Isolating feedforward, lateral, and feedback mechanisms underlying perceptual and attentional impairments of conscious access

Samuel Noorman[1,2,*], Timo Stein[1,2], Johannes J. Fahrenfort[1,2,3,**], Simon van Gaal[1,2,**]

[1]Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands
[2]Amsterdam Brain and Cognition, University of Amsterdam, Amsterdam, The Netherlands
[3]Department of Applied and Experimental Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[*]Correspondence: sgnoorman@gmail.com
[**]Shared senior authorship

This research was supported by a grant from the H2020 European Research Council (ERC STG 715605, SVG).

Abstract

This study investigates failures in conscious access resulting from either weak sensory input (perceptual blindness) or unattended input (attentional blindness). Participants viewed an illusory Kanizsa triangle within a rapid serial visual presentation of distractor stimuli while electroencephalogram (EEG) was recorded. Distinct neural patterns associated with feedforward, lateral, and local versus global feedback processes were identified by training and testing classifiers on specific stimulus features. Perceptual performance was equated between the perceptual (masking) and attentional (attentional blink) manipulation to circumvent common confounds related to conditional differences in task performance. Decoding analyses revealed that lateral and local feedback processes were impaired by masking but spared by the attentional blink, with feedforward processing left largely unaffected by either manipulation. Global feedback processes were directly related to perceptual and metacognitive performance (conscious access), independent of the manipulation. These findings contribute to a comprehensive understanding of four distinct neural stages leading to conscious access.

Introduction

Conscious access to sensory input can be impaired in two distinct ways (Dehaene et al., 2006; Lamme, 2010; Mashour et al., 2020; Northoff & Lamme, 2020). Sensory input may lack sufficient bottom-up strength, or top-down attention may be directed elsewhere. Despite both cases resulting in a failure to perceive a stimulus, their underlying neural mechanisms are thought to be remarkably different. Influential theories of consciousness such as global neuronal workspace and recurrent processing theory propose four stages of neural information processing associated with distinct levels of bottom-up signal strength and top-down attention. These four stages can be investigated empirically by crossing "perceptual" manipulations that degrade the strength of sensory input (e.g., reducing stimulus contrast, masking, continuous flash suppression) with "attentional" manipulations that affect top-down attention (e.g., attentional blink, inattentional blindness, **Fig. 1A**).

According to these theoretical models, all stimuli elicit feedforward information transfer from lower- to higher-level brain regions (**Fig. 1A**, bottom row), but recurrent interactions are initiated only for stimuli with sufficient bottom-up strength (**Fig. 1A**, top row). If stimuli are sufficiently strong and top-down attention is available, neural processing crosses a threshold, triggering a process termed global ignition, facilitating widespread recurrent interactions between frontal, parietal and sensory cortices, yielding conscious access (**Fig. 1A**, top left). Crucially, when top-down attention is lacking, frontoparietal network ignition is prevented, while local recurrent interactions within sensory brain regions remain relatively intact ("attentional blindness", **Fig. 1A**, top right) (Dehaene et al., 2003; Marti et al., 2015; Sergent et al., 2005; Zivony & Lamy, 2022). Weak stimuli result in the absence, or a substantial reduction, of local recurrent interactions ("perceptual blindness", **Fig. 1A**, bottom left) (Fahrenfort et al., 2007; Joglekar et al., 2018; van Gaal et al., 2008; van Vugt et al., 2018).

Although this framework is at the heart of influential theories of consciousness, the four stages of the model and their underlying neural mechanisms have rarely been investigated simultaneously within the same study (for an exception see Fahrenfort et al., 2017). One challenge with comparing results across different studies, or even within a study, is that perceptual manipulations tend to impair overall task performance more than attentional manipulations, so that it may not be surprising to find that perceptual manipulations interrupt recurrent interactions to a greater extent than attentional manipulations. Given the right parameter settings, perceptual manipulations can be used to induce chance-level performance, while it is not possible to use attentional manipulations to drive behavioral performance down to chance, even when they are optimized fully. For this reason, attentional manipulations are often combined with post-hoc selection of a subset of "blind" trials (e.g., attentional blink) or subjects (e.g., inattentional blindness), a methodologically questionable practice that introduces criterion confounds, sampling bias, and regression to the mean (Peters & Lau, 2016; T. Schmidt, 2015; Shanks, 2017). Thus, when comparing perceptual to attentional manipulations, any (neural) effect could reflect differences in task performance rather than genuine differences between manipulations and hence stages in the model depicted in **Figure 1A**. While these issues with comparing conditions that differ in task performance in consciousness research have been acknowledged (Lau, 2022), they have rarely been addressed experimentally (Kanai et al., 2010; Lau & Passingham, 2006; Meuwese et al., 2014). Another major challenge for testing the neural

underpinnings of the four-stage model is to isolate feedforward, local, and global recurrent processes in humans using neuroimaging techniques. Although recent studies suggest that it is feasible to isolate these processes through appropriate stimulus protocols and analysis techniques (Fahrenfort et al., 2007, 2017; Kok et al., 2016; Kok & de Lange, 2014; Vandenbroucke et al., 2014), consciousness research has not yet fully capitalized on these advancements.

To test and further refine the four-stage model of consciousness in humans, we compared all stages within the same experimental setup, matching task performance between a perceptual manipulation (masking) and an attentional manipulation (attentional blink). Matching task performance was crucial to our design to control for confounds due to performance differences between perceptual and attentional failures of awareness. Further, we isolated different neural processes by combining a novel visual stimulus whose features allowed targeting distinct stages of visual processing with time-resolved decoding of these visual features from electroencephalogram (EEG) data (**Fig. 1B**). The target stimulus differed along three dimensions (illusory triangle, non-illusory triangle, and local contrast) that were independently manipulated. First, “Pac-Man” stimuli could create either the perception of an illusory surface in the shape of a Kanizsa triangle when aligned, or not, when misaligned. Second, additional “two-legged white circles” could form either a non-illusory triangle when their line segments were aligned, or not, when the legs were misaligned. Third, for the local contrast manipulation, the whole stimulus was rotated by 180 degrees, so that the same retinotopic positions had high contrast in one spatial configuration and low contrast when flipped 180 degrees.

Decoding the stimulus conditions of the illusory triangle, non-illusory triangle, and local contrast at different points in time, in combination with the associated topography, served as markers of distinct neural processes (**Fig. 1B**). To test the four-stage model, we collected markers of feedforward, local recurrent, and global recurrent processing. Local contrast decoding at early points in time served as a marker for feedforward processing, because the differences in local contrast are processed early in the visual system and are resistant to masking (Fahrenfort et al., 2007, 2017; Kandel et al., 2000; Lamme & Roelfsema, 2000). Leveraging the well-established reliance of the Kanizsa illusion on recurrent processing (Halgren et al., 2003; Kok et al., 2016; Kok & de Lange, 2014; Lee & Nguyen, 2001; Pak et al., 2020; Wokke et al., 2013), illusory triangle decoding at earlier vs. later points in time served as markers for local vs. global recurrent processing (Fahrenfort et al., 2017). Local recurrent processing comprises both feedback and lateral interactions. We therefore attempted to distinguish between these sub-components of local recurrent processing. Both the illusory and the non-illusory triangle involved processing the alignment of collinear line segments of equal length and equal distance between them, from now on referred to as collinearity. Collinearity processing primarily relies on lateral connections (Bosking et al., 1997; Gilbert & Wiesel, 1979; Li, 1998; Liang et al., 2017; K. E. Schmidt et al., 1997; Stettler et al., 2002). By subtracting non-illusory triangle decoding (supported by lateral connections) from illusory triangle decoding (supported by lateral and feedback connections), we aimed to isolate feedback processing, thus effectively subtracting out the contribution of lateral interactions. Armed with these EEG markers of different neural processes, we tested whether the effects of masking and the attentional blink followed the predictions of the four-stage model of consciousness.

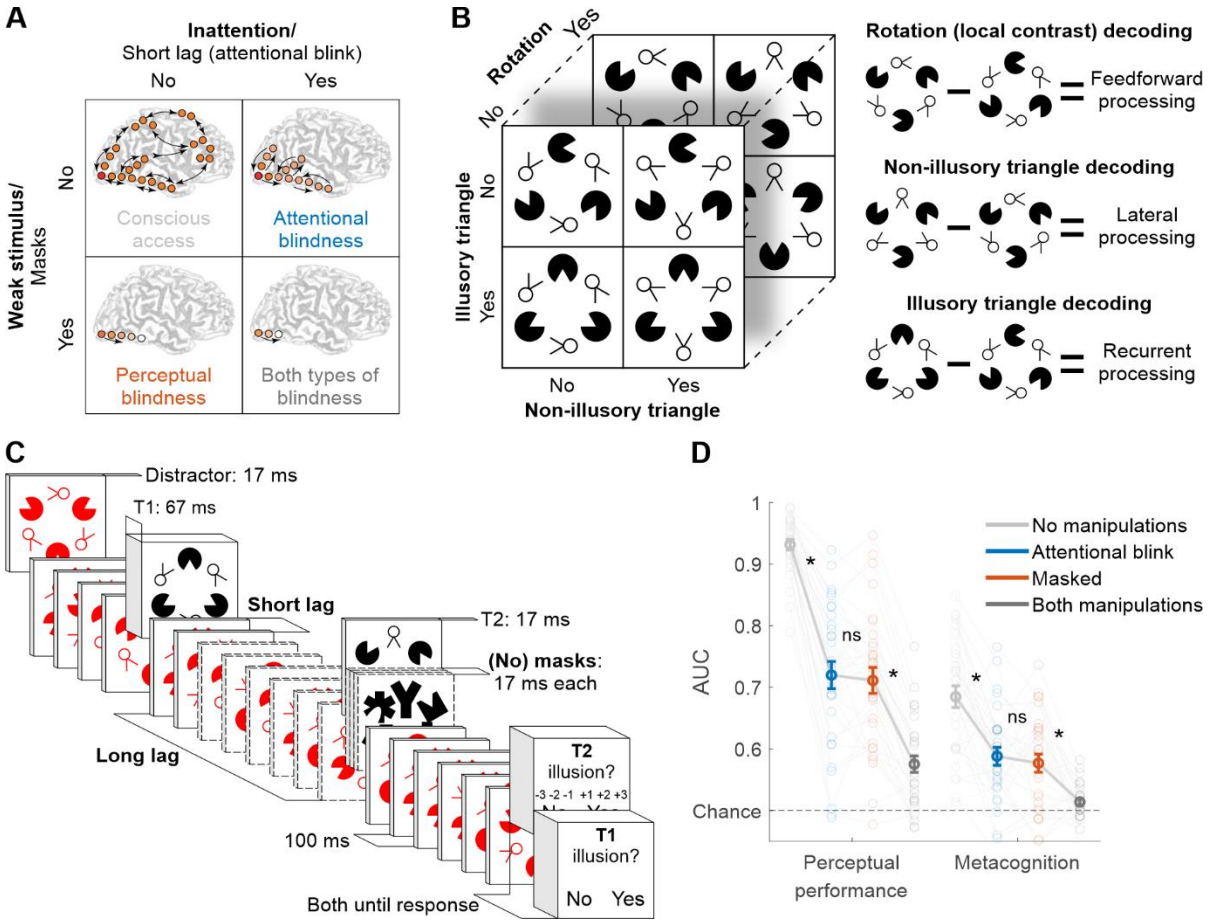

Figure 1. Experimental design and behavior. *(A) Perceptual vs. attentional blindness in the four-stage model. A stimulus with low bottom-up strength (masked) is thought to interrupt local recurrent processing in sensory areas while leaving feedforward processing largely intact, while inattention (induced by the attentional blink) is thought to interrupt global recurrent processing between frontoparietal areas and sensory areas, while leaving local recurrent processing within sensory areas largely intact. Reprinted from Dehaene et al.* (2006) *with permission from Elsevier. (B) Target stimulus set and schematic of the markers for the different types of processing: local contrast decoding (supported by feedforward connections), non-illusory triangle decoding (supported by lateral connections), and illusory triangle decoding (supported by lateral and feedback connections) (C) Trial design. (D) Perceptual performance refers to participants' ability to detect the Kanizsa illusion. Metacognition refers to participants' ability to evaluate their own performance using confidence judgments. Both perceptual performance and metacognition are measured as the area under the receiver operating characteristic curve (AUC). Error bars are mean ± standard error of the mean. Individual data points are plotted using low contrast. Ns is not significant ($P \geq 0.477$, $BF_{01} \geq 4.05$). *$P \leq 0.001$.*

Results

Masking and the attentional blink were matched in perceptual performance and metacognition

We recorded the EEG signal of 30 participants who identified the presence or absence of an illusory surface (triangle) in two black target stimuli (T1 and T2) that were presented amongst red distractors in a rapid serial visual presentation task (**Fig. 1C**). We manipulated the visibility of T2 in two ways: masking the stimulus and manipulating attention, resulting in a 2×2 factorial design (**Fig. 1A**). Specifically, T2 could be either masked or unmasked (perceptual manipulation), and T2 could be presented at either a long interval (900 ms) or a short interval (200 or 300 ms) after T1, inducing an attentional blink (AB) effect for the short T1-T2 intervals (Raymond et al., 1992). This design resulted in four conditions, which we from now on refer to as the masked condition (T2 masked at the long T1-T2 interval), AB condition (T2 unmasked at the short T1-T2 interval), no manipulations condition (T2 unmasked at the long T1-T2 interval), and both manipulations condition (T2 masked at the short T1-T2 interval). At the end of a trial, participants indicated whether each target (T1 and T2) contained an illusory surface or not. Importantly, mask contrast in the masked condition was individually adjusted using a staircasing procedure to match participants' performance in the AB condition, ensuring comparable perceptual performance in the masked and the AB condition (see Methods for more details).

Conscious access can be assessed not only based on perceptual performance but also through metacognitive sensitivity, the ability to evaluate one's own performance (Brown et al., 2019; Dienes, 2007; Fleming & Lau, 2014; Lau & Passingham, 2006; Merikle et al., 2001; Seth et al., 2008). Participants in our study provided confidence ratings on a 3-point scale (low, medium, high) for their responses to T2. To ensure that the distribution of confidence ratings was not influenced by overall differences in perceptual performance between conditions, conditions that were matched in perceptual performance (masked and AB condition) were presented in the same experimental block, while the other block type included the unmatched conditions (no and both manipulations condition).

We used area under the receiver operating characteristic (ROC) curve (AUC) as a shared metric for perceptual performance (detection of the Kanizsa illusion), metacognitive sensitivity, and EEG decoding (see Methods for details on the calculation of these measures). Repeated-measures (rm) ANOVA with the factors masking (present/absent) and T1-T2 lag (short/long) revealed, as expected, that both masking and the short T1-T2 lag impaired perceptual performance ($F_{1,29}$=344.24, $P<10^{-15}$ and $F_{1,29}$=427.54, $P<10^{-15}$) as well as metacognitive sensitivity ($F_{1,29}$=50.78, $P<10^{-7}$ and $F_{1,29}$=47.83, $P<10^{-6}$). Importantly, paired t-tests showed that we successfully matched the key conditions, the masked condition (masked, long lag) and the AB condition (unmasked, short lag) for perceptual performance (t_{29}=0.62, P=0.537, BF_{01}=4.30, **Fig. 1D**, left) as well as for metacognitive sensitivity (t_{29}=0.72, P=0.477, BF_{01}=4.05; **Fig. 1D**, right, see **Fig. S1** for signal detection theory related measures of performance). Thus, the two performance matched conditions were indistinguishable from each other in both measures of conscious access.

Masking and the attentional blink leave feedforward processing largely intact

To derive our markers of the different neural processes from our EEG data, for each stimulus feature we trained linear discriminant classifiers on the T1 data and tested them on the T2 data. Classifiers used raw EEG activity across all electrodes. To leverage the similarities between T1 and T2 in task and stimulus context, all main analyses used T1 training data for T2 decoding. This approach minimized possible differences in conscious access and working memory demands between the training and test datasets.

For the marker of feedforward processing, the classifier categorized stimuli as either pointing upwards or pointing downwards, thereby effectively decoding the stimuli's local differences in contrast at the top vs. bottom of the stimulus. Classification performance (AUC) over time was obtained, with peak decoding accuracy in a 75-95 ms time window (**Fig. 2A** and **Fig. S2A**, top). The peak in decoding accuracy was occipital in nature (see the covariance/class separability map of **Fig. 2A**) (Haufe et al., 2014), consistent with our previous findings (Fahrenfort et al., 2017). We focused our analyses on the averages of this time window. An rm ANOVA with the factors masking (present/absent) and T1-T2 lag (short/long) revealed only a marginal effect of masking on feedforward processing ($F_{1,29}$=6.51, P=0.016), while the T1-T2 lag had no significant effect ($F_{1,29}$=0.32, P=0.578). A paired t-test yielded no evidence for a difference between the performance matched conditions (masked vs. AB; t_{29}=1.42, P=0.166, BF_{01}=2.08; **Fig. 2C**, "Local contrast: 75-95 ms"). These results are in line with theoretical proposals and empirical findings that suggest limited effects of masking and attentional manipulations on feedforward processes (Dehaene et al., 2006; Fahrenfort et al., 2007, 2017; Lamme, 2010).

Stronger effect of masking than the attentional blink on local but not global recurrent processing

For the marker of recurrent processing, we trained a linear classifier on the T1 data to discriminate between the absence and presence of the Kanizsa illusion and tested it on each of the four conditions of the T2 data. The average of all four conditions revealed two prominent peaks in decoding accuracy, consistent with previous research (**Fig. S2C**, top) (Fahrenfort et al., 2017). Based on this previous study, our analyses focused on the averages of the two time windows that encompassed these two peaks: specifically, from 200 to 250 ms and from 375 to 475 ms after target stimulus onset. The covariance/class separability maps (**Fig. 2B**) indicated that during the earlier time window (200–250 ms) classification mainly relied on occipital electrodes. Considering its timing, topology and previous findings, this neural event likely reflects sensory processes and served as the marker for local recurrent processing (Fahrenfort et al., 2017; Kok et al., 2016; Roelfsema, 2006; Wokke et al., 2013; Wyatte et al., 2014). The timing and topology of the later neural event (375–475 ms) overlapped with the event-related potential component P300, which is associated with conscious access (Fahrenfort et al., 2017; Sergent et al., 2005; Weaver et al., 2019) and served as the marker for global recurrent processing.

We tested how the consciousness manipulations affected these markers of local and global recurrent processing (**Fig. 2C**), again conducting rm ANOVAs with the factors masking (present/absent) and T1-T2 lag (short/long) that we followed-up on with paired t-tests comparing

the matched conditions (masked vs. AB). Importantly, we observed a distinct difference between the performance matched conditions in the first decoding peak, the marker of local recurrent processing. Local recurrent processing was significantly impaired by both masking ($F_{1,29}$=162.62, P<10^{-12}) and the T1-T2 lag ($F_{1,29}$=78.07, P<10^{-8}), but importantly, it was more affected by masking than the T1-T2 lag ($F_{1,29}$=18.67, P<0.001) (**Fig. 2C**, "Illusory triangle: 200-250 ms"). Directly comparing the performance matched conditions, local recurrent processing was more strongly impaired in the masked condition than in the AB condition (t_{29}=4.66, P<10^{-4}, BF_{01}=0.003).

The pattern of results of the second peak, the marker of global recurrent processing, was notably different. Global recurrent processing was impaired by both masking ($F_{1,29}$=49.75, P<10^{-7}) and the T1-T2 lag ($F_{1,29}$=78.48, P<10^{-9}), and the matched conditions (masked and AB condition) did not differ significantly from each other (t_{29}=0.21, P=0.837, BF_{01}=5.04) (**Fig. 2C**, "Illusory triangle: 375-475 ms"). Furthermore, another rm ANOVA comparing local and global recurrent processing between the matched conditions (masked/AB) revealed a significant interaction, reflecting a larger difference between the AB and the masked condition in local recurrent processing than in global recurrent processing ($F_{1,29}$=31.53, P<10^{-5}). Across the four conditions, the pattern of behavioral results, both for perceptual performance and metacognitive sensitivity, closely resembled global recurrent processing (**Fig. 2C**), indicating that global recurrent processing reflected conscious access to the Kanizsa illusion.

Additional rm ANOVAs assessing the effect of the consciousness manipulations on the different neural processes showed that, compared to feedforward processing, both manipulations had stronger effects on both local recurrent processing (masking: $F_{1,29}$=99.35, P<10^{-10}; AB: $F_{1,29}$=38.95, P<10^{-6}) and global recurrent processing (masking: $F_{1,29}$=22.25, P<10^{-4}; AB: $F_{1,29}$=49.60, P<10^{-7}). This shows that masking and the AB specifically influenced local and global recurrent processing respectively, while early feedforward processing was less affected.

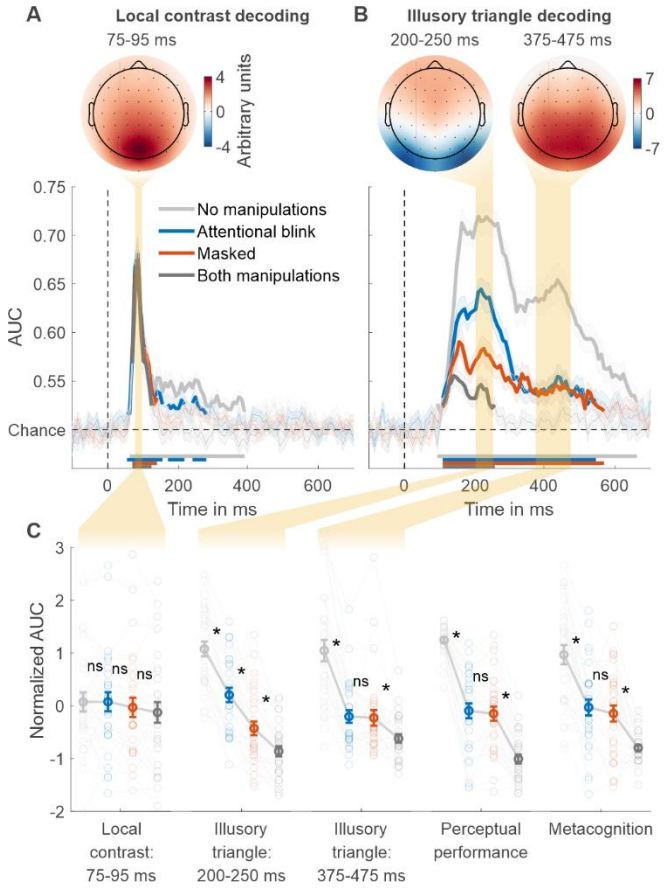

Figure 2. Local contrast and illusory triangle decoding using first targets as training data. *(A) Local contrast decoding. (B) Illusory Kanizsa triangle decoding. For both features, covariance/class separability maps reflecting underlying neural sources are shown. Below these maps: mean decoding performance, area under the receiver operating characteristic curve (AUC), over time ± standard error of the mean (SEM). Thick lines differ from chance: P<0.05, cluster-based permutation test. (C) Normalized (Z-scored) AUC for every measure: mean decoding time windows and two types of behavior. Each measure is Z-scored separately. Perceptual performance refers to participants' ability to detect the Kanizsa illusion. Metacognition refers to participants' ability to evaluate their own performance using confidence judgments. See Figure S3 for the same analyses but then for off-diagonal decoding profiles. Error bars are mean ± SEM. Individual data points are plotted using low contrast. Ns is not significant (P≥0.166, BF_{01}≥2.07). *P≤0.002.*

Distinguishing lateral vs. feedback connections in local recurrent processing

The performance matched masked and AB condition differed only in local recurrent processing, which was markedly more impaired for the masked than the AB condition. To determine whether this effect was specific to (local) feedback connections or also involved lateral interactions, we distinguished the components of local recurrent processing: lateral and feedback connections (Lamme et al., 1998; Roelfsema, 2006) (**Fig. 1B**). In our target stimulus, collinearity was present when the Pac-Man stimuli aligned, inducing the illusory Kanizsa triangle. Notably, collinearity was also present when the line segments of the "two-legged white

circles" of the stimulus aligned, forming the non-illusory triangle. Note that the line segments making up the triangle were equally long, and the spaces between them equally large, for the illusory and non-illusory triangles. Collinearity processing primarily relies on lateral connections (Bosking et al., 1997; Gilbert & Wiesel, 1979; Li, 1998; Liang et al., 2017; K. E. Schmidt et al., 1997; Stettler et al., 2002), while processing of the Kanizsa illusion involves both lateral and feedback connections (Halgren et al., 2003; Kok et al., 2016; Kok & de Lange, 2014; Lee & Nguyen, 2001; Pak et al., 2020; Wokke et al., 2013). Thus, by comparing non-illusory triangle decoding to illusory triangle decoding we can in principle isolate illusion-specific feedback processing from the influence of lateral interactions.

However, the main RSVP task required participants to focus on the illusory triangle, making it task-relevant, while non-illusory triangles were always task-irrelevant. This difference could influence the direct comparison in decoding accuracy between the two types of triangles. Therefore, to equate the effect of task-relevance in the comparison, classifiers were trained on an independent training set in which each relevant stimulus feature was task-relevant. Specifically, in different experimental blocks, participants either focused on local contrast, the non-illusory triangle, or the illusory triangle (**Fig. S4**). We trained a classifier to distinguish between the absence and presence of the task-relevant non-illusory triangle (collinearity-only) and the same was done for the task-relevant illusory triangle (collinearity-plus-illusion). Then, both classifiers were used to decode the presence vs. absence of the illusory triangle in the main RSVP task (cross-task-decoding approach), which ensured that both training and testing were always performed on task-relevant stimuli. By comparing Kanizsa decoding performance in the RSVP task based on the collinearity-only classifier with decoding performance based on the collinearity-plus-illusion classifier, we effectively subtract out the contribution of lateral connections to illusion decoding and isolate illusion-specific feedback processing.

Preserved lateral and local feedback connections during the attentional blink

To determine a time window for (the start of) lateral processing, related to collinearity, we first trained and tested classifiers to distinguish present vs. absent non-illusory triangles (training and testing on non-illusory triangles only). We trained two classifiers, one on the T1 in the RSVP task and one on the independent training set and tested their performance in decoding the non-illusory triangle in the T2 data, where this triangle was always task-irrelevant. The results of both classifiers converged and these analyses revealed a peak in decoding accuracy at ~164 ms, suggestive of an effect of lateral processing, right before the 200-250 ms time window of local recurrent processing (**Fig. S2B**). This peak was also evident when these classifiers were used to categorize the presence vs. absence of the Kanizsa illusion (**Fig. S2C**, first time window), as well as in previous research (Fahrenfort et al., 2017), suggesting that lateral processes also contribute to Kanizsa decoding. Next, we tested that hypothesis in more detail.

We examined how decoding the presence vs. absence of the Kanizsa illusion in the RSVP task was affected by the consciousness manipulations, while training classifiers either on the illusory (collinearity-plus-illusion) or non-illusory (collinearity-only) triangle from the independent training set. **Figure 3A** shows the decoding accuracies of these analyses across the entire time-window (purple and green lines). Follow-up analyses were performed using the

140-190 ms window, observed to reflect the peak of lateral processing (see the previous paragraph). An rm ANOVA with the factors masking (present/absent), T1-T2 lag (short/long), and training set (illusory/non-illusory triangle) revealed that both masking and the short T1-T2 lag impaired decoding accuracy (masking: $F_{1,29}$=58.95, P<10^{-7}; T1-T2 lag: $F_{1,29}$=24.90, P<10^{-4}). Furthermore, paired t-tests comparing the matched conditions (masked vs. AB condition) confirmed that decoding accuracy was more impaired in the masked than AB condition, both when training was done on the illusory (t_{29}=2.26, P=0.031, BF_{01}=0.58) and non-illusory triangle (t_{29}=2.78, P=0.009, BF_{01}=0.21; **Fig. 3B**, "140-190 ms"). Focusing on the performance matched conditions and the role of the training set, an rm ANOVA with the factors condition (masked/AB) and training set (illusory/non-illusory triangle) on T2 illusory triangle decoding revealed no significant effect of training set ($F_{1,29}$=0.09, P=0.766), i.e., no evidence for illusion-specific feedback processing, and no significant interaction ($F_{1,29}$=0.04, P=0.837). This demonstrates that neural processing in the 140-190 ms window indeed reflects lateral rather than feedback connections.

Interestingly, the marker for illusion-specific feedback emerged later, in the 200-250 ms time window reflecting local recurrent processing. As can be seen in **Figure 3A**, when no consciousness manipulations were applied (unmasked, long T1-T2 lag), there was significant illusion-specific feedback processing, i.e., T2 illusory triangle decoding was better after training a classifier on the Kanizsa illusion (collinearity-plus-illusion, green line) than after training a classifier on the non-illusory triangle (collinearity-only, purple line) (t_{29}=4.22, P<0.001, BF_{01}=0.008). Turning to the performance matched conditions, an rm ANOVA with the factors condition (masked/AB) and training set (illusory/non-illusory triangle) on T2 illusory triangle decoding yielded a significant effect of condition ($F_{1,29}$=16.59, P<0.001), with overall better decoding for the AB than for the masked condition, and importantly, a significant interaction ($F_{1,29}$=4.65, P=0.039). **Figure 3A** shows that decoding after training on illusory triangles (collinearity-plus-illusion) was better than after training on non-illusory triangles (collinearity-only) for the AB (t_{29}=2.51, P=0.018, BF_{01}=0.36, **Fig. 3A**, top right) but not for the masked condition (t_{29}=-0.02, P=0.982, BF_{01}=5.14, **Fig. 3A**, bottom left). Thus, the marker for illusion-specific feedback processing was still present in the AB condition, whereas this marker was fully abolished in the masked condition. Feedback processing was not even affected by the AB condition, as an rm ANOVA with the factors (masking is absent) condition (no manipulations/AB) and training set (illusory/non-illusory triangle) revealed no significant interaction ($F_{1,29}$=1.33, P=0.259). The classifier trained on non-illusory triangles (collinearity-only) also performed better during the AB than masking (t_{29}=2.85, P=0.008, BF_{01}=0.18; **Fig. 3B**, "200-250 ms", purple line), hence both lateral and feedback processes were most strongly impaired by masking. Control analyses presented in the **Supplementary information (Fig. S6)** demonstrate that cross-feature-decoding can indeed isolate illusion-specific feedback processes and does not reflect other, e.g., task- or attention-related, processes.

Having established that, we next aimed to replicate the observation that the two consciousness manipulations left feedforward processing largely intact by training the classifier on the independent training set in which local contrast was task-relevant (**Fig. 3A**, light blue lines). As for our previous main analyses, we averaged the 75-95 ms time window, as this again contained the decoding peak with occipital topography (**Fig. S2A**, bottom). Similar to our main analyses, another rm ANOVA with the factors masking (present/absent) and T1-T2 lag

(short/long) showed that neither masking ($F_{1,29}$=0.97, P=0.334) nor the T1-T2 lag ($F_{1,29}$=0.72, P=0.403) had a significant effect on feedforward processing, and a paired t-test revealed no evidence of a significant difference between the two performance matched conditions (masked vs. AB condition, t_{29}=1.20, P=0.240, BF_{01}=2.68; **Fig. 3B**, "75-95 ms"). For completeness, we report local contrast decoding as a function of task-relevance in the **Supplementary information** and **Figure S7**. In short, local contrast decoding did not vary systematically with task-relevance of the stimulus feature.

Finally, we focused on the late 375-475 ms window, the marker for global recurrent processing directly linked to behavioral performance (**Fig. 3A**, last time window). Similarly as above, Illusory triangle decoding was now based on training the decoder on the illusory triangles from the independent training set. Replicating our main analysis, classifier performance was impaired by both masking ($F_{1,29}$=6.01, P=0.020) and T1-T2 lag ($F_{1,29}$=10.10, P=0.004) with no significant differences between the two performance matched conditions (t_{29}=-0.63, P=0.531, BF_{01}=4.27) (**Fig. 3B**, "375-475 ms").

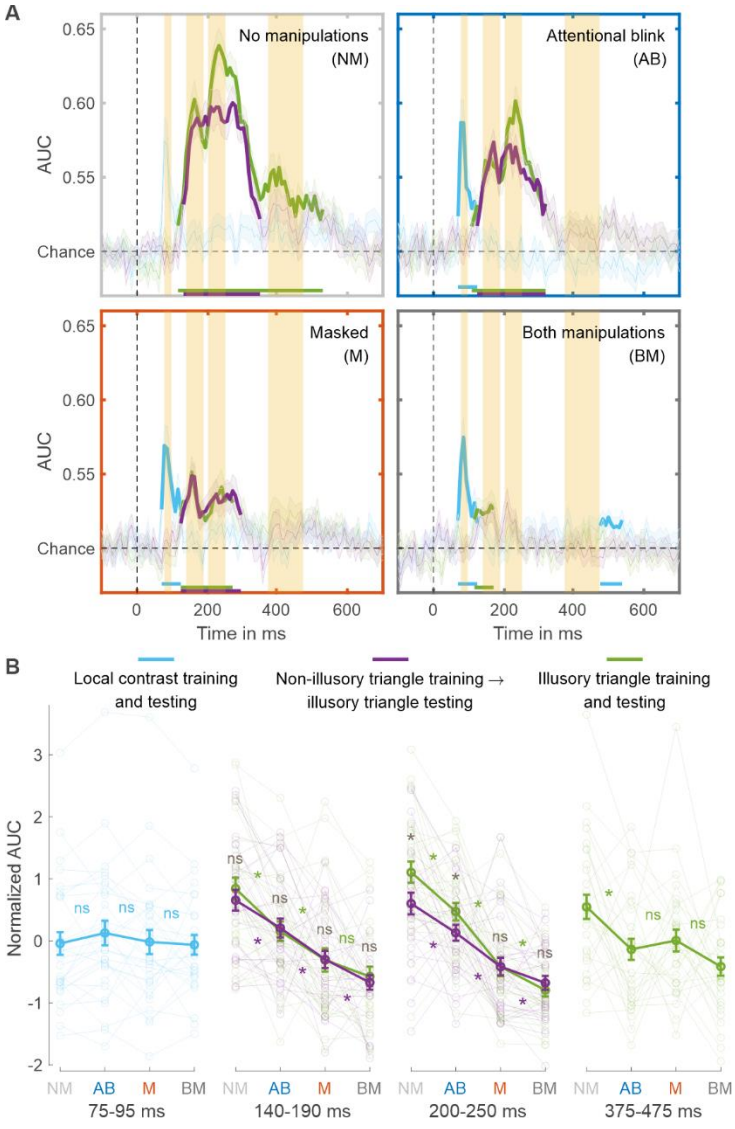

Figure 3. Separating lateral and feedback processes using the independent training dataset. *(A) Illusory triangle decoding, after training classifiers on the independent training set on either the non-illusory (collinearity-only, purple lines) or illusory triangle (collinearity-plus-illusion, green lines). For comparison, training and testing on local contrast is shown in light blue. Mean decoding performance, area under the receiver operating characteristic curve (AUC), over time ± standard error of the mean (SEM) is shown. Thick lines differ from chance: P<0.05, cluster-based permutation test. The highlighted time windows are 75-95, 140-190, 200-250, and 375-475 ms, corresponding to separate panels in (B), which shows normalized (Z-scored) mean AUC for every time window. Each window is Z-scored separately. Error bars are mean ± SEM. Individual data points are plotted using low contrast. Ns is not significant (P≥0.084, BF_{01}≥1.26). *P≤0.048.*

Discussion

We demonstrate that perceptual and attentional manipulations, despite similarly impairing conscious access, exhibit distinct neural profiles in the brain. To investigate this difference, we decoded different visual features targeting distinct stages of visual processing from human EEG activity, while carefully matching a masked condition and an attentional blink (AB) condition in perceptual and metacognitive performance. Decoding the illusory Kanizsa triangle served as a marker for recurrent processing, revealing both global (late and centroparietal) and local (early and occipital) recurrent processing. Global recurrent processing was similarly impaired by the perceptual manipulation (masked condition) and the attentional manipulation (AB condition), closely resembling their matched effects on behavioral performance. However, local recurrent processing was markedly more impaired in the masked than the AB condition (Fahrenfort et al., 2017), even though task performance was matched. This key neural difference was specific to local recurrent processing, as the marker for feedforward processing (local contrast decoding) was barely affected by the two consciousness manipulations. Furthermore, we further analyzed the components of local recurrent processing and differentiated illusion-specific feedback from processing by lateral connections by subtracting non-illusory (collinearity-only) from illusory triangle (collinearity-plus-illusion) decoding. Both feedback and lateral processing were more strongly impaired by masking than the AB. Notably, the marker for illusion-specific feedback processing was unaffected by the AB, but completely abolished by masking. These findings confirm and enrich empirical and theoretical work on perceptual vs. attentional mechanisms of consciousness (Block, 2005; Dehaene et al., 2006; Hatamimajoumerd et al., 2022; Lamme, 2010; Pitts et al., 2018; Sergent & Dehaene, 2004), clearly distinguishing and specifying the neural profiles of each processing stage of the influential four-stage model of conscious experience.

To our knowledge, this is the first study to examine the neural mechanisms underlying conscious access in which behavioral measures of conscious perception are carefully matched between the attentional blink and masking within a single experimental design. Previous investigations have typically employed separate paradigms for perceptual and attentional manipulations, often using different stimuli associated with distinct neural mechanisms, which complicates direct comparisons between manipulations and across studies. Further, inattention approaches generally use stronger sensory input (e.g., stimuli of longer duration, higher contrast) than perceptual manipulations (Stein et al., 2021). Here, we introduced a novel

stimulus that allowed us to isolate four distinct stages of visual processing by decoding different features while holding visual stimulation and task context constant. Furthermore, measurement of conscious perception often differs between perceptual and attentional manipulations. In particular, inattention approaches, which have previously tended to reveal more extensive neural processing, frequently involve post-hoc selection of “blind” trials or participants based on subjective awareness reports, which is susceptible to criterion confounds and introduces selection and sampling biases (Peters & Lau, 2016; T. Schmidt, 2015; Shanks, 2017). In contrast, our study analyzed all trials and included all participants while carefully matching perceptual performance and metacognition between masking and the AB. Therefore, any observed neural difference between masking and the AB can be unequivocally attributed to differences between attentional and perceptual manipulations of conscious access.

Compared to masking, the AB left local recurrent processing intact, while feedforward processing did not differ between the two manipulations. Local recurrent processing plays a critical role in perceptual integration, facilitating the organization of fragmented sensory information, such as lines, surfaces, and objects, into a coherent whole (Roelfsema, 2023). Our EEG decoding results support this notion, demonstrating that the AB allows for greater processing of collinearity and the illusion specifically within a time window spanning 140 to 250 ms after stimulus onset, likely reflecting sparing of local recurrent processes in visual cortex (Fahrenfort et al., 2017; Kok et al., 2016). This aligns with established models of the AB phenomenon, in which the AB reflects a late post-perceptual central bottleneck characterized by limited attentional capacity (Shapiro et al., 1997), so that sensory information presented during the AB can nevertheless undergo extensive processing, allowing for perceptual integration, possibly even leading up to semantic analysis (Luck et al., 1996).

Our finding of preserved local recurrent processing during the AB is also consistent with classic load theory (Lavie & Dalton, 2014), where increasing perceptual load (Lavie & de Fockert, 2003) more strongly reduces distractor processing than increasing cognitive load (e.g., by engaging working memory, as in our AB condition). According to this theory, perceptual and attentional manipulations serve as early and late filters for incoming sensory information, respectively, resulting in more extensive processing under inattention. Indeed, one of the few neuroimaging studies that included both manipulations found that only perceptual but not cognitive (working memory) load decreased fMRI activity in the parahippocampal place area in response to distractor scenes (Yi et al., 2004). However, not all neuroimaging evidence is consistent with a stronger effect of perceptual than cognitive load (Brockhoff et al., 2022). Furthermore, previous research has shown that the impact of inattention vs. masking can depend on the neural architecture required for the task at hand. For example, processes related to the detection of conflicting response tendencies, a hallmark of cognitive control and strongly associated with the prefrontal cortex (Ridderinkhof et al., 2004), are more susceptible to inattention, which reduces the depth of stimulus processing (Nuiten et al., 2021) than to masking, restricting recurrent interactions, but allowing for deep feedforward processing (all the way up to prefrontal cortex) (Jiang et al., 2018; van Gaal et al., 2008). Thus, the preservation of local recurrent interactions appears to be particularly important for perceptual integration, aligning with the influential notion that perceptual segmentation and organization may represent the mechanism of conscious experience (Lamme, 2020).

Local recurrent interactions in visual cortex encompass both lateral and feedback connections. The distinct roles of lateral and feedback connections to visual function have received limited attention in human cognitive neuroscience and remain unaddressed in theories of consciousness. Here we distinguished a marker of lateral processing reflecting basic collinearity processing and a marker of feedback processing reflecting illusion-specific processing. We observed that lateral processing occurred earlier (between 140 and 190 ms after target stimulus onset) than illusion-specific feedback processing (between 200 and 250 ms), in line with animal research (Angelucci & Bressloff, 2006; Lamme et al., 1998; Roelfsema, 2006). Both lateral and feedback processing were more strongly affected by masking than by the AB, indicating that the “attentional blindness” stage of the four-stage model of consciousness (**Fig. 1A**) involves both lateral and feedback connections. Interestingly, masking had a stronger effect on illusion-specific feedback processing than on lateral processing. Along with the distinct temporal and spatial EEG decoding patterns associated with each marker, this suggests a processing sequence from feedforward processing to local recurrent interactions encompassing lateral-to-feedback connections, ultimately leading to global recurrency and conscious report.

Having delineated these distinct stages of feedforward, lateral, feedback and global recurrent processing, one important avenue for future research is to distinguish between unconscious and conscious perceptual processes at these stages. Having opted to equate performance across manipulations in our study, behavioral performance was above chance level for both consciousness manipulations. Follow-up research investigating perceptual integration of fully unconscious stimuli could address ongoing debates between influential theories of consciousness (Cogitate Consortium et al., 2023; Mudrik et al., 2014). The global neuronal workspace theory suggests a durable, yet unconscious processing stage (referred to as preconscious), where the input is not globally available, and amplification through top-down attention is required for conscious access and report (Dehaene et al., 2006). In contrast, others have argued that already local recurrent interactions reflect subjective phenomenal experience (Block, 2005; Lamme, 2010). Moreover, markers like the P300 and ours for global recurrent processing may reflect functions not directly related to conscious experience, like report or decision-making (Alilović et al., 2023; Canales-Johnson et al., 2023; Pitts et al., 2018). Another way forward therefore consists in combining no-report paradigms (Sergent et al., 2021; Tsuchiya et al., 2015) with our EEG markers to examine whether local or global recurrent processing more accurately reflects consciousness in the absence of report.

Methods

Participants

Thirty-three participants took part in the first two sessions (independent EEG training set and practice). Three of them met the practice session’s pre-established criteria for exclusion (see “Procedure”). The remaining 30 participants (22±3 years old, 10 men, 2 left-handed) took part in the final (main experimental) session. They all had normal or corrected-to-normal vision. The

491 study was approved by the local ethics committee. Participants gave informed consent and
492 received research credits or 15 euros per hour.

493 ## Stimuli

494 The target stimulus set had a 2 (illusory Kanizsa triangle: present/absent) × 2 (non-illusory
495 triangle: present/absent) × 2 (rotation: present/absent) design, resulting in eight stimuli (**Fig.**
496 **1B**). Three aligned Pac-Man elements induced the Kanizsa illusion. The non-illusory triangle
497 was present when the stimuli's three other elements (the "two-legged white circles") were
498 aligned. The controls for both the illusory and non-illusory triangle were created by rotating their
499 elements by 90 degrees. Differences in local contrast were created by rotating the entire
500 stimulus by 180 degrees. The targets spanned 7.5 degrees by 8.3 degrees of visual angle. The
501 distance between the three Pac-Man stimuli as well as between the three aligned two-legged
502 white circles was 2.8 degrees of visual angle. Although neuronal responses to collinearity in
503 primary visual cortex are most robust when this distance is smaller (Kapadia et al., 1995, 2000),
504 longer-range lateral connections between neurons with similar orientation selectivity can span
505 distances corresponding to visual angles considerably greater than 2.8 degrees (Bosking et al.,
506 1997; Stettler et al., 2002).
507 The distractor stimulus set was the same as the target stimulus set, with two exceptions.
508 First, the distractors were red instead of black. Second, the distractors' six elements were
509 rotated by 180 degrees relative to the targets', so neither the illusory nor non-illusory triangle
510 was ever present in the distractors. Masks consisted of six differently shaped elements, all
511 capable of covering the targets' elements. Six masks were created by rotating the original mask
512 five times by 60 degrees. They spanned 8.5 degrees by 9.1 degrees of visual angle. The
513 fixation cross, which was always present, was adapted from Thaler et al. (2013).

514 ## Procedure

515 The experiment consisted of three separate sessions conducted on different days: a three-hour
516 session to collect EEG data for the independent training set, a 1.5-hour practice session, and a
517 three-hour experimental session. Tasks were programmed in Presentation software
518 (Neurobehavioral Systems) and displayed on a 23-inch, 60 Hz, 1920×1080 pixels monitor. On
519 each trial of the experimental session, participants were shown two targets (T1 and T2) within a
520 rapid serial visual presentation (RSVP) of distractors (**Fig. 1C**). The targets and distractors had
521 a stimulus onset asynchrony of 100 ms. T2 and distractors were presented for 17 ms each. To
522 improve the decoding analyses' training dataset, T1 was presented for 67 ms. The longer
523 presentation duration facilitated attending to T1, which should result in greater deployment of
524 attentional resources and thereby increase the size of the AB. T1 was preceded by five
525 distractors and T2 was followed by six distractors.
526 T2 visibility was manipulated in two ways, using a perceptual and an attentional
527 manipulation (**Fig. 1A**). The perceptual manipulation consisted in masking T2 with three masks,
528 each presented for 17 ms with an interstimulus interval of 0 ms. The three masks were selected
529 randomly, but all differed from each other. Half of the T2s were masked; for the other half no
530 masks were presented (unmasked condition). The attentional manipulation consisted in
531 shortening the T1–T2 lag from a long interval of eight distractors (900 ms) to a short interval of

one or two distractors (200 or 300 ms). Half the trials had a long lag, the other half had a short lag. The short lag duration was determined for each participant individually during the training session. Short lags were expected to result in an AB. Participants were instructed to fixate on the fixation cross. After the RSVP, they indicated for each target whether it contained the Kanizsa illusion or not. For T2, participants simultaneously reported their confidence in their response: low, medium, or high, resulting in six response options. To get accurate ratings, participants first responded to T2 and then to T1. Response screens lasted until the response. In short, the experimental session had an 8 (T1 stimulus conditions) × 8 (T2 stimulus conditions) × 2 (masked/unmasked) × 2 (short/long T1–T2 lag) task design, resulting in 256 conditions. Each condition was presented four times, totaling 1024 trials.

The experimental session was preceded by the practice session, in which participants were familiarized with the task. To proceed to the experimental session, participants had to score above 80% correct for both T1 and unmasked, long lag T2. One participant was excluded for failing to achieve this. The training session was also used to determine for each participant the duration of the short lag (200 or 300 ms T1–T2 interval) that induced the largest AB (lowest T2 accuracy) and that was used in the subsequent experimental session. Two participants were excluded due to their AB size falling below the predetermined criterion. Specifically, their T2 accuracy at both short lags did not exhibit a decrease of more than 5% compared to long lags.

One of the main goals of this study was to match perceptual performance between the perceptual and the attentional manipulation. We did this in two ways. First, during the training session, the matching was done by staircasing mask contrast using the weighted up-down method (Kaernbach, 1991). Contrast levels ranged from 0 (black) to 255 (white). Mask contrast started at level 220. Each correct response made the task more difficult: masks got darker by downward step size S_{down}. Each incorrect response made the task easier: masks got lighter by upward step size S_{up}. Step sizes were determined by $S_{up} \times p = S_{down} \times (1 - p)$, where p is the accuracy at short lags. The smallest step size was always nine contrast levels. A reversal is making a mistake after a correct response, or vice versa. The staircase ended after 25 reversals. The mask contrast with which the experimental session started was the average contrast level of the last 20 reversals. Second, during the experimental session, after every 32 masked trials, mask contrast was updated in accordance with our goal to match performance over participants, while also matching performance within participants as well as possible.

To ensure that confidence ratings for these matched conditions (masked, long lag and unmasked, short lag) were not contaminated by differences in perceptual performance, one type of block only contained the matched conditions, while the other block type contained the two remaining, unmatched conditions (masked, short lag and unmasked, long lag). To ensure every confidence rating would have enough trials for creating receiver operating characteristic curves, participants were instructed to distribute their responses evenly over all ratings within a block. Participants received feedback about the distribution of their responses. The mask contrasts from a performance matched block were used in the subsequent non-performance matched block to ensure that masking remained orthogonal to the AB manipulation. The experimental session therefore always started with a performance matched block.

We wanted to compare the marker for recurrent processing, illusory triangle decoding, to the marker for lateral processing, non-illusory triangle decoding. However, during the experimental session, the non-illusory triangle was never task-relevant, only the illusory one

was. During the independent EEG classification training session, we therefore made each visual feature, one after the other, task-relevant. A target was presented for 33 ms every 900-1100 ms (**Fig. S4**). Participants had to fixate on the fixation cross and indicate whether the current task-relevant feature was absent or present. For each feature, each target was presented 64 times, totaling 512 trials. The order of the task-relevant features was counterbalanced over participants. For all sessions, response button mapping was counterbalanced within tasks.

Behavioral analysis

To quantify perceptual performance, we constructed receiver operating characteristic (ROC) curves by plotting objective hit rates against objective false alarm rates. We used the six response options to get five inflection points (Green & Swets, 1966). We also quantified metacognitive sensitivity: the ability to know whether you were right or wrong. Performance is high when you are confident in objectively correct responses and not confident in objectively incorrect responses. We again constructed ROC curves, now by plotting the rate of high-confidence correct responses (subjective hit rates) against the rate of high-confidence incorrect responses (subjective false alarm rates). We used the three confidence ratings to get two inflection points. To ensure that T1 was attended, trials with incorrect T1 responses were excluded. Repeated measures ANOVAs and Bayesian t-tests were used to test the differences between experimental conditions.

EEG recording and preprocessing

EEG was recorded at 1024 Hz using a 64 channel ActiveTwo system (BioSemi). Four electrooculographic (EOG) electrodes measured horizontal and vertical eye movements. The data were analyzed with MATLAB (MathWorks). For most of the preprocessing steps, EEGLAB was used (Delorme & Makeig, 2004). The data were re-referenced to the earlobes. Poor channels were interpolated. High-pass filtering can cause artifacts in decoding analyses; we therefore removed slow drifts using trial-masked robust detrending (van Driel et al., 2021). Each target was epoched from -250 to 1000 ms relative to target onset. To improve the results from the independent component analysis (ICA), baseline correction was applied using the whole epoch as baseline (Groppe et al., 2009). ICA was used to remove blinks. Blink components were removed manually. Baseline correction was applied, now using a -250 to 0 ms window relative to target onset. Trials with values outside of a -300 to 300 microvolts range were removed. We used an adapted version of FieldTrip's ft_artifact_zvalue function to detect and remove trials with muscle artifacts (Oostenveld et al., 2011). As in the behavioral analyses, trials with incorrect T1 responses were excluded. Finally, the data were downsampled to 128 Hz.

Multivariate pattern analyses

To establish the markers for the different neural processes of interest (feedforward, lateral, and recurrent), the processing of the different visual features (local contrast, non-illusory triangle, and illusory Kanizsa triangle, respectively) were decoded (**Fig. 1B**) using the Amsterdam Decoding and Modeling (ADAM) toolbox (Fahrenfort et al., 2018). For each participant and each visual feature, a linear discriminant classifier was trained on the T1 data and tested on each

condition of the T2 data. The classifier was trained to discriminate between the feature's (e.g., the illusory triangle's) absence and presence based on the raw EEG activity across all electrodes. AUC was again used as the performance measure. This procedure was executed for every time sample in a trial, yielding classification performance over time. For the time samples from -100 to 700 ms relative to target onset, we used a two-sided t-test to evaluate whether classifier performance differed from chance. We used cluster-based permutation testing (1000 iterations at a threshold of 0.05) to correct for multiple comparisons (Maris & Oostenveld, 2007). To obtain topographic maps showing the neural sources of the classifier performance, we multiplied the classifier weights with the data covariance matrix, yielding covariance/class separability maps (Haufe et al., 2014).

In the decoding analyses described in the results, we applied "diagonal decoding": classifiers were tested on the same time sample they were trained on. We did the same analyses again, now by applying "off-diagonal decoding": classifiers trained on a particular time point are tested on all time points (King & Dehaene, 2014). Off-diagonal decoding allowed us to investigate whether patterns of activity during the time windows of interest were stable over time (**Fig. S3**). For the illusory triangle, classifiers were trained on the 200-250 ms window and then averaged. The same was done for the local contrast 75-95 ms window.

To distinguish between lateral and feedback connections in local (early and occipital) recurrent processing, we trained classifiers on independent data based on collinearity-only (the non-illusory triangle was task-relevant) or collinearity-plus-illusion (the illusory triangle was task-relevant) and then decoded the Kanizsa illusion in T2s of the main RSVP task. The rationale for this analysis is that collinearity is present both when the Pac-Man stimuli align to form the illusory Kanizsa triangle and when the two-legged white circles align to from a non-illusory triangle, but only in the case of the Kanizsa triangle do participants experience an illusion. The comparison of T2 illusion decoding between the classifiers trained on the illusion and on collinearity-only in the training set may therefore isolate illusion-specific feedback processing from basic collinearity processing involving lateral connections (**Fig. 3**). In the **Supplementary information (Fig. S5)**, we compare the independent training set to the training set used for the main analyses, the T1 data from the RSVP task.

As described in the **Supplementary information** as well, a tenfold cross-validation scheme was applied to the data from the independent training set to decode local contrast. Individual participants' data were split into ten equal-sized folds after randomizing the task's trial order. A classifier was then trained on nine folds and tested on the tenth one, ensuring independence of the training and testing sets. This procedure was repeated until each fold served as the test set once. Classifier performance, AUC, was averaged across all ten iterations (**Fig. S7**).

As in the behavioral analyses, repeated measures ANOVAs and Bayesian t-tests were used to test the differences between experimental conditions.

References

Alilović, J., Lampers, E., Slagter, H. A., & van Gaal, S. (2023). Illusory object recognition is either perceptual or cognitive in origin depending on decision confidence. *PLoS Biology*, *21*(3), e3002009.

Angelucci, A., & Bressloff, P. C. (2006). Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. *Progress in Brain Research*, *154*, 93–120.

Block, N. (2005). Two neural correlates of consciousness. *Trends in Cognitive Sciences*, *9*(2), 46–52.

Bosking, W. H., Zhang, Y., Schofield, B., & Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *17*(6), 2112–2127.

Brockhoff, L., Schindler, S., Bruchmann, M., & Straube, T. (2022). Effects of perceptual and working memory load on brain responses to task-irrelevant stimuli: Review and implications for future research. *Neuroscience and Biobehavioral Reviews*, *135*, 104580.

Brown, R., Lau, H., & LeDoux, J. E. (2019). Understanding the Higher-Order Approach to Consciousness. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2019.06.009

Canales-Johnson, A., Beerendonk, L., Chennu, S., Davidson, M. J., Ince, R. A. A., & van Gaal, S. (2023). Feedback information transfer in the human brain reflects bistable perception in the absence of report. *PLoS Biology*, *21*(5), e3002120.

Cogitate Consortium, Ferrante, O., Gorska-Klimowska, U., Henin, S., Hirschhorn, R., Khalaf, A., Lepauvre, A., Liu, L., Richter, D., Vidal, Y., Bonacchi, N., Brown, T., Sripad, P., Armendariz, M., Bendtz, K., Ghafari, T., Hetenyi, D., Jeschke, J., Kozma, C., … Melloni, L. (2023). An adversarial collaboration to critically evaluate theories of consciousness. In *bioRxiv* (p. 2023.06.23.546249). https://doi.org/10.1101/2023.06.23.546249

678 Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious,
679 preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive*
680 *Sciences*, *10*(5), 204–211.
681 Dehaene, S., Sergent, C., & Changeux, J.-P. (2003). A neuronal network model linking
682 subjective reports and objective physiological data during conscious perception.
683 *Proceedings of the National Academy of Sciences of the United States of America*,
684 *100*(14), 8520–8525.
685 Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial
686 EEG dynamics including independent component analysis. *Journal of Neuroscience*
687 *Methods*, *134*(1), 9–21.
688 Dienes, Z. (2007). Subjective measures of unconscious knowledge. In R. Banerjee & B. K.
689 Chakrabarti (Eds.), *Progress in Brain Research* (Vol. 168, pp. 49–269). Elsevier.
690 Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2007). Masking disrupts reentrant
691 processing in human visual cortex. *Journal of Cognitive Neuroscience*, *19*(9), 1488–
692 1497.
693 Fahrenfort, J. J., van Driel, J., van Gaal, S., & Olivers, C. N. L. (2018). From ERPs to MVPA
694 Using the Amsterdam Decoding and Modeling Toolbox (ADAM). *Frontiers in*
695 *Neuroscience*, *12*, 368.
696 Fahrenfort, J. J., van Leeuwen, J., Olivers, C. N. L., & Hogendoorn, H. (2017). Perceptual
697 integration without conscious access. *Proceedings of the National Academy of Sciences*
698 *of the United States of America*, *114*(14), 3744–3749.
699 Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human*
700 *Neuroscience*, *8*, 443.
701 Gilbert, C. D., & Wiesel, T. N. (1979). Morphology and intracortical projections of functionally
702 characterised neurones in the cat visual cortex. *Nature*, *280*(5718), 120–125.
703 Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. *455*.

https://psycnet.apa.org/fulltext/1967-02286-000.pdf

Groppe, D. M., Makeig, S., & Kutas, M. (2009). Identifying reliable independent components via split-half comparisons. *NeuroImage*, *45*(4), 1199–1211.

Halgren, E., Mendola, J., Chong, C. D. R., & Dale, A. M. (2003). Cortical activation to illusory shapes as measured with magnetoencephalography. *NeuroImage*, *18*(4), 1001–1009.

Hatamimajoumerd, E., Ratan Murty, N. A., Pitts, M., & Cohen, M. A. (2022). Decoding perceptual awareness across the brain with a no-report fMRI masking paradigm. *Current Biology: CB*. https://doi.org/10.1016/j.cub.2022.07.068

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, *87*, 96–110.

Jiang, J., Correa, C. M., Geerts, J., & van Gaal, S. (2018). The relationship between conflict awareness and behavioral and oscillatory signatures of immediate and delayed cognitive control. *NeuroImage*, *177*, 11–19.

Joglekar, M. R., Mejias, J. F., Yang, G. R., & Wang, X.-J. (2018). Inter-areal Balanced Amplification Enhances Signal Propagation in a Large-Scale Circuit Model of the Primate Cortex. *Neuron*, *98*(1), 222-234.e8.

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, *49*(3), 227–229.

Kanai, R., Walsh, V., & Tseng, C.-H. (2010). Subjective discriminability of invisibility: a framework for distinguishing perceptual and attentional failures of awareness. *Consciousness and Cognition*, *19*(4), 1045–1057.

Kandel, E. R., Schwartz, J. H., Jessell, T. M., & Siegelbaum, S. (2000). *Principles of neural science*. academia.edu. https://www.academia.edu/download/30536508/neuroscience_syllabus.pdf

Kapadia, M. K., Ito, M., Gilbert, C. D., & Westheimer, G. (1995). Improvement in visual

sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. *Neuron*, *15*(4), 843–856.

Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (2000). Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *Journal of Neurophysiology*, *84*(4), 2048–2062.

King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210.

Kok, P., Bains, L. J., van Mourik, T., Norris, D. G., & de Lange, F. P. (2016). Selective Activation of the Deep Layers of the Human Primary Visual Cortex by Top-Down Feedback. *Current Biology: CB*, *26*(3), 371–376.

Kok, P., & de Lange, F. P. (2014). Shape perception simultaneously up- and downregulates neural activity in the primary visual cortex. *Current Biology: CB*, *24*(13), 1531–1535.

Lamme, V. A. F. (2010). How neuroscience will change our view on consciousness. *Cognitive Neuroscience*, *1*(3), 204–220.

Lamme, V. A. F. (2020). Visual Functions Generating Conscious Seeing. *Frontiers in Psychology*, *11*, 83.

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*(11), 571–579.

Lamme, V. A. F., Supèr, H., & Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Current Opinion in Neurobiology*, *8*(4), 529–535.

Lau, H. C. (2022). *In Consciousness we Trust: The Cognitive Neuroscience of Subjective Experience*. Oxford University Press.

Lau, H. C., & Passingham, R. E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(49), 18763–18768.

Lavie, N., & Dalton, P. (2014). Load theory of attention and cognitive control. *The Oxford*

Handbook of Attention, 56–75.

Lavie, N., & de Fockert, J. W. (2003). Contrasting effects of sensory limits and capacity limits in visual selective attention. *Perception & Psychophysics*, *65*(2), 202–212.

Lee, T. S., & Nguyen, M. (2001). Dynamics of subjective contour formation in the early visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(4), 1907–1911.

Li, Z. (1998). A neural model of contour integration in the primary visual cortex. *Neural Computation*, *10*(4), 903–940.

Liang, H., Gong, X., Chen, M., Yan, Y., Li, W., & Gilbert, C. D. (2017). Interactions between feedback and lateral connections in the primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(32), 8637–8642.

Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature*, *383*(6601), 616–618.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190.

Marti, S., King, J.-R., & Dehaene, S. (2015). Time-Resolved Decoding of Two Processing Chains during Dual-Task Interference. *Neuron*, *88*(6), 1297–1307.

Mashour, G. A., Roelfsema, P., Changeux, J.-P., & Dehaene, S. (2020). Conscious Processing and the Global Neuronal Workspace Hypothesis. *Neuron*, *105*(5), 776–798.

Merikle, P. M., Smilek, D., & Eastwood, J. D. (2001). Perception without awareness: perspectives from cognitive psychology. *Cognition*, *79*(1–2), 115–134.

Meuwese, J. D. I., van Loon, A. M., Lamme, V. A. F., & Fahrenfort, J. J. (2014). The subjective experience of object recognition: comparing metacognition for object detection and object categorization. *Attention, Perception & Psychophysics*, *76*(4), 1057–1068.

Mudrik, L., Faivre, N., & Koch, C. (2014). Information integration without awareness. *Trends in Cognitive Sciences*, *18*(9), 488–496.

782 Northoff, G., & Lamme, V. A. F. (2020). Neural signs and mechanisms of consciousness: Is
783 there a potential convergence of theories of consciousness in sight? *Neuroscience and*
784 *Biobehavioral Reviews*, *118*, 568–587.
785 Nuiten, S. A., Canales-Johnson, A., Beerendonk, L., Nanuashvili, N., Fahrenfort, J. J.,
786 Bekinschtein, T., & van Gaal, S. (2021). Preserved sensory processing but hampered
787 conflict detection when stimulus input is task-irrelevant. *ELife*, *10*.
788 https://doi.org/10.7554/eLife.64431
789 Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software
790 for advanced analysis of MEG, EEG, and invasive electrophysiological data.
791 *Computational Intelligence and Neuroscience*, *2011*, 156869.
792 Pak, A., Ryu, E., Li, C., & Chubykin, A. A. (2020). Top-Down Feedback Controls the Cortical
793 Representation of Illusory Contours in Mouse Primary Visual Cortex. *The Journal of*
794 *Neuroscience: The Official Journal of the Society for Neuroscience*, *40*(3), 648–660.
795 Peters, M. A., & Lau, H. (2016). Correction: Human observers have optimal introspective access
796 to perceptual processes even for visually masked stimuli. *ELife*, *5*.
797 https://doi.org/10.7554/eLife.16332
798 Pitts, M. A., Lutsyshyna, L. A., & Hillyard, S. A. (2018). The relationship between attention and
799 consciousness: an expanded taxonomy and implications for 'no-report' paradigms.
800 *Philosophical Transactions of the Royal Society of London. Series B, Biological*
801 *Sciences*, *373*(1755), 20170348.
802 Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual
803 processing in an RSVP task: an attentional blink? *Journal of Experimental Psychology.*
804 *Human Perception and Performance*, *18*(3), 849–860.
805 Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the
806 medial frontal cortex in cognitive control. *Science*, *306*(5695), 443–447.
807 Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annual Review of*

Neuroscience, *29*, 203–227.

Roelfsema, P. R. (2023). Solving the binding problem: Assemblies form when neurons enhance their firing rate-they don't need to oscillate or synchronize. *Neuron*, *111*(7), 1003–1019.

Schmidt, K. E., Goebel, R., Löwel, S., & Singer, W. (1997). The perceptual grouping criterion of colinearity is reflected by anisotropies of connections in the primary visual cortex. *The European Journal of Neuroscience*, *9*(5), 1083–1089.

Schmidt, T. (2015). Invisible Stimuli, Implicit Thresholds: Why Invisibility Judgments Cannot be Interpreted in Isolation. *Advances in Cognitive Psychology / University of Finance and Management in Warsaw*, *11*(2), 31–41.

Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience*, *8*(10), 1391–1400.

Sergent, C., Corazzol, M., Labouret, G., Stockart, F., Wexler, M., King, J.-R., Meyniel, F., & Pressnitzer, D. (2021). Bifurcation in brain dynamics reveals a signature of conscious processing independent of report. *Nature Communications*, *12*(1), 1149.

Sergent, C., & Dehaene, S. (2004). Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blink. *Psychological Science*, *15*(11), 720–728.

Seth, A. K., Dienes, Z., Cleeremans, A., Overgaard, M., & Pessoa, L. (2008). Measuring consciousness: relating behavioural and neurophysiological approaches. *Trends in Cognitive Sciences*, *12*(8), 314–321.

Shanks, D. R. (2017). Regressive research: The pitfalls of post hoc data selection in the study of unconscious mental processes. *Psychonomic Bulletin & Review*, *24*(3), 752–775.

Shapiro, K. L., Raymond, J. E., & Arnell, K. M. (1997). The attentional blink. *Trends in Cognitive Sciences*, *1*(8), 291–296.

Stein, T., Kaiser, D., Fahrenfort, J. J., & van Gaal, S. (2021). The human visual system differentially represents subjectively and objectively invisible stimuli. *PLoS Biology*,

19(5), e3001241.

Stettler, D. D., Das, A., Bennett, J., & Gilbert, C. D. (2002). Lateral connectivity and contextual interactions in macaque primary visual cortex. *Neuron*, *36*(4), 739–750.

Thaler, L., Schütz, A. C., Goodale, M. A., & Gegenfurtner, K. R. (2013). What is the best fixation target? The effect of target shape on stability of fixational eye movements. *Vision Research*, *76*, 31–42.

Tsuchiya, N., Wilke, M., Frässle, S., & Lamme, V. A. F. (2015). No-Report Paradigms: Extracting the True Neural Correlates of Consciousness. *Trends in Cognitive Sciences*, *19*(12), 757–770.

van Driel, J., Olivers, C. N. L., & Fahrenfort, J. J. (2021). High-pass filtering artifacts in multivariate classification of neural time series data. *Journal of Neuroscience Methods*, *352*, 109080.

van Gaal, S., Ridderinkhof, K. R., Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2008). Frontal cortex mediates unconsciously triggered inhibitory control. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *28*(32), 8053–8062.

van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., & Roelfsema, P. R. (2018). The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science*, *360*(6388), 537–542.

Vandenbroucke, A. R. E., Fahrenfort, J. J., Sligte, I. G., & Lamme, V. A. F. (2014). Seeing without knowing: neural signatures of perceptual inference in the absence of report. *Journal of Cognitive Neuroscience*, *26*(5), 955–969.

Weaver, M. D., Fahrenfort, J. J., Belopolsky, A., & van Gaal, S. (2019). Independent Neural Activity Patterns for Sensory- and Confidence-Based Information Maintenance during Category-Selective Visual Processing. *ENeuro*, *6*(1). https://doi.org/10.1523/ENEURO.0268-18.2018

Wokke, M. E., Vandenbroucke, A. R. E., Scholte, H. S., & Lamme, V. A. F. (2013). Confuse

your illusion: feedback to early visual cortex contributes to perceptual completion. *Psychological Science*, *24*(1), 63–71.

Wyatte, D., Jilk, D. J., & O'Reilly, R. C. (2014). Early recurrent feedback facilitates visual object recognition under challenging conditions. *Frontiers in Psychology*, *5*, 674.

Yi, D.-J., Woodman, G. F., Widders, D., Marois, R., & Chun, M. M. (2004). Neural fate of ignored stimuli: dissociable effects of perceptual and working memory load. *Nature Neuroscience*, *7*(9), 992–996.

Zivony, A., & Lamy, D. (2022). What processes are disrupted during the attentional blink? An integrative review of event-related potential research. *Psychonomic Bulletin & Review*, *29*(2), 394–414.