Perceptual choice boosts network stability: effect of neuromodulation?

Christopher M. Warren[1], Sander Nieuwenhuis[1], and Tobias H. Donner[2]

[1] Institute of Psychology and Leiden Institute for Brain and Cognition, Leiden University, Leiden, The Netherlands
[2] Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

A recent paper demonstrates that conscious perceptual decisions are characterized by a hallmark of attractor states in recurrent cortical networks: increased stability of cortex-wide activity patterns. We propose that this global cortical state change may be caused by a transient gain modulation through ascending brainstem systems.

Most of us have had the experience of waiting for a friend or family member to pick us up at the airport. A car appears in the distance: *Is that them? It's blue, and their car is blue. Is it an SUV or a minivan? They drive a minivan.* As the car gets closer and time passes, the evidence accumulates until you are sure it's them, and you wave. During a perceptual decision like this, your brain moves from a labile, input-sensitive state to a stable, input-insensitive state, often referred to as an attractor state [1]. This attractor state reflects the brain's commitment to one categorical interpretation of the incoming sensory data (*it's them*)—often followed by the commitment to a behavioral response (*wave*). In simulations of the underlying cortical network dynamics, this attractor state is characterized by increased stability (i.e., decreased variability) of neural activity [1]. A recent article in *Proceedings of the National Academy of Sciences of the United States of America* by Schurger and colleagues provides the first direct evidence for this hypothesis [2].

Schurger and colleagues asked their human participants to view simple line drawings of faces and houses. The visibility of these stimuli was systematically varied over a range of five levels around the psychophysical detection threshold. The participants' task was to indicate the category of each stimulus, guessing if necessary, and then to report whether the stimulus had been 'seen' or 'unseen.' This protocol allowed the researchers to compare trials that differed in subjective visibility ('seen' versus 'unseen'), but were matched in terms of objective stimulus strength (sensory evidence). Neural activity was recorded with simultaneous magnetoencephalography (MEG) and electroencephalography (EEG). As a proxy of cortical network stability, the authors quantified the dissimilarity between successive spatial patterns of MEG activity across

sensors over the course of a trial (Figure 1A). For brevity, we henceforth refer to this metric as 'pattern variance.'

The MEG data revealed a period from approximately 500 to 800 milliseconds after stimulus onset during which the spatial pattern of activity was more stable on 'seen' trials than on 'unseen' trials. Importantly, the decrease in MEG pattern variance during 'seen trials' was accompanied by a decrease in overall signal strength, ruling out the concern that the suppression of pattern variance (i.e., increased stability) may reflect an increased signal-to-noise ratio under high visibility (predicting higher signal strength). These findings suggest that the brain's stable commitment to one interpretation of the sensory data is characteristic of conscious perception. In a re-analysis of EEG data from a large pool of patients with disorders of consciousness (of varying degree) and healthy controls, the authors showed that within-trial neural stability does not only reflect the contents of consciousness in fully conscious subjects, but it also reflects people's overall levels of consciousness. Schurger and colleagues were careful to exclude a number of alternative explanations for their findings. For example, they showed that the suppression of MEG pattern variance for 'seen' trials did not just reflect a suppression of oscillatory brain activity in the alpha band, which had been reported in earlier studies and was also found in the present one. Repeating the stability analysis after removing alpha-band activity from the MEG data did not eliminate the reported pattern variance effects.

Intriguingly, Schurger and colleagues also identified an EEG signal, the peak of which closely preceded the suppression in MEG pattern variance, that may reflect an underlying process that drives the emergence of stability: a centro-parietal late positive potential that is often referred to as the P300. In line with previous research, the P300 had significantly higher amplitude on 'seen' trials than on 'unseen' trials. As the authors point out, these findings match with the proposal that the P300 reflects the sudden transition of the cortical network into the global attractor state that culminates in conscious perception [3]. This state transition is often referred to as 'ignition.' The new cortical network state is characterized by the activation of neurons in association cortex engaging in recurrent interactions with sensory cortex (Figure 1B).

Here, we propose an alternative mechanistic scenario for the reported results: a transient boost in the gain of cortical interactions due to decision-related neuromodulator release (Figure 1C). The variability in cortical neural activity, both across trials [4] and within trials [5], is
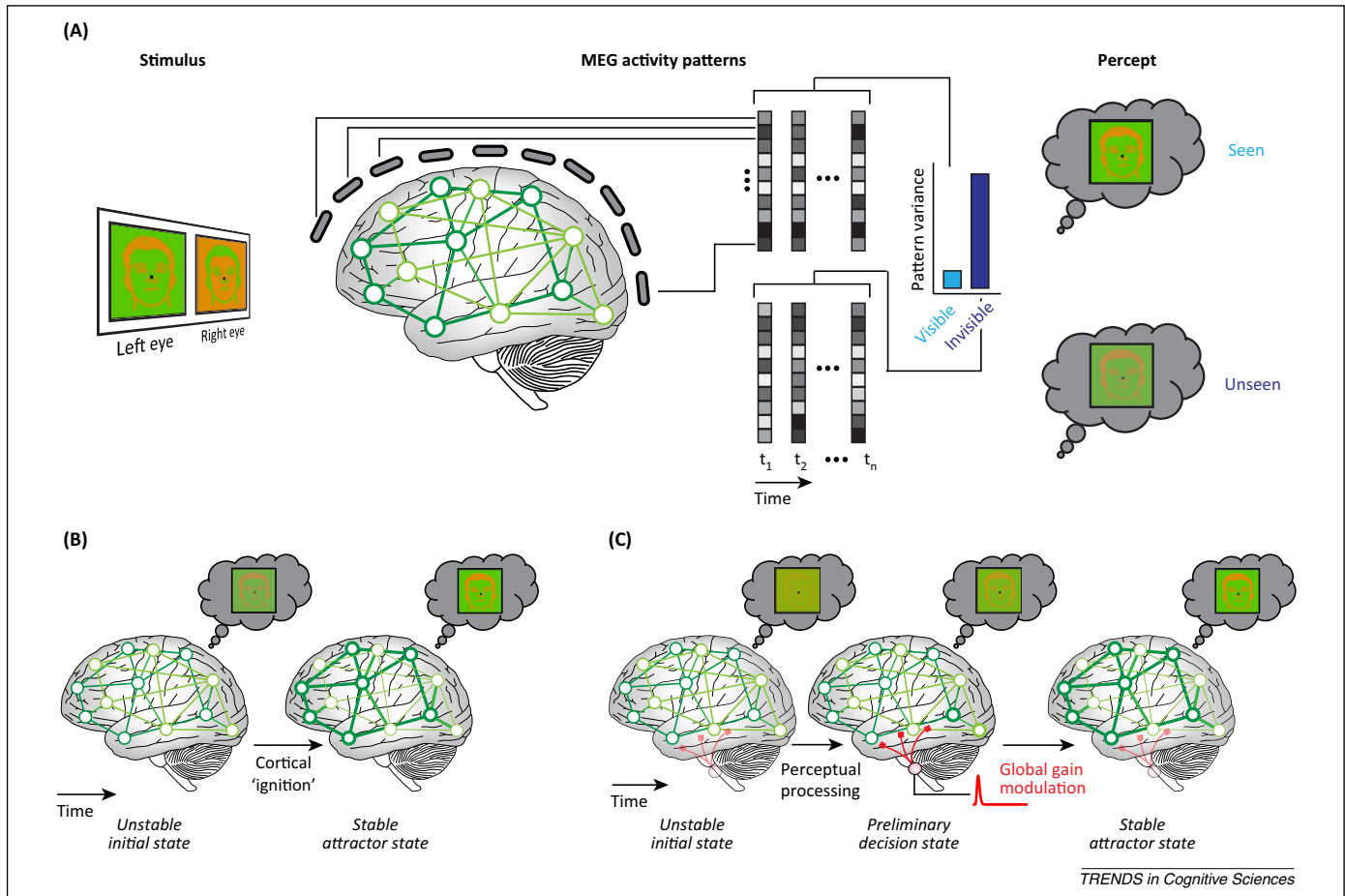
Figure 1. Suppression in neural variability during conscious perception. **A.** Schematic illustration of the approach and result of Schurger and colleagues [2]. Spatial patterns of MEG activity are extracted from multiple time points ($t_1, \ldots t_n$) during a stimulus presentation trial (faces or houses rendered difficult to see via dichoptic fusion). The dissimilarity of these successive patterns in the N-dimensional space spanned by the N MEG sensors is quantified in terms of the directional variance (here called 'pattern variance') and then compared between trials in which the stimulus was seen and those in which the stimulus was not seen. 'Seen trials' are associated with smaller MEG pattern variance, reflecting a stable cortical attractor state. **B, C.** Alternative mechanisms leading from the unstable initial state to the stable attractor state. B. 'Cortical ignition' refers to the recruitment of a large-scale cortical network through sufficient recurrent activation (i.e., the 'global workspace'). C. 'Global gain modulation' involves a cortex-wide neuromodulatory increase in the responsivity of neural networks, which suppresses 'neural noise' and stabilizes selected cortical activity patterns.

regulated by neuromodulators. Brainstem neuromodulatory systems, such as the norepinephrine system, are activated transiently during perceptual decisions [6]. Global transient signals consistent with decision-related neuromodulation have been observed in visual cortex during the report of changes in perception [7] and predict the stability of subsequent perceptual states [8]. One possibility is that the cortical neuromodulatory effects described above are mediated by changes in neural gain (input-to-output multiplier). Intriguingly, previous work has suggested that the P300 reflects a transient increase in neural gain due to the phasic release of norepinephrine [9]. This mechanistic interpretation of the P300 awaits definitive evidence, yet it fits nicely with the findings of Schurger and colleagues. In sum, the increased cortical network stability associated with a conscious perceptual decision observed by Schurger and colleagues may result from a transient modulatory boost in neural gain (Figure 1C).

The innovative study by Schurger and colleagues contributes to mounting evidence suggesting that the study of neural variability will be instrumental in unraveling the mechanisms of cognition (e.g., [10]). The results of the present study are exciting and point to a number of interesting avenues for future work. For example, future empirical and simulation studies should shed light on the mechanistic relationship between the change in neural stability and the P300 component (or other event-related potential components), which are overlapping in time. Another intriguing question is what exactly is becoming more stable during the formation of a decision: is it the cortical representation of the perceived stimulus only, or the entire cortical network? Future studies could restrict the stability analysis to those sensors that carry (or do not carry) information about the experimental stimuli.

References

1 Wang, X.J. (2008) Decision making in recurrent neuronal circuits. *Neuron* 60, 215–234

2 Schurger, A. *et al.* (2015) Cortical activity is more stable when sensory stimuli are consciously perceived. *Proc. Natl. Acad. Sci. U.S.A.* 112, http://dx.doi.org/10.1073/pnas.1418730112

3 Dehaene, S. *et al.* (2003) A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proc. Natl. Acad. Sci. U.S.A.* 100, 8520–8525

4 Reimer, J. *et al.* (2014) Pupil fluctuations track fast switching of cortical states during quiet wakefulness. *Neuron* 84, 355–362

5 Polack, P.O. *et al.* (2013) Cellular mechanisms of brain state-dependent gain modulation in visual cortex. *Nat. Neurosci.* 16, 1331–1339

6 Aston-Jones, G. and Cohen, J.D. (2005) An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* 28, 403–450

7 Donner, T.H. *et al.* (2008) Opposite neural signatures of motion-induced blindness in human dorsal and ventral visual cortex. *J. Neurosci.* 28, 10298–10310

8 Kloosterman, N.A. *et al.* (2015) Top-down modulation in human visual cortex predicts the stability of a perceptual illusion. *J. Neurophysiol.* 113, 1063–1076

9 Nieuwenhuis, S. *et al.* (2005) Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychol. Bull.* 131, 510–532

10 Churchland, A.K. *et al.* (2011) Variance as a signature of neural computations during decision making. *Neuron* 69, 818–831

Predicting the other in cooperative interactions

Alan G. Sanfey[1,2], Claudia Civai[1], and Peter Vavra[1,2]

[1] Donders Institute for Brain, Cognition and Behavior, Radboud University Nijmegen, Kapittelweg 29, 6525 EN Nijmegen, The Netherlands
[2] Behavioral Science Institute, Radboud University Nijmegen, Postbus 9104, 6500 HE Nijmegen, The Netherlands

Recent research has shown that a collection of neurons in dorsal anterior cingulate cortex of rhesus monkeys may specifically encode the choice selection of an interaction partner. This raises interesting and important questions as to the nature of Theory of Mind processes in social interactive decision-making, with potential societal implications.

One notable aspect of human decision-making is the ubiquity of our cooperative interactions [1], both with specific others and with societal institutions more broadly. We generally return the favors of others, and thus might help a friend move with the expectation of future help in return. We also cooperate on a larger scale, such as paying our taxes, when we could potentially avoid doing so. Many of these social choices are risky in that we can't be certain if our positive acts will indeed be reciprocated in the future. A key factor in our decisions to cooperate is the degree to which we can predict that our partner in the exchange will be willing to commit to cooperation. Therefore one extremely important aspect of understanding the motivations and mechanisms underlying these important choices is how we represent the likely decisions of others.

In a recent compelling paper, Haroush and Williams [2] outline the case for a grouping of neurons in the primate brain that appear to specifically encode the choice selection of an interaction partner. These neurons, located in the dorsal anterior cingulate cortex (dACC) of rhesus monkeys, were observed using single-unit recording while the monkeys played a variant of the oft-studied iterative Prisoner's Dilemma game, wherein players must decide to either cooperate with a partner for a potential joint positive gain or defect to guarantee themselves a payoff at the expense of their partner. Using these signals, the monkey's own choice could be correctly predicted on over 65% of rounds. However, the same signals could be used to predict the other, physically present monkey's unobserved choices with even higher accuracy, namely 79%. In other words, these dACC neurons encoded information enabling the monkey to, at least in principle, predict the other's future behavior with high accuracy.

The concept of Theory of Mind (ToM) refers to the ability to understand and predict the behavior of others. By identifying cells that appear to represent the yet unknown intentions of a game partner, this study supports the idea that ToM is a fundamental and specific process. It also raises intriguing questions on three distinct levels. On the computational level: under which circumstances do these neurons get engaged, and how does this impact strategic decisions? On the neural level: how do these other-encoding dACC neurons fit into a larger ToM network that implements these computations? Finally, on a socio-behavioral level: to what extent does the physical presence of others modulate the perception of social context?

Firstly, these results shed light on the circumstances under which ToM is engaged. In a control experiment, when the first monkey defected and this choice was explicitly shown to the second monkey, the latter defected in turn on over 90% of the rounds; that is, the second monkey successfully avoided exploitation. Notably, however, when the monkeys made their choice without directly observing the decision of the other, they cooperated substantially more often. Given that the neural predictions were very accurate, and so presumably should not lead to different decisions than observation, what underlies this difference in cooperation rates? Does revealing one's intentions explicitly change how the ToM network of others is engaged, thus altering the tendency to cooperate by impacting the certainty of beliefs about the other's behavior? It may require a revision of current models of strategic behavior to account for these different levels of cooperation. Recent computational models based on human experiments with similar two-player games suggest that people adapt their decisions based on how they expect others will behave [3,4] as well as how they believe others expect them to behave [5]. One possible extension based on these results could therefore be to explicitly model how certain people are about such beliefs and how this (un)certainty affects their choices.

In terms of the broader neural basis of ToM, in humans this network encompasses posterior areas such as temporo-parietal junction (TPJ) and superior temporal sulcus