodally, and subjects continuously reported their percepts during the experiments. In Experiment 2, we introduced spatial and temporal coincidence between the two modalities in order to create strong cross-modal coherence (Calvert, Spence, & Stein, 2004). Auditory stimuli were again sequences of tones but presented over spatially separated loudspeakers. Visual stimuli were flashing lights placed on the speakers and synchronous with the tones, which could be experienced as either apparent motion across the lights (grouped) or independent flicker (split). We also performed a control experiment to verify that subjects could monitor percept changes simultaneously in both modalities.

Experiment 1: Plaids and streaming

The stimuli in each modality were the same as those used in a previous study (Pressnitzer & Hupé, 2006). Demonstrations are available online at the following address: http://www.cognition.ens.fr/Audition/sup/index.html.

The auditory stimuli were presented over headphones. A high-frequency pure tone A alternated with a low-frequency pure tone B in an ABA− pattern. The frequency of A was 587 Hz and that of B was 440 Hz (5 semitones difference). The duration of each tone was 120 ms. The silence notated "−" that completed the ABA− pattern was also 120 ms long. Listeners report either hearing the sequence as one stream ABA-ABA ("horse" rhythm, grouped percept) or as two independent streams A-A-A-A and −B-B− with isosynchronous rhythm ("Morse" rhythm, split percept). The visual stimuli were two rectangular-wave gratings presented through a 4-deg radius circular aperture on a computer screen 57 cm away. The gratings comprised thin dark stripes (duty cycle = 0.3, spatial frequency = 0.5 cycle/deg) on a lighter background and appeared as figures moving over the background. The intersecting regions were darker than the gratings (multiplicative transparency). The gratings were moving at 1.2 deg/s in directions 120 deg apart. A red fixation point over a 1-deg circular gray mask was added in the middle of the circular aperture, and subjects were instructed to fixate this point throughout stimulus presentation. The stimulus can be perceived either as a single plaid moving in a given direction or as two gratings sliding in opposite directions on top of each other.

Eight observers (4 males and 4 females), all right-handed, participated in the experiment (average age: 23) with no self-reported hearing problem and normal or corrected-to-normal eyesight. They gave informed consent

for their participation in experiments. They were instructed to report their conscious perception of each stimulus continuously during 4-min periods by pressing the right or the left mouse button for split percepts and by releasing the buttons for group percepts. Unimodal (6 repetitions each of the visual and auditory stimuli) and bimodal (6 repetitions) presentations were presented in a randomized order. Half the subjects were instructed to use the right button for visual percepts and the left for auditory percepts (with the opposite association for the other 4 subjects). Observers were given a few practice trials, in particular for the bimodal task. All subjects reported that they were able to perform the task.

Experiment 2: Apparent motion and streaming

Auditory stimuli were similar to those in Experiment 1, but with tone duration twice as long (240 ms) and with a different spatial location for tones A and tones B. The sequences were presented over loudspeakers with centers separated by 30 cm. Tones A came from the left loudspeaker, and tones B came from the right loudspeaker. Subjects were seated in a low reverberation doubled-walled soundproof booth (IAC™) 120 cm away from the loudspeakers. Visual stimuli were flashing red LEDs placed on the centers of the speakers and synchronous with the tones. The LEDs were 14 deg apart. The flash sequence can be experienced as either apparent motion across the lights (grouped) or independent flicker (split). A continuous, low intensity green LED was placed between both speakers for fixation. The room was otherwise completely dark.

Eight observers (4 males and 4 females), all right-handed, participated in the experiment (average age: 24) with no self-reported hearing problem and normal or corrected-to-normal eyesight. They gave informed consent for their participation in the experiments. Six of them had participated in the first experiment. The procedure was the same as for the first experiment. The stimuli were designed to have a high audiovisual coherence, with identical spatial and temporal cues across modalities as well as a relatively long duration to further enhance fusion. During practice trials, observers reported compelling audiovisual association between sounds and lights, as if the lights were making sounds (a phenomenon resembling cross-modal dynamic capture; Soto-Faraco, Lyons, Gazzaniga, Spence, & Kingstone, 2002). This cross-modal perceptual illusion was further enhanced by the fact that both percepts were initially systematically "grouped" percepts and therefore consistent with each other.

Control experiment: Apparent motion and auditory tracking task

Subjects had to simultaneously track changes of percept in the visual modality (apparent motion vs. flicker, like in Experiment 2) and physical changes of frequency

modulation in the auditory modality. The stimulus was a 440-Hz pure tone, which was amplitude-modulated at 5 or 17 Hz (4 subjects, "easy" control) or at 5 or 7 Hz (4 subjects, "difficult" control). Amplitude modulation (AM) depth was 20% in both cases. Subjects had to indicate the modulation frequency (low/high) by pressing or releasing the mouse button, like for the bistable audio stimulus. The time of changes of AM frequency "replayed" the audio bistable judgments of each subject obtained in the previous dual-task experiment. Subjects accurately reported their percepts in this dual task situation, for both task difficulties (Figure 2). The control experiment shows that the main experimental results are not compromised by the use of an audiovisual dual task.

Data analysis
Long-term statistics

For each 4-min trial, we computed the number of perceptual switches and the proportion of time observers experienced the stimuli as grouped. We compared these values for each subject in unimodal and bimodal conditions and for both visual and auditory stimuli. For each subject, we summarized the total amount of perturbation of long-term statistics with a single "mean effect" measure. The mean effect is the mean of the changes (in absolute values) observed between unimodal and bimodal presentation for

the number of switches and the proportion of time spent in each percept in the visual and auditory modalities (4 measures). For switches, changes were expressed as percentage. A mean effect close to zero means that bimodal presentation had no effect on the long-term statistics.

Coincidences

We estimated the probability that switches co-occurred in the two modalities during bimodal presentation. We first computed peri-switch time histograms that collect the number of switches in one modality around the time of a switch observed in the other modality. Histograms were computed for each run, for each modality as reference and for each type of switch (switch to the split percept or switch to the group percept). Twenty-four histograms were thus obtained for each subject. Note that some switches would be computed twice, once as reference, once as measure, but with this method we could use all switches (total 4623 switches in Experiment 1 and 2757 switches in Experiment 2). We estimated the chance level of coincidences by computing shuffled bimodal trials: The series of switches in one modality of a given trial were analyzed together with the series of switches in the other modality but measured in each of the five other bimodal trials. We constructed in this manner 30 shuffled trials (and therefore 120 histograms) for each subject. We found no difference in outcome for the type of switch (group to
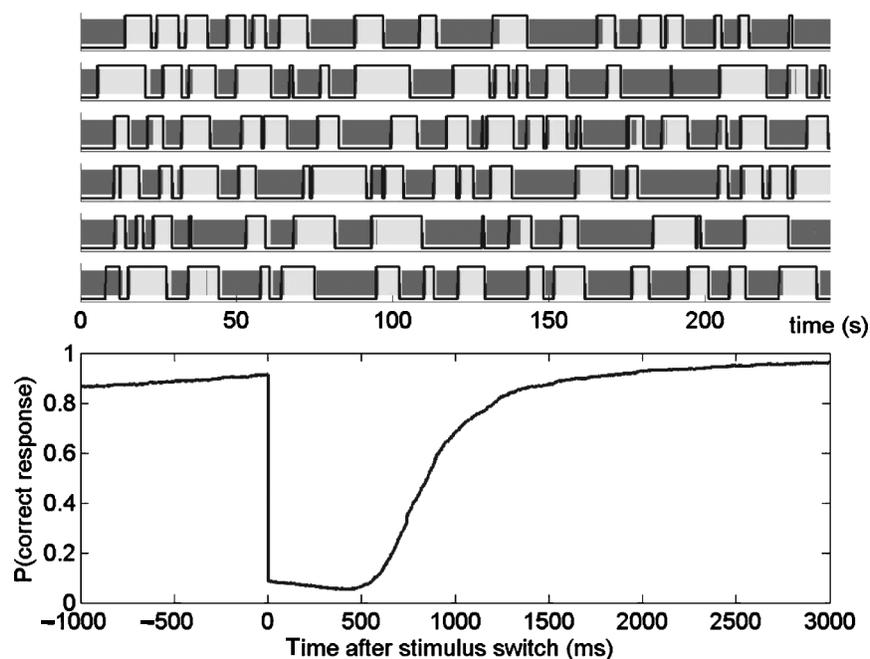

Figure 2. Tracking of physical changes in a dual bistable/objective task. Top: tracking records for subject S3. Each row represents a 4-min trial. Dark lines indicate small physical changes in the auditory stimulus ("difficult" condition; changes in amplitude-modulation frequency from 5 to 7 Hz, 20% modulation depth). The times of the changes were the times of the auditory perceptual switches for this subject in the main bistable experiment. Light/gray background indicates the reports of the subject. Tracking is overall very accurate. Bottom: average of correct reports around the physical transitions (*N* = 8 subjects).

split or split to group) or the modality (audio or visual) taken as reference, so results are presented with all cases averaged. Percept reports were down-sampled to 2 Hz in order to collect enough switches per interval. Averaging coincidences over durations shorter than 500 ms would generate bins including many zero values, precluding reliable statistical testing. However, we verified by using shorter intervals that we were not smoothing out any shorter component of the coincidence patterns (not shown). Our coincidence analysis is similar to a cross-correlation on the times of switches convolved with a smoothing window, with the additional benefit that we have time bins large enough to allow for statistical testing. We verified however that no effect with a time course of less than our chosen binwidth was apparent in the cross-correlation of switches convolved with Gaussian envelopes (not shown).

Common time

The proportion of time spent reporting the analogous percept in both modalities ("split" or "group") was computed for bimodal trials. This measure is similar to the "joint predominance" measure used by Alais and Blake (1999) in their study of contextual effects in binocular rivalry or the "common time" measure used by Grossmann and Dobbins (2003) to study the effect of presenting multiple copies of ambiguously rotating objects. We did not compare the observed values with a hypothetical 50% value (Grossmann & Dobbins, 2003) but rather with common time values obtained in shuffled trials, in order to account for possible biases.

Local temporal dynamics

We computed peri-switch time histograms of the probability of perceiving the same percept in the two modalities. For instance, if we choose audition as the reference modality and select grouped-to-split switches, the histogram would average perceptual states (group or split) in the visual modality following grouped-to-split switches in the auditory modality. We estimated chance by performing the same analysis on shuffled trials, i.e., by matching the auditory responses in one trial to the visual responses in another trial. All four possible histograms were computed for each run, one for each modality and each type of switch. We also computed a statistical index summarizing the strength of the local temporal dynamics interaction (Appendix A). This measure allows us to study the dynamics of cross-modal influences.

Statistical analysis

We ran statistical analyses for each experiment and each measured value. Statistics were computed over the variable

"subject," considered as a random factor. For long-term statistics, we computed twelve measures each time for each subject, six in the unimodal condition and six in the bimodal condition. We performed within subject paired comparisons with several independent measures by subject (mixed model). Tables are available in Appendix B.

Results

Long-term statistics

The switching rate and the proportion of time spent in each perceptual state are displayed in Figure 3. Overall, and for both experiments, no difference was found between the unimodal and bimodal presentation modes, suggesting that there was no interference between the overall dynamics of both competitions (Figures 3A and 3B). We found no significant main effect (Appendix B, Tables B1–B4). There were however significant interactions between subject and task for the number of switches in the apparent motion experiment for both modalities, especially for the auditory modality. Therefore, we examined individual data further, looking for any possible convergence effect, i.e., whether the difference between statistics was reduced by bimodal presentation. This analysis is relevant because even though we adjusted the stimulus parameters for the alternative interpretations to be close to 50% dominance on average, individual values varied from 24% to 69% in the unimodal condition. We found that switching rate and proportion of the "grouped" percept were remarkably stable in both unimodal and bimodal conditions for every subject, even when these values were different for auditory and visual stimuli (Figures 3C and 3D).

Coincidences

The probabilities of concomitant auditory and visual switching during bimodal presentation are presented in Figure 4. Total independence between the two tasks would not necessarily be reflected by flat histograms because of the temporal dynamics intrinsic to each modality. We thus compared histograms for the real bimodal data and for "shuffled" data (see Methods). Independence between tasks would be reflected by an overlap between real and shuffled analyses. If the tasks are not independent, there are two possible outcomes. On the one hand, reporting a switch in one modality may impede the report of a switch in the other modality, causing a dip in the histograms around time zero. On the other hand, switching in one modality may instigate a switch in the other modality, causing a peak around time zero. We found no systematic effect of the double task on coincidences (Figure 4A). Table B5 in
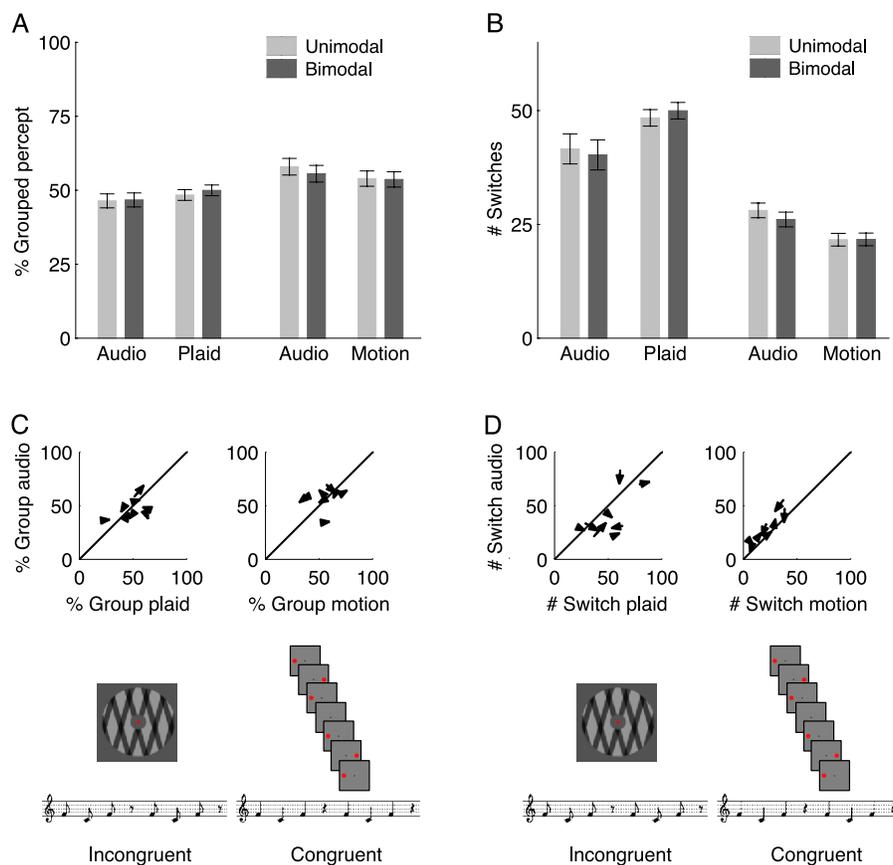
Figure 3. (A, B) Audiovisual presentation does not affect the overall statistics of auditory and visual bistability. The relative dominance of each type of percept (A) and the number of switches (B) were the same in unimodal and bimodal presentations for both plaid (4 leftmost columns) and apparent motion (4 rightmost columns) experiments. Here and later, vertical bars denote 0.95 confidence intervals, while statistical analyses performed on paired comparisons (see Methods and Tables B1–B4 in Appendix B). (C, D) For each subject, we plot the long-term statistical values obtained in both modalities against each other for the unimodal condition and trace an arrow to the values obtained in the bimodal condition. If the values are more similar for bimodal presentation, the arrows should point toward the equi-value line. This is clearly not the case in the plaid (left panel in C and D) nor in the apparent motion (right panel in C and D) experiment, indicating that bimodal presentation did not cause convergence of long-term statistics in individual observers.

Appendix B details the results of the statistical testing that confirms this absence of overall effect. Significant differences between subjects were observed, however. Figures 4B and 4C illustrate these inter-individual differences by showing peri-switch time histograms for each subject and experiment. There are differences between the observed and the shuffled curves for a few subjects only. In addition, for a given subject, there could be a dip in coincidence for one experiment but a peak in the other (subject S3 for instance). We suspect the influences of attentional and motor strategies as well as possible criteria changes for these unreliable differences (see Discussion).

Common time and local temporal dynamics

The overall interaction between auditory and visual percepts, regardless of exact switching time, was estimated by examining the statistical coupling between percepts. The proportion of "common time" spent reporting the equivalent percept in both modalities is shown in Figure 5A. The duration of equivalent percepts across modalities was higher than chance, especially in the apparent motion experiment. We found significant effects in both experiments as well as interactions with subjects (Table B6). For plaids, the effect was weak for three subjects and absent for one. In the apparent motion experiment, the effect was clear for all subjects. The local temporal dynamics analysis shows how this coupling was established (Figure 5B). By computing peri-switch time histograms, we observed that reporting a given percept in one modality progressively increased the probability of reporting the analogous percept in the other modality. We quantified the amount of biasing by computing a "cross-modal effect" measure, detailed in Appendix A. The measure amounts to subtracting the real and shuffled data shown on Figure 5B and then computing the remaining area. The amount of cross-modal effect found was larger
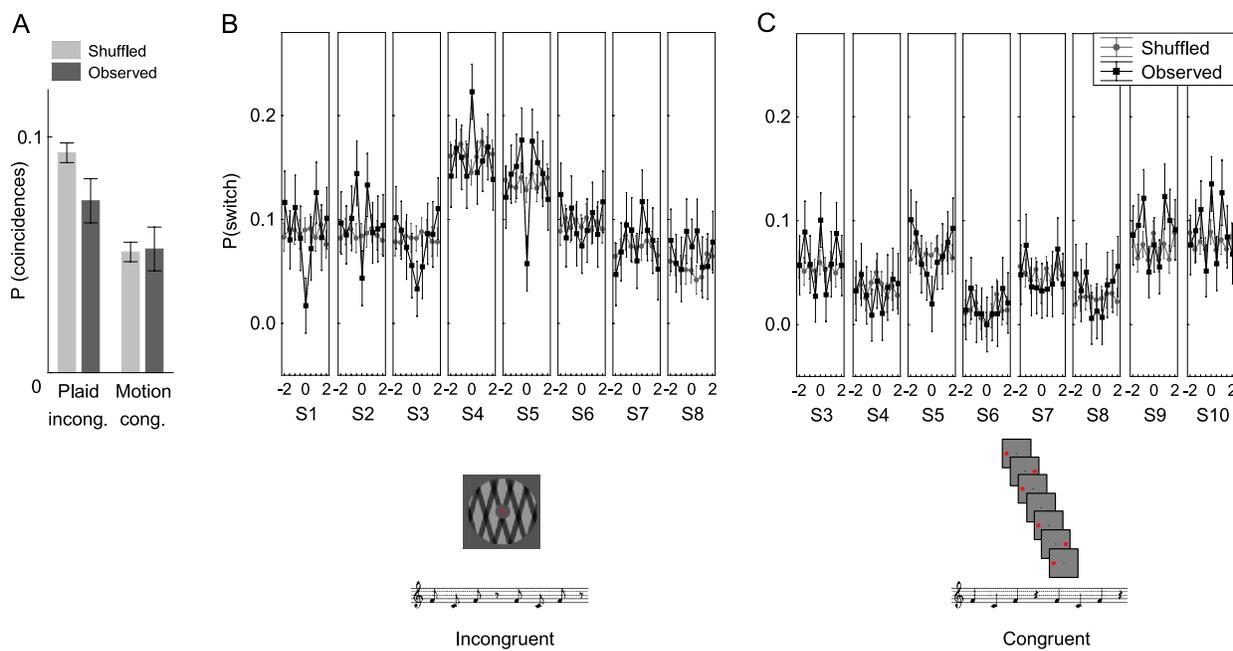
Figure 4. No effect of the bimodal task on the probability of coincidences. (A) The probabilities of observed coincidences of switches (reports of an auditory and a visual switch within 500 ms) were not significantly different from chance level, estimated by shuffling audio and visual report series across trials (see Table B5 in Appendix B). (B, C) Peri-switch time histograms of the probability of a percept switch in the other modality in the plaid experiment (B) and in the apparent motion experiment (C) for each observer. The zero interval computes the mean number of auditory (visual) switches within 0.5 s around a visual (auditory) switch. *X*-axis is time in seconds.

for congruent audiovisual stimuli (apparent motion) as compared with incongruent audiovisual stimuli (plaids).

The cross-modal effect observed in local temporal dynamics is independent of other measures of interaction between the auditory and the visual tasks. First, the effect is relatively slow, so it was observed even though there were no more coincident switches than expected by chance (Figure 4). Second, this cross-modal effect was fairly variable across subjects, and importantly, this variability was unrelated to the amount of change in long-term statistics observed when comparing unimodal

and bimodal presentation (Figure 5C). Similarly, we found that intersubject variability of the cross-modal effect was not related to the amount of change in coincidences or convergence measured for each subject (not shown; this latter measure corresponds to the direction of the arrows plotted in Figures 3C and 3D). There was also no correlation between the strength of the effect and the mean duration of percepts (such a correlation could have confounded the apparent motion experiment results since percepts lasted longer in this case). The susceptibility of the biasing effect to cross-modal coherence as well as its
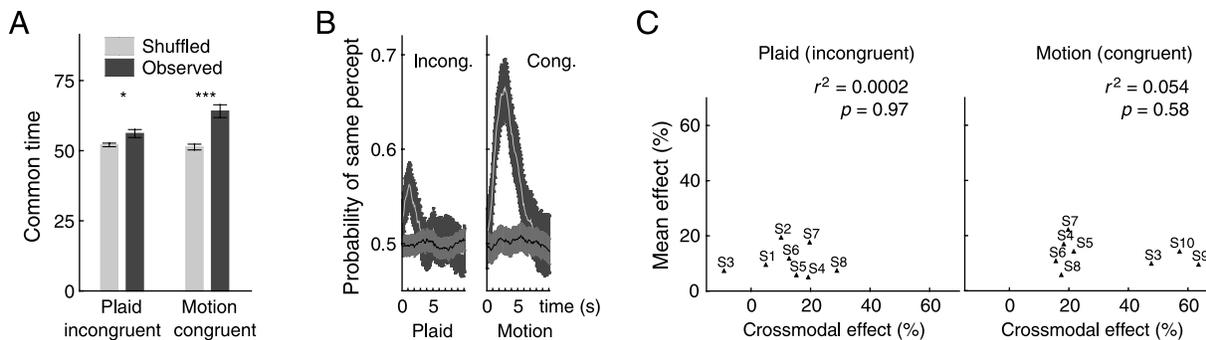

Figure 5. Cross-modal interactions produce contextual effects. (A) The "common time" spent reporting similar percepts (split or group) in both modalities was slightly higher than chance in the plaid experiment (*p = 0.035, see Table B6 in Appendix B) and clearly higher in the apparent motion experiment (***p = 0.0006). Cross-modal effects have short time-dynamics, as shown by peri-switch time histograms (B, light gray curve for shuffled trials). The strength of the cross-modal effect (C) computed for each subject (Appendix A) was not related to any modification of long-term statistics of bistable perception ("mean effect": mean of the relative changes in the audio and visual task, for the proportion of grouped percept and the number of switches; see the Data analysis section).

independence from overall switching statistics strongly suggests a contextual effect, unrelated to the specific competition mechanisms recruited by bistability.

Discussion

Bistability has been used for more than a century as a powerful tool to investigate the neural mechanisms of perceptual organization. In spite of such sustained scrutiny, a long-standing controversy remains pertaining to the neural origins of perceptual switches. Our findings show that bistable switches can co-occur in two modalities with only minimal interaction. As we now discuss, distributed models of bistable perception accommodate this finding far more easily than models relying on a central origin for the switches.

Multiple bistable stimuli to test central vs. distributed hypotheses

Flugel (1913) was one of the first investigators to study the simultaneous perception of several visual bistable stimuli, presenting to subjects multiple Necker cubes. He observed a tendency for cubes to reverse synchronously, but independent switches were still possible. These observations were confirmed for both the Necker cube (Long & Toppino, 2004) and other ambiguous stimuli (e.g., Alais & Blake, 1999; Alais, Lorenceau, Arrighi, & Cass, 2006; Grossmann & Dobbins, 2003, 2006). As noted by Long and Toppino (2004), the possibility of independent switching rules out the extreme model that relies exclusively on a central origin for the switches. The question remains however whether there is *any* involvement of a central mechanism in perceptual decision, or whether perceptual organization is fully resolved at the level(s) where the stimulus is represented.

The partial synchronization of switches observed in the experiments cited above may signal the involvement of a supramodal structure, but they can also reflect contextual effects. When multiple ambiguous objects share at least one property, mechanisms of local cooperativity will bias the competition in favor of a common interpretation (Freeman & Driver, 2006). This is in particular obvious for multiple copies of the apparent motion quartet, which tend to move in the same direction and thus all switch synchronously (Ramachandran & Anstis, 1983). In binocular rivalry, Alais and Blake (1999) and Alais, Lorenceau, et al. (2006) manipulated grouping cues of two rivaling gratings or Gabor patches and observed that joint predominance and cross-correlation values depended on the recruitment of cortical lateral connections.

In order to control for possible contextual effects, we presented stimuli in two modalities, vision and audition.

We had shown previously that bistable perception in audition followed the same rules as in vision (Pressnitzer & Hupé, 2006). We used stimuli with weak or strong cross-modal congruence, which would change the efficiency of the local cooperativity mechanisms. The crucial assumption of our experiment is that a single central, supramodal mechanism for bistability should manifest itself in both cases, with weak or strong cross-modal congruence. Unfortunately, there is no quantitative model of a central, supramodal mechanism for the initiation of perceptual switches. We therefore need to make assumptions about its mode of operation. One simplified description for such a mechanism is that somewhere in the brain, neurons fire at some point and cause the perceptual reorganization involved in perceptual switches, whatever the stimulus or the modality involved. The mandatory involvement of a single group of neurons for triggering switches is crucial to the argument. If two groups of neurons located outside of the sensory cortices were involved in the initiation of switches, one group for each modality, they should not be considered as a single supramodal switching mechanism but rather as a form of distributed mechanisms that happen to occur at a central level. Given such a framework, a central supramodal mechanism must produce changes in perceptual dynamics between unimodal and bimodal presentation. Either the mechanism alternates between modalities and a reduction of coincidences is observed or the mechanism groups together the modalities and auditory and visual percepts tend to switch together. We did not observe either of these effects. There is one final alternative consistent with the central mechanism model: a completely multiplexed decision mechanism. This model might be difficult to implement neurally and would amount conceptually to having different sources for the initiation of switches for distinct bistable stimuli. On the other hand, the absence of mandatory interaction would follow naturally from models that posit a distributed origin for the switches. Cross-modal cooperation mechanisms are also expected in such a class of model, as the different perceptual organization would be decided on the basis of incoming sensory evidence and of the multimodal context. Bistability should occur independently within each modality only when there is no cross-modal congruence, which is what we observed.

Limitations inherent to testing an absence of interaction

Obviously, failure to observe an effect does not prove the effect's absence. However, we think it unlikely that our observations are due to lack of statistical power or to inadequate statistical data analysis. First, we were able to measure reliable and significant effects in the bimodal task, thanks to local temporal dynamics analysis. We could however attribute these effects to cross-modal

contextual influences, as their strength was related to the degree of cross-modal coherence. The fast interactions predicted by a putative central bistability mechanism did not appear using the same analysis. Even if it could never be demonstrated that interactions due to bistability are totally absent, our results show that if they exist, they are negligible compared to cross-modal interactions. To accommodate this with a single supramodal central mechanism responsible for the initiation of perceptual switches seems problematic.

In addition, we observed independence between cross-modal effects and long-term statistics. Even when strong cross-modal influences were present (up to 60% cross-modal effect, Figure 5C), there was no modification of long-term statistics (relative dominance or number of switches, Figures 3A and 3B). In binocular rivalry, Alais and Blake (1999) and Alais, Lorenceau, et al. (2006) found similar independence between contextual effects and average percept duration. Even when there were large differences in relative dominance or number of switches for unimodal presentation, there was no convergence of these values for bimodal presentations (Figures 3C and 3D). Such robustness of long-term statistics and therefore of the dynamics of bistable perception again strongly challenges the possibility of a central switching mechanism.

Furthermore, we observed no synchronization between switches due to bimodal presentation (Figure 4). As suggested by a reviewer, a model with a central component could explain this finding if some switches showed synchronization across modalities while others showed desynchronization, producing a null result overall. The different behavior of switches might be related to distinct neural processes: Nakatani and van Leeuwen (2006) found that a restricted number of distinct patterns of EEG prior to button presses could be identified for the bistable perception of a Necker Cube. However, our results require that the proportion of synchronous and desynchronous switching would be exactly that required to produce the chance level, and that it would vary to parallel the changes in chance level observed with the different subjects and type of experiments (Figure 4A). We thus propose that the interpretation of no synchronization above or below chance is more parsimonious.

Finally, we did observe individual differences with significant modulations of synchronization visible for some subjects (in 8 of 16 graphs, see Figure 4 and Table B5 in Appendix B). However, effects were found in opposite directions (5 decreases vs. 3 increases of synchronization). It seems unlikely that some subjects had a central mechanism whereas others had a distributed mechanism. Subject-specific strategies such as attention, motor strategies, or criteria changes are much more plausible explanations. For example, we would expect fewer coincidences with limited attentional resources. This could happen even if vision and audition do not use a common attentional resource when dealing with stimulus

processing, as shown by Alais, Morrone, and Burr (2006) for signal discrimination. Deciding which perceptual channel should be attended to could still influence the timing of responses without affecting processing within each channel. This is in fact what we observed in a control experiment (described in the Methods section). In this control experiment, we asked subjects to simultaneously track objective auditory and bistable visual percept changes. We observed a small but significant decrease of coincidences for most subjects, indicating that our control task was probably more difficult than the bimodal bistable task, and that simultaneous reports of percept changes in two modalities could indeed induce apparent changes in the coincidence analysis (not shown). Tracking of the objective changes was nevertheless very accurate, which shows that even in a more difficult dual task subjects could report their percepts accurately (Figure 2). All these elements indicate that subjects could perform the bistable dual task appropriately and that the analysis techniques we used could detect interactions if they were present.

Top-down effects

Bistable perception is probably never completely free of any top-down influence—attention, intention, memories, or imagination (which may summon the suppressed interpretation) are certain to fluctuate over time. It is therefore reasonable to assume that these top-down fluctuations alone could cause percept switches (Leopold & Logothetis, 1999; Von Helmholtz, 1925). Our claim that the mechanisms of perceptual bistability are distributed does not rule out the role of top-down effects. These top-down effects—just like cross-modal effects—can be understood within the context of the distinction between what competes and how competition takes place. In our framework, top-down effects like attention influence what competes and, as a consequence, various characteristics of the perceptual reversals. Top-down effects, however, are independent from the cause of the switches, the "how" mechanisms. Several pieces of evidence, which we detail now, are consistent with such an interpretation.

A switching mechanism ("how") can produce perceptual alternations between interpretations even when the stimulus stays constant. However, the "weights" of the different interpretations can determine some characteristics of the competition, such as the average time spent experiencing each interpretation (see e.g., Hupé & Rubin, 2003, 2004). There are many ways to change the weights of competing interpretations by altering the physical stimulus: an example in binocular rivalry would be to change the contrast of the stimulus presented to one eye. Nevertheless, manipulations independent of the stimulus can have effects similar to physical manipulations. Directing attention away from the stimulus, a top-down manipulation, slows down perceptual alternations in a way

Hupé, Joffo, & Pressnitzer

similar to contrast reduction for binocular rivalry (Paffen, Alais, & Verstraten, 2006; for plaids, see also Pastukhov & Braun, 2007). Similarly, observers without prior knowledge of ambiguous figures such as uninformed subjects or children show little spontaneous reversals (Rock, Gopnik, & Hall, 1994; Rock & Mitchener, 1992). This can be explained if the unknown alternative interpretation has very little weight, again a "what" issue. Note that other ambiguous stimuli like plaids or apparent motion do not require prior knowledge—bistable perception of the latter having been observed in pigeons (Vetter, Haynes, & Pfaff, 2000). In our bimodal experiments, cross-modal effects can also be accommodated in the "what" context. When both stimuli share the same perceptual content (motion vs. flicker), perceiving one interpretation in one modality gives more weight to the corresponding interpretation in the other modality: this is what we term a contextual effect. This link may be established in a bottom-up or top-down fashion by means of feedback mechanisms.

The distinction between "what competes" and "how it competes" allows the reconciliation of seemingly contradictory results about the role of top-down mechanisms in bistable perception. On the one hand, computational models that reproduce the specifics of alternation dynamics, like stochasticity or specific relationships between weight and average percept duration (Brascamp, van Ee, Noest, Jacobs, & van den Berg, 2006; Levelt, 1968; Rubin & Hupé, 2005), do not require top-down mechanisms to initiate the switches (Laing & Chow, 2002; Moreno-Bote et al., 2007; Noest et al., 2007; Shpiro et al., 2007; Wilson, 2003). Results from psychophysics also indicate that top-down mechanisms are not necessary to cause perceptual alternations (for attention, see Pastukhov & Braun, 2007). On the other hand, observers can voluntarily trigger some switches (Von Helmholtz, 1925). Using fMRI, Sterzer and Kleinschmidt (2007) showed activation in the prefrontal cortex that preceded perceptual alternations and concluded that this was the trace of the switching mechanism. The two sets of results can be accommodated if we hypothesize that top-down and attentional mechanisms can bias the weights of different interpretations, so that the biasing is expected to eventually trigger some reversals. An alternative explanation of the Sterzer and Kleinschmidt (2007) data is thus that their subjects exercised some degree of intentional control when looking at ambiguous stimuli—in other words, prefrontal activation would be a correlate of voluntary shifts of attention that do trigger phenomenal reversals in some trials. However, it is not the obligatory source of percept switching, as shown by Pastukhov and Braun (2007) and this study. The model of Noest et al. (2007) formalizes this possibility, as attentional biases could be modeled as modifying the underlying perceptual representations but not the decision mechanisms. Within this theoretical framework, perceptual representations of different ambiguous stimuli do not need to be equally biased

by attention or intention. Meng and Tong (2004) as well as van Ee, van Dam, and Brouwer (2005) found different outcomes of intention for different ambiguous stimuli. This had been taken as an argument against a common top-down mechanism involved in switching for all bistable perception. However, different gain factors could easily produce those different outcomes and therefore do not disprove the central processing hypothesis. Rather, these results may show different penetrability by high-level manipulations of the neural processes coding these competing percepts—and thus indicate at which neural level these processes take place (for a similar interpretation, see Long & Toppino, 2004).

Central oscillator?

It has also been argued that a supramodal structure (like a central oscillator) would easily account for the observation that observers tend to be fast or slow switchers for different ambiguous stimuli (Carter & Pettigrew, 2003; Sheppard & Pettigrew, 2006). However, this was shown only within the visual modality, while individual switching rates were not correlated for auditory streaming and visual plaids (Pressnitzer & Hupé, 2006). In the present data, there appears to be a correlation between auditory streaming and apparent motion for the limited number of observers tested (Figure 3D). We have no interpretation for this trend that would require further testing. The global similarity of perceptual dynamics observed for bistable stimuli also seems compatible with a central oscillator account. Note however that quantitative differences do exist between different visual bistable stimuli—see, e.g., van Ee (2005). In any case, both correlations and similarities can point toward the interesting possibility that perceptual organization relies on a distributed but ubiquitous neuronal architecture in charge of resolving conflicting sensory cues (Pressnitzer & Hupé, 2006; Rubin & Hupé, 2005).

Neural correlates of bistable perception

The distributed competition hypothesis predicts that correlates of perceptual decision will be found at different stages of neural processing, corresponding to the loci where the competing features are encoded. This is consistent with current views of visual binocular rivalry, where neural correlates of bistable perception have been found at different stages of the visual pathway, starting at the level of the LGN (Haynes, Deichmann, & Rees, 2005; Wunderlich, Schneider, & Kastner, 2005). Which level depends on the complexity of the rival stimuli (for reviews, see Blake & Logothetis, 2002; Tong et al., 2006), as confirmed by TMS experiments (Pearson, Tadin, & Blake, 2007). Similarly, correlates of switches for ambiguous motion stimuli have been observed in motion

sensitive area MT, both for apparent motion (Muckli et al., 2002; Sterzer, Eger, & Kleinschmidt, 2003) and plaids (Castelo-Branco et al., 2002).

In the auditory modality, correlates of streaming based on frequency differences between tones have also been found at several neural processing stages: in multi-modal area such as the intraparietal sulcus (Cusack, 2005), but also in primary sensory cortices (Kashino, Okada, Mizutani, Davis, & Kondo, 2007; Micheyl, Tian, Carlyon, & Rauschecker, 2005; Wilson, Melcher, Micheyl, Gutschalk, & Oxenham, 2007), and even in peripheral subcortical structures such as the cochlear nucleus (Pressnitzer, Micheyl, Sayles, & Winter, 2007). These various correlates could reflect competition at different levels: between perceptual objects for multimodal areas, but also between acoustical features for lower stages. Interestingly, for bistable verbal transformation (spontaneous changes in perceived meaning for repeated words), frontal activations were observed during perceptual transitions but no involvement of sensory cortices was found, consistent with the idea that speech forms are coded above the primary auditory cortex (Kondo & Kashino, 2007). When the verbal transformation effect was measured in a mental repetition task, without actual presentation of the sound stimulus, the network observed overlapped with regions involved in verbal working memory (Sato et al., 2004).

All these different neurophysiological correlates, both in vision and in audition, suggest networks of distributed competition at the appropriate level of representation for each given stimulus and task, probably involving large modulatory feedforward and feedback connections. Such architecture is fully compatible with the present psycho-physical data to account both for the absence of mandatory interaction and the possibility of contextual effects when two ambiguous stimuli are presented simultaneously.

Conclusions

When simultaneously presenting auditory and visual ambiguous stimuli, we observed little influence of perceptual switching in one modality on the other. Cross-modal effects were present but were related to the degree of audiovisual congruence between stimuli and had no influence on switching statistics. Independent bistability for audiovisual stimuli strongly supports the hypothesis that bistable perceptual switches are not caused by a supra-modal mechanism but rather by independent competition within modalities. This competition can be biased by many factors including cross-modal influences. Such findings are consistent with distributed competition in local neuronal structures in order to achieve perceptual organization and cooperation between loci of perceptual processing to bias the outcome of the competition as proposed in models of visual attention (Duncan, Humphreys, & Ward, 1997).

Appendix A

Method of computation of the cross-modal effect

We constructed an index of the strength of the cross-modal effect. This index was designed to capture most of the information visible in Figure A1. A value of 100% would mean that as soon as there is a switch of percept in one modality, there is also a switch in the other modality. In that case, the "common time" measure would also be 100%. Common time and cross-modal measures are closely related to each other, and performing the analyses with the common time measure gives results very similar to those presented in Figure 5C.
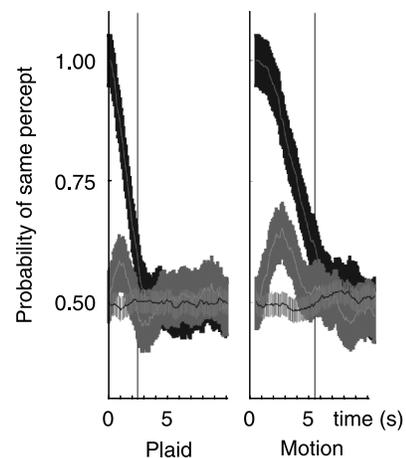

Figure A1. Method of computation of the cross-modal effect. The dark gray curve shows a typical peri-switch histogram, observed here for subject S5. The black curve superimposed is a peri-switch histogram within a single modality, equivalent to an autocorrelo-gram of the responses. It is closely related to the median duration of the percepts, indicated by a vertical line. The light gray curve displays the peri-switch histogram for shuffled trials. Vertical bars denote 0.95 confidence intervals. As seen in the dark gray curve, the probability of reporting equivalent percepts in the two modalities increased after a switch in one modality and then decreased as time approached the average percept duration. This was the case for all subjects. We thus computed an index of the strength of the cross-modal effect that both captured the amount of significant biasing (area between dark gray curve and light gray curve) and was independent of average percept duration. The probability of reporting equivalent percepts was integrated from the switch (time 0) to the median duration (vertical lines). The value obtained for the shuffled trials was subtracted from the real trials. The observed effect was expressed as a ratio between the observed bimodal effect (dark gray curve) and the maximal possible effect given the percept durations (black curves). The effects computed with such a method for S5 were 15% and 22% for plaid and apparent motion, respectively. The values for other subjects range from −9% to 29% for plaid and 16% to 64% in the apparent motion case (Figure 5C).

Appendix B

Statistical tables

	Effect	Degrees of freedom	F audio	P audio	F visual	P visual
Intercept	Fixed	1	216	$<10^{-5}$	131	$<10^{-5}$
Subject	Random	7	11.1	0.003	31	$<10^{-4}$
Task	Fixed	1	0.03	0.87	1.05	0.34
Subject × Task	Random	7	1.28	0.27	1.42	0.21
Error		80				

Table B1. Proportion of grouped percept in Experiment 1 (streaming and plaids). As can be seen in Figure 3A (4 leftmost columns), there was no difference between unimodal and bimodal presentations of the stimuli (no "task" effect) for all subjects (no "Subject × Task" interaction effect, see also Figure 3C: each arrow is very short both along the *x*- and the *y*-axis).

	Effect	Degrees of freedom	F audio	P audio	F visual	P visual
Intercept	Fixed	1	271	$<10^{-6}$	133	$<10^{-5}$
Subject	Random	7	18.2	0.0005	16.2	$<10^{-3}$
Task	Fixed	1	2.17	0.18	0.02	0.90
Subject × Task	Random	7	0.65	0.71	1.59	0.15
Error		80				

Table B2. Proportion of grouped percept in Experiment 2 (streaming and apparent motion). No significant effect (see also the 4 rightmost columns in Figures 3A and 3C).

	Effect	Degrees of freedom	F audio	P audio	F visual	P visual
Intercept	Fixed	1	30.6	$<10^{-3}$	68.8	$<10^{-4}$
Subject	Random	7	28.1	$<10^{-3}$	20.3	$<10^{-3}$
Task	Fixed	1	0.22	0.65	3.83	0.09
Subject × Task	Random	7	1.14	0.34	1.52	0.17
Error		80				

Table B3. Number of switches in Experiment 1 (streaming and plaids). No significant effect (see also the 4 centermost columns in Figures 3B and 3D).

	Effect	Degrees of freedom	F audio	P audio	F visual	P visual
Intercept	Fixed	1	30.7	$<10^{-3}$	28.4	$<10^{-3}$
Subject	Random	7	9.59	0.004	27.3	$<10^{-3}$
Task	Fixed	1	0.40	0.55	0.002	0.97
Subject × Task	Random	7	7.55	$<10^{-6}$	2.45	0.024
Error		80				

Table B4. Number of switches in Experiment 2 (streaming and apparent motion). There were significant interaction effects both for the audio and the visual task but no main effect of the task (see also the 4 rightmost columns in Figure 3B). This means that some subjects had more perceptual switches during the bimodal task compared with the unimodal task, while other subjects had less switches. The size of these changes, though reliable, was small when compared with intersubject variability (arrows are very small in Figure 3D, rightmost panel).

	Effect	Degrees of freedom	F plaid	P plaid	F motion	P motion
Intercept	Fixed	1	30	$<10^{-3}$	15.56	0.006
Subject	Random	7	2.85	0.095	6.16	0.014
Mode	Fixed	1	1.28	0.29	0.014	0.91
Subject $\times$ Mode	Random	7	11.86	$<10^{-13}$	4.19	$<10^{-3}$
Error		1136				

Table B5. Coincidences in Experiment 1 (plaid and streaming) and Experiment 2 (apparent motion and streaming). The "mode" variable corresponds here and in Table B6 to "observed" vs. "shuffled" trials. Statistics were based on 144 measures for each subject, 24 measures in the observed conditions (6 trials by 4 coincidence type—considering the direction of the switch, group, or split, within each modality), and 120 measures in the shuffled condition (30 "trials"). On average, there were neither more nor less coincidences than expected by chance. However, there was significant variability across subjects, especially for the plaid experiment. This can be observed in Figures 4B and 4C. We computed 16 coincidence graphs (8 in each experiment) and observed a decrease of the number of coincidences in 5 instances and an increase in 3 instances. (This might be difficult to see on the graph: we considered significant increases or decreases when error bars for the observed and simulated data did not overlap—meaning that we did not correct the statistical risk for multiple comparisons.)

	Effect	Degrees of freedom	F plaid	P plaid	F motion	P motion
Intercept	Fixed	1	2426	$<10^{-9}$	790	$<10^{-7}$
Subject	Random	7	2	0.190457	3.5	0.06
Mode	Fixed	1	6.84	0.035	33.9	$<10^{-3}$
Subject $\times$ Mode	Random	7	3.87	$<10^{-3}$	2.98	0.005
Error		272				

Table B6. Common time in Experiment 1 (plaid and streaming) and Experiment 2 (apparent motion and streaming). Statistics were based on 36 measures for each subject (6 observed and 30 shuffled trials). There was more common time than expected by chance (Figure 5A), especially for the apparent motion experiment. The strength of the effect however was significantly variable across subjects.

Acknowledgments

References

Alais, D., & Blake, R. (1999). Grouping visual features during binocular rivalry. *Vision Research, 39,* 4341–4353. [PubMed]

Alais, D., Lorenceau, J., Arrighi, R., & Cass, J. (2006). Contour interactions between pairs of Gabors engaged in binocular rivalry reveal a map of the association field. *Vision Research, 46,* 1473–1487. [PubMed]

Alais, D., Morrone, C., & Burr, D. (2006). Separate attentional resources for vision and audition. *Proceedings of the Royal Society B: Biological Sciences, 273,* 1339–1345. [PubMed] [Article]

Blake, R., & Logothetis, N. K. (2002). Visual competition. *Nature Reviews, Neuroscience, 3,* 13–21. [PubMed]

Brascamp, J. W., van Ee, R., Noest, A. J., Jacobs, R. H., & van den Berg, A. V. (2006). The time course of binocular rivalry reveals a fundamental role of noise. *Journal of Vision, 6*(11):8, 1244–1256, http://journalofvision.org/6/11/8/, doi:10.1167/6.11.8. [PubMed] [Article]

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge: MIT Press.

Calvert, G., Spence, C., & Stein, B. (2004). *The handbook of multisensory processes.* Cambridge: MIT Press.

Carter, O. L., & Pettigrew, J. D. (2003). A common oscillator for perceptual rivalries? *Perception, 32,* 295–305. [PubMed]

Castelo-Branco, M., Formisano, E., Backes, W., Zanella, F., Neuenschwander, S., Singer, W., et al. (2002). Activity

patterns in human motion-sensitive areas depend on the interpretation of global motion. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 13914–13919. [PubMed] [Article]

Cosmelli, D., David, O., Lachaux, J. P., Martinerie, J., Garnero, L., Renault, B., et al. (2004). Waves of consciousness: Ongoing cortical patterns during binocular rivalry. *Neuroimage, 23,* 128–140. [PubMed]

Cusack, R. (2005). The intraparietal sulcus and perceptual organization. *Journal of Cognitive Neuroscience, 17,* 641–651. [PubMed]

Duncan, J., Humphreys, G., & Ward, R. (1997). Competitive brain activity in visual attention. *Current Opinion in Neurobiology, 7,* 255–261. [PubMed]

Flugel, J. C. (1913). The influence of attention in illusions of reversible perspective. *British Journal of Psychology, 5,* 357–397.

Freeman, E. D., & Driver, J. (2006). Subjective appearance of ambiguous structure-from-motion can be driven by objective switches of a separate less ambiguous context. *Vision Research, 46,* 4007–4023. [PubMed]

Grossmann, J. K., & Dobbins, A. C. (2003). Differential ambiguity reduces grouping of metastable objects. *Vision Research, 43,* 359–369. [PubMed]

Grossmann, J. K., & Dobbins, A. C. (2006). Competition in bistable vision is attribute-specific. *Vision Research, 46,* 285–292. [PubMed]

Haynes, J. D., Deichmann, R., & Rees, G. (2005). Eye-specific effects of binocular rivalry in the human lateral geniculate nucleus. *Nature, 438,* 496–499. [PubMed] [Article]

Hupé, J. M., & Rubin, N. (2003). The dynamics of bistable alternation in ambiguous motion displays: A fresh look at plaids. *Vision Research, 43,* 531–548. [PubMed]

Hupé, J. M., & Rubin, N. (2004). The oblique plaid effect. *Vision Research, 44,* 489–500. [PubMed]

Kashino, M., Okada, M., Mizutani, S., Davis, P., & Kondo, H. M. (2007). The dynamics of auditory streaming: Psychophysics, neuroimaging and modeling. In B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Upperkamp, & J. Verhey (Eds.), *Hearing-from basic research to applications* (pp. 275–283). Heidelberg: Springer Verlag.

Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior.* Cambridge: MIT Press.

Kondo, H. M., & Kashino, M. (2007). Neural mechanisms of auditory awareness underlying verbal transformations. *Neuroimage, 36,* 123–130. [PubMed]

Laing, C. R., & Chow, C. C. (2002). A spiking neuron model for binocular rivalry. *Journal of Computational Neuroscience, 12,* 39–53. [PubMed]

Leopold, D. A., & Logothetis, N. K. (1999). Multistable phenomena: Changing views in perception. *Trends in Cognitive Sciences, 3,* 254–264. [PubMed]

Levelt, W. J. M. (1968). *On binocular rivalry.* The Hague: Mouton.

Long, G. M., & Toppino, T. C. (2004). Enduring interest in perceptual ambiguity: Alternating views of reversible figures. *Psychological Bulletin, 130,* 748–768. [PubMed]

Meng, M., & Tong, F. (2004). Can attention selectively bias bistable perception? Differences between binocular rivalry and ambiguous figures. *Journal of Vision, 4*(7):2, 539–551, http://journalofvision.org/4/7/2/, doi:10.1167/4.7.2. [PubMed] [Article]

Micheyl, C., Tian, B., Carlyon, R. P., & Rauschecker, J. P. (2005). Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron, 48,* 139–148. [PubMed] [Article]

Moreno-Bote, R., Rinzel, J., & Rubin, N. (2007). Noise-induced alternations in an attractor network model of perceptual bistability. *Journal of Neurophysiology, 98,* 1125–1139. [PubMed]

Muckli, L., Kriegeskorte, N., Lanfermann, H., Zanella, F. E., Singer, W., & Goebel, R. (2002). Apparent motion: Event-related functional magnetic resonance imaging of perceptual switches and States. *Journal of Neuroscience, 22,* RC219. [PubMed] [Article]

Nakatani, H., & van Leeuwen, C. (2006). Transient synchrony of distant brain areas and perceptual switching in ambiguous figures. *Biological Cybernetics, 94,* 445–457. [PubMed]

Noest, A. J., van Ee, R., Nijs, M. M., & van Wezel, R. J. (2007). Percept-choice sequences driven by interrupted ambiguous stimuli: A low-level neural model. *Journal of Vision, 7*(8):10, 1–14, http://journalofvision.org/7/8/10/, doi:10.1167/7.8.10. [PubMed] [Article]

Paffen, C. L., Alais, D., & Verstraten, F. A. (2006). Attention speeds binocular rivalry. *Psychological Science, 17,* 752–756. [PubMed]

Pastukhov, A., & Braun, J. (2007). Perceptual reversals need no prompting by attention. *Journal of Vision, 7*(10):5, 1–17, http://journalofvision.org/7/10/5/, doi:10.1167/7.10.5. [PubMed] [Article]

Pearson, J., Tadin, D., & Blake, R. (2007). The effects of transcranial magnetic stimulation on visual rivalry. *Journal of Vision, 7*(7):2, 1–11, http://journalofvision.org/7/7/2/, doi:10.1167/7.7.2. [PubMed] [Article]

Pressnitzer, D., & Hupé, J. M. (2006). Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Current Biology, 16,* 1351–1357. [PubMed] [Article]

Pressnitzer, D., Micheyl, C., Sayles, M., Winter, I. M. (2007). Responses to long-duration tone sequences in the cochlear nucleus. ARO 30th midwinter meeting, (p. 131). Denver, USA.

Ramachandran, V. S., & Anstis, S. M. (1983). Perceptual organization in moving patterns. *Nature, 304,* 529–531. [PubMed]

Rock, I., Gopnik, A., & Hall, S. (1994). Do young children reverse ambiguous figures? *Perception, 23,* 635–644. [PubMed]

Rock, I., & Mitchener, K. (1992). Further evidence of failure of reversal of ambiguous figures by uninformed subjects. *Perception, 21,* 39–45. [PubMed]

Rubin, N., & Hupé, J.-M. (2005). Dynamics of perceptual bi-stability: Plaids and binocular rivalry compared. In A. Alais & R. Blake (Eds.), *Binocular rivalry*. Cambridge: MIT Press.

Sato, M., Baciu, M., Loevenbruck, H., Schwartz, J. L., Cathiard, M. A., Segebarth, C., et al. (2004). Multistable representation of speech forms: A functional MRI study of verbal transformations. *Neuroimage, 23,* 1143–1151. [PubMed]

Sheppard, B. M., & Pettigrew, J. D. (2006). Plaid motion rivalry: Correlates with binocular rivalry and positive mood state. *Perception, 35,* 157–169. [PubMed]

Shpiro, A., Curtu, R., Rinzel, J., & Rubin, N. (2007). Dynamical characteristics common to neuronal competition models. *Journal of Neurophysiology, 97,* 462–473. [PubMed] [Article]

Soto-Faraco, S., Lyons, J., Gazzaniga, M., Spence, C., & Kingstone, A. (2002). The ventriloquist in motion: Illusory capture of dynamic information across sensory modalities. *Brain Research, 14,* 139–146. [PubMed]

Sterzer, P., Eger, E., & Kleinschmidt, A. (2003). Responses of extrastriate cortex to switching perception of ambiguous visual motion stimuli. *Neuroreport, 14,* 2337–2341. [PubMed]

Sterzer, P., & Kleinschmidt, A. (2007). A neural basis for inference in perceptual ambiguity. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 323–328. [PubMed] [Article]

Tong, F., Meng, M., & Blake, R. (2006). Neural bases of binocular rivalry. *Trends in Cognitive Sciences, 10,* 502–511. [PubMed]

van Ee, R. (2005). Dynamics of perceptual bi-stability for stereoscopic slant rivalry and a comparison with grating, house-face, and Necker cube rivalry. *Vision Research, 45,* 29–40. [PubMed]

van Ee, R., van Dam, L. C., & Brouwer, G. J. (2005). Voluntary control and the dynamics of perceptual bi-stability. *Vision Research, 45,* 41–55. [PubMed]

Van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences*. Doctoral dissertation, Eindhoven University of Technology.

Vetter, G., Haynes, J. D., & Pfaff, S. (2000). Evidence for multistability in the visual perception of pigeons. *Vision Research, 40,* 2177–2186. [PubMed]

Von Helmholtz, H. (1925). *Treatise on physiological optics*. Dover, NY: Southall, J. P.

Wallach, H. (1935). Uber visuell wahrgenommene Bewegungsrichtung. *Psychologische Forschung, 20,* 325–380.

Wilson, E. C., Melcher, J. R., Micheyl, C., Gutschalk, A., & Oxenham, A. J. (2007). Cortical fMRI activation to sequences of tones alternating in frequency: Relationship to perceived rate and streaming. *Journal of Neurophysiology, 97,* 2230–2238. [PubMed] [Article]

Wilson, H. R. (2003). Computational evidence for a rivalry hierarchy in vision. *Proceedings of the National Academy of Sciences of the United States of America, 100,* 14499–14503. [PubMed] [Article]

Wuerger, S., Shapley, R., & Rubin, N. (1996). "On the visually perceived direction of motion" by Hans Wallach: 60 years later. *Perception, 25,* 1317–1318.

Wunderlich, K., Schneider, K. A., & Kastner, S. (2005). Neural correlates of binocular rivalry in the human lateral geniculate nucleus. *Nature Neuroscience, 8,* 1595–1602. [PubMed] [Article]