

A neural basis for distinguishing imagination from reality

Highlights

- Imagination and perception are intermixed in the brain's perceptual system
- Their combined sensory strength predicts reality judgments
- This sensory strength is tracked by the bilateral fusiform gyrus
- Frontal brain regions encode binary judgments of reality

Authors

Nadine Dijkstra, Thomas von Rein,
Peter Kok, Stephen M. Fleming

Correspondence

n.dijkstra@ucl.ac.uk

In brief

Dijkstra et al. show that our brain distinguishes imagination from reality by monitoring how strongly the fusiform gyrus is activated. When imagined and real images produce similar activity, people can confuse them. This helps explain how we normally tell what is real—and why that sometimes fails.



Article

A neural basis for distinguishing imagination from reality

Nadine Dijkstra,^{1,4,*} Thomas von Rein,¹ Peter Kok,¹ and Stephen M. Fleming^{2,3}

¹Department of Imaging Neuroscience, University College London, WC1 3AR London, UK

²Department of Experimental Psychology, University College London, WC1H 0AP London, UK

³Max Planck UCL Centre for Computational Psychiatry and Ageing Research, University College London, WC1B 5EH London, UK

⁴Lead contact

*Correspondence: n.dijkstra@ucl.ac.uk

<https://doi.org/10.1016/j.neuron.2025.05.015>

SUMMARY

Humans are able to imagine scenarios that are decoupled from the current environment by internally activating perceptual representations. Although an efficient re-use of existing resources, it remains unknown how human observers classify perceptual signals as reflecting external reality, as opposed to internal simulation or imagination. Here, we show that judgments of reality are underpinned by the combined strength of sensory activity generated by either imagery or perception in the fusiform gyrus. Activity fluctuations in this region predict confusions between imagery and perception on a trial-by-trial basis and interact with a frontal brain network encoding binary judgments of reality. Our results demonstrate that a key mechanism through which the brain distinguishes imagination from reality is by monitoring the activity of the mid-level visual cortex. These findings increase our understanding of failures of reality testing and lay the foundations for characterizing a generalized perceptual reality monitoring system in the human brain.

INTRODUCTION

Imagining possible futures requires humans and other animals to engage in internal simulation—a set of brain processes that support anticipation, planning, and sophisticated behavioral control. However, the capacity for internal simulation comes at a cost. In humans, visual mental imagery co-opts brain systems specialized for perception of the external environment^{1–8} (for reviews, see Dijkstra et al.⁹ and Pearson¹⁰). Although an efficient re-use of neural resources, this overlap in brain systems for perception and imagination creates the potential for confusing the two. It remains unclear how, at any given moment, human observers determine whether perceptual signals reflect reality or imagination.

Here, we generated confusions between imagery and perception in healthy human volunteers, allowing us to use functional magnetic resonance imaging (fMRI) to isolate neural processes underlying subjective judgments about reality.¹¹ Twenty-six participants were instructed to detect barely visible left- or right-tilted gratings that gradually appeared within dynamic noise on half of the trials (Figure 1A). The orientation of the to-be-detected gratings, left or right tilted, was manipulated in a blockwise fashion. At the same time, on half the blocks, participants were instructed to imagine a grating with the same orientation (congruent condition) and on the other half they imagined a grating with a perpendicular orientation (incongruent condition). The blockwise manipulation of stimulus and imagery orientation allowed us to determine the relevant orientation for false alarms

within each block. For instance, congruent false alarms are those in which the stimulus is physically absent, the observer is instructed to imagine the same orientation as they are seeking to detect, and an erroneous report of stimulus presence is given. On each trial, participants made a reality judgment (indicating whether a grating had been presented on the screen) and rated the vividness of their mental imagery.

In previous studies, we have found that congruent imagery increases the probability of judging a stimulus as real, relative to both incongruent imagery and no imagery conditions.¹² This effect was best explained by a computational model in which imagery added sensory evidence to perception,^{13,14} suggesting that when imagery and perception are congruent, their sensory signals can be confused. In line with this, we found that participants were more likely to indicate that a stimulus had been presented in the congruent ($M = 0.37$, $SE = 0.02$) compared with the incongruent condition ($M = 0.31$, $SE = 0.3$; $F(1,25) = 10.11$, $p = 0.004$, $\eta^2 = 0.01$), irrespective of whether a stimulus had actually been presented (i.e., an increase in both congruent false alarms and hits; Figure 1B). Furthermore, imagery vividness was judged as significantly higher when an external stimulus was present, compared with when it was absent, only in the congruent condition ($M_{diff} = 0.24$, $SE_{diff} = 0.04$, $t(25) = 5.32$, $p_{holm} = 0.00002$), but this effect did not reach significance in the incongruent condition ($M_{diff} = 0.07$, $SE_{diff} = 0.04$, $t(25) = 1.61$, $p_{holm} = 0.143$, Bayes Factor (BF)₁₀ = 0.89; Figure 1C). Together, these results indicate substantial confusions between imagination and reality under near-threshold conditions. Note that we think this effect is



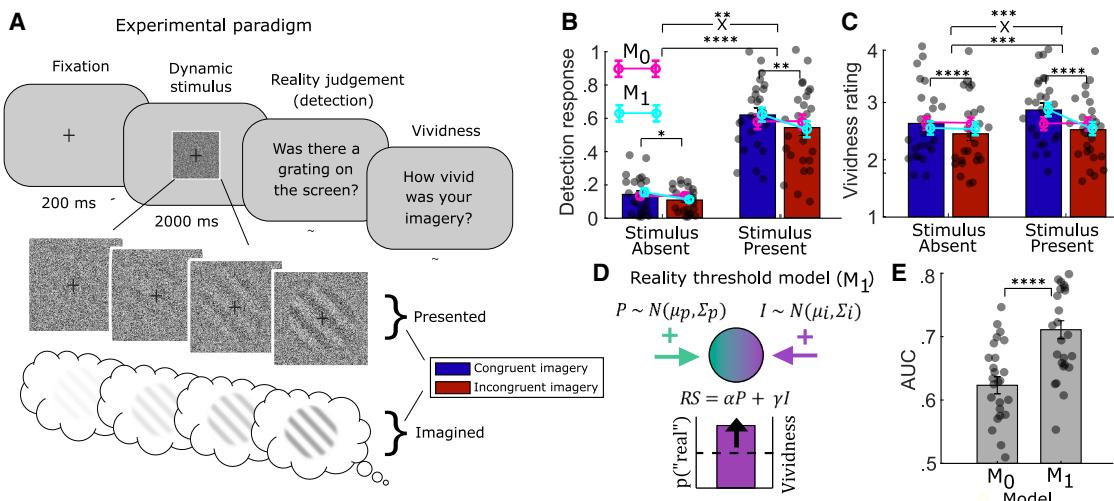


Figure 1. Experimental paradigm and behavioral results

(A) Prior to each block of trials, participants were instructed which gratings they should imagine and detect. During each trial, once the dynamic noise appeared on screen, participants simultaneously imagined and tried to detect the instructed gratings. To accurately infer whether a grating was presented or not, imagined and perceived stimuli have to be kept apart. In the congruent condition, there is a potential for confusing imagery and perception. To ensure participants were complying with instructions, we recorded which grating participants reported imagining during each block.

(B and C) (B) Effect of presented stimulus (absent vs. present) and imagined stimulus (congruent vs. incongruent) on perceptual reality judgments (detection responses; left) and (C) vividness ratings (right). Magenta and cyan lines indicate, respectively, model fits for a null model (M_0), in which detection judgments and vividness are based on separate latent variables, and for the reality threshold model M_1 .

(D) Reality threshold model M_1 assumes that imagery and perception signals are added together and that this mixed signal determines detection and vividness responses, adapted from Dijkstra and Fleming.¹¹ I , imagery signal; P , perception signal; RS , reality signal; N , normal distribution.

(E) Model fit is measured as the area under the curve (AUC) between simulated and observed behavioral responses for the null and reality threshold models. Dots reflect individual participants ($n = 26$) and error bars reflect standard errors of the mean (SEM). * $p < 0.05$; ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$.

unlikely to be driven by a relative increase in conflict during the incongruent condition, as in previous studies we have found a similar profile of reality judgments for both incongruent imagery and no imagery conditions.^{12–14}

This combination of behavioral effects can be explained by a “reality threshold” model (Figure 1D).¹¹ The reality threshold model assumes that imagery and perception activate the same neural populations, such that any imagined and perceived signals are summed together, creating a combined “reality signal,” RS . The strength of this intermixed signal in turn determines both the experienced vividness of the percept and, depending on whether it crosses a reality threshold, whether it is inferred to reflect reality.¹¹ In other words, the reality threshold model proposes that perceptual reality monitoring is based on comparing the sensory strength of imagined and perceived signals against a reality threshold.

We compared the fit of this model (M_1) against a null model (M_0) that assumes separate latent variables underpinning imagery vividness and reality judgments (cf. magenta and cyan fits in Figure 1B) and show that a reality threshold model provides a better fit, both qualitatively (Figures 1B and 1C) and quantitatively (Figure 1E; $t(25) = 5.37$, $p = 0.0001$). We also established that a linear combination of imagery and perception provides a more parsimonious explanation of the data than either a multiplicative or a sublinear interaction (Figure S1). Finally, the reality threshold model predicts that in the congruent condition, reality judgments should covary with imagery vividness ratings because they are both underpinned by the same latent variable.

This should not occur in the incongruent condition, where imagery and sensory stimulation activate non-overlapping sensory representations. In line with this prediction, we found a significant interaction between congruency and response on vividness ratings ($F(23,1) = 20.04$, $p = 0.00001$), such that vividness ratings were higher for presence compared with absence responses in the congruent ($F(23,1) = 38.75$, $p = 0.000002$), but not in the incongruent ($F(23,1) = 0.56$, $p = 0.46$, $BF_{10} = 0.32$), condition (Figure S1D).

RESULTS

FG tracks latent variable underpinning reality judgments

To identify brain areas encoding a hypothesized RS , we first performed a whole-brain analysis testing for parametric modulations by reality and imagery vividness judgments. In line with our reality threshold model, we found striking similarities in the neural networks underlying imagery vividness and reality judgments (Figures 2A and 2B), with significant modulation by both of these factors in the bilateral fusiform gyrus (FG), bilateral intraparietal sulcus (IPS), and bilateral precentral sulcus (PCS; Figures 2A and 2B; Tables S1 and S2). Note that an overlap between regions modulated by perceptual reality judgments and imagery vividness ratings cannot be ascribed to collinearity between these variables, as the correlation between these regressors was low ($M = 0.1$, $SD = 0.2$).

To determine which of these regions showed signatures of an RS , we extracted the latent variable from our model fits

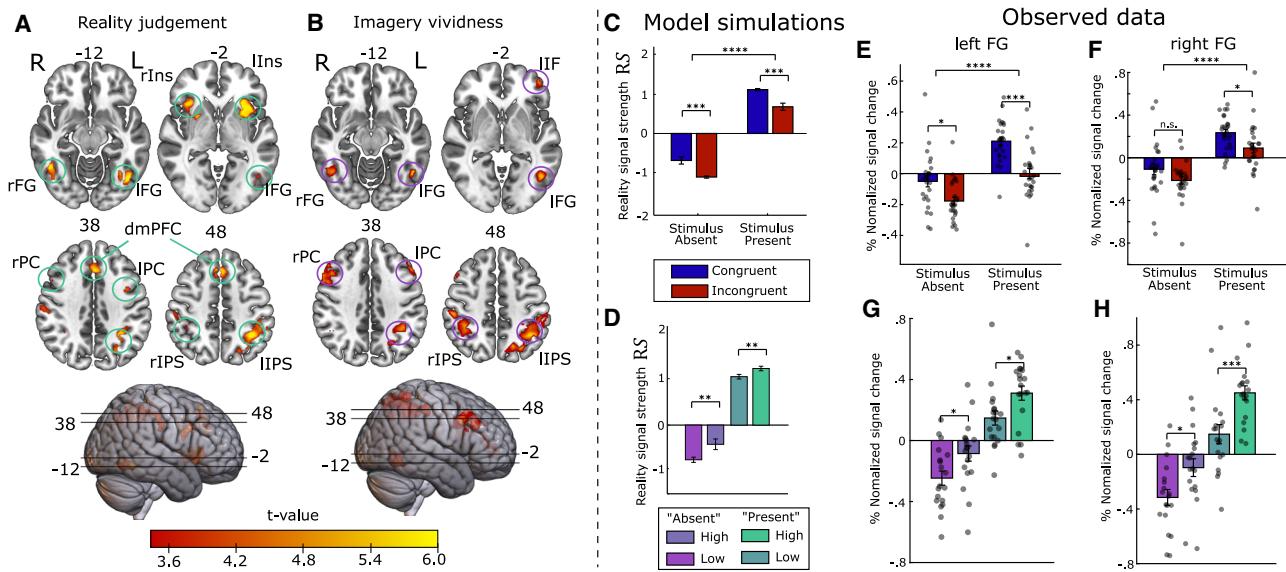


Figure 2. Univariate neural basis of a reality signal

(A and B) (A) Brain areas showing significant positive modulations by reality judgment response and by (B) imagery vividness rating. Ins, anterior insula; PC, precentral; dmPFC, dorsomedial prefrontal cortex; IPS, intraparietal sulcus; FG, fusiform gyrus; IF, inferior frontal. Images are thresholded at $p < 0.05$, family-wise error (FWE) corrected at the cluster level, with a cluster-forming threshold of $p < 0.001$, uncorrected.

(C) Simulated reality signal strength in each condition obtained from fits of subject-specific imagery vividness and detection sensitivity parameters.

(D) Simulated signal per vividness (high vs. low) and detection ("present" vs. "absent") response type.

(E and F) (E) Effect of stimulus presence and congruency in left and (F) right FG.

(G and H) (G) Influence of response type on activity in left and (H) right FG ($p < 0.05$ FWE cluster-level corrected). Dots reflect individual participants ($n = 26$) and error bars reflect standard errors of the mean (SEM). * $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$; **** $p < 0.001$.

separately for each experimental condition. Besides covarying with reality judgments and vividness ratings, an *RS* should show both a main effect of congruency as well as a main effect of stimulus presence, such that activity is highest for congruent-present trials (Figure 2C). Furthermore, the *RS* should track both reality judgments and vividness ratings, such that for both absence and presence responses, the signal is modulated by vividness (Figure 2D).

We next performed region of interest (ROI) analyses within the regions sensitive to reality judgments (Figure 2A) to test for computational signatures of a *RS* (Figures 2C and 2D). We found that both the left and right FG exhibited a pattern of effects consistent with an *RS*, although the main effect of congruency was only significant in the left ($F(25,1) = 15.8$, $p = 0.0005$; Figure 2E) and not the right FG ($F(25,1) = 4.13$, $p = 0.053$; Figure 2F). In the right FG, the effect of congruency was modulated by individual differences in behavior (median split: $t(24) = 2.43$, $p = 0.02$), such that only participants who showed a stronger behavioral effect of imagery congruency on perceptual reality judgments also showed a significant congruency effect in the right FG ($t(12) = 2.81$, $p = 0.01$). In addition, and in line with the predictions of a reality threshold model, we observed modulations by imagery vividness that remained significant when the data were divided into presence and absence responses (Figures 2G and 2H; left FG, absence: $t(35.9) = 2.33$, $p = 0.026$; presence: $t(35.9) = 2.36$, $p = 0.024$; right FG, absence: $t(35.7) = 2.63$, $p = 0.012$; presence: $t(35.7) = 3.65$, $p = 0.0008$).

Notably, we did not observe this pattern of effects in any of the other ROIs: although we found main effects of stimulus presence in all ROIs (all p values < 0.0007), only the FG showed a main effect of congruency—a key requirement for a candidate neural basis of a *RS*. We found a significant interaction between stimulus presence and congruency in the left IPS ($F(25,1) = 8.75$, $p = 0.007$; Figure S2A), driven by a significant negative effect of congruency only during stimulus-present trials. This finding suggests that in IPS, in contrast to FG, congruent external input is suppressed rather than enhanced. A whole-brain analysis additionally revealed a main effect of congruency in early visual cortex (EVC; Figure S2C). However, unlike the pattern observed for FG, EVC showed a significant interaction between congruency and stimulus presence ($F(25,1) = 8.307$, $p = 0.008$, $\eta_p^2 = 0.029$), such that stimulus presence had a positive effect during congruent conditions but a negative effect during incongruent conditions—at odds with the pattern predicted for an *RS* (Figure 2C). There was also no effect of reality judgment responses on EVC activity ($t(25) = -0.46$, $p = 0.62$, $BF_{10} = 0.22$), and the effect of imagery vividness ratings was negative ($t(25) = -2.41$, $p = 0.023$; Figure S2E)—again at odds with the positive relationship predicted for an *RS* (Figure 2D). Finally, although we were able to effectively decode grating orientation from EVC activity both when the stimulus was present (perception) and absent (imagery), the extent of decoding did not relate to perceptual reality judgments (Figure S2F). Taken together, these results are consistent with EVC supporting mental imagery by suppressing irrelevant signals rather than enhancing relevant signals,¹⁵ similar

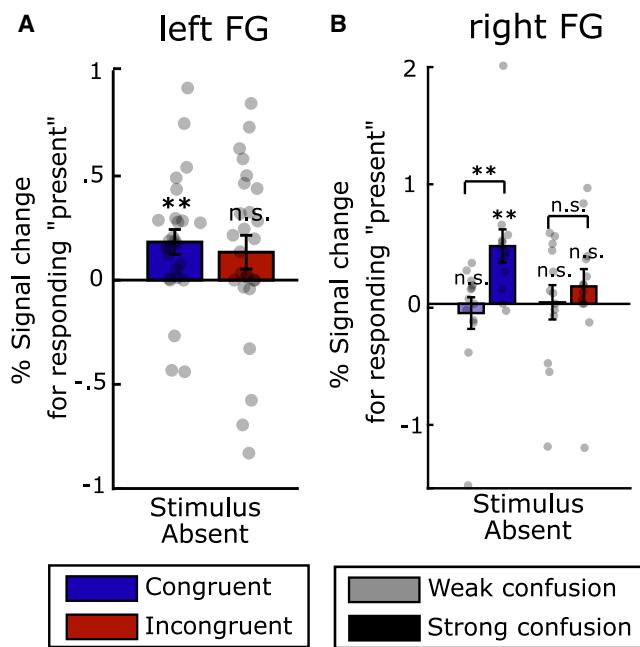


Figure 3. Predicting confusions between imagery and perception from FG

(A) The influence of reality false alarms (presence responses during stimulus-absent trials) on neural activity in the congruent condition in the left FG and (B) right FG, separately for participants with greater and fewer reality monitoring failures. Dots reflect individual participants ($N = 26$) and error bars reflect standard errors of the mean (SEM). In the cross-decoding analysis, only participants with enough trials in each condition were included ($n = 16$). Dots reflect individual participants, and error bars reflect standard errors of the mean (SEM). * $p < 0.05$; ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$.

to effects of attention. Importantly, these results indicate that although EVC activity is modulated by mental imagery, its activity profile is distinct from the RS signature identified in mid-level visual area FG (Figure 2), and it is therefore unlikely to be a candidate neural basis for distinguishing imagination from reality.

Predicting confusions between imagery and perception

If the FG indeed reflects an RS, we should be able to use activity in this region to predict when participants confuse their imagery for perception. In our paradigm, this would be reflected by participants claiming that a stimulus that was only imagined and not presented was in fact real—a reality false alarm. This confusion between imagery and perception should occur specifically in congruent, but not incongruent, imagery trials. To pursue this hypothesis, we tested for a parametric modulation of neural activity by reality judgment specifically during stimulus-absent trials, separately for congruent and incongruent conditions. In the left FG, there was indeed a significant modulation during false alarms for the congruent condition ($t(25) = 3.05, p = 0.005$) which did not reach significance in the incongruent condition ($t(25) = 1.66, p = 0.11, BF_{10} = 0.69$; Figure 3A). Again, in the right FG the effect was modulated by individual differences in the strength of the imagery congruency effect (median split: $t(24) = -2.91, p = 0.007$) such that reality monitoring errors

only significantly modulated right FG activity in participants who also showed stronger behavioral confusion ($t(12) = 3.45, p = 0.005$) but did not reach significance in participants who showed weaker confusion ($t(12) = -0.59, p = 0.56, BF_{10} = 0.37$; Figure 3B).

Taken together, our findings reveal FG as a candidate neural basis for a perceptual RS: activity here tracks both reality judgments as well as imagery vividness ratings, is sensitive to external stimuli as well as internal imagery congruency, and predicts confusions between imagery and perception, especially in participants who show strong behavioral confusions between the two.

A network for perceptual reality decisions

Both our model fits and analyses of FG activity patterns reveal that the latent variable underpinning an RS varies along a continuum. This continuous latent variable needs to be classified into a binary reality judgment in order to effectively navigate and control behavior. To explore neural substrates supporting binary reality judgments, we sought to identify brain areas in a whole-brain analysis that were better characterized by binary profiles tracking the trial-by-trial decisions of the participant, as opposed to a graded RS. We found a set of frontal and subcortical regions that showed a stronger modulation by reality judgments compared with imagery vividness: bilateral anterior insula, bilateral caudate nucleus, and the dorsomedial prefrontal cortex (Figure 4A; Table S3). As expected from candidate neural substrates of reality judgments, these regions all showed stronger activity during presence compared with absence responses, irrespective of actual stimulus presence and irrespective of imagery congruency (Figure 4B, all effect of stimulus presence and congruency $p > 0.19$, all $BF_{10} < 0.61$). Intriguingly, although the univariate signals in these regions were better captured by a binary reality judgment, we found that signatures of a graded RS were also maintained in the multivariate activity patterns in the bilateral anterior insula and dmPFC (Figure S4). This suggests that these regions may act to transform a RS originating in visual areas into an explicit reality judgment—a hypothesis we sought to explicitly test by examining functional connectivity.

If one or more of these brain areas support a transformation of a continuous RS in the FG into a binary reality judgment, we would expect significant functional coupling between these regions and the FG. Model simulations show that although stimulus presence and imagery congruency both influence mean levels of the RS and reality judgments, within-condition covariation between the RS variable and thresholded reality judgments remains consistently high across all conditions (Figure 4C). We found that this predicted pattern of covariation was uniquely expressed in the bilateral anterior insula, which showed significant connectivity with the FG in all conditions (Figure 4D; Figure S3). This connectivity profile was not observed in the caudate or dmPFC (Figure 4D), despite these regions showing a binary activity profile (Figure 4B). Together, these results reveal anterior insula as a candidate neural substrate for downstream evaluation of an FG RS signal when forming binary reality judgments, consistent with higher-order contributions to distinguishing reality from imagination via monitoring the activity of sensory cortices.^{16–19}

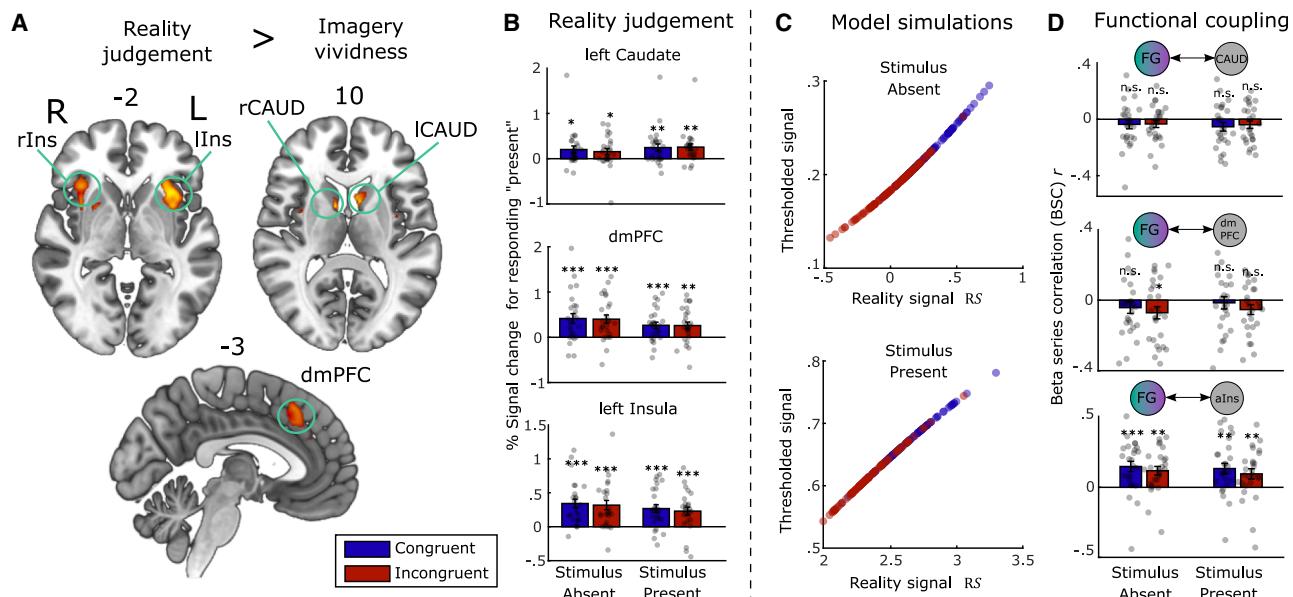


Figure 4. Frontal brain areas supporting binary reality judgments

(A) Brain areas showing a stronger parametric modulation by reality judgments than by imagery vividness ratings. Images are thresholded at $p < 0.05$, FWE corrected at the cluster level, with a cluster-forming threshold of $p < 0.001$, uncorrected.

(B) Effect of binary reality judgment regressor in each ROI showing significant effects in all conditions.

(C) Model simulations characterizing a relationship between reality signal (RS) and reality judgments, revealing that the covariation between the RS and judgment remains constant within each condition, despite mean shifts between conditions.

(D) Functional coupling between the left FG and the left caudate, dmPFC, and left anterior insula, separated by condition. Significant coupling in all conditions was observed with the anterior insula but not for any other ROI. Ins, anterior insula; CAUD, caudate; dmPFC, dorsomedial prefrontal cortex. Dots reflect individual participants, and error bars reflect standard errors of the mean (SEM). * $p < 0.05$; ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$.

DISCUSSION

In recent years, human neuroscience has revealed substantial overlap between the brain systems recruited for perception and imagination.^{2–4,7,9,10,20} Beyond the earliest sensory processing stages, neuronal activity is a mixture of internal and external signals. Given that downstream regions supporting higher-order cognition do not have direct access to these earliest processing stages, there is a potential for confusing internal and external signals. Here, we discover a remarkably simple mechanism for keeping track of reality in the face of this overlap. We show that a latent variable sensitive to combinations of both imagery and perceptual signals was uniquely tracked in the activity of the FG, enabling downstream brain regions to efficiently distinguish reality from imagination.

Due to the absence of bottom-up activity during imagery, activity in the visual cortex tends to be weaker than during perception.^{10,21} Therefore, tracking the combined strength of sensory signals may be an efficient way of leveraging existing perceptual systems for internal simulation, while still maintaining a grip on external reality. One striking feature of our results is that a RS was underpinned by activity in higher- rather than lower-level visual regions, despite the stimulus features (grating orientations) being first represented earlier in the visual pathway (see Figure S2 for analyses of EVC).^{3,22} It is possible that perceptual reality monitoring may function more effectively at the level of objects rather than low-level stimulus features,²³ as objects are central

to both navigation and (internal) simulation and planning.^{24–26} This view is also in line with a consistent role for the FG in supporting a range of visual mental imagery contents.²² Future research should further investigate this by documenting the neural substrates of RSs across different types of stimuli.

The additive effect of imagery on perception reported here and elsewhere^{12,13,27,28} suggests that the brain does not fully discount self-generated signals during imagery. Previous accounts have suggested that the volitional control associated with self-generated signals is used by the brain to "tag" self-generated signals as internal.^{18,29,30} In turn, it has been suggested that hallucinations are caused by disruptions in this mechanism, leading to the erroneous ascription of self-generated signals to external reality.^{31–33} However, if the control signals associated with imagery led subjects to ascribe congruent sensory signals to imagination, we would have expected a suppressive, not an additive, effect of congruent imagery on perception. Furthermore, in previous studies we found that congruent imagery also increased perceptual reality judgments relative to a condition in which both imagery and associated control signals were absent.^{12,14} Together, these results suggest that control signals associated with self-generated perception are not enough to distinguish imagination from reality. However, it is possible that these control signals will play a role in different experimental scenarios and/or in naturalistic settings. By explicitly manipulating the degree of control associated with imagery, future research could investigate its effect on perceptual reality monitoring.

Our model and data indicate that a *RS* is thresholded to allow for a binary classification of reality (vs. imagination). These binary classifications were manifest in the activity profiles of a frontal network of brain regions, including dmPFC and the anterior insula. These regions showed a global increase in univariate activation when a reality judgment was made and also contained multivariate signatures of a graded *RS*. Furthermore, we found significant functional connectivity between the FG and the anterior insula cortex, in line with the idea that the anterior insula “reads out” a *RS* in the FG. Intriguingly, a similar frontal network has been observed in relation to auditory hallucinations^{34,35} (Figure S3C)—consistent with dysfunction in this network leading to reality monitoring failures and with a broader role in forming an internal model of perceptual reality.^{18,36,37} In our study, a latent *RS* in the visual cortex best distinguished imagination from reality. In other contexts—including those using virtual reality techniques—other inputs may drive inferences on reality within this network. Our study lays the foundations for further research characterizing a generalized perceptual reality monitoring system in the human brain.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to, and will be fulfilled by, the lead contact, Nadine Dijkstra (n.dijkstra@ucl.ac.uk).

Materials availability

Experimental code for this experiment can be downloaded from <https://github.com/ImagineRealityLab/NIMADET/tree/main/Experiment>.

Data and code availability

- All behavioral data are openly available at <https://github.com/ImagineRealityLab/NIMADET>.
- All normalized and anonymized fMRI data are openly available at <https://osf.io/5x7be/> (first half) and <https://osf.io/ypwq9/> (second half).
- All code for the experiment and analyses is openly available at: <https://github.com/ImagineRealityLab/NIMADET>.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

ACKNOWLEDGMENTS

The authors would like to thank Alexane Leclerc for her piloting work on the experimental design and AI Powers for sharing the contrast map depicted in Figure S4. N.D. was supported by the European Union’s Horizon 2020 research and innovation programme under the Marie-Skłodowska Curie grant agreement no. 882832 and a European Research Council (ERC) starting grant (587992). P.K. was supported by a Wellcome/Royal Society Sir Henry Dale Fellowship (218535/Z/19/Z) and an ERC starting grant (948548). S.M.F. is a CIFAR Fellow in the Brain, Mind and Consciousness Program and is supported by UK Research and Innovation (UKRI) under the UK Government’s Horizon Europe funding guarantee (selected as ERC Consolidator, grant number 101043666). The Wellcome Centre for Human Neuroimaging was supported by core funding from the Wellcome Trust (203147/Z/16/Z).

AUTHOR CONTRIBUTIONS

Conceptualization, N.D., P.K., and S.M.F.; data curation, N.D. and T.v.R.; formal analysis, N.D. and T.v.R.; funding acquisition, N.D. and S.M.F.; investigation, N.D. and T.v.R.; methodology, N.D., P.K., and S.M.F.; project administration, N.D. and T.v.R.; software, N.D.; resources, N.D., P.K., and S.M.F.; validation, N.D.; visualization, N.D.; writing – original draft, N.D.; writing – review and editing, N.D., P.K., and S.M.F.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
 - Human participants
- **METHOD DETAILS**
 - Experimental procedure and design
 - Stimuli
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Behavioral analyses
 - Computational models
 - fMRI pre-processing
 - General linear models
 - Multivariate analyses
 - Functional connectivity analyses

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.neuron.2025.05.015>.

Received: December 5, 2024

Revised: April 15, 2025

Accepted: May 14, 2025

Published: June 5, 2025

REFERENCES

1. Dijkstra, N., Bosch, S.E., and van Gerven, M.A.J. (2017). Vividness of visual imagery depends on the neural overlap with perception in visual areas. *J. Neurosci.* 37, 1367–1373. <https://doi.org/10.1523/JNEUROSCI.3022-16.2016>.
2. Albers, A.M., Kok, P., Toni, I., Dijkerman, H.C., and De Lange, F.P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Curr. Biol.* 23, 1427–1431. <https://doi.org/10.1016/j.cub.2013.05.065>.
3. Lee, S.-H., Kravitz, D.J., and Baker, C.I. (2012). Disentangling visual imagery and perception of real-world objects. *NeuroImage* 59, 4064–4073. <https://doi.org/10.1016/j.neuroimage.2011.10.055>.
4. Ragni, F., Tucciarelli, R., Andersson, P., and Lingnau, A. (2020). Decoding stimulus identity in occipital, parietal and inferotemporal cortices during visual mental imagery. *Cortex* 127, 371–387. <https://doi.org/10.1016/j.cortex.2020.02.020>.
5. Naselaris, T., Olman, C.A., Stansbury, D.E., Ugurbil, K., and Gallant, J.L. (2015). A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *NeuroImage* 105, 215–228. <https://doi.org/10.1016/j.neuroimage.2014.10.018>.
6. Kamitani, Y., and Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nat. Neurosci.* 8, 679–685. <https://doi.org/10.1038/nn1444>.
7. Reddy, L., Tsuchiya, N., and Serre, T. (2010). Reading the mind’s eye: decoding category information during mental imagery. *NeuroImage* 50, 818–825. <https://doi.org/10.1016/j.neuroimage.2009.11.084>.
8. Senden, M., Emmerling, T.C., van Hoof, R., Frost, M.A., and Goebel, R. (2019). Reconstructing imagined letters from early visual cortex reveals tight topographic correspondence between visual mental imagery and perception. *Brain Struct. Funct.* 224, 1167–1183. <https://doi.org/10.1007/s00429-019-01828-6>.

9. Dijkstra, N., Bosch, S.E., and van Gerven, M.A.J. (2019). Shared neural mechanisms of visual perception and imagery. *Trends Cogn. Sci.* 23, 423–434. <https://doi.org/10.1016/j.tics.2019.02.004>.
10. Pearson, J. (2019). The human imagination: the cognitive neuroscience of visual mental imagery. *Nat. Rev. Neurosci.* 20, 624–634. <https://doi.org/10.1038/s41583-019-0202-9>.
11. Dijkstra, N., and Fleming, S.M. (2023). Subjective signal strength distinguishes reality from imagination. *Nat. Commun.* 14, 1627. <https://doi.org/10.1038/s41467-023-37322-1>.
12. Dijkstra, N., Mazor, M., Kok, P., and Fleming, S. (2021). Mistaking imagination for reality: Congruent mental imagery leads to more liberal perceptual detection. *Cognition* 212, 104719. <https://doi.org/10.1016/j.cognition.2021.104719>.
13. Dijkstra, N., Kok, P., and Fleming, S.M. (2022). Imagery adds stimulus-specific sensory evidence to perceptual detection. *J. Vis.* 22, 11. <https://doi.org/10.1167/jov.22.2.11>.
14. Dijkstra, N., Mazor, M., and Fleming, S.M. (2024). Confidence ratings do not distinguish imagination from reality. *J. Vis.* 24, 13. <https://doi.org/10.1167/jov.24.5.13>.
15. Pace, T., Koenig-Robert, R., and Pearson, J. (2023). Different mechanisms for supporting mental imagery and perceptual representations: modulation versus excitation. *Psychol. Sci.* 34, 1229–1243. <https://doi.org/10.1177/09567976231198435>.
16. Simons, J.S., Garrison, J.R., and Johnson, M.K. (2017). Brain mechanisms of reality monitoring. *Trends Cogn. Sci.* 21, 462–473. <https://doi.org/10.1016/j.tics.2017.03.012>.
17. Lau, H. (2019). Consciousness, Metacognition, & Perceptual Reality Monitoring. Preprint at PsychArxiv. <https://doi.org/10.31234/osf.io/ckbyf>.
18. Dijkstra, N., Kok, P., and Fleming, S.M. (2022). Perceptual reality monitoring: Neural mechanisms dissociating imagination from reality. *Neurosci. Biobehav. Rev.* 135, 104557. <https://doi.org/10.1016/j.neubiorev.2022.10.4557>.
19. van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., and Roelfsema, P.R. (2018). The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science* 360, 537–542. <https://doi.org/10.1126/science.aar7186>.
20. Zeidman, P., and Maguire, E.A. (2016). Anterior hippocampus: The anatomy of perception, imagination and episodic memory. *Nat. Rev. Neurosci.* 17, 173–182. <https://doi.org/10.1038/nrn.2015.24>.
21. Koenig-Robert, R., and Pearson, J. (2021). Why do imagery and perception look and feel so different? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 376, 20190703. <https://doi.org/10.1098/rstb.2019.0703>.
22. Spagna, A., Hajhajate, D., Liu, J., and Bartolomeo, P. (2021). Visual mental imagery engages the left fusiform gyrus, but not the early visual cortex: A meta-analysis of neuroimaging evidence. *Neurosci. Biobehav. Rev.* 122, 201–217. <https://doi.org/10.1016/j.neubiorev.2020.12.029>.
23. Passingham, R.E., and Lau, H.C. (2019). Acting, seeing, and conscious awareness. *Neuropsychologia* 128, 241–248. <https://doi.org/10.1016/j.neuropsychologia.2017.06.012>.
24. Prinz, J. (2011). Is Attention Necessary and Sufficient for Consciousness? In *Attention: Philosophical and Psychological Essays* (Oxford University Press), pp. 174–204.
25. Prinz, J. (2007). Accessed, accessible, and inaccessible: Where to draw the phenomenal line. *Behav. Brain Sci.* 30, 521–522. <https://doi.org/10.1017/S0140525X07003020>.
26. Prinz, J. (2012). The Conscious Brain (Oxford University Press). <https://doi.org/10.1093/acprof:oso/9780195314595.001.0001>.
27. Farah, M.J. (1985). Psychophysical evidence for a shared representational medium for mental images and percepts. *J. Exp. Psychol. Gen.* 114, 91–103. <https://doi.org/10.1037/0096-3445.114.1.91>.
28. Moseley, P., Smailes, D., Ellison, A., and Fernyhough, C. (2016). The effect of auditory verbal imagery on signal detection in hallucination-prone individuals. *Cognition* 146, 206–216. <https://doi.org/10.1016/j.cognition.2015.09.015>.
29. Haggard, P. (2017). Sense of agency in the human brain. *Nat. Rev. Neurosci.* 18, 196–207. <https://doi.org/10.1038/nrn.2017.14>.
30. Waters, F., Barnby, J.M., and Blom, J.D. (2021). Hallucination, imagery, dreaming: reassembling stimulus-independent perceptions based on Edmund Parish's classic misperception framework. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 376, 20190701. <https://doi.org/10.1098/rstb.2019.0701>.
31. Fletcher, P.C., and Frith, C.D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat. Rev. Neurosci.* 10, 48–58. <https://doi.org/10.1038/nrn2536>.
32. Frith, C.D., Blakemore, S.J., and Wolpert, D.M. (2000). Abnormalities in the awareness and control of action. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 355, 1771–1788. <https://doi.org/10.1098/rstb.2000.0734>.
33. Frith, C. (2005). The neural basis of hallucinations and delusions. *C. R. Biol.* 328, 169–175. <https://doi.org/10.1016/j.crvi.2004.10.012>.
34. Zmigrod, L., Garrison, J.R., Carr, J., and Simons, J.S. (2016). The neural mechanisms of hallucinations: A quantitative meta-analysis of neuroimaging studies. *Neurosci. Biobehav. Rev.* 69, 113–123. <https://doi.org/10.1016/j.neubiorev.2016.05.037>.
35. Powers, A.R., Mathys, C., and Corlett, P.R. (2017). Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. *Science* 357, 596–600. <https://doi.org/10.1126/science.aan3458>.
36. Nguyen, V.T., Breakspear, M., Hu, X., and Guo, C.C. (2016). The integration of the internal and external milieu in the insula during dynamic emotional experiences. *NeuroImage* 124, 455–463. <https://doi.org/10.1016/j.neuroimage.2015.08.078>.
37. Michel, M. (2017). A role for the anterior insular cortex in the global neuronal workspace model of consciousness. *Conscious. Cogn.* 49, 333–346. <https://doi.org/10.1016/j.concog.2017.02.004>.
38. Marks, D.F. (1973). Visual imagery differences in the recall of pictures. *Br. J. Psychol.* 64, 17–24. <https://doi.org/10.1111/j.2044-8295.1973.tb01322.x>.
39. Mazor, M., Friston, K.J., and Fleming, S.M. (2020). Distinct neural contributions to metacognition for detecting, but not discriminating visual stimuli. *eLife* 9, e53900. <https://doi.org/10.7554/elife.53900>.
40. Morales, J., Lau, H., and Fleming, S.M. (2018). Domain-general and domain-specific patterns of activity supporting metacognition in human prefrontal cortex. *J. Neurosci.* 38, 3534–3546. <https://doi.org/10.1523/JNEUROSCI.2360-17.2018>.
41. Andersson, J.L.R., Hutton, C., Ashburner, J., Turner, R., and Friston, K. (2001). Modeling Geometric Deformations in EPI Time Series. *NeuroImage* 13, 903–919. <https://doi.org/10.1006/nimg.2001.0746>.
42. Sladky, R., Friston, K.J., Tröstdl, J., Cunningham, R., Moser, E., and Windischberger, C. (2011). Slice-timing effects and their correction in functional MRI. *NeuroImage* 58, 588–594. <https://doi.org/10.1016/j.neuroimage.2011.06.078>.
43. Ashburner, J., and Friston, K.J. (2005). Unified segmentation. *NeuroImage* 26, 839–851. <https://doi.org/10.1016/j.neuroimage.2005.02.018>.
44. Mumford, J.A., Poline, J.-B., and Poldrack, R.A. (2015). Orthogonalization of Regressors in fMRI Models. *PLoS One* 10, e012625. <https://doi.org/10.1371/journal.pone.0126255>.
45. Mumford, J.A., Turner, B.O., Ashby, F.G., and Poldrack, R.A. (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *NeuroImage* 59, 2636–2643. <https://doi.org/10.1016/j.neuroimage.2011.08.076>.
46. Rissman, J., Gazzaley, A., and D'Esposito, M. (2004). Measuring functional connectivity during distinct stages of a cognitive task. *NeuroImage* 23, 752–763. <https://doi.org/10.1016/j.neuroimage.2004.06.035>.
47. Cisler, J.M., Bush, K., and Steele, J.S. (2014). A comparison of statistical methods for detecting context-modulated functional connectivity in fMRI. *NeuroImage* 84, 1042–1052. <https://doi.org/10.1016/j.neuroimage.2013.09.018>.

STAR★METHODS

KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|--------------------------------|---|---|
| Deposited data | | |
| Behavioural data | Database: https://github.com/ImagineRealityLab/NIMADET | N/A |
| fMRI data | Database: https://osf.io/5x7be/ (first half) and https://osf.io/ypwq9/ (second half) | N/A |
| Software and algorithms | | |
| Custom code | Nadine Dijkstra | https://doi.org/10.5281/zenodo.15348809 |
| SPM | Wellcome Centre for Human Neuroimaging, Institute of Neurology, UCL | RRID:SCR_007037 |

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Human participants

Thirty-six participants gave written informed consent and participated in the study. Ten participants were excluded from the final analyses: two due to repeatedly failing an imagery check (see below), two due to excessive movement in the scanner, five due to repetition of the same vividness rating, and one due to withdrawing from the study. Our final sample consisted of 26 participants (mean age = 25.92, SD = 3.88). The study was approved by the University College London ethics committee (approval number 8213_001). Participants were compensated £45 for their participation. In line with the GDPR principle of Data Minimisation (only necessary data should be collected and processed), we refrained from collecting information on participants' sex, gender, socioeconomic status, ancestry, race and ethnicity.

METHOD DETAILS

Experimental procedure and design

Prior to entering the scanner, participants completed a behavioural practice session lasting 45 minutes. Participants started by filling out the consent form and the Vividness of Visual Imagery Questionnaire (VVIQ2³⁸). The task was then explained in stages. For each participant, two random grating orientations out of 135°, 75° or 15° were chosen to represent grating A and grating B throughout the experiment. However, due to randomization issues, 29/31 participants received 15° as grating A and 135° as grating B. First, participants were asked to practice detecting these gratings. A separate staircase was then run for each of the two gratings to determine the 70% detection threshold. Accuracy was evaluated after each block of 6 trials for 8 staircase steps in total. If accuracy dropped below 60%, visibility was increased, and if it exceeded 80%, visibility was decreased. The step change was relative to the discrepancy in accuracy to threshold. Visibility levels were averaged over the two staircases and the same level was used for both gratings in the main experiment. Next, participants practiced imagining the gratings within dynamic noise. During this practice task, participants were asked to imagine one orientation per block. After each trial, they indicated how vivid their imagery was and performed an orientation discrimination task, comparing the imagined grating orientation with a probe orientation. Participants were only allowed to continue after they had completed five correct trials in a row. Finally, the main task was practiced for 24 trials. After participants entered the scanner, and prior to starting the main task, another staircase was run to ensure that the visibility level was correctly calibrated to the scanner environment.

The main experimental paradigm is depicted in Figure 1A. Participants were asked to detect gratings that gradually appeared within dynamic noise until threshold while simultaneously imagining either the same grating (congruent condition) or a grating orthogonal to the one they were detecting (incongruent condition). In order to avoid visual priming, no trial-wise cues were delivered. Instead, the identity of the grating that should be imagined was cued in a block-wise fashion. Within each imagery block, the identity of the to-be-detected grating changed halfway (i.e. two detection mini-blocks per imagery block), which was instructed separately via an additional instruction screen. This mini-block structure was implemented to ensure that our effect of interest (congruency between imagery and detection) fluctuated at a frequency that remained detectable following low-pass filtering of the fMRI BOLD signal. The starting imagery and detection orientations were randomized, after which the orientations alternated over blocks.

Each trial started with a 200ms fixation cross followed by 2s of either pure dynamic white noise (grating-absent trial) or dynamic white noise within which a gradually appearing stimulus was embedded (grating-present trial). A grating was present in 50% of the trials. After the stimulus, participants first indicated whether or not a grating had been presented on the screen and then how vivid

their imagery was. To accurately infer whether a grating had been presented on the screen, imagined and perceived signals have to be kept apart, which is especially difficult in the congruent condition. To ensure that participants imagined the correct grating, after each imagery block, participants indicated which grating they had imagined. Incorrectly answered blocks were removed prior to analysis and participants with less than 75% correct blocks were excluded completely from analysis. On average, 0.9 (SD = 1.1) out of 16 blocks were removed.

After the main experiment, participants performed a functional localizer task inside the scanner. The localizer task consisted of gratings flickering at 2Hz, presented at 100% contrast. Each block contained gratings with a fixed orientation (135°, 75° or 15°). The orientations were presented in a pseudorandom order followed by an ~14.3 s blank screen containing only a fixation bull's-eye. Participants were tasked with responding when the black fixation dot briefly dimmed to ensure central fixation. All participants were presented with 16 localizer blocks, which totalled ~15 min.

Stimuli

The stimuli were generated in MATLAB (version R2012b) and consisted of sinusoidal gratings tilted at an orientation of 135°, 75° or 15°, masked with an annulus and embedded in white noise. The visibility of the stimuli was manipulated by changing the probability that a given pixel was replaced by a random value. On each stimulus-present trial, stimulus visibility followed a logistic ramp function to threshold by generating a new image at a different visibility level at 30Hz (60 images over 2 seconds). A similar procedure was used for stimulus-absent trials, but in this case each image consisted of pure noise.

QUANTIFICATION AND STATISTICAL ANALYSIS

Behavioral analyses

We first tested whether imagery influenced perception by performing a repeated-measures ANOVA with imagery condition (congruent vs incongruent) and stimulus (absent vs present) as independent variables and perceptual detection rate as the dependent variable. To test whether perception influenced imagery responses we performed another repeated-measures ANOVA with condition (congruent vs incongruent) and stimulus (absent vs present) as independent variables and imagery vividness rating as the dependent variable.

Computational models

Computational models were adapted from Dijkstra and Fleming.¹¹ We assume that the observer's reality judgement RJ and their imagery vividness V are both a function of two random variables, a perceptual sample P and an imagery sample I :

$$RJ = f(P, I)$$

$$V = g(P, I)$$

In keeping with standard signal detection theory (SDT) approaches, the perceptual sample is drawn from a bivariate Gaussian with mean μ_P and covariance Σ_P :

$$P \sim N(\mu_P, \Sigma_P)$$

where μ_P is set to the participant's detection sensitivity d' specific to that condition such that $\mu_P = [d' \ 0]$ for stimulus-present trials in the congruent condition, $\mu_P = [0 \ d']$ for stimulus-present trials in the incongruent condition, and $\mu_P = [0 \ 0]$ for stimulus-absent trials.

The imagery sample is drawn from a bivariate Gaussian with mean μ_I and covariance Σ_I :

$$I \sim N(\mu_I, \Sigma_I)$$

where $\mu_I = [V_m \ 0]$. V_m reflects the participant's mean vividness rating, averaged over all trials. Σ_P and Σ_I are both set to the identity matrix.

According to a null-model (M_0) that assumes a neural separation between imagery and perception, reality judgements RJ and imagery vividness ratings V are each based solely on the perception and imagery samples, respectively:

$$\begin{aligned} M_0 \quad RJ &= f(\alpha P) \\ V &= g(\gamma I) \end{aligned} \tag{Equation 1}$$

In contrast, according to the reality threshold model (M_1), perception P and imagery I samples are combined to create a reality signal RS which determines both RJ and V :

$$\begin{aligned} M_1 \quad RS &= \alpha P + \gamma I \\ RJ &= f(RS) \\ V &= g(RS) \end{aligned} \tag{Equation 2}$$

where the α and γ parameters scale the contribution of the perception and imagery signals respectively. We also consider two alternative models in which imagery are intermixed in non-linear ways. M_2 is a multiplicative model:

$$\begin{aligned} M_2 \quad RS &= \alpha P \times \gamma I \\ RJ &= f(RS) \\ V &= g(RS) \end{aligned} \quad (\text{Equation 3})$$

where imagery and perception signals are multiplied, creating supra-linear effects, and M_3 is a sublinear model:

$$\begin{aligned} M_3 \quad RS &= (\alpha P + \gamma I + c)^{\theta} \\ RJ &= f(RS) \\ V &= g(RS) \end{aligned} \quad (\text{Equation 4})$$

where imagery and perception signals are intermixed in a sub-linear way. c is a constant that ensures that the RS is positive and prevents complex numbers when raising the sum to the power θ . In the sublinear model (M_3), θ is restricted between 0 and 1, leading to sublinear growth.

For all models, reality judgments RJ are made by comparing the underlying latent variable to a reality threshold T , such that if $f(x) > T$ then a trial is classified as real. Vividness ratings V are assumed to be linearly related to the underlying random variable via $g(x)$. To facilitate model fitting and simulation, we implemented $f(x)$ as a logistic regression $\frac{1}{1+e^{(x-T)/s}}$ giving the probability that x leads to a binary reality judgement, and $g(x)$ as a multinomial ordinal regression $\log\left(\frac{P(Y>k|x)}{P(Y\leq k|x)}\right) = \theta_k - x^T \beta$ giving the cumulative probability that x leads to a vividness rating of k or lower. α , γ and the parameters of both regression equations are found through fitting the models to the observed behavioural data. Model comparison showed that the additive reality threshold model M_1 provided the best fit for the data (Figures 1 and S1).

For neuroimaging analyses, RS signals were simulated per participant based on model parameters fitted to the behavioural data, and extracted per condition and per trial type as the intermixed signal along the dimension relevant for the reality judgement in that specific condition.

fMRI pre-processing

Data pre-processing followed the procedures described in earlier studies from our lab.^{11,39,40} fMRI analysis was performed using SPM12 (Statistical Parametric Mapping: www.fil.ion.ucl.ac.uk/spm). The first five volumes of each run were discarded to allow for T1 stabilization. Functional images were realigned and unwarped using local field maps⁴¹ and then slice-time corrected.⁴² Each participant's structural image was segmented into gray matter, white matter, CSF, bone, soft tissue, and air/background images using a nonlinear deformation field to map it onto template tissue probability maps.⁴³ This mapping was applied to both structural and functional images to create normalized images in Montreal Neurological Institute (MNI) space. For univariate analyses, normalized images were additionally spatially smoothed using a Gaussian kernel (6 mm FWHM). We set a within-run 4 mm affine motion cut-off criterion. Pre-processing and construction of first- and second-level models used standardized pipelines and scripts available at <https://github.com/metacoglab/MetaLabCore/>.

General linear models

The presented stimulus orientation as well as the congruency between the presented and imagined orientation fluctuated every ~200s (i.e. at around 1/200Hz). This means that a standard high pass filter (HPF) of 1/128Hz would remove some of the effects of interest. Therefore, for all first-level general linear models (GLMs), we lowered the HPF to 1/256Hz. For all GLMs, we included the following nuisance regressors: the detection screen with a duration until response, the detection response as a delta function, the vividness screen until response, the vividness response as a spike regressor, and the stimulus presentation during the incorrect imagery blocks.

We first defined a GLM based on the manipulated factors of our experimental design with one regressor per congruency condition (congruent vs incongruent) x stimulus presentation (present x absent). Regressors were defined per run and contrasts were concatenated over runs. To test for effects of imagery-perception congruency, we contrasted all congruent regressors against all incongruent regressors. Incorrect imagery trials were removed from condition-specific regressors of interest and added as nuisance regressors.

Next, to investigate which brain regions were modulated by participants' behavioural responses, we defined a GLM that also included the vividness ratings and detection judgements as parametric modulators. To maximize power to detect variation in behavioural responses, we concatenated the runs for this GLM using SPM's 'spm_fmri_concatenate' function. This function adds single-session GLM block regressors, and adjusts the high-pass filter and temporal non-sphericity calculations to account for the original session length. The values for the parametric regressors were mean centred per experimental cell to account for the effects of congruency and stimulus presentation on the behavioural responses but were not orthogonalized with respect to each other to prevent a bias in ascribing variation to either one or the other predictor.⁴⁴

Multivariate analyses

Multivariate analyses were performed on single-trial betas based on the normalized, non-smoothed data. Betas were estimated per trial for the localizer and task data separately through a GLM which included one regressor for that trial and one regressor for all other trials.⁴⁵ To identify potential multivariate signatures of the reality signal *RS* we employed a step-wise approach. We first identified brain regions coding for graded vividness ratings via a whole-brain searchlight decoding analysis. The searchlight radius was set to 4 voxels, leading to a searchlight size of 257 voxels. To decode the graded vividness ratings, we used a linear regression classifier with ridge regression regularization. We trained the classifier on data from all runs but one as follows:

$$\beta = \frac{(X^T \times X + \lambda \times I_k)}{(X^T \times V)}$$

In this equation, X is a $n \times k$ matrix where n is the number of trials and k is the number of voxels, λ is a regularization parameter (set to 10^5), I_k is the identity matrix over voxels, and V is a $n \times 1$ vector containing the vividness ratings corrected for main effects of congruency and stimulus presence. This classifier was tested on the data from the held-out run to generate predicted vividness V' :

$$V' = X \times \beta$$

To evaluate decoding accuracy, we calculated the Spearman correlation between the predicted and observed vividness ratings for each searchlight. Significance was determined via a one-sample t-test on the Spearman correlation over participants.

To further test which brain regions reflected a multivariate *RS*, we asked which of the significant vividness decoding searchlights ($p < 0.001$ uncorrected) also demonstrated other effects that activity patterns tracking an *RS* should show according to our model simulations (Figure 2C). Specifically, we asked whether these regions carried a signal covarying with a main effect of imagery-perception congruency, a main effect of stimulus presence, and a main effect of reality judgement. For each searchlight, we calculated the cross-validated predicted signal V' based on the vividness representation and then tested whether this predicted signal showed each of these main effects. We concluded that a region displayed a multivariate *RS* if it a) contained at least 50 searchlights with a significant vividness representation and b) showed a main effect of congruency AND a main effect of stimulus presence AND a main effect of reality judgements, all at $p < 0.05$ FDR corrected. These results are presented in Figure S4.

Functional connectivity analyses

We used a beta series correlation (BSC) approach to investigate the task-specific effects on functional connectivity between specified regions of interest,⁴⁶ which has been shown to have greater sensitivity to detect functional coupling in event-related designs.⁴⁷ Single-trial betas were estimated using the one GLM per trial approach (see above). Per ROI, activation time-series were created by sorting the trial betas into the different conditions and then calculating the mean beta value over voxels, per trial, creating a beta series per ROI. Functional connectivity was then estimated by computing the pair-wise correlation between the beta-series of two ROIs.