Directed information sharing in the brain distinguishes perception from its consequences during perceptual bistability

Andrés **Canales-Johnson**[1,2,3,4]∞, Lola **Beerendonk**[1,2]∞, Srivas **Chennu**[5], Matthew J. **Davidson**[6], Robin A.A. **Ince**[7], Simon **van Gaal**[1,2]

[1] Conscious Brain Lab, Department of Psychology, University of Amsterdam, Nieuwe Achtergracht 129-B, 1018 WT, Amsterdam, The Netherlands.

[2]Amsterdam Brain & Cognition, University of Amsterdam, Nieuwe Achtergracht 129-B, 1018 WT Amsterdam, The Netherlands.

[3] Cambridge Consciousness and Cognition Lab, Department of Psychology, University of Cambridge, Downing Street, CB2 3EB Cambridge, United Kingdom.

[4] Vicerrectoría de Investigación y Posgrado, Universidad Católica del Maule, Talca, Chile.

[5] School of Computing, University of Kent, Canterbury CT2 7NF, United Kingdom.

[6] School of Psychology, University of Sydney, Sydney, Australia

[7] Institute of Neuroscience and Psychology, University of Glasgow, Scotland G12 8QB, United Kingdom.

∞ Equal contribution

Corresponding author: Andres Canales-Johnson (afc37@cam.ac.uk), Simon van Gaal (s.vangaal@uva.nl)

Abstract

In the search for the neural basis of conscious experience, perception and its cognitive consequences are typically confounded as neural activity is recorded while participants explicitly report what they experience. Here we present a novel way to disentangle perception from report using eye-movement analysis techniques based on convolutional neural networks and neurodynamical analyses based on information theory. We use a bistable visual stimulus that instantiates two well-known properties of conscious perception: integration and differentiation. At any given moment, observers either perceive the stimulus as one integrated unitary object or as two differentiated objects that are clearly distinct from each other. Using electroencephalography, we show that measures of integration and differentiation based on information theory closely follow participants perceptual experience of those contents when switches were reported. We observed increased information integration between anterior to posterior electrodes (front to back) prior to a switch to the integrated percept, and higher information differentiation of anterior signals leading up to reporting the differentiated percept. Crucially, information integration was closely linked to perception and even observed in a no-report condition when perceptual transitions were inferred from eye-movements alone. In contrast, the link between neural differentiation and perception was observed solely in the active report condition. Our results therefore suggest that perception and report require distinct amounts of anterior-posterior network communication and anterior information differentiation: while front to back directed information is associated with changes in the content of perception, regardless of report, frontal information differentiation was absent in no-report conditions, and therefore has a different role not directly linked to perception *per se*.

INTRODUCTION

Consciousness is subjective experience, the 'what it is likeness' of our experience, for example, when we perceive a certain scene or endure pain. It can be argued that having such experiences is the main reason why life matters to us, and what may set us apart from non-living but otherwise smart "things", such as robots, your phone or the internet (1). But what makes us conscious? Conscious experience is a truly subjective and private phenomenon and it cannot be observed directly from the outside. To objectively study and understand it, we have to get access to the inner subjective experience of others, for instance via self-report tasks or via behavioral tasks that we believe can capture conscious content (2). In this way, when combined with neuroimaging tools, the neural correlates of consciousness (NCCs) are sought (3). However, while doing so, what we observe in measures of brain activity may not be as pure as we hoped for. Our measurements of "consciousness" may be cofounded with several cognitive factors, e.g., the act of reporting, attention, surprise or decision making, arising after conscious experience has emerged. This severely complicates our attempt to isolate the neural basis of conscious experience (4–9).

Perceptual ambiguity is a key phenomenon to study the brain mechanisms of conscious perception (10, 11), often experimentally elicited by using 'multistable stimuli' (12). Multistability can be induced in several ways or using several ambiguous stimuli, for instance using binocular rivalry, structure from motion, the Necker cube and motion induced blindness. The common feature of such paradigms is that an ambiguous stimulus can be interpreted in two, or multiple, ways, without changing the sensory input that reaches the senses. Confronted with this ambiguity, observers experience spontaneous fluctuations between interpretations of the stimulus. Single cell recordings in monkeys revealed that the strongest perceptual modulations of neuronal firing occur in higher association cortices, including inferotemporal cortex (ITC) and dorsolateral prefrontal cortex (DLPFC), and much less so in earlier sensory regions (13–16). However, previous human fMRI studies using report-based paradigms, in which observers indicate, in one way or another, when their perception switches, have associated both early and higher sensory regions with perceptual transitions (16). For example, when face and house stimuli compete for perceptual dominance, category specific regions in ITC activate stronger for the dominant percept, even before observers indicate a perceptual switch. BOLD related perceptual modulations are observed even lower in the cortical hierarchy, for example in the lateral geniculate nucleus (17). Another common observation in human fMRI studies is the association of a large network of parietal and frontal brain areas, traditionally associated with attentional and cognitive functions, during the report of perceptual transitions (for an overview see 12). Targeting specific nodes in this network using transcranial magnetic stimulation (TMS), e.g. inferior parietal cortex (IPS), or inferior frontal cortex (IFC), shapes the rate of perceptual transitions, suggesting their causal influence in resolving perceptual ambiguity (12, 18, 19). One central and unresolved issue in consciousness science in general (8), and multistable perception in particular, is what processes these large clusters of activations in frontoparietal cortex reflect. The *feedback* account states that frontal regions actively exert top-down influence on sensory brain regions to resolve perceptual ambiguity. The opposite *feedforward* account links frontal activity to processing of the consequences of perceiving, and thus processes occurring *after* perceptual ambiguity was resolved by posterior brain regions (19–21).

By studying conscious contents in the presence and absence of explicit report during visual bistable perception, we here investigate the neural mechanisms underlying perceptual transitions, while dissociating perception from report (or other unknown factors). The bistable stimulus we use consists of two overlapping gratings, known as ambiguous plaids or moving plaids (22, 23). This ambiguous stimulus can be perceived as *one* plaid moving coherently in a vertical direction, or *two* plaids sliding across one another horizontally. The plaid stimulus was thus perceived as perceptually integrated (one object) or perceptually differentiated (two objects), respectively. This stimulus allows us to track perception in the absence of report by capitalizing on the occurrence of optokinetic nystagmus (OKN) (24–27). OKN is a visually induced reflex, comprised of a combination of slow-phase and fast-phase eye movements that allow the eyes to follow objects in motio for instance when looking at the trees alongside the road whilst moving past them in a car. Importantly, OKN follows perception when viewing bistable visual stimuli, which makes it useful for assessing perception in the absence of reports (24, 25, 27).

This stimulus allows us to directly link perception to neural metrics inspired by information theory (28). In a recent report-based study on perceptual ambiguity it was shown that frontoparietal information integration, computed as the amount of information sharing between frontal and parietal EEG signals, increased when participants reported an ambiguous auditory stimulus as perceptually integrated compared to when it was reported as perceptually differentiated. On the contrary, information differentiation, computed as the amount of information diversity within frontal or parietal EEG signals separately, showed the opposite pattern within the same frontoparietal electrodes: it

increased when participants reported the bistable stimuli as perceptually differentiated compared to perceptually integrated. This suggests that information integration and information differentiation go hand in hand with observers' phenomenology of an integrated or differentiated ambiguous stimulus. However, as observers had to explicitly report the perceptual switches by pressing buttons, it remains an open question whether the observed neural information patterns are specifically related to changes in perception or to other task-related factors, or even post-perceptual decisional processes (4, 6). Further, because this study did not test for directionality of neural information flow, it cannot arbitrate between feedback and feedforward theories of perceptual ambiguity. Hence, dissociating content-related neural activity from the post-perceptual neural activity requires "no-report" paradigms, in which participants do not make and explicit decision to report their conscious experience (27, 29–31), as well as measures of directed information, to compare the feedforward and feedback contributions to perception in the absence of report.

RESULTS

Oculomotor signals

There were two experimental conditions that were performed in alternating runs. During the report runs (**Figure 1A**), observers pressed one button when the percept changed from vertical movement to horizontal movement and another button to indicate changes from horizontal movement to vertical movement. During the no-report runs (**Figure 1B**), observers were instructed to remain focused and just passively view the stimulus, and therefore perceptual transitions were rendered task-irrelevant. Participants were not informed about the relationship between perception and oculomotor signals and not about our goal to infer perception from their eye signals.

We first characterized changes in oculomotor signals in both directions of perceptual change (vertical/horizontal) in the report condition time-locked to the button response. We computed the slow phase of the OKN (see Methods) as it distinguishes between visual percepts during binocular rivalry (6, 27, 31, 32). OKN is particularly useful as positive and negative zero-crossings tend to precede the perceptual changes indicated by button presses. We observed OKN zero crossing going from positive to negative, before the visual stimulus was reported as integrated, and zero crossing going from negative to positive, when the stimulus was reported as differentiated (**Figure 1C**). In the left panel of **Figure 1C** (Report: BP) we depicted the raw OKN time series time-locked to the button press (-200 to 500 ms) in microvolts. Next, for both the report (Report: OKN; **Figure 1C**) and no-report (No report: OKN; **Figure 1C**) conditions, after computing the OKN crossings in the continuous oculomotor signal, we labeled the data with their corresponding button press labels, and created epochs locked to the OKN crossings, which as expected, revealed a clear moment for the perceptual transitions (**Figure 1C**). In the middle and right panels of **Figure 1C** we plotted the slow phase of the OKN signal locked to zero-crossing in arbitrary units, that was computed as described in the Methods. It is important to note the similarity of OKN data in both report and no-report conditions, demonstrating its potential to capture the contents of perception in both cases.

To quantify the predictive value of the OKN crossings for identifying perceptual contents, we performed a decoding analysis using a convolutional neural network (CNN). To measure generalization performance of the CNN, we split the OKN epochs and their corresponding button press labels into 3 parts: training, validation and test. 70% of the entire dataset was allocated to the training set. The remaining 30% were further split equally to create validation and test datasets, each containing 15% of the original data and labels (see Methods for details).

First, we trained a CNN to decode the button presses from the oculomotor signal using a cross-classification procedure in the report condition. We obtained a classification accuracy of 85% (**Figure 1D**). Next, in the same report condition, a second CNN was trained to decode the labels based on the OKN crossings, reaching a classification accuracy of 79%, indicating that OKN changes could reliably decode changes in perception. We finally estimated OKN crossings in the no-report data, labelled each crossing accordingly, and trained a third CNN to decode the labels, obtaining a classification accuracy of 80% (**Figure 1D**). Finally, another indication that our CNN accurately marks the occurrence of perceptual switches is that the number of switches in the report condition strongly correlates when based on button presses versus CNN performance (across subject Pearson's r: 0.88; p <0.001, see (24), for a similar analysis).

Histograms

We next characterized the distributions of perceptual switches in the report and no-report condition, locked to button presses and OKN crossings (**Figure 1E**). All distributions were approximated well with a gamma distribution (p <0.001), a common observation in report-based rivalry paradigms (33). That the no-report dominance durations are well approximated by a typical gamma function further indicates that the variability (and phenomenology) in perceptual switches during the passive viewing condition is similar to active report conditions (see also (24, 33).

To investigate whether perceptual ambiguity is resolved differently when the stimulus is reported as integrated or differentiated, we computed the time delay between OKN crossings and button presses in the report condition. We observed a longer delay between OKN crossing and the button press when the stimulus was reported as integrated versus differentiated ($t_{1,39}$ =2.23; p = 0.032; Cohen's d=0.352, **Figure 1F**). This increase in reaction time after a change in percept has occurred (as confirmed via OKN) suggests an increase in cognitive demand when reporting an integrated percept (transitioning from two objects to one) than when reporting a differentiated percept (going from one to two objects). We return to these features in *Time-frequency characteristics of perceptual switches.*

Overall, it is important to note the similarity of OKN data in both report and no-report conditions, demonstrating its ability to capture the content of perception in both cases. We observed the classic gamma functions for the distribution of percepts in each case. As a result, comparing our measures of integration and differentiation when time-locked to OKN crossings presents a powerful opportunity to assess the impact of report-based paradigms on the neural correlates of perception.

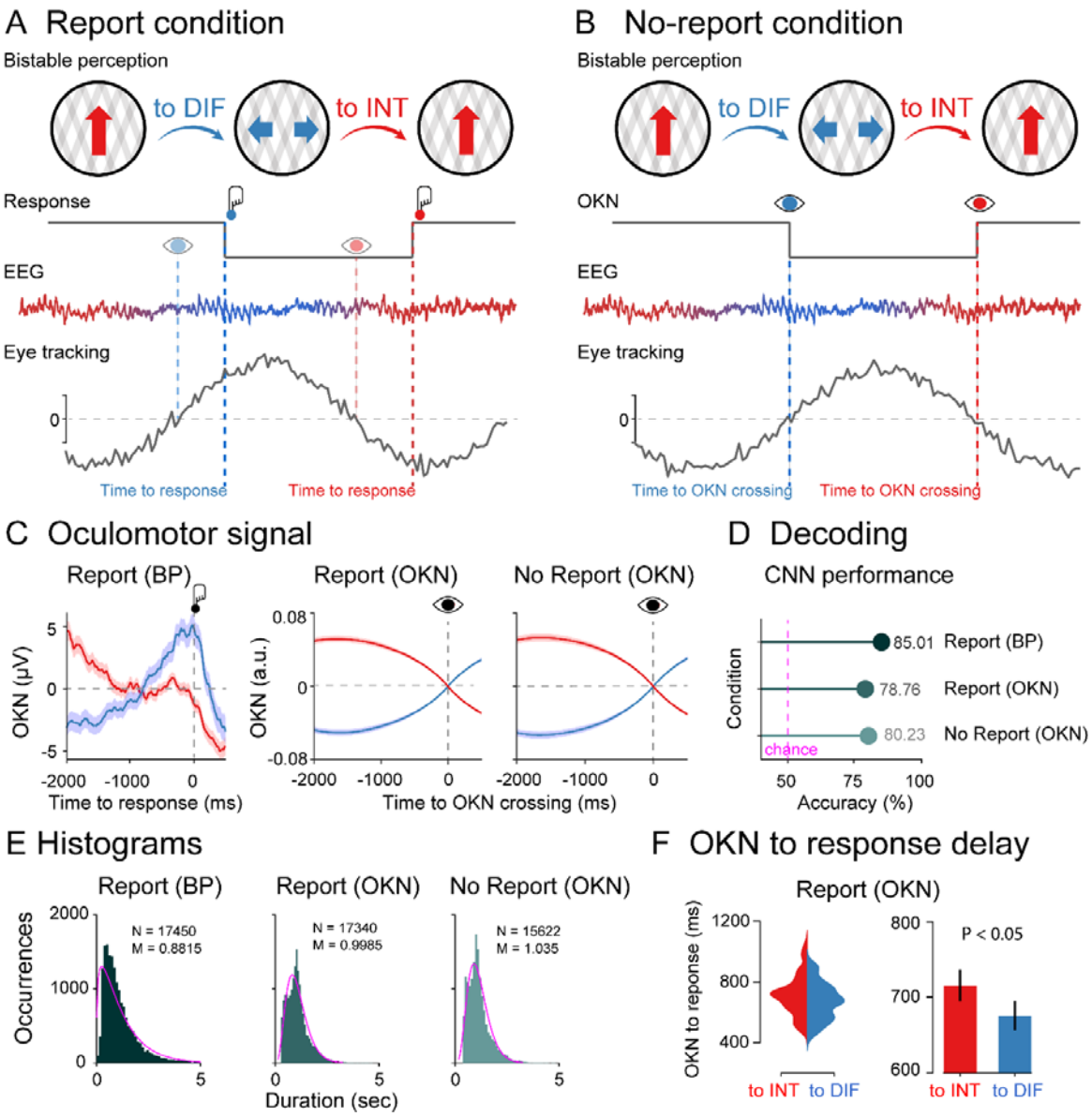


Figure 1. Experimental design, oculomotor decoding and behavior. *(A) Phenomenology during visual bistability in the report condition. Participants observed an ambiguous stimulus (moving plaids) that are experienced either as one plaid moving vertically (integrated percept; red arrow) or as two plaids moving horizontally (differentiated percept; blue arrows). Perceptual transitions occur either in integrated to differentiated direction (to DIF; blue) or in the differentiated to integrated direction (to INT; red). Middle row: Behavioral responses during the task. Participants pressed one button when perceiving that the integrated percept had fully changed into the differentiated percept (red button) and another button when perceiving that the differentiated percept had fully changed into the integrated percept (blue button). Bottom rows: Dynamical analyses for EEG and oculomotor (eye tracking) signals. From the oculomotor response, OKN zero-crossings were estimated to infer perceptual switches in the absence of explicit report. (B) In the no-report condition observers passively viewed the bistable stimulus (no button presses) and perceptual alternations were inferred from oculomotor signals (to INT: red eye; to DIFF: blue eye). (C) Raw oculomotor signals locked to the button press (Report BP) and locked to the OKN crossings (Report: OKN; No report: OKN) after computing the slow phase of the OKN. Decoding accuracies (D), and histograms (E) for perceptual switches in the report condition locked to the button response, in the report condition locked to the OKN crossings, and in the no-report condition locked to the OKN crossings. (F) Delay between OKN crossings and button responses in the report condition (left panel: single-subject distributions; right panel: group-level analysis).*

Report condition locked to button presses

We first investigated the neural dynamics of information integration when participants reported perceptual switches by pressing buttons. To this end, we computed a metric of information integration known as Directed Information (DI) which quantifies directional connectivity between neural signals (34, 35). Compared to traditional causality

detection methods based on linear models (e.g., Granger causality), DI is a model-free measure and can detect both linear and nonlinear functional relationships between brain signals. We took advantage of previous work that made this measure statistically robust when applied to neural data (34, 36–38). DI quantifies functional connectivity by measuring the degree to which the past of a "sender signal" *X* (e.g., EEG traces of anterior electrodes) predicts the future of another "receiver signal" *Y* (e.g., EEG traces of posterior electrodes), conditional on the past of the receiver signal Y. Thus, if there is significant DI between EEG signal *X* at one time, and EEG signal *Y* at a later time, this shows that signal *X* contains information about the future signal *Y*. Conditioning out the past of signal *Y* ensures the delayed interaction is providing new information over and above that available in the past of signal *Y*. For all DI analyses, we tested multiple delays from 0 ms to 500 ms (in steps of 4 ms) between sender and receiver signal, which allows us to investigate the characteristic time delay of directed information transfer during perceptual switches between anterior and posterior signals. For all analyses reported here we lock data to perceptual switches (either marked by a button press or eye-movement analysis) and inspect the EEG dynamics leading up to this perceptual switch. We have excluded all trials in which the previous perceptual switch occurred less than 2 seconds before the switch of interest (at time 0).

In the case of the report condition, **Figure 2A** shows DI between frontal and parietal signals, in both directions, so in the feedback direction (anterior to posterior electrodes) and in the feedforward direction (posterior to anterior electrodes, **Supplementary Figure 1A** for electrode location). We plot feedforward and feedback directed information as a function of signal delay between the two electrode sets, both when the moving plaids are reported as integrated (to INT; red color) as well as differentiated (to DIF; blue color). Signal delays are plotted on the y-axis, and testing time before button-press is plotted on the x-axis. A cluster-based permutation test performed on the difference between DI leading up to the integrated percept versus differentiated percept shows two significant clusters (cluster P<0.01). One cluster was observed indicating an increase in DI for the switch to integrated condition compared to the differentiated condition, at ~50 ms delay between the time-period 1050 to 900 ms prior to button press. A second cluster indicated an increase in DI prior to a switch to the integrated condition ~250 ms delay approximately 1200 to 500 ms before the button press–(**Figure 2**). Importantly, these significant clusters were observed in the feedback direction only. Note that there are no button presses in the time-window of interest, due to our trial exclusion procedure.

Next, we tested for an interaction between the direction of perceptual change (to integrated, to differentiated) and information direction (front to back, back to front). We performed a cluster-based permutation test on the difference between perceptual switch (to INT minus to DIFF trials) and information direction (feedback minus feedforward). We observed two significant clusters (cluster $p < 0.01$) showing stronger DI when perception switched to integrated as compared to differentiated in the front to back direction, but not in the back to front direction (**Figure 2C**). Finally, to test for the spatial specificity of the DI effect (i.e., the anterior-posterior direction), we computed DI in the right-left direction using temporal electrodes (**Supplementary Figure 1B**). No DI differences between perceptual switches were observed in the right to left direction, nor in the left to right direction (**Supplementary Figure 2**).

We next analyzed the dynamics of information differentiation (K-complexity) within frontal and parietal signals separately. Neural differentiation metrics quantify the diversity of information patterns within brain signals and it has been useful for distinguishing between conscious states (39–41) and conscious contents previously (28). We performed cluster-based permutation testing on the difference between the Information Differentiation (ID) time series leading up to the integrated percept versus the differentiated percept for the same set of anterior (front) and posterior (back) electrodes. As expected, significantly increased ID was observed before the bistable stimulus was reported as differentiated as compared to when was reported as integrated. This effect was observed approximately -1300 to 250 ms around the button press indicating the upcoming perceptual switch (**Figure 2E**, note that in the 2000 ms before the perceptual switch indicated by the response no previous switches are incorporated). Importantly, no such effect was observed for the posterior electrodes (**Figure 2D**).

We next tested for an interaction between the direction of perceptual change and information differentiation by subtracting between to DIF and to INT trials within frontal and back ROIs **(Figure 2E**). As predicted, a cluster-based permutation test revealed an interaction between perceptual change, perceptual direction and ROI based on information differentiation ($p < 0.01$), showing higher ID in the anterior region prior to a change to seeing the differentiated percept, without changes to ID in the posterior regions (**Figure 1E**). Finally, we computed ID in the

right and left ROI using the temporal electrodes (**Supplementary Figure 1B**). No ID differences between perceptual switches were observed in the right nor left ROI (**Supplementary Figure 3**).

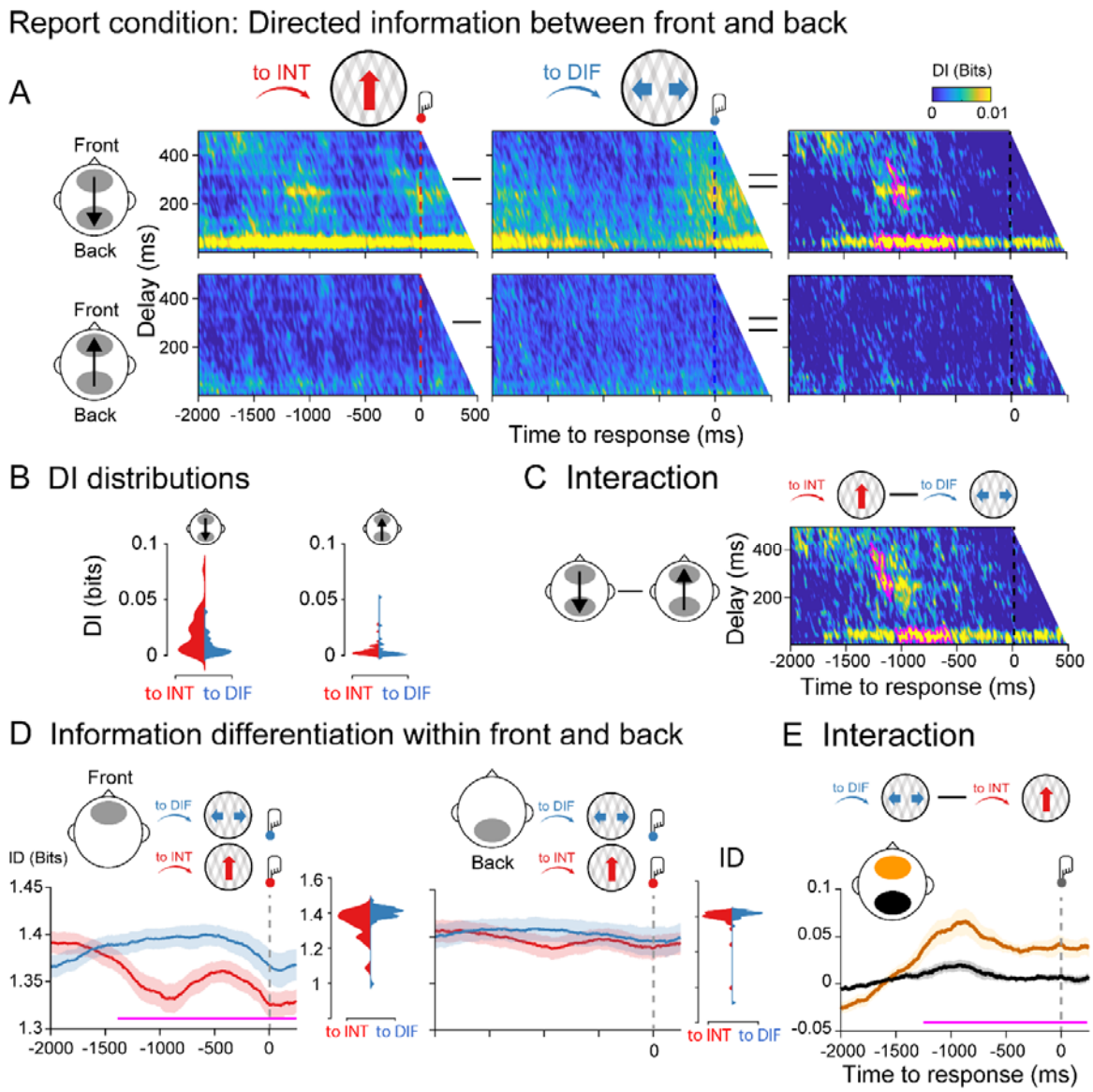


Figure 2. Information dynamics of the report condition locked to button responses. *(A) Group-level DI between front and back ROIs (upper row) and between back and front ROIs (lower row) when moving plaids are reported as integrated (to INT; red color), reported as differentiated (to DIF; blue color), and the cluster-based permutation tests between the two. (B) Distribution of single-participant DI values for the frontal to parietal direction (upper row) and for the parietal to frontal direction (lower row). Values were extracted based on the significant clusters obtained in the frontal to parietal contrast (purple blobs). (C) Group-level DI statistical interaction effect computed as the difference between to INT and to DIFF trials between front and back ROIs (i.e., a double-subtraction). (D) Group-level ID within the frontal ROI and the corresponding single-subject ID values (left panel) when moving plaids are reported as differentiated and reported as integrated, and the same for the parietal ROI (right panel) (E) Group-level ID statistical interaction effect computed as the difference between to DIF and to INT trials between front (yellow color) and back (black color) ROIs.*

Taken together, and consistent with previously reported effects in the auditory domain (28), these results show that having an integrated versus a differentiated percept, goes hand in hand with front to back neural directed information integration and frontal differentiation brain measures. Next, we aim to establish to what extent these effects depend on the necessity for reporting perceptual transitions (task-relevance of switches) and we further specify the timing of these effects by locking our analyses more closely to the perceptual transitions based on OKN crossings.

Report condition locked to OKN crossings

DI analyses were performed in the same way as in the report condition but this time locked to OKN crossings instead of button presses. We observed significant clusters of increased DI when the visual stimulus was perceived as integrated as compared to differentiated, again only in the feedback direction (**Figure 3A**). Significant clusters spanned delays between 50 to 450 ms and occurring roughly -1600 to 100 ms around the moment OKN crossed. A similar interaction analysis between switch direction and integration direction (feedback vs feedforward) was performed. The difference in directed information (DI) between integrated and differentiated percepts was stronger for the feedback direction than the feedforward direction (interaction cluster p <0.01; **Figure 3C**).

For information differentiation however, locked to the OKN crossings in the report condition, no robust clusters were observed (**Figure 3D**), neither for the front (left panel) nor the back (right panel) ROIs.

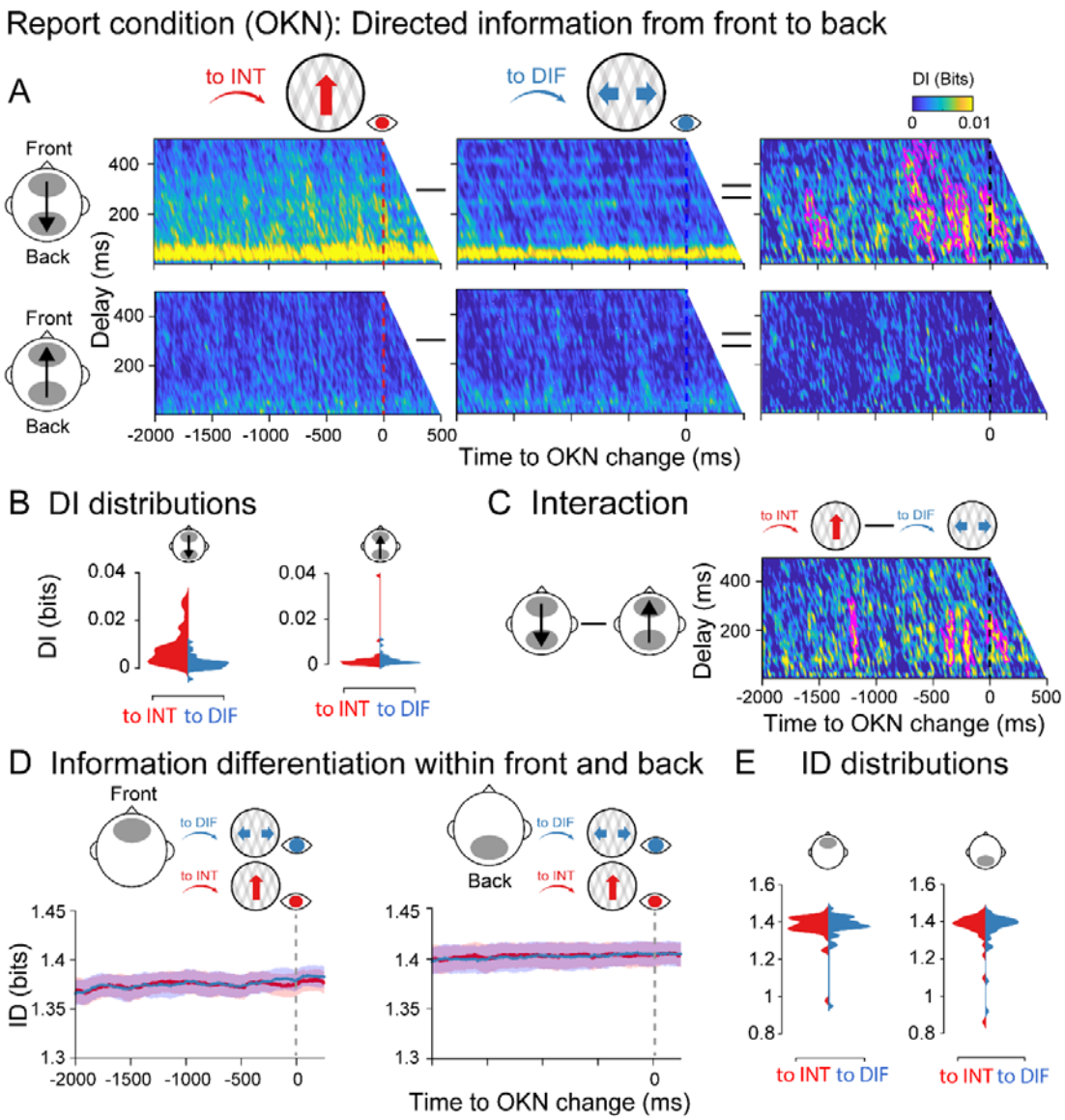


Figure 3. Information dynamics of the report condition locked to OKN crossings. *Specifics between A-D are identical to figure 2. (E) Single participant ID values were extracted using as a reference the significant time window observed in Figure 2D.*

No report condition

After establishing the dynamics of information integration and differentiation in the report condition, and finding that differentiation measures, but not integration measures, were dependent on report, we next analyzed both during the no-report condition (i.e., locked to the OKN crossings). Similar to the report condition, a cluster-based

permutation test revealed significant clusters of increased DI when the visual stimulus was perceived as integrated as compared to differentiated, but uniquely in the feedback direction (**Figure 4A**). The observed clusters showed a similar range of delays (around 50 to 400 ms) and they were observed in similar time windows (-600 to 0 ms around the OKN crossings) as in the report condition when time-locked to the OKN (**Figure 3**). We also observed significant clusters showing an interaction between perceptual switch and information direction (p <0.01), showing stronger integration when the percept switched to integrated as compared to differentiated, and more so in the feedback than feedforward direction (**Figure 4C**).

Finally, we analyzed the dynamics of information differentiation during the no-report condition. Again, no robust ID temporal clusters were observed when testing the entire time-window -2000 to 500 ms around the OKN crossing between perceptual switches neither in the front nor in the back ROIs (**Figure 4D**). Taken together, these results indicate that directed information distinguishes visual contents even in the absence of explicit report, while information differentiation in anterior locations is dependent upon report.

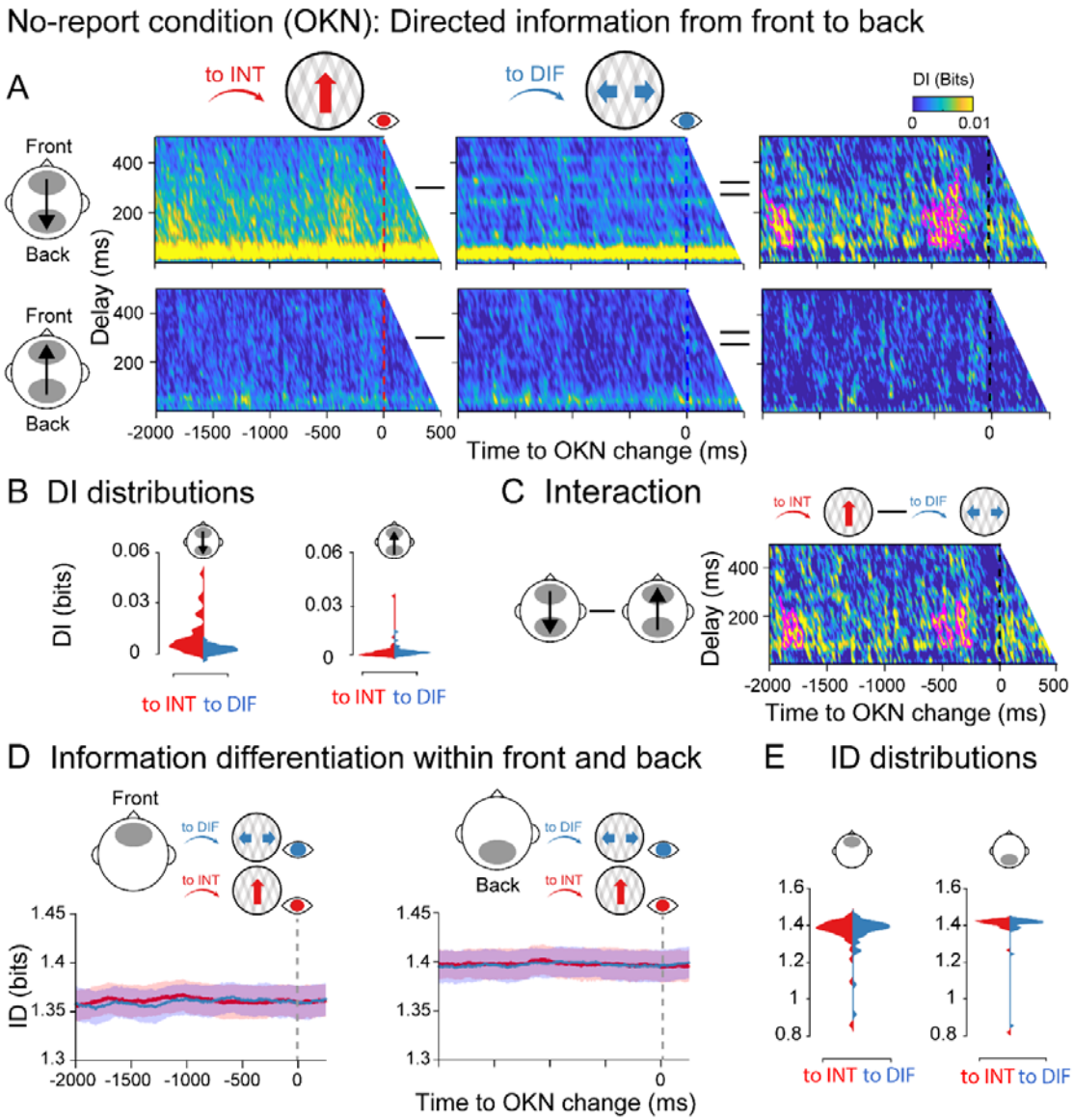


Figure 4. Information dynamics of the no-report condition locked to OKN crossings. *Specifics between A-D are identical to figure 2. (E) Single participant ID values were extracted using as a reference the significant time window observed in Figure 2D.*

Eye movements cannot explain the observed neural effects

We would like to note that the observed EEG effects are unlikely to be the result of differences in eye-movement signals between conditions for several reasons. First, measures of neural differentiation were specific to the report-locked analyses, although similar eye-movement patterns were present in all conditions. Second, the difference in directionality of the DI effects (presence of a DI effect in the feedback but not feedforward direction, no effects for temporal electrodes) indicates that general eye-movement confounds are also not likely. Third, OKN signals gradually change over time over a period of 2 seconds leading up to the perceptual switch, with highly similar time-courses for both types of percepts (see **Figure 1C**). The neural effects we observed do not reflect such a gradual build-up, and differ based on the perceived content which differs to what one would expect based on the pattern of the OKN signals. Fourth, it is also unlikely that eye-related signals (e.g. muscle activity) get somehow embedded in the EEG recordings and drive our anterior to posterior effects. The idea would be that if this signal is first measured on frontal and then on posterior electrodes, transfer entropy or source conduction would increase measures of directed information flow. This interpretation is unlikely given that we did not observe any differences in OKN signals when comparing to INT and to DIFF periods in the time window 1200 to 500 ms before the button press ,where we observed the differences in our neural measures in the report condition (**Figure 1C**; dependent-samples t-test (to INT, to DIF): $t_{1,39} = 0.093$; $P = 0.926$; $BF_{01} = 5.77$). Finally, when the eyes change their direction maximally, that is at the zero crossings where the derivative of the slope is largest (**Figure 1C**), or when the eyes most strongly indicate a certain perceptual state (at the peak of the OKN signal in **Figure 1C**), DI in the feedback direction was not maximal. If the eyes drive the neural effects, a peak in DI would be expected. Together, these considerations indicate that the results observed are unlikely to be caused by incidental differences in eye-movement patterns between conditions or perceptual states.

Time-frequency characteristics of perceptual switches

To compare the integration and differentiation results with more traditional electrophysiological features of multistable perception, we computed the time frequency profile of perceptual switches to the differentiated and integrated percepts in **Figure 5** (using cluster-based corrections for multiple comparisons). We observed increased oscillatory power in a broad frequency band from 1 to 18 Hz at frontal electrode sites, preceding a perceptual transition to an integrated compared to a differentiated percept. This was only observed in the button press locked analyses of the report condition, not in any of the other conditions. Interestingly, the same pattern of results was observed in the frontal complexity measure (ID) **(Figure 2D,E**). Complexity measures have been linked to changes in frequency dynamics (42). We observed that in the time-window that overlaps both measures (-1050 to the response), the decrease in complexity negatively correlated with time-frequency power (Pearson's $r = -0.39$; $p = 0.013$). This is to be expected, because an increase in power, reflecting a more repetitive and synchronized signal, should be inversely related to neural indices of complexity, which their increment reflects the lack of repetition (i.e., more information redundancy) within a signal (42).

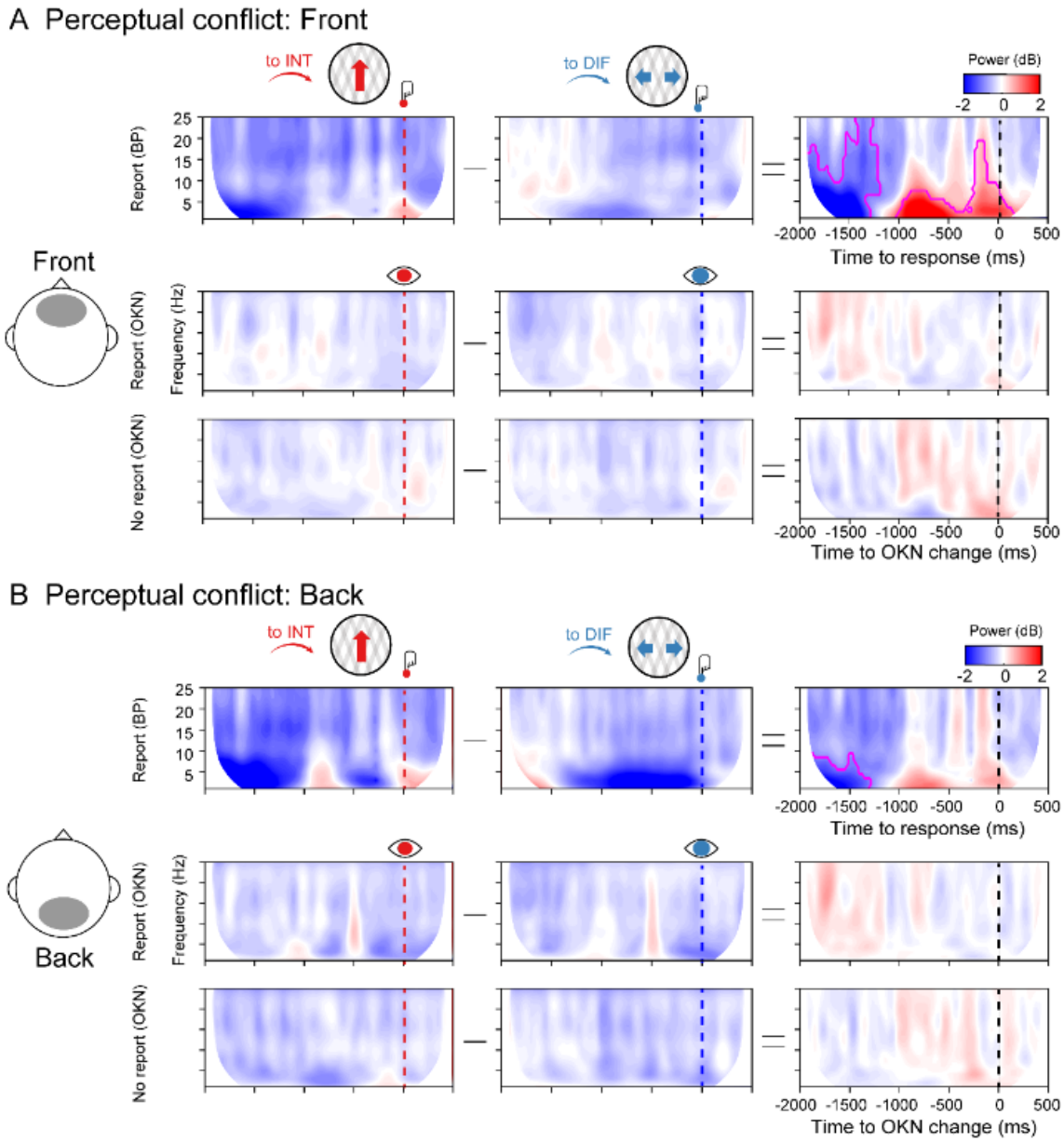


Figure 5. Spectral power in the report and no report conditions. *(A) Group-level spectral power in the front ROI for the report condition (upper row), report condition locked to OKN crossings (middle row), and no report condition (lower row), when moving plaids are reported as integrated (to INT; red color), reported as differentiated (to DIF; blue color), and the cluster-based permutation tests between the two. (B) Same as (A) but for the back ROI.*

DISCUSSION

Perceptual rivalry is the phenomenon that the perceptual interpretation of a bistable visual stimulus alternates over time in the absence of physical changes in the presented sensory input. The use of perceptual rivalry, as a tool, has provided fundamental insight in what neural processes may reflect the content of our conscious experience. However, although great progress has been made in unravelling neural processes underlying the competition between neural representations and the perceptual change between alternative percepts, it has proven notoriously difficult to separate neural correlates of consciousness, from processes that have another origin and are associated, for instance, with report, attention, surprise or commitment to a decision (6, 8). Here we aimed to tackle this issue using an experimental set-up that contained four crucial ingredients. First, we experimentally introduced conditions in which perceptual switches were task-relevant and had to be reported about, versus conditions in which perceptual transitions were irrelevant and bistable stimuli just had to be viewed passively (24, 27, 43). Second, we relied on neural measures with millisecond temporal resolution, allowing us to pinpoint the relevant neural

processes as they evolve over time. Third, we capitalized on novel information based measures that have shown to reflect the phenomenology of conscious perception (28, 44), allowing us to track specific perceptual content over time, while it naturally alternates. Fourth and finally, we introduced a novel way to analyze eye-tracking data to be able to pinpoint when precisely in time perceptual interpretations alternate, allowing us to time-lock our analysis to focus on processing leading up to the perceptual change (times prior to OKN crossing), from processing involved in translating perception into action, as well as other cognitive confounds arising after the perceptual switch (times after OKN, prior to button-press). This effort has led to several novel results and conclusions, which are summarized below.

We showed that perceptual switches can systematically be inferred from eye-movement measurements during passive viewing of a bistable visual stimulus by capitalizing on OKN measures in combination with deep neural network modelling. This confirms previous work that has shown that OKN and pupil measures can act as reliable indicators of the dynamics of perceptual and binocular rivalry (25, 27, 32). In fact, percept durations varied commonly during experimental blocks in which perceptual switches had to be reported, and were therefore task-relevant, and blocks in which observers passively viewed the bistable stimulus (perceptual switches are task-irrelevant). In both conditions, observer's percept durations followed a right-skewed gamma distribution as has been observed previously across a wide range of bistability paradigms (33), suggesting that the phenomenology of perception was similar during active report and passive viewing. Next, we showed that when perception of a bistable stimulus was explicitly reported, directed information from anterior to posterior signals was increased before the observers reported to perceive the integrated stimulus (one plaid moving upwards) compared to when the differentiated percept was reported. On the other hand, before the percept was reported as differentiated (two stimuli, each moving sideways in a different direction), we observed higher information differentiation of anterior signals, as compared to integrated percepts, before the perceptual switch indicated by the button press. This effect was not observed on posterior sensors. These results indicate that our information-based measures derived from electrophysiological activity capture and track the phenomenology of the percept during perceptual ambiguity. These results are in agreement with our previous report-based study on auditory bistable perception showing a correspondence between perceptual integration and differentiation and neural metrics of information integration and differentiation (28). During report, when DI is high an integrated percept is likely to be perceived, whereas when ID is high, a differentiated percept is more likely to be perceived.

When perceptual changes were inferred from eye-movements during passive viewing, we observed that our directed integration measure still tracked the perceptual state of the observer, but uniquely so in the feedback direction (from anterior to posterior signals). However, information differentiation measures no longer tracked the evolving percept, suggesting that our differentiation measures may reflect other processes not directly linked to perception, such as differences in task-set or for example differences in attention due to the necessity to report. In sum, the relative strength of directed information in the feedback direction indicated which percept likely dominated during passive viewing, whereas neural differentiation does not. We would like to note that this means that directed information is thus a "symmetrical measure" in the sense that when it is high it reflects that an integrated percept is likely to be dominant, whereas when it is low, a differentiated stimulus is likely dominating perception.

It is important to stress that we used information-specific EEG measures, which allowed us to track the contents of rivalrous states while quantifying the directionality of integration for our observers. This approach is similar to previous binocular rivalry studies using for instance alternating face/house stimuli and isolating (the competition between) neural activity in category-specific brain regions in ventral temporal cortex, either using electrophysiology (21) or fMRI (17, 45). The use of these information-based measures enabled us to track the neural signatures of the content of perception before perception alternated, instead of only focusing on transition-related neural changes, which has often been the case in previous fMRI studies, due to the low temporal resolution of the BOLD signal. A previous fMRI study has shown that the typical dorsolateral prefrontal cortex (PFC) activation pattern observed when perceptual switches are locked to button presses (27) disappear when perceptual switches are inferred from eye movements during passive viewing and therefore PFC involvement may potentially be related to report, or other consequences of perceptual switches. Similarly, Brascamp and colleagues have shown that prefrontal activation in fMRI may be driven by the fact that perceptual switches often draw attention (possibly because they are surprising and relevant) but are not related to perceptual transitions *per se*. They have done so by designing a task in which switching perception between the stimuli of each could not be noticed by observers, as if these switches were preconscious (potentially accessible but not actually accessed) (46). Interestingly, our information differentiation (ID)

results also suggest that frontal information differentiation may reflect post-perceptual processing rather than perception *per se*. In the report condition, we found that reporting the stimulus as integrated took longer than reporting it as differentiated, indicative of a task-dependent increase in the cognitive demand of disambiguating the stimulus identity. It may be that these differences can explain the frontal information differentiation effect observed during the report of perceptual switches, compared to the absence of this effect during the passive viewing condition. Knapen and colleagues (9) showed that large parts of the prefrontal network, again especially dorsolateral PFC, activates stronger to perceptual transitions with longer durations, so when the system takes longer to settle in a particular perceptual state (the transition phase between interpretations is longer) than when the transition phase is shorter. Finally, a recent study (19) has shown that transcranial stimulation of the right inferior frontal cortex (IFC) reduced the occurrence of perceptual changes for bistable stimuli. The authors argued that the IFC may register perceptual conflicts (or the mismatch/prediction error signal) between two possible perceptual interpretations, gradually building up towards the perceptual switch. Therefore, the IFC may be influencing the competition between pools of neurons coding for the different percepts in visual cortex, thereby "steering" conscious perception (20).

The experiment designed here can be regarded as a so-called "no-report paradigm" and not the stricter "no-cognition paradigm", because in principle when one eliminates any type of explicit report, this does not necessarily remove all possible post-perceptual cognitive processes that could happen throughout the experiment (4). Observers may still reason, think or reflect about the presented bistable stimulus and perceptual alternations. This issue is similar to previous attempts to minimalize cognitive processes during perceptual rivalry (24, 27, 43). In our experiment, the crucial aspect is that perceptual switches were task-irrelevant during the passive viewing condition and task-relevant during the report condition (30). We cannot and did not control for what observers were doing during passive viewing and therefore do not know whether and how observers were reasoning about the perceived stimulus, or the occurrence of perceptual switches. Therefore, we only intended to arbitrate between task/report-related and perception related information-based measures of perceptual transitions. Interestingly, using a similar set-up, combining OKN measures with a manipulation of task-relevance of perceptual switches, it has recently been shown that also in the pupil response, task/report and perception related dilations and constrictions can be separated (24).

Frontal theta-band power increase has often been related to the processing of conflict, both at the level of stimuli as well as for conflicting stimulus-response mappings (47–49). In the time-frequency response-locked analyses we observed increased theta-band power leading up to the integrated compared to differentiated percept, as well as longer reaction times between OKN crossings and report. As frontal theta band power increases were only observed in the response locked analyses, these time-frequency results suggest that the difference in power between integrated and differentiated percepts was contingent on the requirement to report.

Despite much debate in the field, influential theories of consciousness (3), such as Global Neuronal Workspace theory, Recurrent Processing theory, and Integrated Information Theory, agree that the common property of conscious experience relates to the brain's capacity to integrate information through recurrent processing (the combination of lateral and feedback interactions), enhancing cortico-cortical interactions at a local scale (between nearby regions) or global scale (between distant regions) (50–52). In agreement with these theories, our DI results support the interpretation that top-down information flow resolves the perceptual ambiguity of a bistable stimulus, both in the presence and absence of explicit report. Although most consciousness theories are mostly concerned with explaining which neural processes can be observed when sensory information "crosses the threshold" from subliminal (unconscious) to phenomenal or access consciousness, we here show that the amount of directed information flow (i.e., information integration) between anterior to posterior regions also reflects the likelihood of perceiving the integrated version of an ambiguous stimulus, and even in a context where report is not required. This finding may unite assumptions of Global Neuronal Workspace and Recurrent Processing theory (feedback is important for consciousness) with assumptions from Integrated Information Theory (information integration is important). Future studies however are needed to provide a more mechanistic explanation of the neural processes DI reflects.

In conclusion, our results suggest that the relevant dynamical mechanism for perceiving different contents during visual bistability, controlling for many post-perceptual factors, is the directed information between frontal and posterior signals, rather than the isolated information differentiation contained within the front or the back of the brain.

METHODS

Participants

Forty-two participants (31 females, 2 left-handed) aged between 18 and 35 (M = 20 years, SD = 3.73), recruited from the University of Amsterdam (Amsterdam, the Netherlands) participated in this study for monetary compensation. All participants had normal or corrected-to-normal vision. This study was approved by the ethics committee of the Psychology Department of the University of Amsterdam, and written informed consent was obtained from all participants after the explanation of the experimental protocol. One participant was excluded during data collection for not following the task instructions. Another participant was excluded during the analyses because of very strong artefacts in the EEG data.

Stimulus

The stimulus consisted of two overlapping semi-transparent half-wave gratings, known as ambiguous plaids or moving plaids (22, 23, 53). Each grating was a sinusoid (0.33 cycles per degree of visual angle), clipped to include only positive contrast luminance, positioned at the centre of an aperture (diameter = 36.03). Peak contrast was set to 0.025 relative to the uniform grey background. Encoding gamma was set to 2. The orientation of each grating was 15° and -15°. Motion was created by changing the phase of the grating on each frame (6 Hz motion).

We selected this specific stimulus for three reasons. Firstly, the stimulus is ambiguous, meaning that perception alternates whilst the sensory input remains constant. This approach is considered to be one of the most powerful methodologies to study the neural underpinnings of phenomenal consciousness (6). Secondly, the stimulus can be perceived as one grating moving coherently or two gratings sliding across one another. In other words, gratings are either perceptually integrated or perceptually differentiated, respectively. Note that the ambiguous plaids stimulus is considered to be tristable (rather than bistable) because it has one integrated percept (the gratings moving together as a single pattern) and two differentiated percepts (the gratings sliding across one another) with alternating depth order (which grating is perceived as foreground and which as background) (53). However, we treat the stimulus as if it were bistable, because we instruct participants to respond to changes from the integrated percept to the differentiated percept and vice versa. The last reason for using this stimulus is that it allows us to track perception in the absence of responses by exploiting the occurrence of optokinetic nystagmus (OKN) (25–27).

Experimental design and procedure

There were two experimental conditions (i.e. report and no-report) that were performed in alternating runs of four minutes each (Figure 1a). Participants performed twenty runs in total, split up in five blocks consisting of two runs of each condition (Figure 1a). The order within the blocks was always the same and started with a no-report run (i.e. no-report, report, no-report, report). Note that the same stimulus was used for both conditions and that the stimulus was constant throughout each run. Only the direction of motion of the stimulus (i.e. upwards or downwards) was changed every two runs in order to exclude the direction of the stimulus as a confounding factor (Figure 1A). During the report runs, participants were instructed to press one button when the percept changed from vertical movement to horizontal movement and another button to indicate changes from horizontal movement to vertical movement (Figure 1B and C) (i.e. 2AFC). The buttons were operated with the index fingers of both hands and the contingencies of the buttons were counterbalanced across participants. During the no-report runs, participants were instructed to remain focused on the stimulus at all times and to relax their hands on the desk in front of them (i.e. passive viewing of the stimulus).

The experiment was programmed and executed using the Psychophysics Toolbox (version 3.0.14; Brainard, 1997) and Eyelink Toolbox extensions (Cornelissen, Peters, & Palmer, 2002) for MATLAB (R2016a, MathWorks, Inc., Natick, MA, US). Stimuli were presented on an Asus VG236H LCD monitor (23” diagonal, 1920 x 1080 pixel resolution; 120 Hz refresh rate) at a viewing distance of 50 cm.

Electroencephalography (EEG) recording and preprocessing

EEG signals were acquired through a 64-channel Biosemi ActiveTwo system (Biosemi, Amsterdam, the Netherlands) with two online references (CMS and DRL) placed according to the international 10-20 system. Two external electrodes were placed on the earlobes for possible use as offline references. Four additional external electrodes were used to record vertical and horizontal eye-movements, adding up to 72 channels in total. Data were sampled at 512 Hz. Preprocessing was done by means of custom-made MATLAB (R2016a, The MathWorks, Inc.) scripts

supported by EEGLAB (54). Continuous EEG data were first down-sampled to 250 Hz and filtered between 1-100 Hz. Data from the 64 channels over the scalp surface (i.e. reference electrodes and external electrodes excluded) were retained for further analyses. All twenty runs were extracted and subsequently appended in order to eliminate the time periods between runs in which no stimulus was present. Channels with a variance smaller than -2 or larger than 2 standard deviations (SD) of the mean activity of all channels were rejected and a notch filter of 50 Hz was applied to remove the line noise.

Independent Component Analysis (ICA) (54) was performed over all remaining channels. Independent components representing eye blinks, eye movements, muscle artefacts, and other types of noise were removed from the EEG signal after which the signals from the previously rejected channels were replaced with the weighted average activity in all remaining channels by spherical spline interpolation. Subsequently, the data were separated into separate datasets containing the report and no-report runs for each participant. The report dataset was segmented into epochs from -2000 to 500 ms around responses. Epochs with response repetitions and epochs that contained more than one response were rejected (22.30% of epochs on average). Finally, epochs were rejected if they exceeded certain thresholds for amplitude (< -150 uV or > 150 uV) or slope (> 60 uV/epoch) and the data were referenced to the average activity in all channels.

Eye-tracking recording

Eye movements were recorded at 500 Hz with an EyeLink 1000 (SR Research) infrared eye tracker, calibrated using a 6-point calibration procedure at the start of every block (i.e. five times throughout data collection). During the blocks, participants' heads were positioned on a chin rest in order to minimize head movements.

Optokinetic nystagmus (OKN) signal analysis

OKN is a combination of slow-phase and fast-phase eye movements that allows the eyes to follow objects in motion when the head remains stationary, for instance when looking at the trees alongside the road whilst moving past them in a car. The slow-phase movements try to match the stimulus speed to keep the retinal image stable and are interrupted by fast eye movements that reset the eye in orbit. Thus, the velocity of slow-phase OKN provides a continuous and robust estimate of conscious perception rather than the actual visual input, which makes it useful for assessing perception in the absence of reports (25, 27, 32, 55–57)

From the eye movement data, we obtained mean slow-phase OKN velocity and classification accuracy of percepts based on this signal. We performed the following pre-processing steps to obtain the velocity of slow-phase OKN.

First, periods of blinks and saccades were detected using the manufacturer's standard algorithms with default settings. The subsequent data analyses were performed using custom-made Python software. The following steps were applied to each pupil recording: (i) linear interpolation of values measured just before and after each identified blink (interpolation time window, from 150 ms before until 150 ms after blink), (ii) temporal filtering (third-order Butterworth, low-pass: 10 Hz), (iii) removal of pupil responses to blinks and to saccades, by first estimating these responses by means of deconvolution, and then removing them from the pupil time series by means of multiple linear regression (58), and (iv) conversion to units of modulation (percent signal change) around the mean of the pupil time series from each block.

Second, we smoothed the integrated OKN with a 100 ms Gaussian kernel. We then computed instantaneous velocity of integrated OKN as the difference between neighboring two time points (2 ms difference). To obtain a velocity of the slow-phase OKN, we further smoothed the instantaneous velocity with the 100 ms Gaussian kernel. Finally, we segmented the time course of the velocity of slow-phase OKN from 1 s before to 2 s after the onset of stimuli. We did not include the first trial of each block in the analysis as we did not record fixation position before the first trial.

Convolutional Neural Network (CNN) classification on button press and OKN

First, to quantify discriminability of perceptual report in OKN and button press in the report condition, we employed a convolutional neural network (CNN) with the instantaneous velocity of slow-phase OKN as a feature (Figure 1C and 1D).

The CNN consisted of the following layers in sequence:

A 1D input layer of dimension 1 X 625 units, matching the shape of an epoch of smoothed OKN sampled at 500 Hz.

A convolutional layer consisting of 8 convolutional filters, each with the shape 1 x 25. Units in this layer used ReLU activation functions. The stride step size of the convolution over the inputs was set to 1.

A max pooling layer with the shape 1 x 5, and a stride step size of 2.

A second convolutional layer with 16 convolutional filters, each with the shape 2 x 50. Units used ReLU activation functions and the stride step size of the convolution over the inputs was set to 1.

Another max pooling layer with the shape 1 x 5 and stride step size of 2.

A third (and final) convolutional layer with 32 convolutional filters, each with the shape 2 x 75. Units used ReLU activation functions and the stride step size of the convolution over the inputs was set to 1.

A fully connected dense layer of shape 1 x 2, consisting of units with softmax activation functions.

To prevent overfitting and measure generalisation performance of the CNN, we first split the OKN data epochs and their corresponding button press labels into 3 parts: training, validation and test. 70% of the entire dataset consisting of 13,406 epochs was allocated to the training set. The remaining 30% were further split equally to create validation and test datasets, each containing 15% of the original data and labels. We used stratified sampling when creating this splits, to ensure that the relative proportion of samples of each class was the same in each split.

The CNN was then trained using the training set, with a mini-batch size of 128, over 30 epochs. We randomly shuffled the training dataset at the beginning of each epoch and evaluated the CNN's performance on the validation dataset once every 10 mini-batches. The stochastic gradient descent algorithm with a momentum of 0.9 was used to learn the weights that minimised the cross-entropy loss over the training dataset. At the end of the training procedure, we recorded the accuracy with which the trained CNN classified OKN epochs in the held-out test dataset.

The above procedure – including CNN training, validation, and testing— was repeated, after defining epoch labels based on the OKN crossings in the report condition. Finally, we re-trained and evaluated the CNN by defining epoch labels based on the OKN crossings in the no-report condition (**Figure 1**).

Directed Information (DI)

In order to quantify the directed functional connectivity between different EEG signals, we used Directed Information (DI), also known as Transfer Entropy, an information theoretic measure of Wiener-Granger causality (59–62). Compared to traditional causality detection methods based on linear models (e.g. Granger causality), DI is a model-free measure and can detect both linear and nonlinear functional relationships between brain signals. We took advantage of previous work that made this measure statistically robust when applied to neural data (34, 36–38).

DI quantifies functional connectivity by measuring the degree to which the past of a signal *X* predicts the future of another signal *Y*, conditional on the past of *Y*, defined at a specific lag or delay τ:

$$DI = I\,(Y_t; X_{t-\tau} \mid Y_{t-\tau})$$

Thus, if there is significant DI between EEG signal *X* at one time, and EEG signal *Y* at a later time, this shows that signal *X* contains information about the future signal *Y*. Conditioning out the past of signal *Y* ensures the delayed interaction is providing new information over and above that available in the past of signal *X*. For all DI analysis, we tested delays from 0 ms to 500 ms in steps of 4 ms.

Information Differentiation (ID)

We computed Kolmogorov-Chaitin complexity (K complexity) as a metric of information differentiation (63, 64). K complexity quantifies the algorithmic complexity (i.e. the diversity of information patterns) of an EEG signal by measuring its degree of redundancy: from a highly redundant signal (i.e. less differentiated signal) to a slightly redundant one (i.e. highly differentiated signal) (28, 39–41). Algorithmic complexity of a given EEG sequence can be described as the length of the shortest computer program that can generate it. A short program corresponds to a less complex sequence. K complexity was estimated by quantifying the compression size of the EEG using the Lempel-Ziv zip algorithm (65).

Algorithmic information theory has been introduced by Andreï Kolmogorov and Gregory Chaitin as an area of interaction between computer science and information theory. The concept of algorithmic complexity or Kolmogorov-Chaitin complexity (K complexity) is defined as the shortest description of a string (or in our case a signal *X*). That is to say, K complexity is the size of the smallest algorithm (or computer program) that can produce that particular time series. However, it can be demonstrated by *reductio ad absurdum* that there is no possible algorithm that can measure K complexity (66). To sidestep this issue, we can estimate an upper-bound value of K complexity(*X*). This can be concretely accomplished by applying a lossless compression of the time series and quantifying the compression size. Capitalizing on the vast signal compression literature, we heuristically used a classical open-source compressor gzip (67) to estimate K complexity(*X*). It is important to standardize the method of representation of the signal before compression in order to avoid irrelevant differences in complexity. Specifically, to compute K complexity(*X*):

First, the signals were transformed into sequences of symbols. Each symbol represents, with identical complexity, the amplitude of the corresponding channel for each time point. The number of symbols was set to 32 and each one corresponds to dividing the amplitude range of that given channel into 32 equivalent bins. Similar results have been obtained with binning ranging from 8 to 128 bins (39).

Next, time series were compressed using the compressLib library for Matlab, this library implements the gzip algorithm to compress Matlab variables.

Finally, K complexity(*X*) was calculated as the size of the compressed variable with time series divided by the size of the original variable before compression. Our premise is that the bigger the size of the compressed string, the more complex the structure of the time series, thus indexing the complexity of the electrical activity recorded at an electrode.

For each trial and ROI, K complexity was estimated using a 100 ms sliding window with 4 ms time step.

EEG time-frequency analysis

Epochs were grouped based on conditions: report, report (OKN), and no report (OKN). Then, EEG-traces were decomposed into time-frequency charts from 2 Hz to 26 Hz in 13 linearly spaced steps (2 Hz per bin). The power spectrum of the EEG-signal (as obtained by the fast Fourier transform) was multiplied by the power spectra of complex Morlet wavelets ($e^{i2\pi tf}e^{-t^2/(2\sigma^2)}$) with logarithmically spaced cycle sizes ranging from 3 to 12. The inverse Fourier transform was then used to acquire the complex signal, which was converted to frequency-band specific power by squaring the result of the convolution of the complex and real parts of the signal ($real[z(t)]^2 + imag[z(t)]^2$). The resulting time-frequency data were then averaged per subject and trial type. Finally, time-frequency traces were transformed to decibels (dB) and normalized to the length of the entire trial (baseline: -2000ms to 500 ms), according to: $dB = 10 * log10\frac{power}{baseline}$.

EEG electrode selection (ROI)

Canonical bilateral frontal (n = 6), parietal (n = 6), as well as right temporal (n = 6) and left temporal (n = 6) electrode clusters were selected for DI and ID analyses (Supplementary Figure 1). Values within each region of interest (ROI) were averaged before computing DI and ID metrics, and subsequently averaged per condition and participant.

Statistical analysis

To test for significant time-delay DI (Figure 2 and 3), and time-frequency spectral power (Figure 5) a cluster-based nonparametric statistical test implemented in FieldTrip (68) was used. In brief, time-delay DI charts (-2000 to 500 ms) were compared in pairs of experimental conditions (perceptual switches: to integrated vs. to differentiated). For each such pairwise comparison, epochs in each condition were averaged subject-wise. These averages were passed to the analysis procedure of FieldTrip, the details of which are described elsewhere (68). In short, this procedure compared corresponding temporal points in the subject-wise averages using dependent samples t-tests for within-subject comparisons. Although this step was parametric, FieldTrip uses a nonparametric clustering method to address the multiple comparisons problem. t values of adjacent temporal points whose P values were lower than 0.01 were clustered together by summating their t values, and the largest such cluster was retained. This whole procedure, i.e., calculation of t values at each temporal point followed by clustering of adjacent t values, was then repeated 1000 times, with recombination and randomized resampling of the subject-wise averages before each

repetition. This Monte Carlo method generated a nonparametric estimate of the p-value representing the statistical significance of the originally identified cluster. The cluster-level t value was calculated as the sum of the individual t values at the points within the cluster.

In the case of DI (Figure 2 and 3), in order to test a possible interaction effect between information direction and perceptual switch, a separate cluster-based permutation test was calculated as the difference between perceptual switch (to integrated, to differentiated) and direction of change (front-to-back, back-to-front).

Similarly, for ID analysis, we performed a cluster-based permutation test across the ID time series (-2000 to 500 ms) between perceptual switches for the front and back ROIs separately. Next, in order to test for an interaction between perceptual switch and direction of change, we performed another cluster-based permutation test on the difference between perceptual switch (to integrated, to differentiated) and ROI (front, back).

In case of null findings, we performed a Bayesian RANOVA with identical parameters and settings on the same data, to test if there was actual support of the null hypothesis. When reported, BF_{01} refers to the Bayes Factor in favor of the null hypothesis. Statistical analyses were performed using MATLAB (2019a), Jamovi (Version 0.8.1.6) [Computer Software] (Retrieved from https://www.jamovi.org) (open source), and JASP Team (2018; JASP; version 0.8.4 software) statistical software.

Funding

This research was supported by an ABC Talent Grant of the University of Amsterdam (ACJ, SVG) and a grant from the H2020 European Research Council (ERC STG 715605, SVG).

Author's contributions

Conceived and designed the experiments: ACJ, SVG. Performed the experiments: ACJ, LB. Analyzed the data: ACJ, MD. Contributed reagents/materials/analysis tools: RI, SC. Wrote the paper: ACJ, LB, SC, RI, MD, SVG.

Acknowledgments

We thank Dr. William J. Harrison, Dr. Maartje de Jong, Dr. William Gilpin and Dr. Tomas Knapen for contributing to valuable discussion and insights and proofreading the manuscript. This manuscript is dedicated to the memory of Prof. Walter J. Freeman (1927 - 2016) whose pioneering work on Neurodynamics has inspired and ignited countless meaningful insights during the execution of this project.

Conflict of Interest

None declared.

REFERENCES

1. S. Dehaene, H. Lau, S. Kouider, What is consciousness, and could machines have it? *Science (80-.).* **358**, 486–492 (2017).

2. J. C. Francken, *et al.*, An academic survey on theoretical foundations, common assumptions and the current state of the field of consciousness science. PsyArXiv: doi:10.31234/osf.io/8mbsk.

3. F. Crick, C. Koch, Towards a neurobiological theory of consciousness. *Semin. Neurosci.* **2**, 263–275 (1990).

4. N. Block, What Is Wrong with the No-Report Paradigm and How to Fix It. *Trends Cogn. Sci.* **23**, 1003–1013 (2019).

5. J. Aru, T. Bachmann, W. Singer, L. Melloni, Distilling the neural correlates of consciousness. *Neurosci. Biobehav. Rev.* **36**, 737–746 (2012).

6. N. Tsuchiya, M. Wilke, S. Frässle, V. A. F. Lamme, No-Report Paradigms: Extracting the True Neural Correlates of Consciousness. *Trends Cogn. Sci.* **19**, 757–770 (2015).

7. J. F. Storm, *et al.*, Consciousness regained: Disentangling mechanisms, brain systems, and behavioral responses. *J. Neurosci.* **37**, 10882–10893 (2017).

8. C. Koch, M. Massimini, M. Boly, G. Tononi, Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* **17**, 307–321 (2016).

9. T. Knapen, J. Brascamp, J. Pearson, R. van Ee, R. Blake, The role of frontal and parietal brain areas in bistable perception. *J. Neurosci.* (2011) https:/doi.org/10.1523/JNEUROSCI.1727-11.2011.

10. R. Blake, A primer on binocular rivalry, including current controversies. *Brain Mind* (2001) https:/doi.org/10.1023/A:1017925416289.

11. P. Sterzer, A. Kleinschmidt, G. Rees, The neural bases of multistable perception. *Trends Cogn. Sci.* **13**, 310–318 (2009).

12. J. Brascamp, P. Sterzer, R. Blake, T. Knapen, Multistable Perception and the Role of the Frontoparietal Cortex in Perceptual Inference. *Annu. Rev. Psychol.* **69**, 77–103 (2018).

13. D. A. Leopold, N. K. Logothetis, Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature* (1996) https:/doi.org/10.1038/379549a0.

14. D. L. Sheinberg, N. K. Logothetis, The role of temporal cortical areas in perceptual organization. *Proc. Natl. Acad. Sci. U. S. A.* (1997) https:/doi.org/10.1073/pnas.94.7.3408.

15. T. I. Panagiotaropoulos, G. Deco, V. Kapoor, N. K. Logothetis, Neuronal Discharges and Gamma Oscillations Explicitly Reflect Visual Consciousness in the Lateral Prefrontal Cortex. *Neuron* **74**, 924–935 (2012).

16. F. Tong, Primary visual cortex and visual awareness. *Nat. Rev. Neurosci.* **4**, 219–229 (2003).

17. F. Tong, M. Meng, R. Blake, Neural bases of binocular rivalry. *Trends Cogn. Sci.* **10**, 502–511 (2006).

18. D. Carmel, V. Walsh, N. Lavie, G. Rees, Right parietal TMS shortens dominance durations in binocular rivalry. *Curr. Biol.* (2010) https:/doi.org/10.1016/j.cub.2010.07.036.

19. V. Weilnhammer, *et al.*, An active role of inferior frontal cortex in conscious experience. *Curr. Biol.* **31**, 2868-2880.e8

20. A. Bartels, Consciousness: What is the role of prefrontal cortex? *Curr. Biol.* (2021) https:/doi.org/10.1016/j.cub.2021.05.012.

21. M. C. de Jong, *et al.*, Intracranial Recordings Reveal Unique Shape and Timing of Responses in Human Visual Cortex during Illusory Visual Events. *Curr. Biol.* **30**, 3089-3100.e4 (2020).

22. J.-M. Hupé, C. Lamirel, J. Lorenceau, Pupil dynamics during bistable motion perception. *J. Vis.* **9**, 10 (2009).

23. J. M. Hupé, N. Rubin, The dynamics of bi-stable alternation in ambiguous motion displays: A fresh look at plaids. *Vision Res.* (2003) https:/doi.org/10.1016/S0042-6989(02)00593-X.

24. J. W. Brascamp, G. de Hollander, M. D. Wertheimer, A. N. DePew, T. Knapen, Separable pupillary signatures of perception and action during perceptual multistability. *Elife* **10**, e66161 (2021).

25. M. Naber, S. Frassle, W. Einhäuser, Perceptual Rivalry: Reflexes Reveal the Gradual Nature of Visual Awareness. *PLoS One* **6** (2011).

26. B. Cohen, V. Matsuo, T. Raphan, Quantitative analysis of the velocity characteristics of optokinetic nystagmus and optokinetic after-nystagmus. *J. Physiol.* (1977) https:/doi.org/10.1113/jphysiol.1977.sp011955.

27. S. Frassle, J. Sommer, a. Jansen, M. Naber, W. Einhauser, Binocular Rivalry: Frontal Activity Relates to Introspection and Action But Not to Perception. *J. Neurosci.* **34**, 1738–1747 (2014).

28. A. Canales-Johnson, *et al.*, Dissociable Neural Information Dynamics of Perceptual Integration and Differentiation during Bistable Perception. *Cereb. Cortex* **30**, 4563–4580 (2020).

29. I. Schlossmacher, T. Dellert, M. Pitts, M. Bruchmann, T. Straube, Differential effects of awareness and task relevance on early and late ERPs in a no-report visual oddball paradigm. *J. Neurosci.* **40**, 2906–2913 (2020).

30. M. A. Cohen, K. Ortego, A. Kyroudis, M. Pitts, Distinguishing the Neural Correlates of Perceptual Awareness and Postperceptual Processing. *J. Neurosci.* (2020) https:/doi.org/10.1523/JNEUROSCI.0120-20.2020.

31. V. Kapoor, *et al.*, Decoding the contents of consciousness from prefrontal ensembles. *bioRxiv* (2020) https:/doi.org/10.1101/2020.01.28.921841.

32. M. Fujiwara, *et al.*, Optokinetic nystagmus reflects perceptual directions in the onset binocular rivalry in Parkinson's disease. *PLoS One* (2017) https:/doi.org/10.1371/journal.pone.0173707.

33. J. W. Brascamp, P. C. Klink, W. J. M. Levelt, The "laws" of binocular rivalry: 50 years of Levelt's propositions. *Vision Res.* **109**, 20–37 (2015).

34. R. A. A. Ince, *et al.*, A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula. *Hum. Brain Mapp.* (2017) https:/doi.org/10.1002/hbm.23471.

35. R. A. A. Ince, S. Panzeri, S. R. Schultz, "Summary of Information Theoretic Quantities" in *Encyclopedia of Computational Neuroscience*, (2014) https:/doi.org/10.1007/978-1-4614-7320-6_306-1.

36. M. Besserve, S. C. Lowe, N. K. Logothetis, B. Schölkopf, S. Panzeri, Shifts of Gamma Phase across Primary Visual Cortical Sites Reflect Dynamic Stimulus-Modulated Information Transfer. *PLoS Biol.* (2015) https:/doi.org/10.1371/journal.pbio.1002257.

37. B. L. Giordano, *et al.*, Contributions of local speech encoding and functional connectivity to audio-visual speech perception. *Elife* (2017) https:/doi.org/10.7554/eLife.24763.

38. H. Park, R. A. A. Ince, P. G. Schyns, G. Thut, J. Gross, Representational interactions during audiovisual speech entrainment: Redundancy in left posterior superior temporal gyrus and synergy in left motor cortex. *PLoS Biol.* (2018) https:/doi.org/10.1371/journal.pbio.2006558.

39. J. D. Sitt, *et al.*, Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. *Brain* **137**, 2258–2270 (2014).

40. M. M. Schartner, R. L. Carhart-Harris, A. B. Barrett, A. K. Seth, S. D. Muthukumaraswamy, Increased spontaneous MEG signal diversity for psychoactive doses of ketamine, LSD and psilocybin. *Sci. Rep.* **7**, 1–12 (2017).

41. M. Schartner, *et al.*, Complexity of multi-dimensional spontaneous EEG decreases during propofol induced general anaesthesia. *PLoS One* **10** (2015).

42. P. A. M. Mediano, F. E. Rosas, A. B. Barrett, D. Bor, Decomposing Spectral and Phasic Differences in Nonlinear Features between Datasets. *Phys. Rev. Lett.* **127**, 124101 (2021).

43. E. D. Lumer, G. Rees, Covariation of activity in visual and prefrontal cortex associated with subjective visual perception. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 1669–1673 (1999).

44. R. A. A. Ince, *et al.*, A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula. *Hum. Brain Mapp.* **38**, 1541–1573 (2017).

45. F. Tong, K. Nakayama, J. T. Vaughan, N. Kanwisher, Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron* (1998) https:/doi.org/10.1016/S0896-6273(00)80592-9.

46. S. Dehaene, J. P. Changeux, Experimental and Theoretical Approaches to Conscious Processing. *Neuron* **70**, 200–227 (2011).

47. J. Jiang, Q. Zhang, S. Van Gaal, EEG neural oscillatory dynamics reveal semantic and response conflict at difference levels of conflict awareness. *Sci. Rep.* **5**, 12008 (2015).

48. M. X. Cohen, S. van Gaal, Subthreshold muscle twitches dissociate oscillatory neural signatures of conflicts from errors. *Neuroimage* **86**, 503–513 (2014).

49. J. Jiang, Q. Zhang, S. van Gaal, Conflict Awareness Dissociates Theta-band Neural Dynamics of the Medial Frontal and Lateral Frontal Cortex during Trial-by-trial Cognitive Control. *Neuroimage* **116**, 102–111 (2015).

50. V. a F. Lamme, Towards a true neural stance on consciousness. *Trends Cogn. Sci.* **10**, 494–501 (2006).

51. G. A. Mashour, P. Roelfsema, J. P. Changeux, S. Dehaene, Conscious Processing and the Global Neuronal Workspace Hypothesis. *Neuron* **105**, 776–798 (2020).

52. G. Tononi, M. Boly, M. Massimini, C. Koch, Integrated information theory: from consciousness to its physical substrate. *Nat Rev Neurosci* **17**, 450–461 (2016).

53. G. Huguet, J. Rinzel, J. M. Hupé, Noise and adaptation in multistable perception: Noise drives when to switch, adaptation determines percept choice. *J. Vis.* (2014) https:/doi.org/10.1167/14.3.19.

54. A. Delorme, S. Makeig, EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **134**, 9–21 (2004).

55. D. A. Leopold, J. C. Fitzgibbons, N. K. Logothetis, The role of attention in binocular rivalry as revealed through optokinetic nystagmus. *Multidiscip. Res.* (1995).

56. R. Fox, S. Todd, L. A. Bettinger, Optokinetic nystagmus as an objective indicator of binocular rivalry. *Vision Res.* (1975) https:/doi.org/10.1016/0042-6989(75)90265-5.

57. P. Enoksson, BINOCULAR RIVALRY AND MONOCULAR DOMINANCE STUDIED WITH OPTOKINETIC NYSTAGMUS. *Acta Ophthalmol.* (1964) https:/doi.org/10.1111/j.1755-3768.1964.tb03642.x.

58. T. Knapen, *et al.*, Cognitive and ocular factors jointly determine pupil responses under equiluminance. *PLoS One* **11**, e0155574. (2016).

59. J. Massey, Causality, feedback and directed information. *Proc. Int. Symp. Inf. Theory Applic.(ISITA-90)* (1990).

60. T. Schreiber, Measuring Information Transfer. *Phys. Rev. Lett.* **85**, 461–464 (2000).

61. M. Wibral, *et al.*, Transfer entropy in magnetoencephalographic data: Quantifying information flow in cortical and cerebellar networks. *Prog. Biophys. Mol. Biol.* (2011) https:/doi.org/10.1016/j.pbiomolbio.2010.11.006.

62. R. Vicente, M. Wibral, M. Lindner, G. Pipa, Transfer entropy-a model-free measure of effective connectivity for the neurosciences. *J. Comput. Neurosci.* (2011) https:/doi.org/10.1007/s10827-010-0262-3.

63. A. N. Kolmogorov, Three approaches to the definition of the concept of quantity of information. *IEEE Trans. Inf. Theory* **14**, 662–669 (1965).

64. G. J. Chaitin, Information-theoretic computation complexity. *IEEE Trans. Inf. Theory* **20**, 10–15 (1974).

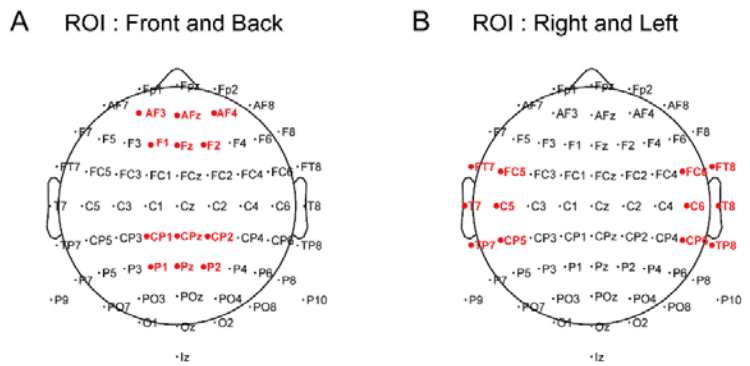
65. A. Lempel, J. Ziev, On the Complexity of Finite Sequences. *IEEE Trans. Inf. Theory* **22**, 75–82 (1976).

66. G. J. Chaitin, The berry paradox. *No TitleComplex Syst. Bin. Networks* **461**, 23–31 (1995).

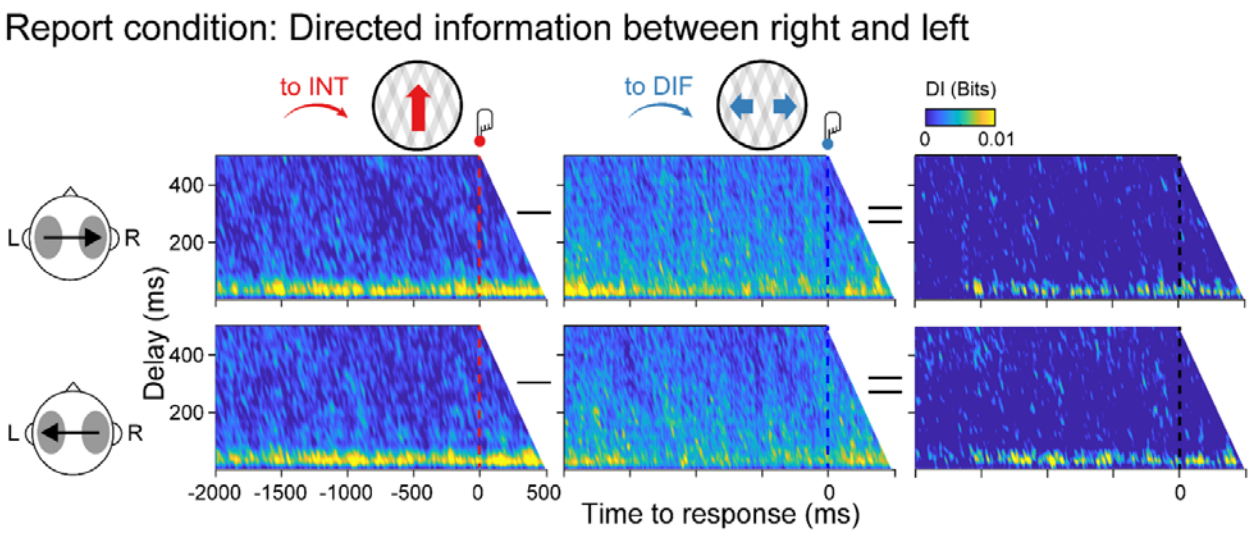
67. D. Salomon, *Data compression: the complete reference* (Springer Science & Business Media, 2004).

68. E. Maris, R. Oostenveld, Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).

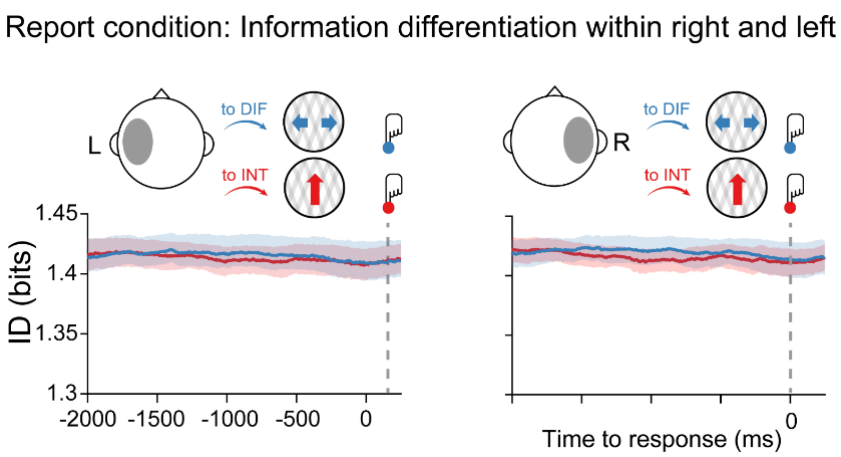
Supplementary Figures



Supplementary Figure 1. *Electrodes included in the (A) Front and Back ROIs, and (B) Right and Left ROIs.*



Supplementary Figure 2. Directed information of the report condition locked to button responses. *(A) Group-level DI between right and left ROIs (upper row) and between left and right ROIs (lower row) when moving plaids are reported as integrated (to INT; red color), reported as differentiated (to DIF; blue color). No significant clusters were found.*



Supplementary Figure 3. Information differentiation of the report condition locked to button responses. *Group-level ID within the left temporal ROI (left panel) and right temporal (right panel) when moving plaids are reported as differentiated and reported as integrated. No significant clusters were found.*