eNeuro

Research Article: New Research | Cognition and Behavior

No evidence for neural overlap between unconsciously processed and imagined stimuli

https://doi.org/10.1523/ENEURO.0228-21.2021

Cite as: eNeuro 2021; 10.1523/ENEURO.0228-21.2021

Received: 20 May 2021
Revised: 1 September 2021
Accepted: 2 September 2021

1. Manuscript Title

No evidence for neural overlap between unconsciously processed and imagined stimuli

2. Abbreviated Title

ImaginationUnconscious

3. List all Author Names and Affiliations in order as they would appear in the published article

Nadine Dijkstra[1,2], Simon van Gaal[3], Linda Geerligs[1], Sander E. Bosch[1] & Marcel A.J. van Gerven[1]

1. Donders Institute for Brain, Cognition & Behaviour, Radboud University, Nijmegen, the Netherlands
2. Wellcome Centre for Human Neuroimaging, University College London, United Kingdom
3. Department of Psychology, Brain & Cognition, University of Amsterdam, the Netherlands

4. Author Contributions: Each author must be identified with at least one of the following:

ND, SB, SvG, MvG Designed research, ND, SB Performed research, ND, LG Analyzed data, ND Wrote the paper.

5. Correspondence should be addressed to (include email address)

Nadine Dijkstra (n.dijkstra@ucl.ac.uk)

6. Number of Figures 5

7. Number of Tables 2

8. Number of Multimedia 0

9. Number of words for Abstract 198

10. Number of words for Significance Statement 66

11. Number of words for Introduction 739

12. Number of words for Discussion 1798

13. Acknowledgements

N. Dijkstra was supported by a Rubicon grant 019.192SG.003, M.A.J. van Gerven & S.E. Bosch were supported by VIDI grant 639.072.513 and L. Geerligs was supported by VENI grant 451.1.013, all from the Netherlands Organization for Scientific Research.

14. Conflict of Interest A. No (State 'Authors report no conflict of interest')

15. Funding sources

Netherlands Organization for Scientific Research

Abstract. Visual representations can be generated via feedforward or feedback processes. The extent to which these processes result in overlapping representations remains unclear. Previous work has shown that imagined stimuli elicit similar representations as perceived stimuli throughout the visual cortex. However, while representations during imagery are indeed only caused by feedback processing, neural processing during perception is an interplay of both feedforward and feedback processing. This means that any representational overlap could be due to overlap in feedback processes. In the current study we aimed to investigate this issue by characterizing the overlap between feedforward- and feedback-initiated category-representations during imagery, conscious perception and unconscious processing using fMRI in humans of either sex. While all three conditions elicited stimulus representations in left lateral occipital cortex (LOC), significant similarities were only observed between imagery and conscious perception in this area. Furthermore, connectivity analyses revealed stronger connectivity between frontal areas and left LOC during conscious perception and imagery compared to unconscious processing. Together, these findings can be explained by the idea that long-range feedback modifies visual representations, thereby reducing representational overlap between purely feedforward and feedback-initiated stimulus representations measured by fMRI. Neural representations influenced by feedback, either stimulus-driven (perception) or purely internally-driven (imagery), are however relatively similar.

Significance statement: Previous research has shown substantial neural overlap between imagery and perception, suggesting overlap between bottom-up and top-down processes. However, because conscious perception also involves top-down processing, this overlap could instead reflect similarity in feedback processes. In this study, we showed that the overlap between perception and imagery disappears when stimuli are rendered unconscious via backward masking, suggesting reduced overlap between purely bottom-up and top-down generated representations.

Introduction. Visual experience relies on neural representations in visual cortex, which can be activated in two different ways. Externally, by light bouncing off of objects and hitting the retina, from which signals are sent via feedforward connections to early visual cortex and areas further up in the visual hierarchy (e.g. lateral occipital cortex). Or internally, via

feedback signals from high-level brain areas, such as areas in prefrontal cortex, for example during mental imagery and dreaming (Dentico et al., 2014; Dijkstra, Zeidman, Ondobaka, van Gerven, & Friston, 2017; Mechelli, Price, Friston, & Ishai, 2004). It remains unclear to what extent activation patterns in visual cortex caused by feedforward and feedback signals are similar.

Previous work has compared neural representations during perception and imagery, revealing convincing evidence that there is neural representational overlap between perception and imagery throughout large parts of visual cortex (Albers, Kok, Toni, Dijkerman, & de Lange, 2013; Cichy et al., 2012; Dijkstra, Bosch, & van Gerven, 2017; Horikawa & Kamitani, 2017; Johnson & Johnson, 2014; Lee, Kravitz, & Baker, 2012; O'Craven & Kanwisher, 2000; Reddy, Tsuchiya, & Serre, 2010; Stokes, Thompson, Cusack, & Duncan, 2010; Thirion et al., 2006). The strongest overlap between perception and imagery is typically observed in high-level visual areas (Lee et al., 2012; Reddy et al., 2010b; Stokes et al., 2010), whereas the overlap in low-level areas seems to depend on the required detail of the imagery task (Kosslyn & Thompson, 2003) and the experienced imagery vividness (Albers et al., 2013; Dijkstra et al., 2017; Lee et al., 2012).

However, while activation in visual cortex during mental imagery indeed only relies on feedback signals (Dijkstra et al., 2017; Dijkstra et al., 2020; Mechelli et al., 2004), visual activation during perception reflects an interplay between feedforward and feedback processes (Bastos et al., 2012; Bastos et al., 2015; Dijkstra et al., 2017; Dijkstra et al., 2020; Muckli, 2010; Lamme & Roelfsema 2000). To determine whether visual representations activated by feedforward and feedback signals do indeed activate similar neural populations, one needs to investigate a situation in which visual representations are caused by feedforward signals only and compare those to events that include feedback processing as well.

Backward masking has been hypothesized to disrupt feedback from high-level visual cortex to early visual cortex (Del Cul, Baillet, & Dehaene, 2007; Fahrenfort, Scholte, & Lamme, 2007; Lamme, Zipser, & Spekreijse, 2002; Roelfsema, Lamme, Spekreijse, & Bosch, 2002; van Gaal & Lamme, 2012). In a backward masking paradigm, a briefly presented target stimulus is rapidly followed by a second, masking stimulus. Appropriate backward masking renders the target stimulus invisible. Several studies have shown that masking leaves the feedforward sweep relatively unaffected, which can still cause activation in high-level visual

cortex (Jiang & He, 2006; Sterzer, Haynes, & Rees, 2008), while feedback processing is disrupted (Fahrenfort, Scholte, & Lamme, 2007; Lamme et al., 2002; Mashour, Roelfsema, Changeux, & Dehaene, 2020; van Gaal & Lamme, 2012). These and other observations have led to the idea that the feedforward sweep is unconscious and that recurrent processing is an important factor in achieving conscious awareness (Lamme, 2015; Mashour et al., 2020; Tononi, 2008). However, the exact relationship between feedback processing and conscious awareness is still debated (see e.g. Boly et al., 2017).

In the current study we investigated to what extent visual representations in visual cortex are modified by feedback, by comparing conditions in which stimuli are consciously perceived, not consciously perceived and imagined. We rely on the assumption that unconscious processing contains less or no feedback processing, and that therefore comparing unconscious to conscious and imagined representations will provide insight into the effects of feedback processing. However, it is important to note that this is an assumption based on previous research which will not be tested in the current study. Therefore, the exact implications of our results need to be inferred with caution. More elaborate and nuanced interpretations will be given in the discussion. We quantified the representational overlap between the different conditions by training a classifier on one condition and testing it on another condition ('cross-decoding'; Albers et al., 2013; Dijkstra et al., 2018; Lee et al., 2012). The only difference between the conscious and unconscious condition was the stimulus onset asynchrony (SOA) between the target and the mask. To cue visual imagery in a way that does not induce an informative cue-signal that can be picked up by a classifier, we used a retro-cue paradigm (Harrison & Tong, 2009; see Fig. 1B).

Materials and Methods

Participants. Thirty-seven participants with normal or corrected-to-normal vision gave written informed consent and participated in the study. All participants were naïve to the aim of the experiment and most participants were familiar with similar visual perception fMRI studies. Two participants were excluded from the final analyses: one because they quit the experiment prematurely and one because they had misunderstood the task. Due to an accidental change in the refresh rate of the monitor (from 60 Hz to 75 Hz) the timing was slightly different for for 6/35 participants: presentation from 17ms to 13ms, ISI conscious from 66ms to 80ms, so that the presentation times were slightly shorter for the unconscious

condition and slightly longer for the conscious condition. Because this error did not change visibility ratings (unconscious: 1.37 (SD = 0.27) versus 1.35 (SD = 0.58); $t(33) = 0.079$, $p = 0.94$ – conscious: 2.92 (SD = 0.37) vs 2.98 (SD = 0.61); $t(33) = -0.25$, $p = 0.80$) or discrimination sensitivity (unconscious: 0.19 (SD = 0.28) versus 0.03 (SD = 0.18); $t(33) = 1.9$, $p = 0.07$ – conscious: 3.33 (SD = 0.61) vs 3.82 (SD = 0.90); $t(33) = -1.26$, $p = 0.22$) we decided not to remove these participants. Thirty-five participants were included in the main analyses (mean age 25.9, *SD* = 5.9).

Experimental design. Prior to the experiment, participants filled out the Vividness of Visual Imagery Questionnaire 2 (VVIQ2; Marks, 1973), which is a 16-item questionnaire that measures the general vividness of a participant's imagery. The experiment consisted of two tasks, a perception and an imagery task, which were executed in interleaved blocks and whether participants started with the imagery or perception task was counterbalanced over participants. The perception task consisted of conscious and unconscious trials, which only differed in ISI between the stimulus and the mask: 0ms for the unconscious trials and 66ms (4 frames) for the conscious condition. We chose to operationalize conscious versus unconscious processing via experimental manipulation (strong versus weak masking) and not via post-hoc trial selection based on visibility reports, because this latter approach has been shown to violate statistical assumptions and may lead to spurious unconscious effects (for more details, see Shanks, 2017). During the perception task, a stimulus was presented very briefly (17ms), followed by a backward mask. Participants subsequently indicated whether the presented stimulus was animate or inanimate and rated the visibility of the stimulus on a scale from 1 (not visible at all) to 4 (perfectly clear; Fig. 1A). To prevent motor preparation, the response mapping for both the animacy and visibility ratings were randomized over trials. During the imagery task, two stimuli were each successively presented for 500ms, followed by a retro-cue indicating which of the two the participant should imagine. The participant then imagined the cued stimulus and subsequently indicated the animacy and the visibility of the imagined stimulus (Fig. 1B). The button-response mapping for the animacy task and the visibility rating was randomized over trials to prevent motor preparation.

There were 184 conscious and 184 unconscious trials, 46 repetitions per stimulus, divided over 4 blocks. Each conscious-unconscious block lasted approximately 9 minutes.

There were 144 imagery trials, 36 repetitions per stimulus, divided over 4 blocks. Each imagery block lasted approximately 7 minutes. The order of the different stimuli and SOAs (unconscious versus conscious trials) within the perception task and the stimuli and retro-cue combinations during imagery was fully counterbalanced within participants and which task (imagery or perception) was executed first, was randomized between participants. In total, there were 8 blocks, leading to an experimental time of approximately 65 minutes per participant. Including breaks and an anatomical scan, this added up to 90 minutes of fMRI scanning time.

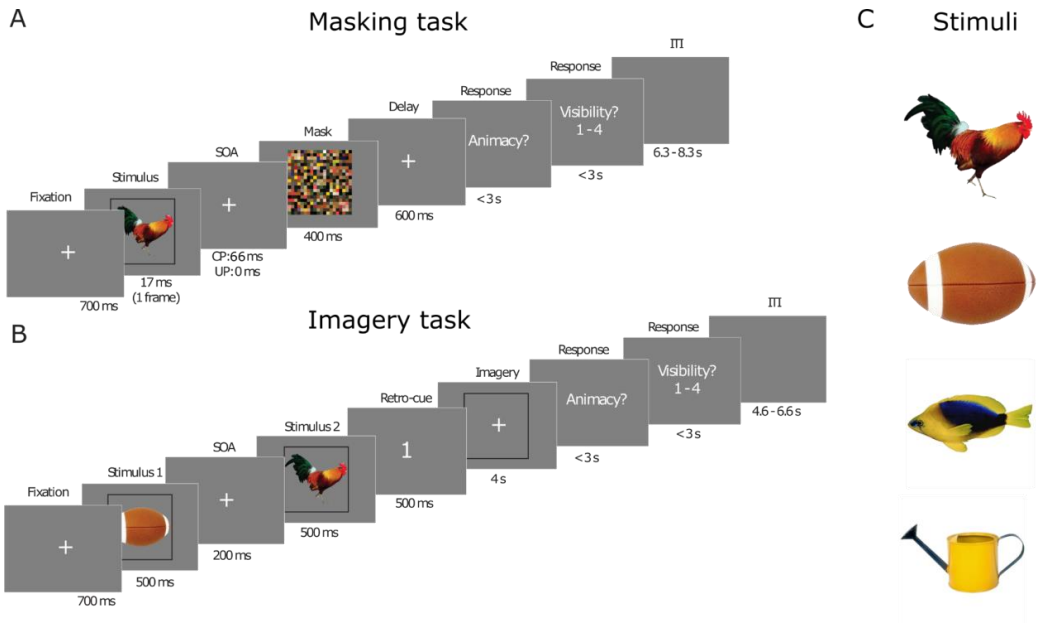

Figure 1. Experimental paradigm. (A) Masking task. A stimulus is presented for 17ms, followed by a mask (duration 400ms) after 0ms (unconscious condition) or 66ms (conscious condition). Participants had to indicate whether the stimulus was animate or inanimate and rate the visibility. (B) Visual imagery task. Participants were presented with two stimuli after each other followed by a cue indicating whether to imagine the first or the second stimulus, as vividly as possible. After the imagery, participants had to indicate whether the imagined stimulus was animate or inanimate and rate the visibility of their imagery. (C) Stimuli used: a rooster, a football, a fish and a watering can from the POPORO stimulus data set (Kovalenko, Chaumon, & Busch, 2012). The neural analyses focused on pairwise comparisons between all possible combinations of stimuli.

Stimuli. We used stimuli from the POPORO stimulus data set (Kovalenko, Chaumon, & Busch, 2012), which contains colour images of everyday objects and animals. From these stimuli we selected four (two animate and two inanimate) for the final study. The stimuli were selected based on (a) familiarity and visual difference, such as to maximise classification performance and on (b) accuracy and visibility scores calculated in a pilot

experiment. The stimuli were presented at 50% contrast on a grey background screen. They encompassed a 4 by 4 cm square which corresponded to a visual angle of 2.81 degrees. The stimuli were relatively small to prevent large eye-movements, which would affect our fMRI analyses. The mask was created by randomly scrambling the pixel values of all stimuli taken together and was also 4 by 4 cm in order to fully mask the presented stimuli.

Behavioural analysis. To characterize performance on the discrimination animacy task we calculated d' as the distance between the signal and the signal plus noise, calculated as the difference between the hit-rate and the false alarm rate (Macmillan & Creelman, 1990). A high d' value indicates better performance and a d' of zero indicates chance-level performance.

fMRI acquisition. Each block was scanned in a separate fMRI run, adding up to 8 runs in total. In between runs, the researcher checked in with the participant and asked whether they needed a break. The experiment continued when the participant said they were ready to continue. fMRI data were recorded on a Siemens 3T Skyra scanner with a Multiband 6 sequence (TR: 1 s; voxel size: 2 x 2 x 2 mm; TE: 34 ms) and a 32-channel head coil. For all participants, the field of view was tilted -25° from the transverse plane, using the Siemens AutoAlign Head software, resulting in the same tilt relative to the individual participant's head position. T1-weighted structural images (MPRAGE; voxel size: 1 x 1 x 1 mm; TR: 2.3 s) were also acquired for each participant.

fMRI pre-processing. Prior to decoding analyses, data were pre-processed using SPM12 (RRID: SCR_007037). All functional imaging data were motion-corrected (realignment) and co-registered to the T1 structural scan. The scans were then normalized to MNI space using DARTEL normalisation and smoothed with a 6 mm Gaussian kernel, which has been shown to improve group-level decoding accuracy (Gardumi et al., 2016; Hendriks, Daniels, Pegado, & Op de Beeck, 2017; Misaki, Luh, & Bandettini, 2013). A high-pass filter of 128s was used to remove slow signal drift.

Multivariate pattern analysis. Multivariate analyses were performed using Matlab version 2018a (RRID: SRC_001622). We used linear discriminant analysis to decode the stimulus

identity per searchlight based on the beta estimates per trial. All individual trial beta estimates were obtained from one general linear model (GLM) which contained a separate regressor for each trial set at the onset of the image (or imagery frame for imagery with a duration of 0 (spike) for the conscious and unconscious conditions and a duration of 4 for the imagery condition (Dijkstra et al., 2017; Bosch et al., 2014). Additional regressors in this GLM were (1) the animacy response screen onsets, duration set to the time until response; (2) animacy response button presses, duration 0 (spike); (3) the visibility response screen onsets; duration set to the until response; (4) visibility response button presses, duration 0 (spike); (5) onset of the first stimulus in the retro-cue task, duration 500ms; (6) onset of the second stimulus in the retro-cue task, duration 500ms and (8) a constant value per run to eliminate run-specific changes in mean signal amplitude. Finally, the average signals from the white matter (WM) and cerebral spinal fluid (CSF) (Caballero-Gaudes & Reynolds, 2017; Lund, Nørgaard, Rostrup, Rowe, & Paulson, 2005) as well as the motion parameters were included as nuisance regressors. Decoding within and across conditions was done pairwise between all combinations of the four stimuli, resulting in six decoding pairs, over which the accuracy was then averaged. Searchlights had a radius of 4 voxels, resulting in 257 voxels per searchlight on average. Searchlights moved through the brain based on the center voxel such that voxels participated in multiple searchlights (Allefeld & Haynes, 2014). Leave-one-run-out cross-validation was performed, such that for each fold, a classifier was trained on three runs and tested on the fourth, left-out run. This was done for all comparisons except for imagery-conscious and imagery-unconscious cross-decoding, because these data already came from different task runs (see Fig. 1). Generalization across conditions is often asymmetric which could be due to a variety of reasons such as differences in signal to noise ratio between the two conditions (van den Hurk & Op de Beeck, 2019). Because we did not have a priori hypotheses about asymmetries in cross-decoding directions and because both directions revealed qualitatively similar results, we average across both cross-decoding directions before doing statistics across subjects.

Psychophysiological interaction analysis. After identifying a visual area that contained stimulus information (significant stimulus decoding) in all three conditions, we performed a psychophysiological interaction (PPI) analysis to investigate differences in connectivity between this area and the rest of the brain between the conditions (Friston et al., 1997). Per

participant, the seed-region was defined as an 8 mm sphere centred on the peak averaged univariate activation over the three conditions, within a 16 mm sphere centred around the voxels in which decoding was significant for all three conditions at the group level (Fig. 3, MNI: -54 -65 -10). This approach ensures that approximately the same region was used for every participant while also taking account differences in structural and functional anatomy between participants. This method and size of region of interest (ROI) definition is based on recommendations in the literature for comparable analyses (Zeidman et al., 2019a,b). One participant was excluded because the t-contrast of the averaged activation over the three conditions versus 0 did not reach the statistical threshold of 0.05 in any of the voxels within the group sphere. Two PPI contrasts were calculated: (Conscious perception & unconscious processing) > imagery (feedforward) and (conscious perception & imagery) > unconscious processing (feedback). Connectivity with significant areas was compared in a post-hoc analysis by calculating the difference in connectivity between each two conditions (Fig. 4C; Friston et al., 1997). Note that the connectivity analyses were not stimulus specific; therefore, the first comparison, where we compare conditions that contained a mask (conscious & unconscious) with conditions that did not contain a mask (imagery), might be driven (partly) by processing of the mask instead of the stimuli preceding the mask.

Statistical analysis. The application of standard second-level statistics, including t-tests, to multivariate pattern analysis (MVPA) measures is in many cases invalid due to violations of assumptions. Therefore, we used permutation testing to generate the empirical null-distribution, thereby circumventing the need to rely on assumptions about this distribution. We followed the approach suggested by (Stelzer, Chen, & Turner, 2013) for searchlight MVPA measurements which uses a combination of permutation testing and bootstrapping to generate chance distributions for group studies. Due to the large computational load of searchlight decoding analysis, per participant, 25 permutation maps were generated by permuting the class labels within each run. Group-level permutation distributions were subsequently generated by bootstrapping over these 25 maps, i.e. randomly selecting one out of 25 maps per participant and then averaging over participants. 10000 bootstrapping samples were used to generate the group null-distribution per voxel and per comparison. *P*-values were calculated per voxel as the right-tailed area of the histogram of permutated accuracies from the mean over participants. We corrected for multiple comparisons using

whole-brain FDR-correction with a q-value cut-off of 0.01. Cluster correction was performed, ensuring that voxels were only identified as significant if they belonged to a cluster of at least 50 significant voxels (Dijkstra, Bosch, & van Gerven, 2017).

Data and code availability. All data will be made publicly available upon publication of this manuscript. Analysis code for this study will be made available via the corresponding author upon request.

Results

Behavioural results. To check whether participants indeed did not consciously perceive the stimuli in the unconscious condition, we tested their perceptual sensitivity and visibility scores. Whereas d' was clearly significantly above zero for both the conscious (M = 3.74, SD = 0.87, $t(34)$ = 25.40, $p < 0.0001$) as well as the imagery (M = 3.32, SD = 0.83, $t(34)$ = 23.74, $p < 0.0001$) condition, this was not the case for the unconscious condition (M = 0.05, SD = 0.20, $t(34)$ = 1.57, p = 0.127; BF01 = 0.549; Fig. 2A). Furthermore, d' was significantly higher for both the conscious condition ($t(34)$ = 23.18, $p < 0.0001$) and the imagery condition ($t(34)$ = 20.60, $p < 0.0001$) compared to the unconscious condition. d' in the conscious condition was also slightly higher than in the imagery condition ($t(34)$ = 2.62, p = 0.013). Furthermore, the visibility ratings for both the conscious condition (M = 3.03, SD = 0.54, $t(34)$ = 10.94, $p < 0.0001$) as well as the imagery condition (M = 2.91, SD = 0.38, $t(34)$ = 11.76, $p < 0.0001$) were much higher than for the unconscious condition (M = 1.37, SD = 0.54; Fig. 2B). A few participants rated a proportion of trials in the unconscious condition as high visibility (Fig. 2B), however, all of these participants still had a discrimination accuracy at chance (all < 53.3%). Furthermore, there was no significant relationship between mean visibility rating and d' in the unconscious condition over participants (r = 0.14, p = 0.41). Given the nonsignificant task performance and the potential confusion caused by the randomization of response mapping between trials, these high visibility ratings during the unconscious condition are unlikely to reflect true conscious visibility. Together, these results suggest that the stimuli were indeed strongly masked and therefore we were able to isolate feedforward processing as much as possible (Fahrenfort et al., 2007).

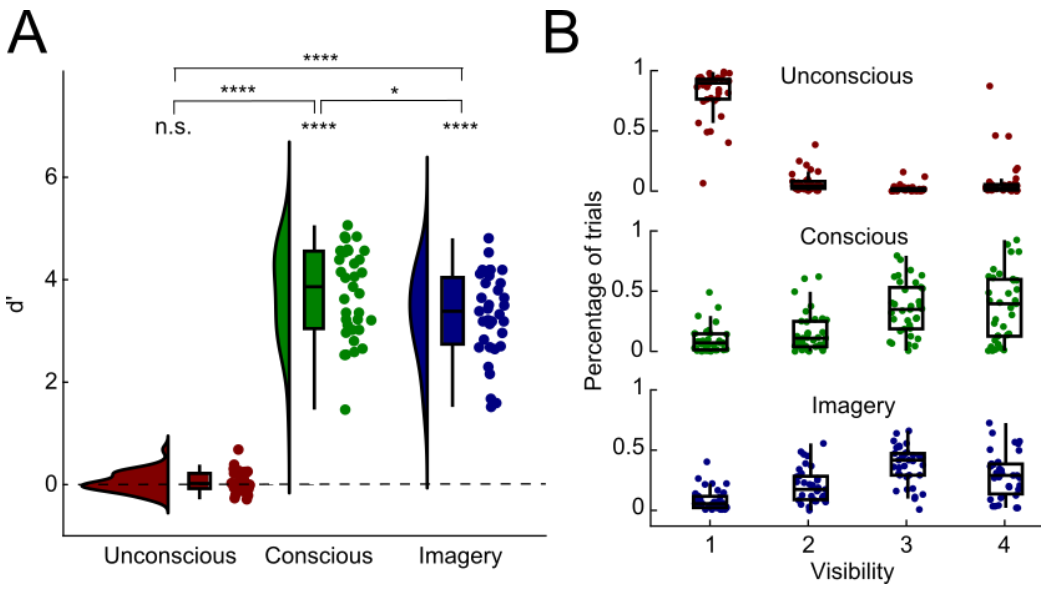

Figure 2. Behavioural results. (A) d' for the animacy task shown separately for each condition. The bell-shaped curves represent the distribution over participants, the boxplots indicate the four quartiles and the dots represent individual participants. d' was significantly higher than zero in the conscious as well as imagery condition, but not in the unconscious condition. P-values: * < 0.05, **** < 0.0001. (B) Percentage of trials of each visibility rating (1-4) separately for the three conditions. Boxplots represent the distributions over participants and dots represent individual participants.

Decoding within conditions. To investigate which areas represented stimulus information during the three conditions, we performed a searchlight decoding analysis separately for each condition (Fig. 3). Statistical tests were performed using group-level permutation testing as described in Stelzer et al. (2013) and corrected for multiple comparisons (see *Methods*). Significant decoding clusters are shown in Figure 3 and listed in Table 1. The cut-off accuracy value for significance was 0.508 for the unconscious and conscious conditions and 0.511 for imagery. The relatively low decoding accuracy of conscious representations (~0.55) compared to other studies (~0.55-0.65) (e.g. Eger et al., 2008; Axelrod & Yovel, 2015) is likely due to the backward mask, which adds noise to the stimulus response. Given the low temporal resolution of fMRI, this means that the BOLD signal at the time of the stimulus will contain a mixture of stimulus response and response to the mask, increasing variance unrelated to the stimulus and thereby decrease decoding performance. In line with previous studies (Dijkstra, Bosch, & van Gerven, 2019; Pearson, Naselaris, Holmes, & Kosslyn, 2015), we could decode stimulus information during conscious perception as well

as imagery in low- and high-level visual areas, intra-parietal sulcus and lateral frontal cortex (Fig. 3B-E). Interestingly, significant decoding of unconscious stimuli was only observed in left high-level visual cortex, temporal pole and lateral frontal cortex (Fig. 3A). There was no significant unconscious decoding in low-level visual areas. All three conditions showed stimulus representations in left lateral occipital cortex (LOC; Fig. 3E).

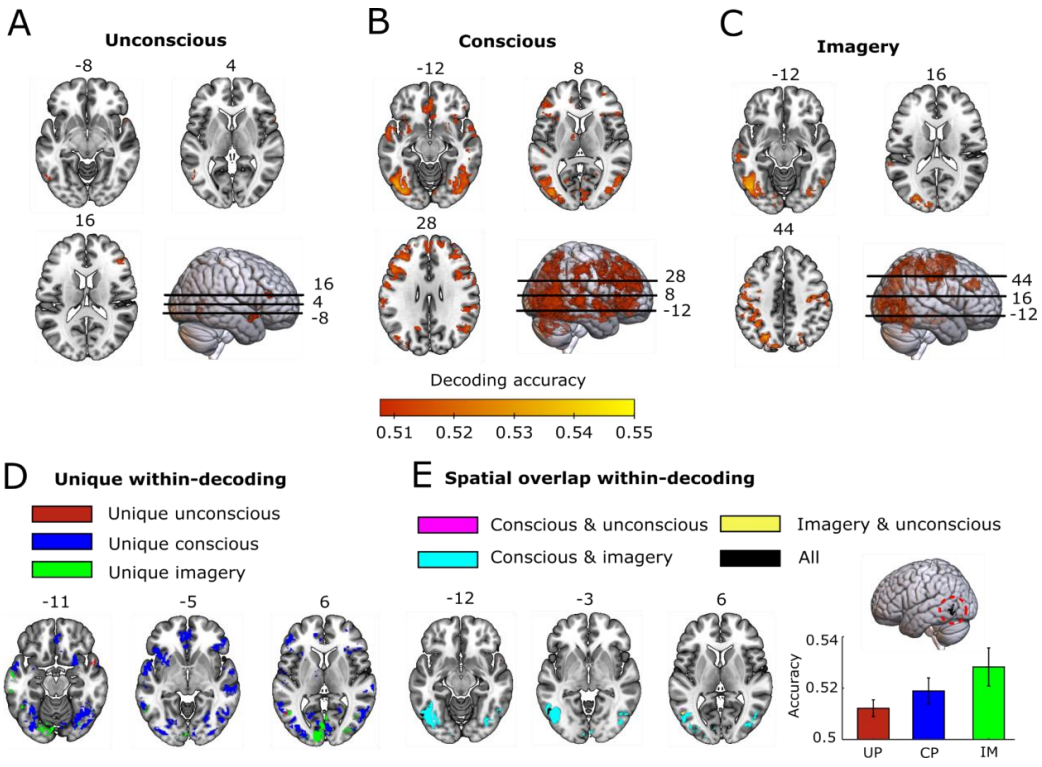

Figure 3. Condition specific neural representations. (A-C). For each condition, significant decoding clusters are shown for various axial slices. The heatmap indicates average decoding accuracy. (D-E) Significant decoding accuracy clusters (D) unique for each condition and (E) spatially overlapping between conditions. Significant decoding accuracy was found in all three conditions (indicated in black, circled in red) around the left lateral occipital cortex (LOC) at MNI coordinates -54 -65 -10. Decoding accuracies for the three conditions (UP = unconscious processing, CP = conscious perception, IM = imagery) within this ROI are plotted, with the error bars indicating the standard error of the mean (SEM).

Table 1. Significant within decoding clusters. Atlas labels determined using the AAL atlas (Tzourio-Mazoyer et al. 2002) on the basis of the MNI coordinates of the peak decoding accuracy.

Lobe	Atlas label	Condition	MNI peak			N voxels	Peak accuracy
			X	Y	Z		
Occipital	Occipital_Sup_R	Conscious	30	-76	46	394	0.52
	Occipital_Inf_L	Conscious	-48	-70	-6	9302	0.54
		Imagery	-42	-66	-6	4922	0.54

	Occipital_Inf_R	Imagery	46	-76	-2	459	0.53
	Cuneus_L	Conscious	0	-72	34	171	0.52
	Calcarine_R	Conscious	12	-60	14	115	0.52
Temporal	Temporal_Sup_L	Conscious	-58	0	-4	951	0.53
	Temporal_Sup_R	Conscious	68	-26	2	395	0.53
			64	-2	-10	220	0.52
	Temporal_Sup_L	Imagery	-64	-38	20	100	0.53
	Temporal_Mid_L	Imagery	-60	-20	-20	182	0.53
	Temporal_Inf_L	Unconscious	-56	-62	-6	86	0.52
	Temporal_Pole_Sup_R	Unconscious	52	14	-12	91	0.52
Parietal	Parietal_Inf_L	Conscious	-32	-36	40	72	0.52
	Parietal_Inf_R	Imagery	40	-40	56	143	0.53
	Precuneus_L	Conscious	-14	-58	68	110	0.52
	Precuneus_R	Imagery	20	-72	46	284	0.53
	SupraMarginal_R	Conscious	52	-30	46	485	0.52
		Imagery	64	-22	40	90	0.52
	Cingulum_Mid_L	Imagery	-4	30	32	263	0.53
	Cingulum_Mid_R	Conscious	8	-34	42	56	0.52
Frontal	Frontal_Sup_Medial_L	Conscious	-6	58	22	468	0.52
	Frontal_Sup_R	Conscious	18	52	26	91	0.52
		Imagery	24	-4	60	172	0.53
	Frontal_Inf_Tri_L	Conscious	-48	18	28	1738	0.53
		Unconscious	44	36	16	62	0.52
	Frontal_Med_Orb_R	Conscious	2	46	-4	575	0.52
	Supp_Motor_Area_L	Imagery	-6	4	68	557	0.63
	Precentral_L	Conscious	-56	-2	26	76	0.52
		Imagery	-56	8	26	59	0.52
Cerebellum	Cerebellum_Crus2_R	Conscious	30	-80	-40	71	0.52

Psychophysiological interaction analysis. The decoding analysis showed that left LOC contained stimulus information during all three conditions (Fig. 3E, lateral view), suggesting that this area might be where feedback and feedforward signals overlap. Before directly investigating the representational overlap between conditions using across-condition decoding generalisation, we first investigated whether this area indeed showed more feedback connectivity during conscious perception and imagery compared to unconscious processing and more feedforward connectivity during conscious and unconscious processing

compared to imagery. To investigate this, we performed a PPI analysis to characterize differences in brain connectivity between the three conditions (Fig. 4, Table 2).

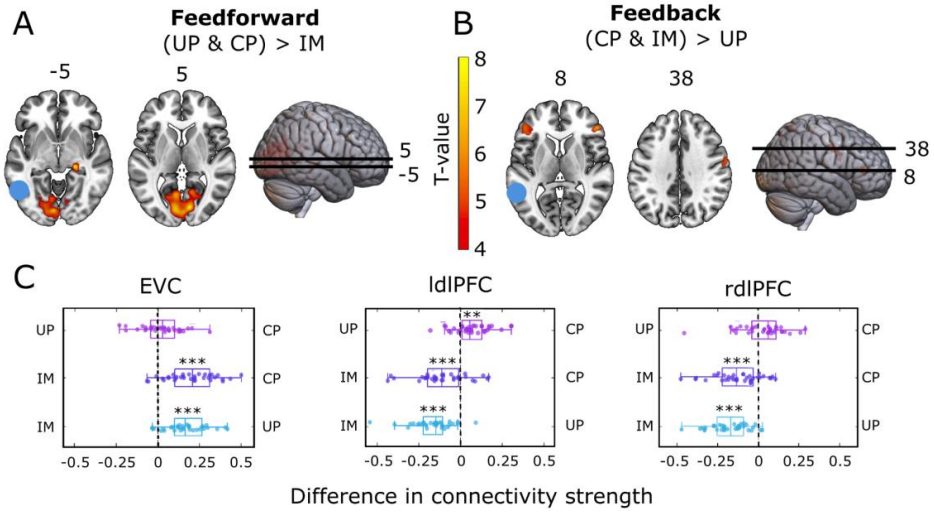

Figure 4. Psychophysiological interactions with left LOC as seed region. (A) The blue dot illustrates the location of the seed region, red-yellow indicates brain areas that showed significantly stronger connectivity with left LOC during conscious perception (CP) and unconscious processing (UP) compared to imagery (IM), i.e. in conditions where feedforward connections were present versus not. (B) The blue dot indicates the location of the seed region, red-yellow indicates brain areas that showed significantly stronger connectivity with left LOC during conscious perception and imagery compared to unconscious processing, i.e. in conditions where feedback connections were present versus not (C) Direct comparisons of connectivity between all conditions for left high-level visual cortex and early visual cortex (EVC; left); left high-level visual cortex and left dorsolateral prefrontal cortex (ldlPFC; middle) and left high-level visual cortex and right dorsolateral prefrontal cortex (rdlPFC). Boxplots indicate variance over participants and dots represent individual participants. ** $p < 0.005$, *** $p < 0.0005$.

In line with the predictions, there was stronger connectivity during conscious perception and unconscious processing compared to imagery between left LOC and early visual cortex (EVC; MNI: -1 -85 9) as well as right LGN (MNI: 24 -29 4; Fig. 4A-C), in line with the idea that during these conditions there was more feedforward processing than during imagery. However, because these conditions also differed in whether a mask was presented (conscious and unconscious) or not (imagery), and the PPI analysis is not stimulus-specific, this feedforward connectivity might partly reflect processing of the mask and not the (unconscious) stimulus before the mask. Furthermore, there was stronger connectivity during conscious perception and imagery compared to unconscious processing between left

LOC and bilateral dorsolateral prefrontal cortex (dlPFC; left MNI: -45 36 9; right MNI: 48 36 9) and right lateral frontal cortex, in line with increased feedback connectivity during these conditions. Post-hoc direct comparisons between conditions of the regions showing significant changes in connectivity (Fig. 4A,B) showed that connectivity between EVC and left LOC was stronger during conscious perception compared to imagery as well as during unconscious processing compared to imagery (Fig. 4C left). Furthermore, coupling between left LOC and left dlPFC was stronger during conscious perception compared to unconscious processing as well as during imagery compared to both conscious and unconscious processing (Fig. 4C middle). Finally, coupling between left LOC and right dlPFC was stronger during imagery compared to both conscious and unconscious processing (Fig. 4C right). These results indicate that, in line with our assumption, long-range feedback processing is indeed stronger during conscious perception and imagery compared to unconscious processing.

Table 2. Clusters connected with high-level within-decoding spatial overlap-cluster. Atlas labels determined using the AAL atlas (Tzourio-Mazoyer et al. 2002) on the basis of the MNI coordinates of the peak T-value for the PPI analysis.

Lobe	Atlas label	Comparison	MNI peak			N voxels	Peak T val
			X	Y	Z		
Occipital	Calcarine_R	(CP & UP) > IM	10	-82	4	3654	8.21
Temporal	Temporal_Inf_L	(CP & IM) > UP	-54	-58	-8	87	5.11
Parietal	Parietal_Sup_L	(CP & IM) > UP	-22	-72	52	50	5.48
	Parietal_Sup_R	(CP & IM) > UP	16	-60	68	62	5.73
	Precuneus_L	(CP & UP) > IM	-10	-52	20	53	4.6
	Postcentral_R	(CP & IM) > UP	62	-4	36	120	5.63
Frontal	Frontal_Inf_Tri_L	(CP & IM) > UP	-46	34	8	219	5.95
	Frontal_Inf_Tri_R	(CP & IM) > UP	46	34	10	149	6.8
	Frontal_Inf_Oper_R	(CP & IM) > UP	48	4	22	60	4.66
Other	Lateral Gen Nuc	(CP & UP) > IM	22	-28	-4	80	9.12

Generalisation across conditions. The above decoding analysis showed that left LOC contained stimulus information during all three conditions (Fig. 3E, lateral view) suggesting that this area might be where feedback and feedforward signals overlap. To directly test

whether the representations between conditions were similar, we performed across-condition decoding, where we trained a classifier to dissociate the stimuli in one condition, and tested it in another condition. In this analysis, above-chance cross-decoding accuracy would indicate that the underlying stimulus representations are to some extent similar. Significant across-condition clusters are shown in Figure 5 and listed in Table 3. In line with previous studies (Dijkstra, Bosch, & van Gerven, 2017a, 2019a; Lee et al., 2012; Pearson & Kosslyn, 2015; Reddy, Tsuchiya, & Serre, 2010c), we found representational overlap between conscious perception and imagery in visual, parietal and frontal areas (Fig. 5A, Table 1). In contrast, there was no significant cross-decoding between the unconscious condition and the other conditions in any brain area, suggesting an absence of representational overlap. Furthermore, despite the significant decoding in left LOC within all conditions (unconscious: *M* = 0.512, *SD* = 0.063; conscious: *M* = 0.519, *SD* = 0.097; imagery: *M* = 0.528, *SD* = 0.098), there was no significant cross-decoding overlap between the unconscious condition and the other conditions in this area, even at lower statistical thresholds (Fig. 5B).

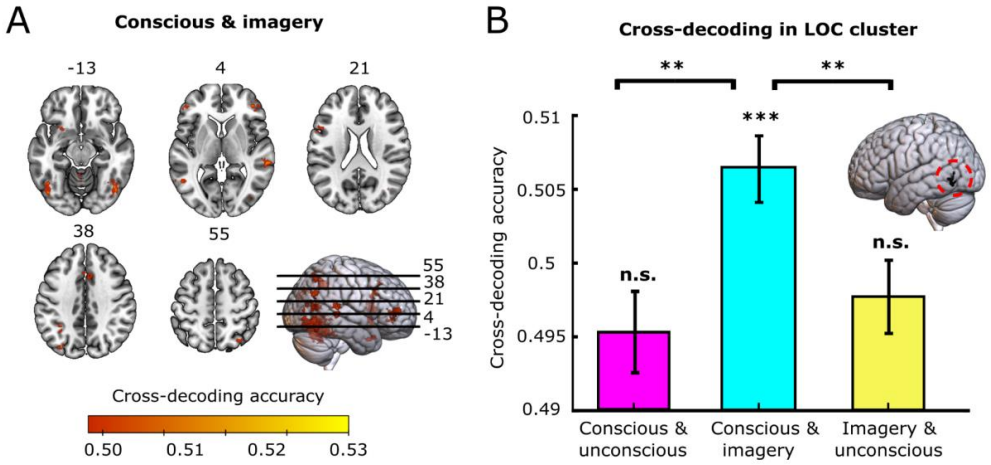

Figure 5. Across condition decoding accuracy. There was only significant representational overlap between conscious perception and mental imagery. (A) Significant cross-decoding clusters are shown for various axial slices. (B) Cross-decoding accuracy within the LOC cluster that had significant within-condition decoding in all three conditions (Fig. 3E), the same voxels were evaluated in all comparisons. Error bars indicate the SEM, n.s. = non-significant, * p < 0.05, ** p < 0.01, *** p < 0.005.

Taken together, these results suggest that there is no representational overlap between unconscious and imagined neural representations. However, it is possible that we did not

observe significant representational overlap here, not because there is no overlap, but because we do not have enough power to reveal this overlap. The results presented in Figure 5B show that cross-decoding accuracy between conscious perception and imagery is significantly higher than the cross-decoding accuracy between the other conditions. This means that while we cannot exclude the possibility of overlap with unconscious representations, we can conclude that representational overlap with unconscious representations is lower than the overlap between conscious and imagined representations. However, this might partly be due to the fact that unconscious representations were less strong compared to the other conditions (see Fig. 3). We discuss this possibility in more detail in the discussion.

Table 3. Significant across condition decoding clusters. Atlas labels determined using the AAL atlas (Tzourio-Mazoyer et al. 2002) on the basis of the MNI coordinates of the peak decoding accuracy. Condition is not indicated here because only imagery-conscious across condition decoding was significant.

Lobe	Atlas label	MNI peak			N voxels	Peak accuracy
		X	Y	Z		
Occipital	Occipital_Mid_L	-38	-80	34	59	0.51
	Occipital_Inf_R	44	-78	-4	261	0.52
	Lingual_R	20	-54	-10	91	0.51
Temporal	Temporal_Mid_R	60	-34	4	122	0.52
	Temporal_Pole_Sup_L	-46	16	-26	72	0.51
	Fusiform_L	-46	-64	-18	641	0.52
Parietal	Parietal_Sup_R	32	-62	50	97	0.51
	Parietal_Inf_L	-32	-52	42	113	0.52
	Cingulum_Mid_R	4	14	30	79	0.52
	Precuneus_L	-16	-56	14	76	0.52
	Angular_R	48	-62	32	60	0.51
Frontal	Frontal_Sup_Orb_L	-26	14	-14	59	0.52
	Frontal_Mid_R	46	52	8	113	0.52
	Frontal_Inf_Oper_L	-50	12	12	183	0.52
	Frontal_Inf_Tri_L	-48	42	0	52	0.51
Cerebellum	Cerebellum_3_R	12	-38	-24	142	0.52

Discussion

In this study we aimed to investigate the overlap between neural representations caused by feedforward versus feedback signals by comparing brain activity during mental imagery, conscious perception and unconscious processing. We found significant stimulus decoding for all three conditions in left high-level visual cortex (LOC). Furthermore, a PPI analysis showed that this area indeed showed more feedback connectivity during conscious perception and imagery compared to unconscious processing. These results suggested that this area might be the place where feedforward and feedback-initiated representations overlap. However, across-condition generalization revealed there was only significant representational overlap in this area between conscious perception and imagery, but not unconscious perception. These findings are in line with the idea that feedback changes the "format" of neural representations, leading to the reduction of overlap between representations caused by feedforward and feedback signals, but the presence of overlap between representations caused by feedback processes associated with perception of external stimuli and feedback processes associated with mental imagery.

The significant decoding of unconscious category-specific stimuli in high-level cortex agrees with previous findings (Axelrod, Bar, & Rees, 2015; Fahrenfort et al., 2012; Jiang & He, 2006; Rees, 2007). Although both conscious and unconscious category-specific representations were present in high-level visual cortex, we did not find representational overlap between the two. This is in line with previous studies using backward masking (Bar et al., 2001) and dichoptic fusion (Schurger, Pereira, Treisman, & Cohen, 2010). These studies also showed conscious and unconscious representations in high-level visual cortex, but no spatial or representational overlap between them. Conscious and unconscious representations may differ in several respects, including their duration, intensity, coherence, stability and reproducibility (Lamme & Roelfsema, 2000; Schurger et al., 2010, 2015; Tononi & Koch, 2008). It has been proposed that long-range feedback may stabilize activity in local neural processors, as if the brain "decides" what specific input it has received. The network's decision, given the input, is what may be reflected in conscious access (Dehaene, 2014; Schurger et al., 2015). The stabilization of neural activity by feedback therefore may change the format of neural category-specific representations (Baria, Maniscalco, & He, 2017; Dehaene, Sergent, & Changeux, 2003; Dijkstra et al., 2018; He, 2018; Weaver, Fahrenfort, Belopolsky, & van Gaal, 2019; Xie, Kaiser and Cichy 2020; King, Pescetelli, & Dehaene, 2016).

Although an intriguing possibility, some previous fMRI studies did report cross-decoding between conscious and unconscious conditions (Fahrenfort et al., 2012; Moutoussis & Zeki, 2002; Sterzer et al., 2008; Sterzer & Rees, 2008). In these studies, awareness of face/house stimuli was either manipulated by dichoptic fusion (Fahrenfort et al., 2012; Moutoussis & Zeki, 2002), Continuous Flash Suppression (CFS; Sterzer et al., 2008) or binocular rivalry (Sterzer & Rees, 2008). Which specific brain areas retain information about unconscious stimuli likely depends on the methods used to render the stimuli invisible (Fogelson, Kohler, Miller, Granger, & Tse, 2014; Axelrod et al., 2015; Izatt et al., 2014). Dichoptic fusion, CFS and binocular rivalry all rely on interactions between inputs from the two eyes and may primarily affect inhibition-adaptation cycles as early as V1, although much is still unclear at present (Axelrod et al., 2015; Rees, 2007; Tong, Meng, & Blake, 2006). In contrast, the neural effects of backward masking have previously been shown to disrupt recurrent interactions between high- and low-level visual regions (Del Cul, Baillet, & Dehaene, 2007; Fahrenfort, Scholte, & Lamme, 2007; Lamme, Zipser, & Spekreijse, 2002; Roelfsema, Lamme, Spekreijse, & Bosch, 2002; van Gaal & Lamme, 2012). Future research is necessary to fully determine the specific effects of each visibility manipulation on neural processing to unravel the discrepancies between studies and to understand why representational overlap between conscious and unconscious representations is sometimes observed and sometimes not.

The idea that feedback processing changes the format of neural representations suggests that the representational overlap between these different modes of perception should change over time. Because of the sluggishness of the BOLD response, fMRI lacks the temporal resolution needed to characterize such dynamics. In contrast, recent studies using methods with higher temporal resolution such as electro-encephalography (EEG) and magneto-encephalography (MEG) do indeed suggest differences in the timing of representational overlap between conscious perception, unconscious processing and imagery. During conscious perception, neural representations first change rapidly over time during early time windows, likely reflecting the feedforward sweep, after which representations stabilize later in time, presumably via recurrent processing (Baria, Maniscalco, & He, 2017; Cichy, Pantazis, & Oliva, 2014; Dijkstra et al., 2018; He, 2018; Mostert, Kok, & de Lange, 2015; Schurger et al., 2015). Recent evidence shows that neural representations of stimuli that were strongly masked or missed during the attentional blink,

only overlap with conscious conditions at early stages of input processing (until ~250ms; Meijs, Mostert, Slagter, de Lange, & van Gaal, 2019; Weaver et al., 2019). Furthermore, a recent MEG study revealed that representations during imagery mostly overlap with representations during later stages of conscious perception (Dijkstra et al., 2018; Xie, Kaiser and Cichy 2020). This supports the idea that neural representations of consciously reported and unreported stimuli are similar during initial feedforward (and likely local recurrent) processing, but that long-range feedback changes the neural representations, which then mimics the representations initiated by imagery-related feedback processing.

It is important to note that the exact relationship between (long-range) feedback processing and conscious awareness is still debated (see e.g. Boly et al., 2017). Some theories suggest that local recurrent processing within sensory systems is sufficient for conscious experience (Lamme, 2015), whereas others propose that communication within a broader network, including fronto-parietal areas, is necessary (Dehaene & Changeux, 2011; Mashour, Roelfsema, Changeux, & Dehaene, 2020) and still others propose that activation of meta- representations is sufficient (Brown, Lau, & LeDoux, 2019; Lau & Rosenthal, 2011). Here, we used perception rendered unconscious via backward masking as a proxy for feedforward visual processing and in line with this assumption, our PPI results suggested that visual activity was only driven in a feedforward fashion in the unconscious condition. However, it is possible that there was still some form of feedback processing present during the unconscious condition, either weaker or more local compared to the conscious condition, that was not picked up by the PPI analysis. This means that the absence of representational overlap between the conscious and unconscious condition might be due to other factors that are affected by awareness besides feedback processing. Future research directly investigating how top-down processing changes neural representations, using methods with a higher temporal resolution, will give more insight into this issue.

Finally, in line with previous studies we not only found significant cross decoding between conscious perception and imagery in several visual areas (Albers et al., 2013; Cichy et al., 2012; Dijkstra et al., 2017; Lee et al., 2012; O'Craven & Kanwisher, 2012; Reddy, Tsuchiya, & Serre, 2010), but also in parietal and frontal areas (Christophel, Klink, Spitzer, Roelfsema, & Haynes, 2017; Dijkstra et al., 2017). Additionally, we observed stronger connectivity between LOC and the dlPFC during imagery and conscious perception than during unconscious perception. The dlPFC has been implicated in numerous studies

investigating the neural mechanisms of conscious reportability (conscious access) of input (Davidson et al., 2010; Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006; Lau & Passingham, 2006; Rees, 2007). These studies, similarly to ours, have all focused on conscious access of an external stimulus, whereas a recent study showed similar feedback connectivity during conscious perception and mental imagery (Dijkstra, Zeidman, Ondobaka, Van Gerven, & Friston, 2017). The current results suggest that dlPFC is important for conscious access, regardless of whether it is internally or externally generated. However, it should be noted that our perception task was not passive; participants actively attended to specific features of the stimulus in order to judge its animacy. Therefore, overlap between imagery and perception reported here might (partly) be due to the employment of similar attentional mechanisms (Dijkstra et al., 2019). During both the perception and imagery task, participants had to attend to specific spatial locations and features in order to correctly execute the animacy task. This means that during both tasks, spatial and feature based top-down attention was employed. Moreover, the increase in dlPFC connectivity during imagery compared to conscious perception might reflect the increased attentional load of generating a sensory representation in the absence of its corresponding input (Dijkstra et al, 2017). Furthermore, the nature of the imagery task used here, in which the imagined image is presented relatively shortly before the imagery, might result in lingering feedforward activity. Several studies using the same paradigm only showed feedback processing during imagery (Dijkstra, Zeidman, Ondobaka, Van Gerven, & Friston, 2017; Dijkstra, Ambrogioni, Vidaurre, & van Gerven, 2020), however, we cannot completely rule out that the imagery also contained some feedforward processing. To fully address this, future research should investigate whether similar patterns are found with conscious but passive perception and with imagery initiated from long-term memory.

An alternative possibility for our findings is that feedback does not change the representational format per se, but that during the conscious condition, feedback enhances representations of feedforward information, for example via gain increase (Reynolds & Heeger, 2009; Wyart, Nobre, & Summerfield, 2012). Our results would then suggest that this kind of feedback-related enhancement is necessary to detect representational overlap between perception and imagery. This would also mean that using more sensitive methods, such as single-cell recordings, might still uncover representational overlap between the

neural populations recruited during imagery and those activated by unconsciously processed stimuli.

Related to this, it is important to note that while we did find significant decoding within unconscious processing, the decoding accuracy in this condition was lower than during both imagery and conscious perception. This means that our power to detect representational overlap with the unconscious condition was lower compared to the other conditions. Therefore, we cannot rule out that our lack of representational overlap with unconscious processing is due to low unconscious decoding. It is theoretically possible that the amount of representational overlap with unconscious conditions is as high as the other conditions, but that the low power within the unconscious condition prevented us from detecting this. Low unconscious decoding may partly reflect an inherent feature of unconscious processes, in the sense that feedforward initiated representations are less strong (especially higher up in the cortical hierarchy) compared to representations that have been stabilized via long-range feedback connections as mentioned above (Lamme & Roelfsema, 2000; Schurger et al., 2010, 2015; Tononi & Koch, 2008), leading to lower decoding accuracy and therefore less power to detect representational overlap (Fahrenfort et al., 2012; van Gaal & Lamme, 2012; Weaver, Fahrenfort, Belopolsky, & Van Gaal, 2019). Furthermore, although this type of masking has been shown to selectively disrupt feedback processing while keeping feedforward activity intact (Fahrenfort et al., 2007; Van Gaal et al., 2011, 2008), due to the low temporal resolution of the BOLD signal we are unable to completely rule out a reduction in feedforward activity due to the masking procedure. To fully rule out this possibility, ideally, the within-decoding accuracy in all conditions is equalized experimentally, for example by lowering the contrast of the stimulus in the conscious condition (see Lau and Passingham, 2006 for a similar approach in behaviour). This is an interesting avenue for future research.

In summary, our results show that neural representations measured by fMRI, triggered by purely feedforward (unconscious processing) or feedback (mental imagery) processes show reduced overlap. This suggests that the large representational overlap between imagery and perception reported in the literature (Dijkstra, Bosch, & van Gerven, 2019; Pearson, 2019) is undetectable for stimulus triggered activation in the absence of feedback processing. Our results suggest that long-range feedback processing alters the format or strength of neural representations, for example through stabilization of the neural

code. More insight into this dynamical process can be gained using methods with higher temporal resolution than fMRI. Future research should explore exactly how feedback changes the format of representations and how different methods of rendering stimuli invisible affect this process.

References

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & de Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology*, *23*(15), 1427–1431. https://doi.org/10.1016/j.cub.2013.05.065

Allefeld, C., & Haynes, J. D. (2014). Searchlight-based multi-voxel pattern analysis of fMRI by cross-validated MANOVA. NeuroImage, 89, 345–357. https://doi.org/10.1016/j.neuroimage.2013.11.043

Axelrod, V., Bar, M., & Rees, G. (2015). Exploring the unconscious using faces. *Trends in Cognitive Sciences*, 19, 35–45. https://doi.org/10.1016/j.tics.2014.11.003

Axelrod, V., & Yovel, G. (2015). Succesful decoding of famous faces in the fusiform face area. *Plos One, 10*(2): e0117126. https://doi.org/10.1371/journal.pone.0117126

Bar, M., Tootell, R. B. H., Schacter, D. L., Greve, D. N., Fischl, B., Mendola, J. D., … Dale, A. M. (2001). Cortical mechanisms specific to explicit visual object recognition. *Neuron*, *29*(2), 529–535. https://doi.org/10.1016/S0896-6273(01)00224-0

Baria, A. T., Maniscalco, B., & He, B. J. (2017). Initial-state-dependent, robust, transient neural dynamics encode conscious visual perception. *PLoS Computational Biology, 13*(11), e1005806. https://doi.org/10.1371/journal.pcbi.1005806

Bastos, Andre M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*, 695–711. https://doi.org/10.1016/j.neuron.2012.10.038

Bastos, AndréMoraes Moraes, Vezoli, J., Bosman, C. A., Schoffelen, J. M., Oostenveld, R., Dowdall, J. R., … Fries, P. (2015). Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron*, *85*(2), 390–401. https://doi.org/10.1016/j.neuron.2014.12.018

Caballero-Gaudes, C., & Reynolds, R. C. (2017). Methods for cleaning the BOLD fMRI signal. NeuroImage, 154, 128–149. https://doi.org/10.1016/j.neuroimage.2016.12.018

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J. D. (2017). The

654 Distributed Nature of Working Memory. *Trends in Cognitive Sciences*, *21*, 111–124.
655 https://doi.org/10.1016/j.tics.2016.12.007
656 Cichy, R.M., Heinzle, J., & Haynes, J.-D. (2012). Imagery and perception share cortical
657 representations of content and location. *Cerebral Cortex, 22*(2): 372-80.
658 https://doi.org/10.1093/cercor/bhr106
659 Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space
660 and time. *Nature Neuroscience*, *17*(3), 455-462. https://doi.org/10.1038/nn.3635
661 Davidson, M., Persaud, N., Maniscalco, B., Mobbs, D., Passingham, R., Cowey, A., & Lau, H.
662 (2010). Awareness-related activity in prefrontal and parietal cortices reflects more than
663 superior performance capacity: A blindsight case study. *Journal of Vision*, *58*(2), 605-
664 611. https://doi.org/10.1167/10.7.897
665 Dehaene, S. (2014). Consciousness and the brain: deciphering how the brain codes our
666 thoughts. In *Penguin Group; New York*.
667 Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious,
668 preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive*
669 *Sciences*, *10*(5), 204–211. https://doi.org/10.1016/j.tics.2006.03.007
670 Dehaene, S., Sergent, C., & Changeux, J.-P. (2003). A neuronal network model linking
671 subjective reports and objective physiological data during conscious perception.
672 *Proceedings of the National Academy of Sciences, 100*(14), 8520–8525.
673 https://doi.org/10.1073/pnas.1332574100
674 Del Cul, A., Baillet, S., & Dehaene, S. (2007). Brain dynamics underlying the nonlinear
675 threshold for access to consciousness. *PLoS Biology*, *5*(10), 2408–2423.
676 https://doi.org/10.1371/journal.pbio.0050260
677 Dentico, D., Cheung, B. L., Chang, J.-Y., Guokas, J., Boly, M., Tononi, G., & Van Veen, B.
678 (2014). Reversal of cortical information flow during visual imagery as compared to
679 visual perception. *NeuroImage*, *100*, 237–243.
680 https://doi.org/10.1016/j.neuroimage.2014.05.081
681 Dijkstra, N., Ambrogioni, A., Vidaurre, D., & van Gerven, M. A. J. (2020). Neural dynamics of
682 perceptual inference and its reversal during imager. *eLife, 9:e53588.*
683 *https://doi.org/10.7554/eLife.53588*
684 Dijkstra, N., Bosch, S. E., & van Gerven, M. A. J. (2017). Vividness of visual imagery depends
685 on the neural overlap with perception in visual areas. *Journal of Neuroscience*, *37*(5).

https://doi.org/10.1523/JNEUROSCI.3022-16.2016

Dijkstra, N., Zeidman, P., Ondobaka, S., van Gerven, M. A. J., & Friston, K. (2017). Distinct Top-down and Bottom-up Brain Connectivity during Visual Perception and Imagery. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-05888-8

Dijkstra, N., Bosch, S. E., & van Gerven, M. A. J. (2019). Shared Neural Mechanisms of Visual Perception and Imagery. *Trends in Cognitive Sciences*, *23*, 18–29. https://doi.org/10.1016/j.tics.2019.02.004

Dijkstra, N., Mostert, P., de Lange, F. P., Bosch, S., & van Gerven, M. A. J. (2018). Differential temporal dynamics during visual imagery and perception. *ELife*, *7*, 1–16. https://doi.org/10.1101/226217

Eger, E., Kell, C. A., & Kleinschmidt, A. (2008). Graded size sensitivty of object exemplar-evoked activity patterns within human LOC subregions. *Journal of Neurophysiology, 100*: 2038-2047. https://doi.org/10.1152/jn.90305.2008.

Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2007). Masking disrupts reentrant processing in human visual cortex. *Journal of Cognitive Neuroscience*, *19*(9), 1488–1497. https://doi.org/10.1162/jocn.2007.19.9.1488

Fahrenfort, J. J., Snijders, T. M., Heinen, K., van Gaal, S., Scholte, H. S., & Lamme, V. A. F. (2012). Neuronal integration in visual cortex elevates face category tuning to conscious face perception. *Proceedings of the National Academy of Sciences*, *109*(52), 21504–21509. https://doi.org/10.1073/pnas.1207414110

Fogelson, S. V., Kohler, P. J., Miller, K. J., Granger, R., & Tse, P. U. (2014). Unconscious neural processing differs with method used to render stimuli invisible. *Frontiers in Psychology*, *5*, 1–11. https://doi.org/10.3389/fpsyg.2014.00601

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage*, *6*, 218-229. https://doi.org/10.1006/nimg.1997.0291

Gardumi, A., Ivanov, D., Hausfeld, L., Valente, G., Formisano, E., & Uludağ, K. (2016). The effect of spatial resolution on decoding accuracy in fMRI multivariate pattern analysis. *NeuroImage*, *132*, 32–42. https://doi.org/10.1016/J.NEUROIMAGE.2016.02.033

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635. https://doi.org/10.1038/nature07832

718 Hendriks, M. H. A., Daniels, N., Pegado, F., & Op de Beeck, H. P. (2017). The Effect of Spatial
719 Smoothing on Representational Similarity in a Simple Motor Paradigm. *Frontiers in*
720 *Neurology*, *8*, 222. https://doi.org/10.3389/fneur.2017.00222
721 He, B. J. (2018). Robust, Transient Neural Dynamics during Conscious Perception. *Trends in*
722 *Cognitive Sciences, 22*, 563–565. https://doi.org/10.1016/j.tics.2018.04.005
723 Horikawa, T., & Kamitani, Y. (2017). Generic decoding of seen and imagined objects using
724 hierarchical visual features. *Nature Communications*, *8*, 15037.
725 https://doi.org/10.1038/ncomms15037
726 Jiang, Y., & He, S. (2006). Cortical Responses to Invisible Faces: Dissociating Subsystems for
727 Facial-Information Processing. *Current Biology*, *16*(20), 2023–2029.
728 https://doi.org/10.1016/j.cub.2006.08.084
729 Johnson, M. R., & Johnson, M. K. (2014). Decoding individual natural scene representations
730 during perception and imagery. *Frontiers in Human Neuroscience*, *8*, 59.
731 https://doi.org/10.3389/fnhum.2014.00059
732 King, J.-R., Pescetelli, N., & Dehaene, S. (2016). Brain Mechanisms Underlying the Brief
733 Maintenance of Seen and Unseen Sensory Information. *Neuron*, *92*(5), 1122–1134.
734 https://doi.org/10.1016/j.neuron.2016.10.051
735 Kosslyn, S. M., & Thompson, W. L. (2003). When is early visual cortex activated during visual
736 mental imagery? *Psychological Bulletin*, *129*(5), 723–746.
737 https://doi.org/10.1037/0033-2909.129.5.723
738 Kovalenko, L. Y., Chaumon, M., & Busch, N. A. (2012). A pool of pairs of related objects
739 (POPORO) for investigating visual semantic integration: Behavioral and
740 electrophysiological validation. *Brain Topography*, *25*(3), 272–284.
741 https://doi.org/10.1007/s10548-011-0216-8
742 Lamme, V. (2015). The Crack of Dawn. In *Open MIND* (Vol. 22).
743 https://doi.org/10.15502/9783958570092
744 Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by
745 feedforward and recurrent processing. *Trends in Neurosciences*, 23, 571–579.
746 https://doi.org/10.1016/S0166-2236(00)01657-X
747 Lamme, V. A. F., Zipser, K., & Spekreijse, H. (2002). Masking interrupts figure-ground signals
748 in V1. *Journal of Cognitive Neuroscience*, *14*(7), 1044–1053.
749 https://doi.org/10.1162/089892902320474490

750 Lau, H. C., & Passingham, R. E. (2006). Relative blindsight in normal observers and the neural
751 correlate of visual consciousness. *Proceedings of the National Academy of Sciences*,
752 *103*(49), 18763–18768. https://doi.org/10.1073/pnas.0607716103
753 Lee, S.-H., Kravitz, D. J., & Baker, C. I. (2012). Disentangling visual imagery and perception of
754 real-world objects. *NeuroImage*, *59*(4), 4064–4073.
755 https://doi.org/10.1016/j.neuroimage.2011.10.055
756 Lund, T. E., Nørgaard, M. D., Rostrup, E., Rowe, J. B., & Paulson, O. B. (2005). Motion or
757 activity: Their role in intra- and inter-subject variation in fMRI. NeuroImage, 26(3), 960–
758 964. https://doi.org/10.1016/j.neuroimage.2005.02.021
759 Macmillan, N. A., & Creelman, C. D. (1990). Detection Theory: A User's Guide. In *Lawrence*
760 *Erlbaum Associates.*
761 Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of*
762 *Psychology*, *64*, 17–24. https://doi.org/10.1111/j.2044-8295.1973.tb01322.x
763 Mashour, G. A., Roelfsema, P., Changeux, J. P., & Dehaene, S. (2020, March 4). Conscious
764 Processing and the Global Neuronal Workspace Hypothesis. *Neuron*, *105*, 776–798.
765 https://doi.org/10.1016/j.neuron.2020.01.026
766 Mechelli, A., Price, C. J., Friston, K. J., & Ishai, A. (2004). Where bottom-up meets top-down:
767 neuronal interactions during perception and imagery. *Cerebral Cortex*, *14*(11), 1256–
768 1265. https://doi.org/10.1093/cercor/bhh087
769 Meijs, E. L., Mostert, P., Slagter, H. A., de Lange, F. P., & van Gaal, S. (2019). Exploring the
770 role of expectations and stimulus relevance on stimulus-specific neural representations
771 and conscious report. *Neuroscience of Consciousness*, 1.
772 https://doi.org/10.1093/NC/NIZ011
773 Misaki, M., Luh, W. M., & Bandettini, P. A. (2013). The effect of spatial smoothing on fMRI
774 decoding of columnar-level organization with linear support vector machine. *Journal of*
775 *Neuroscience Methods*, *212*(2), 355–361.
776 https://doi.org/10.1016/j.jneumeth.2012.11.004
777 Mostert, P., Kok, P., & de Lange, F. P. (2015). Dissociating sensory from decision processes in
778 human perceptual decision making. *Scientific Reports*, *5*, 18253.
779 https://doi.org/10.1038/srep18253
780 Moutoussis, K., & Zeki, S. (2002). The relationship between cortical activation and
781 perception investigated with invisible stimuli. *Proceedings of the National Academy of*

Sciences, *99*(14), 9527–9532. https://doi.org/10.1073/pnas.142305699

Muckli, L. (2010). What are we missing here? Brain imaging evidence for higher cognitive functions in primary visual cortex V1. *International Journal of Imaging Systems and Technology*, *20*(2), 131–139. https://doi.org/10.1002/ima.20236

O'Craven, K.M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Congitive Neuroscience, 12(6):* 1013-23. https://doi.org/10.1162/08989290051137549

Pearson, J. (2019). The human imagination: the cognitive neuroscience of visual mental imagery. *Nature Reviews Neuroscience, 20*, 624-634. https://doi.org/10.1038/s41583-019-0202-9

Pearson, J., & Kosslyn, S. M. (2015). The heterogeneity of mental representation: Ending the imagery debate. *Proceedings of the National Academy of Sciences, 112*(33), 10089–10092. https://doi.org/10.1073/pnas.1504933112

Pearson, J., Naselaris, T., Holmes, E. A., & Kosslyn, S. M. (2015). Mental Imagery: Functional Mechanisms and Clinical Applications. *Trends in Cognitive Sciences*, *19*(10), 590–602. https://doi.org/10.1016/j.tics.2015.08.003

Reddy, L., Tsuchiya, N., & Serre, T. (2010a). Reading the mind's eye: decoding category information during mental imagery. *NeuroImage*, *50*(2), 818–825. https://doi.org/10.1016/j.neuroimage.2009.11.084

Rees, G. (2007). Neural correlates of the contents of visual awareness in humans. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*(1481), 877–886. https://doi.org/10.1098/rstb.2007.2094

Roelfsema, P. R., Lamme, V. A. F., Spekreijse, H., & Bosch, H. (2002). Figure—Ground Segregation in a Recurrent Network Architecture. *Journal of Cognitive Neuroscience*, *14*(4), 525–537. https://doi.org/10.1162/08989290260045756

Schurger, A., Pereira, F., Treisman, A., & Cohen, J. D. (2010). Reproducibility distinguishes conscious from nonconscious neural representations. *Science*, *327*(5961), 97–99. https://doi.org/10.1126/science.1180029

Schurger, A., Sarigiannidis, I., Naccache, L., Sitt, J. D., Dehaene, S., & Goldberg, M. E. (2015). Cortical activity is more stable when sensory stimuli are consciously perceived. *Proceedings of the National Academy of Sciences*, *112*(6), 2083–E2092. https://doi.org/10.1073/pnas.1418730112

814 Shanks, D. (2017). Regressive research: The pitfalls of post-hoc data-selection in the study of
815 unconscious mental process. *Psychonomic Bulletin and Review, 24*(3): 752-775.
816 https://doi.org/10.3758/s13423-016-1170-y
817 Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction
818 in classification-based multi-voxel pattern analysis (MVPA): Random permutations and
819 cluster size control. *NeuroImage*, *65*, 69–82.
820 https://doi.org/10.1016/j.neuroimage.2012.09.063
821 Sterzer, P., Haynes, J. D., & Rees, G. (2008). Fine-scale activity patterns in high-level visual
822 areas encode the category of invisible objects. *Journal of Vision*, *8*(15), 10–10.
823 https://doi.org/10.1167/8.15.10
824 Sterzer, P., & Rees, G. (2008). A neural basis for percept stabilization in binocular rivalry.
825 *Journal of Cognitive Neuroscience*, *20*(3), 389–399.
826 https://doi.org/10.1162/jocn.2008.20039
827 Stokes, M., Thompson, R., Cusack, R., & Duncan, J. (2010). Codes in Visual Cortex during
828 Mental Imagery. *Journal of Neuroscience*, *29*, 1565–1572.
829 https://doi.org/10.1523/JNEUROSCI.4657-08.2009
830 Thirion, B., Duchesnay, E., Hubbard, E., Dubois, J., Poline, J.-B., Lebihan, D., & Dehaene, S.
831 (2006). Inverse retinotopy: inferring the visual content of images from brain activation
832 patterns. *NeuroImage*, *33*(4), 1104–1116.
833 https://doi.org/10.1016/j.neuroimage.2006.06.062
834 Tong, F., Meng, M., & Blake, R. (2006). Neural bases of binocular rivalry. *Trends in Cognitive*
835 *Sciences*, *10,* 502–511. https://doi.org/10.1016/j.tics.2006.09.003
836 Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. *The*
837 *Biological Bulletin*, *215*(3), 216–242. http://www.ncbi.nlm.nih.gov/pubmed/19098144
838 Tononi, G., & Koch, C. (2008). The Neural Correlates of Consciousness: An Update. *Annals of*
839 *the New York Academy of Sciences*, *1124*(1), 239–261.
840 https://doi.org/10.1196/annals.1440.004
841 Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N.,
842 Mazoyer, B., & Joliot, N. (2002). Automated anatomical labeling of activations in SPM
843 using a macroscopic anatomical parcellation of the MNI MRI single-subject brain.
844 *Neuroimage, 15*(1): 273-89. https://doi.org/10.1006/nimg.2001.0978
845 Van Gaal, S., Lamme, V. A. F., Fahrenfort, J. J., & Ridderinkhof, K. R. (2011). Dissociable brain

mechanisms underlying the conscious and unconscious control of behavior. Journal of Cognitive Neuroscience, 23(1), 91–105. https://doi.org/10.1162/jocn.2010.21431

van Gaal, S., & Lamme, V. A. F. (2012). Unconscious High-Level Information Processing. *The Neuroscientist*, *18*(3), 287–301. https://doi.org/10.1177/1073858411404079

Van Gaal, S., Ridderinkhof, K. R., Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2008). Frontal cortex mediates unconsciously triggered inhibitory control. Journal of Neuroscience, 28(32), 8053–8062. https://doi.org/10.1523/JNEUROSCI.1278-08.2008

van den Hurck, J., Op de Beeck, H. (2019). Generalization assymetry in multivariate cross-classification: When representation A generalizes better to representation B than B to A. *BioRXiv.* https://doi.org/10.1101/592410

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

Xie, S., Kaiser, D., & Cichy, R.M. (2020). Visual imagery and perception share neural representations in the alpha frequency band. *Current Biology, 30:* 2621-2627. https://doi.org/10.1016/j.cub.2020.04.074