Brain state and changes of mind: Probing the neural bases of multi-stable perceptual dynamics
Kloosterman, N.A.

Citation for published version (APA):
Chapter 7

Summary and Discussion
The internal state of our brain changes constantly, affecting the way in which the cerebral cortex processes information. Changes of cortical state have traditionally been associated with slow and largely automatic fluctuations of wakefulness and arousal, but they can also occur on a rapid (sub-second to second) time scale and be triggered in a top-down fashion by cognitive acts – for example, detecting perceptual changes in a visual stimulus. In this thesis, I aimed to detail the neurophysiological properties of fast changes in cortical state and their consequences for perception.

Chapter 2 reports an MEG study investigating modulations in cortical population activity during perceptual changes in the MIB bistable visual illusion and their consequences for perception. My colleagues and I observed a transient, retinotopically widespread modulation of beta (12-30 Hz) frequency power over visual cortex that is closely linked to the time of subjects’ behavioral report of the target disappearance. We show that this beta-modulation is a top-down signal, in the sense that it is decoupled from both the physical stimulus properties and the motor response, but contingent on cognitive factors such as behavioral relevance of perceptual changes. Critically, the modulation amplitude predicts the duration of the subsequent illusory target disappearance. Taken together, these findings suggest that the transformation of the perceptual change into a report triggers a top-down mechanism that stabilizes the newly selected perceptual interpretation.

Chapter 3 further pinpoints the factors driving the beta-band modulation by demonstrating that the beta suppression over visual cortex occurs not only when subjects promptly report the target disappearances by button press, but also when they silently count them. This finding establishes behavioral relevance as a key factor underlying the top-down signal and indicates that the beta modulation is not due to the motor act used to report switches. Chapter 4 provides further psychophysical support for a stabilization mechanism in bistable perception that prevents immediate return transitions by showing that the typically observed lack of very brief percepts in bistable perception is not due to an inability to report these short-lived percepts.

Due to their abundant projections to the cerebral cortex, neuromodulatory centers in the brainstem are able to control the global brain state. The activity of these brainstem centers can be measured non-invasively in the phasic dilation of the pupil. We reasoned that if these systems exert such an influence around the time of perceptual decisions,
then there should be pupil dilation during the perceptual switches. The pupillometry study reported in Chapter 5 shows that the pupil indeed transiently dilates during perceptual switches, even in the absence of immediate motor reports. Importantly, we report that the pupil responds differently depending on the type of perceptual switch that is reported during MIB, with stronger dilation following target disappearance than re-appearance report. This finding is consistent with the modulation of fMRI- and MEG-activity around perceptual reports that also reflects perceptual content during MIB and Replay, suggesting that pupil dilation indeed provides an index of fast changes in brain state. In addition, this pupil response scales with the level of surprise about perceptual switches, suggesting that surprise about the timing of perceptual events also affects the internal state of the brain.

Finally, in Chapter 6 we directly correlate this index of brain state to modulations of neural population activity during the perceptual switches. The chapter reports concurrent pupillometry and whole-brain fMRI to characterize the complete cortical and subcortical distribution of activity modulations during perceptual switches in MIB separately for trials associated with large and small pupil dilation. We find modulations of fMRI activity throughout early visual, parietal and anterior cingulate cortices (ACC) around the perceptual events. While visual and parietal cortex are key players in the MIB illusion, ACC is known to send top-down inputs to neuromodulatory brainstem centers such as the locus coeruleus (LC). The stronger suppression of fMRI-activity in visual cortex during MIB disappearance trials with larger phasic pupil dilation provides a new clue for a link between neuromodulatory systems controlling pupil diameter and top-down signals in the brain.

Despite this finding, this study also highlights that the majority of the fMRI-activity modulation in our peripheral region of V1 shows no relationship with the concomitant phasic dilation of the pupil. At least two factors other than pupil-linked neuromodulation could be at play. First, slow fluctuations in attention level that presumably occur throughout the continuous presentation of the MIB stimulus might also affect the trial-to-trial modulation of fMRI-activity. Second, fMRI-activity in the central, stimulated parts of V1 showed strong positive modulations around the perceptual switches in our study, which possibly resulted in suppression of fMRI-activity in the periphery of V1 due to reduced blood flow to this region (Woolsey et al., 1996).
Taken together, the work presented in this thesis establishes the existence of a novel type of top-down signal in visual cortex around perceptual decisions, which has a profound influence on neural activity and the dynamics of perception. In this last chapter, I will elaborate on the possible sources of the top-down signal, focusing on phasic neuromodulation. To accommodate this set of results within a currently prevalent theoretical framework, I will provide an extension of the standard class of models of bistable perception (see Chapter 1) that modulates the activity of stimulus-specific populations of neurons following the commitment to a new percept. Further, I will review existing evidence for top-down signals during other perceptual tasks, speculate about their role in higher-level cognitive tasks and psychopathology and posit research questions to be addressed in future studies.

The origin of fast changes in brain state

What is the source of the top-down signals observed in visual cortex? The signal might originate from higher cortical areas (Nienborg and Cumming, 2009; Siegel et al., 2012), the thalamus (Wilke et al., 2009), neuromodulatory brainstem centers (Aston-Jones and Cohen, 2005; Parikh et al., 2007; Einhauser et al., 2008; Hupe et al., 2009; de Gee et al., 2014), or from a combination of cortical feedback and neuromodulation (Noudoost and Moore, 2011).

A number of characteristics of the top-down signal reported in this thesis suggest that phasic neuromodulation might be the key underlying mechanism. Neuromodulatory brainstem systems, such as the noradrenergic (NA) locus coeruleus (LC) and the cholinergic (Ach) basal forebrain systems, also exhibit transient activity during perceptual reports, which can reflect the content of the report, both during bistable perceptual tasks (Einhauser et al., 2008; Hupe et al., 2009) and detection tasks (Aston-Jones and Cohen, 2005; Parikh et al., 2007; de Gee et al., 2014). In addition, these brainstem systems have strong anatomical connections with the cerebral cortex (Aston-Jones and Cohen, 2005) (Figure 1). Further, beta-band oscillations, as observed in the MEG around perceptual switches in Chapters 2 and 3, are often associated with attentional demands and cognitive tasks (Engel and Fries, 2010; Donner and Siegel, 2011). Moreover, modulations of beta-band power in visual cortex during visual stimulation have been suggested to index changes in neuromodulatory state (Belitski et al., 2008; Donner and Siegel, 2011). Neuromodulatory brainstem systems might be in an ideal position to stabilize bistable
perceptual dynamics, because they can dynamically alter key cortical circuit parameters in profound ways. In particular, neuromodulators suppress cortical variability (Polack et al., 2013) and may amplify inhibitory interactions in cortical circuits (Haider et al., 2012).

Finally, given that the anterior cingulate cortex (ACC) is the hypothesized source of decision-related signals in the LC (Aston-Jones and Cohen, 2005), the finding of fMRI-activation in ACC around perceptual reports during MIB (Chapter 6) also fits well with this explanation.

In spite of this indirect evidence, more research is needed to directly test the role of neuromodulatory systems in the top-down modulation. One fruitful approach could be to manipulate central NA or Ach levels pharmacologically in humans during bistability or other perceptual decision making tasks while measuring neural activity with MEG. If these systems play such a role, then amplifying their level of activity would result in both larger pupil dilation as well as stronger decision-related beta-band modulation in visual cortex. Additionally, by comparing the variability of neural activity during drug and placebo conditions (for instance by computing the trial-to-trial variance), this approach could reveal whether increased neuromodulator levels are related to suppression of cortical variability.

**A simple computational model of bistable perceptual stabilization**

How can a transient top-down signal in visual cortex stabilize a perceptual illusion? As explained in Chapter 2, the movement of a “percept variable” (green ball in Figure 2A) (Moreno-Bote et al., 2007; Braun and Mattia, 2010) across an energy landscape with
two valleys (basins of attraction; in the case of MIB corresponding to target visible and invisible) provides a useful metaphor for understanding this effect (Deco and Romo, 2008). In this scheme, the stabilizing state change can be conceived as an active force (red arrow) transiently deepening the valleys. Only if the sensory input is ambiguous and, consequently, the perceptual interpretation meta-stable (i.e., during MIB), does this transient state change culminate in a perceptual stabilization: the stronger the state change (i.e., the longer the red arrow) during a perceptual transition, the longer the subsequent perceptual illusion. During Replay, the physical removal of the target stimulus instantaneously alters the energy landscape, thus overriding the effect of the internal state change and precluding a link to percept duration.

By simulating an extension of one version of the current neural models of bistable perception (Noest et al., 2007) we established that this stabilization scheme during MIB could be implemented by transiently boosting, via feedback, the strength of mutual inhibition between stimulus-selective neural populations in visual cortex (Figure 2B, C). As explained in the Introduction, the standard model consists of two populations of neurons driven by distinct stimulus components (e.g. static target and moving mask). These populations are subject to slow decay in activity due to adaptation and compete with each other by mutual inhibition. The interaction between adaptation, noise, and mutual inhibition gives rise to spontaneous dominance transitions between the two populations. The dominance transitions are thought to underlie the perceptual switches.

We\textsuperscript{1} extended this model by adding a third, modulatory neural population, which was driven by the dominance transitions in the two competing populations and, in turn, sent feedback to both competing populations (Figure 2B). The feedback was not selective for one of the competing visual populations (as would, for instance, be expected for top-down selective attention (Desimone and Duncan, 1995; Harris and Thiele, 2011), but equally impinged on both visual populations. Importantly, however, the feedback was temporally specific, peaking precisely at the time of the dominance transition (inset in Figure 2B). The feedback influence modulated the gain of the mutual inhibition between the two cortical populations, which, in turn, deepens the valleys in the energy landscape in Figure 2A. The phasic release of modulatory neurotransmitters such as noradrenaline (NA) or acetylcholine (Ach), could mediate such an effect (Aston-Jones and Cohen, 2005;

\textsuperscript{1} The model extension was conceived and implemented by Tobias Donner and Tomas Knapen.
**Figure 2 | Conceptual model of beta-power transient**

**A.** Schematic of dynamical algorithm for cortical state change and perceptual stabilization. In both MIB (top) and Replay-active (bottom), the percept (green ball) is in the “target visible” valley. Adaption gradually flattens this valley before target disappearance (sequence: t₂, t₁, t₀). The ball then hops into the “target invisible” valley (perceptual switch, t₁). Behavioral report of this perceptual event induces a state change that deepens the target invisible valley (t₂, red arrows). This, in turn, stabilizes perception during MIB (t₃: top right): When the state change is strong, the percept variable is less likely to move back to the visible valley some time after the switch (t₃). By contrast, during Replay (t₃: bottom right), the physical target reappearance alters the energy landscape (i.e., eliminates the target invisible valley), and thereby prevents the state change at t₂ from affecting the percept duration. 

**B.** Extension of the bistability model. Two neural populations (X₁ and X₂) are driven by distinct inputs (I₁, I₂), subject to slow adaptation dynamics, and compete via mutual inhibition. In our extension of this class of models, transients of X₁ and X₂ drive a third population (M), which sends non-specific modulatory feedback to the competing populations, boosting their mutual inhibition. Inset: Average time course of M-output, time-locked to switches in dominance between X₁ and X₂. 

**C.** Correlation between M-response during dominance transition and duration of subsequent dominance state across different strengths of the inhibition boost.
Einhauser et al., 2008; Sarter et al., 2009). Specifically, Polack et al. (2013) recently found that NA depolarized membrane potentials of neurons in mouse visual cortex during locomotion, thereby enhancing the gain (multiplicative transformation) of the V1 neurons.

Our simulation confirmed that the stronger the transient response of the “modulatory” (M) population during perceptual switches, the longer the subsequent dominance duration lasted. There was a robust, positive correlation between the strength of the transient M-response around the switch and the subsequent dominance duration (Figure 2C). The magnitude of this correlation increased monotonically with the strength of the relative contribution of the inhibition boosting effect to the overall mutual inhibition (see Supplement 1 for model details). Note that the activity of the modulatory population was positively correlated to the percept duration, whereas the report-related beta-band modulation in visual cortex observed in our MEG measurements in Chapter 1 was negatively correlated to the percept duration. One explanation for the opposite sign could be that the beta-band modulation in visual cortex did not reflect the modulatory population per se (which we assume is located in a brain region outside of visual cortex and shows an increase in activity, see The origin of fast changes in brain state), but rather its modulatory feedback effect on neural (dendritic) population activity in visual cortex (i.e., state change), as measured by MEG. Future work could investigate whether the possible brain regions containing modulatory populations, such as the LC or basal forebrain, indeed increase their activity around the decisions.

Although this model is consistent with an active role of phasic neuromodulation around perceptual switches, there are other possibilities. First, by simulating an alternative feedback model, we found that a global suppression of the input gain in visual cortex yielded similar results as the inhibition boost (data not shown). However, if reflecting global input suppression, the MEG power modulations during MIB disappearance should have been the converse of the stimulus-induced MEG response shown in Chapter 2 (Figure 1C) and Chapter 3 (Figure 1C). This would predict high-frequency suppression and low-frequency enhancement, in sharp contrast with the data (see e.g. Chapter 2, Figure 2D and Chapter 3, Figure 2A). Second, passive adaptation of the neurons representing the target stimulus could also have a stabilizing effect due to a prolonged suppression of this weakened population’s activity by the competing pool of sensory neurons. Future studies could investigate whether adaptation alone could account for a perceptual stabilization
mechanism, for instance by testing whether subjects with stronger adaptation dynamics also show longer median bistable percept durations.

**Concluding remarks**

In this thesis, I have explored how a new and unexpected class of top-down signals shapes neural activity and perception in the human brain. These signals occur around perceptual decisions about the disappearance of a salient stimulus and seem to reflect the transformation of a perceptual change into a behavioral report. One possible underlying mechanism of these signals is a phasic activation of brainstem systems, which diffusely release neuromodulator across the brain around decisions about task-relevant perceptual changes, thereby changing its internal state.

The global (i.e. retinotopically non-specific) state change in visual cortex is likely a general phenomenon. Besides during MIB, functionally analogous modulations of fMRI activity have been observed in human early visual cortex around perceptual reports in various other bistable phenomena: bistable motion binding (de-Wit et al., 2012), plaids (N. Rubin, personal communication), 3D structure from motion, apparent motion, and binocular rivalry (unpublished observations from our lab), as well as during visual detection tasks (Nienborg and Cumming, 2009; Choe et al., 2014).

It has remained unknown which (if any (Sirotin and Das, 2009)) component of electrophysiological activity these global hemodynamic responses reflect, and how the underlying cortical state change affects perception. By characterizing the underlying patterns of electrophysiological population activity in visual cortex, this thesis reveals beta-band modulation as an index of internal state changes, in line with results from LFP recordings in monkey (Wilke et al., 2006; Belitski et al., 2008). Further, the work presented here establishes that the global state change around perceptual reports shapes the stability of bistable perceptual interpretations. Taken together, the thesis provides an initial understanding of this previously neglected class of signals in visual cortex. Future studies could test the scope of these top-down signals across other bistable illusions, such as binocular rivalry (Knapen et al., 2011), 3D structure-from-motion (Klink et al., 2012), or the Necker cube, as well as during non-perceptual decision tasks, such as economic or value-based decision-making.
The findings presented in this thesis have implications for the putative link between the fMRI-signal and neural oscillations, as measured here with MEG. During visual stimulation, both the fMRI-signal and oscillations in the high frequency (gamma) band increase (Niessing et al., 2005; Donner and Siegel, 2011). In contrast, around the perceptual reports in MIB and its replay we observed no modulation in the gamma band but strong modulation in both the low-frequency band and the fMRI signal. This suggests that the coupling between band-limited oscillatory activity and the fMRI-signal is process-dependent, with no fixed mapping between these signals (Kopell et al., 2000; Maier et al., 2008; Donner and Siegel, 2011). Moreover, it suggests that the fMRI-response reflects a mixture of stimulus-related signals and neuromodulatory signals related to top-down factors, such as task instructions. Future studies could employ simultaneous fMRI and EEG during perceptual tasks to study the relative contributions of these stimulus-related and top-down signals to the fMRI-signal on a trial-by-trial basis (Philiastides and Sajda, 2007).

It is tempting to speculate that the stabilization mechanism introduced in this thesis generalizes to psychological phenomena related to decision commitment. Specifically, decision-makers often show a striking tendency to stick to their previous judgments, even in the face of salient counterevidence. For example, a juror could decide whether a defendant is guilty based on preliminary evidence and then fails to reconsider that decision even when strong contradictory evidence is presented at trial – a phenomenon known as confirmation bias (Nickerson, 1998). Confirmation bias can be seen as an excessive stabilization of an initial decision, resulting in extreme commitment to the decision and a bias to confirm it\(^2\). Relatedly, in the phenomenon of cognitive dissonance an individual is confronted with new information that conflicts with existing beliefs, ideas, or values, leading to psychological discomfort (Festinger, 1957). Similarly, these fixed, incongruent beliefs could be the result of an excessive commitment to a previous decision. It would be interesting to investigate whether similar changes in brain state occur during commitment to decisions in these phenomena and if the strength of top-down signals during decisions scales with the strength of decision commitment (Jazayeri and Movshon, 2007; Stocker and Simoncelli, 2008).

\(^2\) Another example is the neuromodulation hypothesis laid out in this thesis.
Finally, understanding fast brain state changes identified in this thesis can help to understand psychopathological conditions in which the dynamics of thought and brain state are disturbed. For instance, depressive patients tend to dwell on the same negative thoughts for a long time, without being able to change their mind. It is possible that these symptoms result from an altered functioning of the stabilization mechanism, possibly mediated by the disturbance of neuromodulatory systems in these patients. Another condition that could possibly involve altered brain state dynamics is schizophrenia. One unresolved question is why delusions (for instance, paranoid convictions) are so persistent in these patients, despite a lack of supporting sensory evidence (Schmack et al., 2013). Again, this symptom can be seen in terms of commitment to a decision or belief, which in this case is strikingly dissociated from reality. Indeed, schizophrenic patients show altered bistable perceptual dynamics (Sanders et al., 2014), suggesting alterations in the neural mechanism underlying the time course of bistable perception. Using MIB and other bistable phenomena, future studies could investigate whether the bistable dynamics in schizophrenia are affected by an altered functioning of top-down signals associated with brain state (Notredame et al., 2014).

In this thesis, I aimed to establish the existence of fast changes in brain state that occur as a consequence of cognitive processes. These changes in brain state not only affect brain activity, but also shape the contents of our subjective experience over time. The mechanisms described here might explain one of the most striking properties of consciousness: the spontaneous and abrupt changes in our everlasting stream of thought.
SUPPLEMENT 1: SIMULATION METHODS FOR NEURAL MODEL OF VISUAL BISTABILITY WITH FEEDBACK MODULATION

We extended an established model of the neural mass dynamics underlying perceptual bistability in visual cortex (Noest et al., 2007). The “standard model” consists of two populations of neurons ($X_1$ and $X_2$), which are driven by two distinct constant stimuli ($I_1$ and $I_2$), and inhibit each other, and which are subject to slow adaptation and noise. In our extension, the following pair of differential equations governed the dynamics of these two populations:

$$
\tau \frac{dX_1}{dt} = I_1 - (1 + A_1)X_1 - (\gamma + \mu S[M])S[X_2] + N(0,\sigma_{X_1})
$$

$$
\tau \frac{dX_2}{dt} = I_2 - (1 + A_2)X_2 - (\gamma + \mu S[M])S[X_1] + N(0,\sigma_{X_2})
$$

where $S[X]$ corresponds to a Naka-Rushton function of the “local fields” in $X_i$ (Noest et al., 2007), $N$ corresponds to normally distributed noise, and $A_i$ describes adaptation. See Table 1 for definitions of the other model parameters, and their selected values. The “standard model” does not contain $M(\mu = 0$, i.e. no effect of $M$). The adaptation dynamics is given by:

$$
\tau_A \frac{dA_i}{dt} = -A_i + \alpha S[X_i],
$$

$i, j \in \{1, 2\}$  

Simulating this model with continuous inputs $X_1$ and $X_2$ yielded spontaneous fluctuations in the activity of $X_1$ and $X_2$, which, in turn, produced spontaneous alternations between dominance of the two populations of visual cortical neurons.

We extended this standard model by means of a “modulatory population” $M$, which was driven by the two cortical populations:

$$
\tau_M \frac{dM}{dt} = -A_i + \nu S[X_1] + \nu S[X_2] + N(0,\sigma_M)
$$

$M$ fed back to the two cortical populations $X_i$ and modulated the strength of their mutual inhibition in an additive fashion, governed by parameter $\mu$ in eq. 1, which controls the effect of $S[M]$ on the two cortical populations. See Table 1 for a complete list of parameters used in the simulations.
We systematically varied the strength of the parameter controlling the relative contribution of the $M$-influence on the mutual inhibition between $X_1$ and $X_2$ across a wide range of values, simulated several hundreds of dominance transitions (minimum: 200) for each of these parameter values, and finally computed Spearman correlations between transient $M$-activity and dominance duration for each of these parameter values (Figure 2C). To this end, we extracted the durations of individual dominance intervals from the simulated dynamics of $X_1$ and $X_2$ as follows. We computed the time course of the difference between $S[X_1]$ and $S[X_2]$ and temporally smoothed this difference time course using a Gaussian kernel ($\sigma$: 25 time steps). Any transgression of 0 was counted as a switch in dominance. Dominance durations were then computed as the intervals between two switches. We averaged the activity of $M$ in the interval of [-50,150] time steps around each transition, to correlate this with the dominance durations. We computed the $M$-activity based on

Table 1, related to Figure 2 | Model parameters used for simulations

| Model parameters, their meanings, and values used in simulations |
|----------------------|------------------|
| $\tau$               | Timescale of $X$ population activities 1.0 |
| $\tau_a$             | Timescale of adaptation 125.0 |
| $\tau_m$             | Timescale of $M$ population activity 15 |
| $\alpha$             | Gain of Adaptation 4.0 |
| $\nu$                | Gain of $M$-population 120 |
| $\gamma$             | Inhibition Gain 3.0 |
| $\mu$                | Gain of influence of $M$ on $X$-populations 40 [0.1,100] (on inhibition) 3 [0.1, 4.0] (on inputs) |
| $\sigma_{x_1}, \sigma_{x_2}$ | Standard deviation of noise in $X$-populations 0.003, [0.0, 0.015] |
| $\sigma_m$           | Standard deviation of noise in $M$-population 0.03, [0.0, 0.05] |

Numbers are (means of) the parameters used for simulations reported in this paper. Numbers in parentheses indicate the range of parameters used. Only the gain of the modulatory $M$-influence on the $X$ populations in the noise level in all three populations was varied across the range indicated.

Analyses of model dynamics and perceptual stabilization

We systematically varied the strength of the parameter controlling the relative contribution of the $M$-influence on the mutual inhibition between $X_1$ and $X_2$ across a wide range of values, simulated several hundreds of dominance transitions (minimum: 200) for each of these parameter values, and finally computed Spearman correlations between transient $M$-activity and dominance duration for each of these parameter values (Figure 2C). To this end, we extracted the durations of individual dominance intervals from the simulated dynamics of $X_1$ and $X_2$ as follows. We computed the time course of the difference between $S[X_1]$ and $S[X_2]$ and temporally smoothed this difference time course using a Gaussian kernel ($\sigma$: 25 time steps). Any transgression of 0 was counted as a switch in dominance. Dominance durations were then computed as the intervals between two switches. We averaged the activity of $M$ in the interval of [-50,150] time steps around each transition, to correlate this with the dominance durations. We computed the $M$-activity based on
the local fields (i.e., before output non-linearity), because these provide a closer proxy of MEG measurements than neural outputs (Donner and Siegel, 2011).