

Visual features drive the category-specific impairments on categorization tasks in a patient with object agnosia

Noor Seijdel^{a,b,*}, H. Steven Scholte^{a,b}, Edward H.F. de Haan^{a,b}

^a Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands

^b Amsterdam Brain & Cognition (ABC) Center, University of Amsterdam, Amsterdam, the Netherlands

ARTICLE INFO

Keywords:

Lesion study
Visual categorization
Category-specific impairments
Deep convolutional neural networks

ABSTRACT

Object and scene recognition both require mapping of incoming sensory information to existing conceptual knowledge about the world. A notable finding in brain-damaged patients is that they may show differentially impaired performance for specific categories, such as for “living exemplars”. While numerous patients with category-specific impairments have been reported, the explanations for these deficits remain controversial. In the current study, we investigate the ability of a brain injured patient with a well-established category-specific impairment of semantic memory to perform two categorization experiments: ‘natural’ vs. ‘manmade’ scenes (experiment 1) and objects (experiment 2). Our findings show that the pattern of categorical impairment does not respect the natural versus manmade distinction. This suggests that the impairments may be better explained by differences in visual features, rather than by category membership. Using Deep Convolutional Neural Networks (DCNNs) as ‘artificial animal models’ we further explored this idea. Results indicated that DCNNs with ‘lesions’ in higher order layers showed similar response patterns, with decreased relative performance for manmade scenes (experiment 1) and natural objects (experiment 2), even though they have no semantic category knowledge, apart from a mapping between pictures and labels. Collectively, these results suggest that the direction of category-effects to a large extent depends, at least in MS’ case, on the degree of perceptual differentiation called for, and not semantic knowledge.

1. Introduction

Object or scene recognition requires mapping of incoming sensory information to existing conceptual knowledge about the world. A notable finding in brain-damaged patients is that they may show differentially impaired knowledge of, most prevalently, living things compared to non-living things (Gainotti, 2000). For many years, researchers have been investigating these category-specific semantic deficits. To date, the debate remains unsettled on how this distinction in breakdown of semantic knowledge along the natural/living versus manmade/non-living axis arises (Capitani et al., 2003; Gainotti, 2000; Young et al., 1989).

Some studies have suggested that evolutionary pressures have led to a specialized, distinct neural mechanism for different categories of knowledge (e.g. animals, plants and artefacts) (Caramazza and Shelton, 1998; Nielsen, 1946), and that category-specific deficits arise from damage to one of these distinct neural substrates. However, the most widespread views hold that they emerge because living and non-living

things have different processing demands (i.e. they rely on different types of information). The first (most dominant) of those theories assumes that the storage of semantic information is divided into parts dominated by different knowledge aspects (e.g. perceptual, functional) and proposes that the dissociation arises from a selective breakdown of perceptual compared to functional associative knowledge. While man-made objects have ‘clearly defined functions’ and are mostly differentiated by their functional qualities, animals have less defining functions and are mostly distinguishable in terms of their visual appearance (Warrington and Shallice (1984). This ‘differential weighting’ of perceptual and associative attributes might underlie the dissociation between living and non-living things. Later on this theory was revised to also include other modality-specific knowledge channels, such as a ‘motor-related’ channel, to support findings indicating greater impairments for certain more ‘motor-related’ or ‘manipulable’ items (such as tools or kitchen utensils) compared to larger manmade objects (such as vehicles) (Warrington and McCarthy, 1987).

A number of studies have emphasized the importance of

* Corresponding author. Psychology Department - Brain & Cognition, University of Amsterdam, Nieuwe Achtergracht 129b, 1018 XE, Amsterdam, the Netherlands.
E-mail address: noor.seijdel@gmail.com (N. Seijdel).

intercorrelations amongst individual semantic features. This intercorrelation theory states that concepts are represented as patterns of activation over multiple semantic properties within a *unitary distributed system*. This intercorrelation theory is appealing in that it does not rely on damage to specific subtypes of attribute (visual, associative, motor) to produce category-specific deficits (Caramazza et al., 1990; Caramazza and Shelton, 1998; Tyler and Moss, 2001). Still another account holds that living items contain a larger number of structurally similar exemplars (e.g. many different types of trees), requiring a more fine-grained visual analysis for successful recognition (Sartori et al., 1993). In other words, it could be inherently more difficult to visually recognize living things compared to non-living things. This view of the structural description system, and their account for category-specific impairments is consistent with work on normal subjects and animal studies (Gaffan and Heywood, 1993).

In line with these findings, a more recent study by Panis et al., (2017) suggested that category-specific impairments may be explained by a deficit in recurrent processing between different levels of visual processing in the inferotemporal cortex. According to them, category-specificity has a perceptual nature, and the direction can shift, depending on perceptual demand. High structural similarity between stored exemplars might be beneficial for integrating local elements and parts into whole representations because the global and local features of these exemplars are more stable and more highly correlated in the real-world than the features from categories with low structural similarity. At the same time however, high structural similarity may be harmful for matching or precise recognition operations, because there may be more competition between the activated representations (Gerlach, 2009).

Here it's important to note that different tasks have been used to evaluate patients' ability to recognize objects from different categories. Category-specific impairments have been established both using semantic memory experiments or visual recognition tasks at different levels (picture naming, picture-word matching, categorization). The differences in perceptual demand for these tasks (i.e. on which perceptual information they depend) might underlie the differences in category-specificity that have previously been found.

In the current study, we investigate the ability of a brain-injured patient, who is believed to have a category-specific impairment of

semantic memory to perform scene- and object-categorization tasks (Fig. 1). This patient, MS, has played a crucial role in the development of theories on category-specificity, showing a very clear category-specific deficit on semantic category fluency tests in previous studies. He has shown to perform better than control participants on non-living categories and significantly worse on living items (Young et al., 1989). A recent study showed that his impairments have remained unchanged for more than 40 years (de Haan et al., 2020). MS' problems with living items relative to non-living ones is apparent across a variety of tasks, including mental imagery, retrieval of information and visual recognition (Mehta et al., 1992). However, there is a striking dissociation between MS' preserved ability to access information about category membership in an implicit test (by priming identification of living and non-living items with related category labels), where there is no difference between the categories, and his severe problems in accessing such information in an explicit test. These findings suggest that it's an "access" rather than a "storage" problem. The question remains as to whether MS can access stored representations of visual stimuli and, if so, what are the relationships between perceptual demand, recognition and semantic memory.

In the current study, two types of questions were addressed. The first - in order to investigate whether or not the category-specific impairment is dependent on semantic category or perceptual factors - concerned MS' ability to categorize visual images as depicting naturalistic vs. manmade scenes (experiment 1) and objects (experiment 2). Our findings show a difference between the two tasks, with better performance for manmade objects, compared to naturalistic objects (as is usually the case), and better performance for natural scenes compared to manmade scenes.

The second question concerned the type of information that might underlie the observed behavior. There is a large body of research looking into semantic category-specific impairments using computational modeling and different types of implementing artificial damage (Guest et al., 2020). For example, studies using connectionist simulations have shown how the distinctiveness of functional features was related to perceptual features varying across semantic categories and how damaging may lead to patterns of impairments for different types of information across varying semantic categories (Tyler et al., 2000). Recently, a class of computational models, termed deep convolutional neural networks (DCNNs), inspired by the hierarchical architectures of

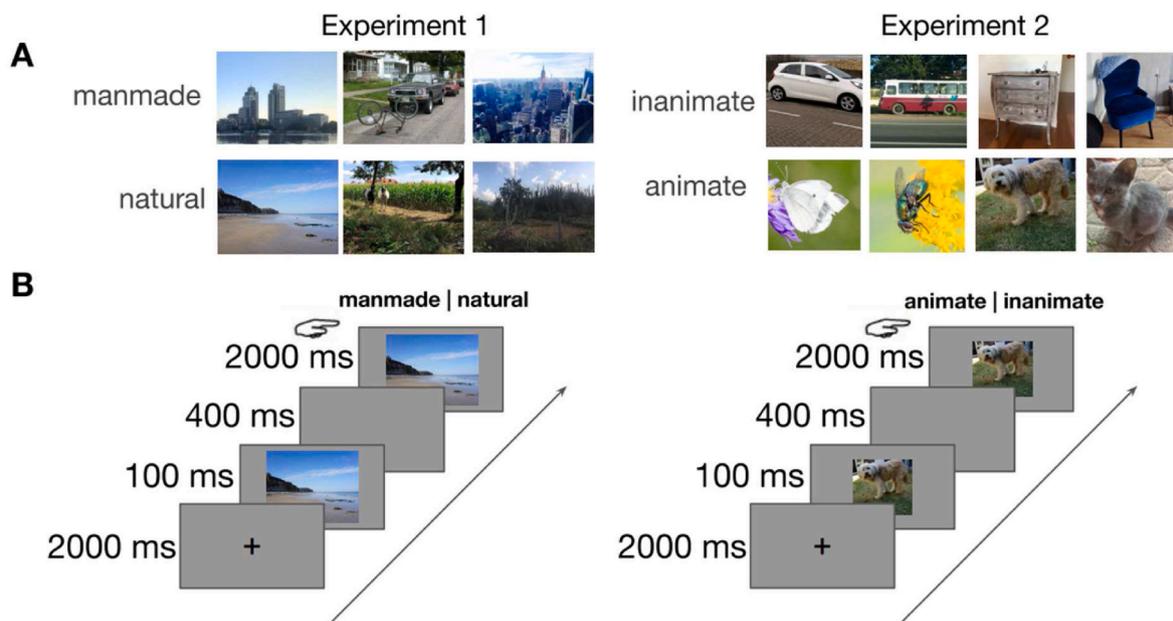


Fig. 1. Stimuli and experimental paradigm. A) Exemplars of the two categories in experiment 1 (manmade vs. natural scenes) and experiment 2 (manmade vs. natural objects). B) Experimental design. After a 2000 ms blank screen, the stimulus was shown for 100 ms, followed by a 400 ms blank screen. Then, the image reappeared, and MS was asked to categorize the stimulus by pressing the corresponding button.

ventral visual streams demonstrated striking similarities with the cascade of processing stages in the human visual system (Cichy et al., 2016; Güçlü and van Gerven, 2015; Khaligh-Razavi and Kriegeskorte, 2014)). In particular, it has been shown that internal representations of these models are hierarchically similar to neural representations in early visual cortex (V1–V3), mid-level (area v4), and high-level (area IT) cortical regions along ventral stream. Therefore, we evaluated performance of different DCNN architectures (all ResNets, with varying depth) and compared it to MS' behavior. Moreover, 'adding lesions' to higher-order layers of a DCNN, by removing certain blocks, resulted in similar response patterns with decreased performance for manmade scenes (experiment 1) and natural objects (experiment 2).

Altogether, results from the current study indicate that, at least in specific cases such as MS, category-specific impairments can be explained by perceptual aspects of exemplars within different categories, rather than semantic category-membership.

2. Materials and methods

2.1. Case history

MS is a former police cadet who contracted herpes encephalitis in 1970 (for a full case description see also (Newcombe and Ratcliff, 1975; Ratcliff and Newcombe, 1982)). Most of the ventral temporal cortex of both hemispheres was destroyed, extending to occipital cortex on the right, leaving him with a complete left homonymous hemianopia. He suffers from achromatopsia (Chadwick et al., 2019; Mollon et al., 1980), has severe object agnosia and prosopagnosia, but is able to read accurately. His comprehension of what he reads is affected by an impairment of semantic memory. His semantic memory impairment is more marked for living than for non-living things (Young et al., 1989; de Haan et al., 2020).

Anatomical scans (Smits et al., 2019) revealed an, at least partially, intact primary visual cortex (V1) in both hemispheres. Further inspection of the anatomical scan suggests that this part of cortex in the right hemisphere, that could consist of parts of V1 to V4, is disconnected from subsequent cortical areas.

2.2. Stimuli - scenes

240 images (640*480 pixels, full-color) of real-world scenes were obtained from a previous unpublished study by Chow-Wing-Bom et al. (2019). Of these 240 images, 80 images were labeled natural (>90% naturalness rating in an independent experiment), 80 images were man-made (<10% naturalness rating) and 80 images were ambiguous (between 10 and 90% naturalness rating). Ambiguous trials were collected for a different purpose and aren't analyzed in the current study. In total, 160 trials were thus used for further analysis. The stimulus set contained a wide variety of different outdoor scenes including beaches, mountains, forests, streets, buildings and parking lots. Most of the scenes (presented at ~12.3°) contained objects. For example, more than 50% of the manmade scenes contained a vehicle, mostly bicycles and cars. For the naturalistic scenes, more than 25% of the scenes contained an animal.

2.3. Stimuli - objects

80 images (512*512 pixels) of natural objects (animals: dogs, cats, butterflies and flies) and manmade objects (cars, busses, cabinets and chairs) were selected from several online databases, including MS COCO (Lin et al., 2014), the SUN database (Xiao et al., 2010), Caltech-256 (Griffin et al., 2007), Open Images V4 (Kuznetsova et al., 2020) and LabelMe (Russell et al., 2008). For each of the subcategories, ten images were shown at ~13°.

2.4. Experimental design

During the experiments, stimuli were presented for 100 ms, followed by a 500 ms blank screen. This rapid presentation was included for a different purpose (to evaluate EEG responses after a first brief presentation). Then, to allow for reliable behavioral responses, the stimulus reappeared for 2000 ms and MS was asked to categorize the image as accurately as possible using one of two corresponding response buttons. For the scenes, MS was asked to categorize images as being "manmade" or "natural". For the objects, the two categories were "animate" vs. "inanimate". Each image was shown only once in both experiments, resulting in 160 trials for experiment 1, and 80 trials in experiment 2. Stimuli were presented in a randomized sequence, at eye-level, in the center of a 23-inch ASUS TFT-LCD display (1920*1080 pixels, at a refresh rate of 60 Hz), while MS was seated approximately 70 cm from the screen. The task was programmed in- and performed using Presentation (Version 18.0, Neurobehavioral Systems Inc., Berkeley, CA, www.neurobs.com). After every 40 trials there was a short break. During the task, EEG was recorded.

2.5. Statistical analysis: behavioral data

Choice accuracy and reaction times were computed for each condition (Fig. 2). Differences between the conditions were tested using two-tailed permutation testing with 5000 permutations. In each iteration, we divided the permuted dataset into two datasets x and y (all items

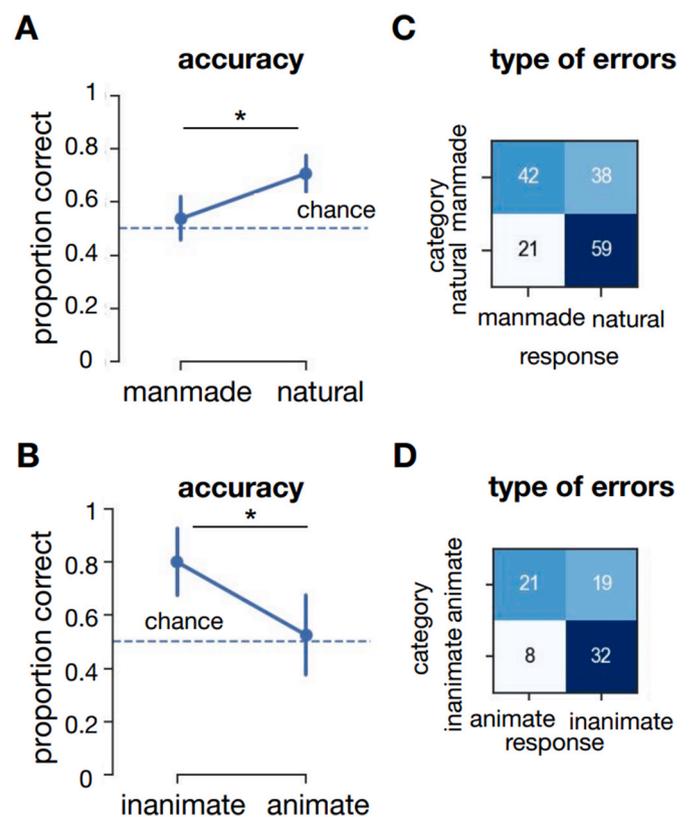


Fig. 2. Behavioral results of patient MS in experiment 1 and 2. A) Results from experiment 1 (manmade vs. natural scenes) indicating accuracy (proportion correct) per condition. B) Results from experiment 2 (manmade vs. natural objects). Horizontal black lines indicate the results of two sample permutation tests, two-tailed using 5000 permutations. Blue dashed line represents chance level. Error bars represent the bootstrap 95% confidence interval. * = $p < 0.05$. C/D) number of instances for each category-response combination, indicating the types of errors for both experiments. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

assigned randomly) and computed the difference (here: mean) of sample x and sample y . We repeated this until all permutations were evaluated and stored the differences. The p -value was computed by taking the number of times the stored differences were at least as extreme as the original difference, divided by the total number of permutations. Here, the p -value is defined as the probability, given that the null-hypothesis (no difference between the conditions) is true, that we obtain results that are at least as extreme as the results we observed. In each iteration, all samples were taken into account (resampling was dependent only on the assignment of values to condition groups). Behavioral data were analyzed in Python using the following packages: Statsmodels, SciPy, NumPy, Pandas and Seaborn (Jones et al., 2001; McKinney and Others, 2010; Oliphant, 2006; Seabold and Perktold, 2010).

2.6. Deep convolutional neural networks (DCNNs)

First, to evaluate how many layers were sufficient to accurately perform the categorization tasks, tests were conducted on four deep residual networks (ResNets (He et al., 2016);) with increasing number of layers; ResNet-6, ResNet-10, ResNet-18 and Resnet-34. We selected ResNets because they contain skip connections. Skip connections in deep architectures skip some layer in the neural network and feed the output of one layer as the input to the next layers (instead of only the next one). This allows us to ‘damage’ or remove one layer while still evaluating the output of the network. Pre-trained networks were fine-tuned (40 epochs) to perform either the manmade vs. natural scene categorization task, or the object categorization task, using PyTorch (Paszke et al., 2019). Each model was initialized five times with different seeds to perform statistical analyses. For ResNet10, the most shallow network that was able to successfully perform the task (>95% accuracy on all conditions), we evaluated categorization performance after ‘lesioning’ higher-order layers. To this end, we removed one of the ‘building blocks’, while keeping the skip connection intact.

3. Results

First, categorization performance (proportion correct) of MS was computed for both categorization tasks. Results from two-sample permutation tests with 5000 permutations indicated higher performance for natural scenes (experiment 1) and manmade objects (experiment 2) images ($p = 0.007$, $p = 0.016$, respectively). Thus, in the scene categorization task, MS was significantly better at classifying visually the natural compared to man-made environments. In contrast, on the object categorization task, he was significantly better at assigning the man-made objects to the correct category compared to the natural objects.

ResNet-10, -18 and -34 all showed virtually perfect performance for both tasks, for all categories (Fig. 3). For the most shallow network, ResNet-6, there was a slight decrease in performance, specifically for manmade scenes (experiment 1) and natural objects ($p = 0.02$, $p = 0.03$, respectively). Overall these results indicate that performance of a shallow ResNet-6 may decrease in a similar fashion as MS. This supports the idea that performance is decreased for specific categories because those stimuli (in our dataset) are more difficult. Still, even for a shallow ResNet-6, the two-option categorization tasks seem too easy.

Finally, we evaluated the performance of ResNet-10 after ‘lesioning’ higher-order layers (Fig. 4A). We focused on ResNet-10, because this network was the most shallow network to perform virtually perfect on both tasks (without lesion), indicating the network’s ability to categorize all conditions. In order to mimic lesions to higher-order areas in the visual processing stream, we removed connections to the final building block of the network (Block 4). Permutation tests with 5000 permutations between ResNet-10 without and with lesion, indicated a decrease in performance after elimination of higher-order layers, specifically for manmade (experiment 1) and natural (experiment 2) images (both $p < 0.001$). For natural scenes, there was a slight increase in performance after the removal of higher order layers ($p = 0.023$).

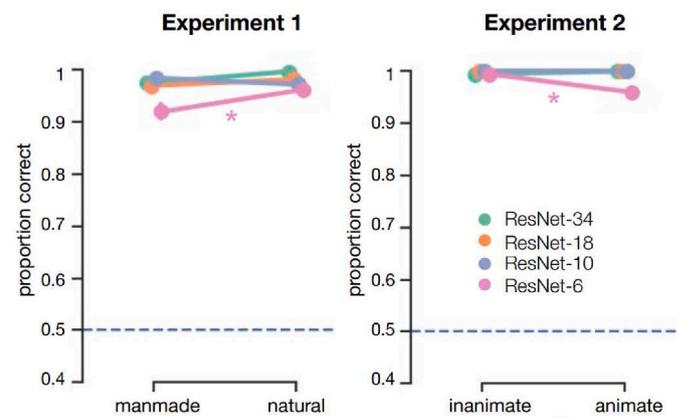


Fig. 3. Performance of ResNets with different depth (number of layers) on the images from experiment 1 and 2. The ResNets were pretrained on ImageNet, and fine-tuned on an independent set of manmade and natural scenes and images containing manmade and natural objects. Blue dashed line represents chance level. Error bars represent the bootstrap 95% confidence interval. * = $p < 0.05$. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

To assess whether the networks also perform worse for trials that MS answered incorrectly, we split the data into two groups (trials that MS answered correctly vs. trials that MS answered incorrectly) and computed the networks’ performance for both groups. Results indicated that for experiment 1 (manmade vs. natural scenes) ResNet-10 made more errors for images that were answered incorrectly by MS, both with and without a lesion. For experiment 2 (objects), although there were more errors for trials answered incorrectly by MS, there was no significant difference. Overall, these results indicate that there is an overlap in the trials that are ‘difficult’ for MS and the networks.

Lesions in earlier layers of the network (blocks 1–3) resulted in a strongly biased response, in which the network generally classified all images as belonging to the same category (Supplementary Figure 1). The direction of this bias was variable across different initializations, suggesting that the earlier layers are crucial to obtain a useful representation, and the bias was not caused by the current stimulus set.

4. Discussion

We evaluated the extent to which MS’ ability to recognize visual information shows selective impairments for semantic categories. Our findings show a dissociation between two associated tasks (categorization of manmade vs. natural scenes and objects), with impaired performance for natural objects (as is usually the case), and better performance for naturalistic scenes compared to manmade scenes. For the objects, performance was primarily impaired for *insects*. Overall, these results indicate that the category-specific effects, at least for patient MS, are better explained as a visual impairment, invalidating the idea that this is a purely semantic disorder (i.e. by category membership only). This is in line with earlier findings from Young et al. (1989), and suggests that, similar to findings by Gerlach (2001) and Låg (2005), the direction of category-effects to a large extent depends on the degree of perceptual differentiation called for. Using Deep Convolutional Neural Networks as ‘artificial animal models’ (Scholte, 2018), we further explored the type of information that might underlie such behavior. Overall, shallow DCNNs and DCNNs with ‘lesions’ in higher order areas showed similar response patterns, with decreased performance for manmade (experiment 1) and natural (experiment 2) categories. While there was some overlap in the type of errors, there were differences as well. On the subordinate level, the deficit in the object recognition task was mostly restricted to the subcategory of insects for MS, whereas lesioned ResNet-10 performance was impaired for both mammals and

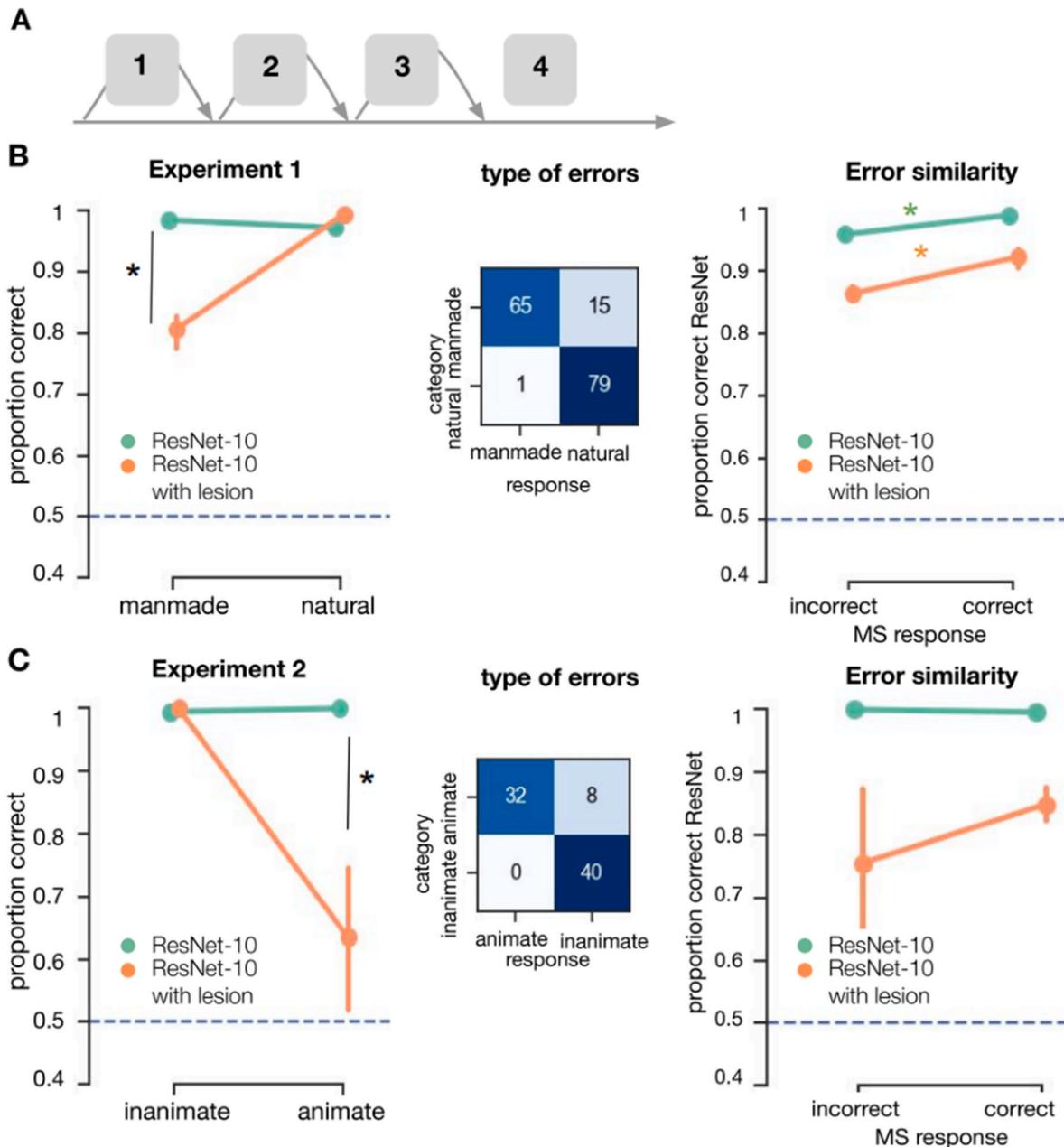


Fig. 4. Performance of ResNet-10 with and without lesion on the images from experiment 1 and 2. A) Schematic representation of ResNet-10. ResNet are built by stacking blocks (containing the convolution, batch normalization and pooling operations). Bypassing the different blocks, skip connections add the input directly to the next block. Here, we simulated a lesion in ResNet-10 by removing block 4. B) Performance of ResNet-10, with and without lesion, on the scene categorization task. Left panel indicates performance on the two different categories. Centre panel indicates the types of errors, by showing the responses for the different categories. Right panel indicates network performance for trials that MS answered correctly vs. incorrectly. C) Performance of ResNet-10, with and without lesion, on the object categorization task. Error bars represent the bootstrap 95% confidence interval, blue dashed lines indicate chance level performance. * = $p < 0.05$. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

insects.

While DCNNs trained to classify images contain mappings from labels to certain objects, they do not contain semantic knowledge about the images and the objects that are depicted. The similarity in response patterns strengthens the notion that the category-specific effects are driven by visual properties. However, it's important to note that in the current study only one architecture type (ResNet) and one approach to modelling damage (removing a block) was evaluated. In previous studies simulating 'lesions' in computational models there has been considerable variability in the way in which damage has been implemented (Guest et al., 2020), in some cases resulting in different response patterns. Whether the current effects hold for different model

architectures or different implementations of damage requires future experimental study. The extension of the damage in MS' brain makes it difficult to model or simulate damage to a specific brain region or stage in the visual processing stream. In the current study, we therefore do not focus on the exact location of the damage. Future research, comparing patients with lesions in different parts of the visual cortex, will be necessary to evaluate a more precise mapping between different types of damage in visual cortex and in ResNet layers.

4.1. Category selectivity in the visual ventral stream

There is an ongoing debate on the emergence of category selectivity

in the visual ventral stream of healthy subjects. A popular view is that observed category effects indicate a high-level representation in which neurons are organised around either object category or correlated semantic or conceptual features (Konkle and Oliva, 2012; Kriegeskorte et al., 2008; Mahon et al., 2009). An alternative view is that categorical responses in the ventral stream are driven by combinations of more basic visual properties that covary with different categories (Andrews et al., 2015; Long et al., 2018). The conflation of visual and categorical properties in object images means that category-selective responses could be expected under both accounts. Results from the current study do not speak to these findings, nor include/exclude the possibility for object category-selective responses driven by categorical or semantic properties. However, these findings do indicate that in object recognition impairments (following brain damage to certain regions), category-selectivity can emerge based on basic visual properties.

4.2. Object representations in IT

A question that remains unresolved in this study is which visual features might be involved in classification of the different categories, i. e. which dimensions in stimulus or object space are utilized by MS. Recent work by Bao et al. (2020) shows that specialization of different categories in certain regions in IT can be explained by two dimensions, progressing from animate to inanimate (dimension 1), and from more stubby to spiky (dimension 3). Following these dimensions, lesions to different parts of IT should lead to agnosias in specific sectors of object space. For example, the observation that MS' specifically does not recognize insects (which are generally more 'spiky' than mammals, Supplementary Figure 2) as being animate might be explained by a disturbed 'spiky animate corner' in object space.

4.3. Effect of typicality on category-membership decisions

The typicality of a target object is known to influence category-membership decisions (Shoben, 1982). For a given semantic category, the more typical members can be accepted as belonging to that category more quickly than less typical members. In earlier studies, MS also showed faster reactions to more typical exemplars (Young et al., 1989). However, on top of this 'typicality effect', MS showed faster responses to non-living things than living things. In the current study, performance on experiment 2 was merely decreased for insects (Supplementary Figure 2). One explanation could be that insects are less typical for the 'natural' condition than mammals, and therefore performance was decreased for these images.

4.4. Objects vs. scene categorization

While object and scene recognition both require mapping of low-level incoming sensory information to high-level representations and semantic knowledge, perceiving a scene may involve different types of information than recognition of objects. Previous work indeed suggests that patients with visual form agnosia may employ different visual properties for the classification of scenes. The unexpected observation in the current study, that performance was worse for manmade scenes compared to natural scenes, has been shown earlier in an individual with visual form agnosia (Steeves et al., 2004). DF, who has a profound deficit in object recognition but spared color and visual texture perception, could classify scenes and was fastest for natural scenes. The fact that natural scenes contained more predictable color, texture information, and spatial structure than non-natural scenes (Oliva and Torralba, 2001; Burton and Moorhead, 1987), could potentially explain why performance was increased for those images.

5. Conclusion

Overall, these results suggest that semantic impairments for certain

categories can, at least in MS' case, be explained by differences in perceptual demand and visual features, rather than by a deficit in semantic memory. The pattern of results seems to indicate that the task and stimuli that one uses has a strong influence on the diagnostic visual features that can be used for a certain task, and that those factors may lead to different, opposing results in terms of semantic category-specific deficits. Because of the differences between the two tasks we cannot pinpoint exactly what visual properties drive our effects. There are several possible explanations. First of all, it could be that MS' performance is impaired for images containing certain visual properties, such as 'spiky' features and straight lines. Secondly, it could be that MS performance is impaired for certain images, because they are perceptually more difficult based on other image properties, such as the image complexity, viewpoint or prototypicality. The stimuli used in the current experiment were not controlled for such factors, and future research is needed to determine the exact underpinnings of the impairment in the specific conditions. The finding that deep neural networks (which have no semantic knowledge, apart from adding a label to a visual image) show the same effects for shallow models and models with lesions in higher layers supports the notion that the effects are driven by visual features and not semantic category.

Data and code availability

Data and code to reproduce the analyses in this article are available at the Open Science Framework (#9h7mf) and at https://github.com/noorsejidel/2020_Object_agnosia.

Declaration of competing interest

The authors declare no competing financial interests.

Acknowledgements

This work is supported by and Advanced Investigator Grant from the European Research Council (ERC) to Edward de Haan (#339374).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.neuropsychologia.2021.108017>.

Credit author statement

N.S., Conceptualization, Investigation, Formal analysis, Data curation, Project administration, Software, Validation, Methodology, Visualization, writing—original draft, writing—review and editing. H.S.S., Conceptualization, Resources, Investigation, Software, Supervision, Writing—review and editing. E.H.F.d.H., Conceptualization, Resources, Investigation, Funding acquisition, Supervision, writing—review and editing.

References

- Andrews, T.J., Watson, D.M., Rice, G.E., Hartley, T., 2015. Low-level properties of natural images predict topographic patterns of neural response in the ventral visual pathway. *Journal of Vision* 15 (7). <https://doi.org/10.1167/15.7.3>, 3–3.
- Bao, P., She, L., McGill, M., Tsao, D.Y., 2020. A map of object space in primate inferotemporal cortex. *Nature* 583, 103–108. <https://doi.org/10.1038/s41586-020-2350-5>.
- Burton, G.J., Moorhead, I.R., 1987. Color and spatial structure in natural scenes. *Applied Optics* 26, 157–170. <https://doi.org/10.1364/AO.26.000157>.
- Capitani, E., Laiacona, M., Mahon, B., Caramazza, A., 2003. What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. *Cogn. Neuropsychol.* 20 (3), 213–261. <https://doi.org/10.1080/02643290244000266>.
- Caramazza, A., Hillis, A.E., Rapp, B.C., Romani, C., 1990. The multiple semantics hypothesis: multiple confusions? *Cogn. Neuropsychol.* 7 (3), 161–189. <https://doi.org/10.1080/02643299008253441>.

- Caramazza, A., Shelton, J.R., 1998. Domain-specific knowledge systems in the brain the animate-inanimate distinction. *J. Cognit. Neurosci.* 10 (1), 1–34. <https://doi.org/10.1162/089892998563752>.
- Chadwick, A., Heywood, C., Smithson, H., Kenridge, R., 2019. Translucence perception is not dependent on cortical areas critical for processing colour or texture. *Neuropsychologia* 128, 209–214.
- Chow-Wing-Bom, H.T., Scholte, S., De Klerk, C., Mareschal, D., Groen, I., Dekker, T.M., 2019. Development of Rapid Extraction of Scene Gist. *PERCEPTION* 48, 40–41.
- Cichy, R.M., Khosla, A., Pantazis, D., Torralba, A., Oliva, A., 2016. Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific Reports* 6, 27755. <https://doi.org/10.1038/srep27755>.
- de Haan, E.H.F., Seijdel, N., Kenridge, R.W., Heywood, C.A., 2020. Plasticity versus chronicity: stable performance on category fluency 40 years post-onset. *J. Neuropsychol.* 14, 20–27. <https://doi.org/10.1111/jnp.12180>.
- Gaffan, D., Heywood, C.A., 1993. A spurious category-specific visual agnosia for living things in normal human and nonhuman primates. *J. Cognit. Neurosci.* 5 (1), 118–128. <https://doi.org/10.1162/jocn.1993.5.1.118>.
- Gainotti, G., 2000. What the locus of brain lesion tells us about the nature of the cognitive defect underlying category-specific disorders: a review. *Cortex: a Journal Devoted to the Study of the Nervous System and Behavior* 36 (4), 539–559. [https://doi.org/10.1016/S0010-9452\(08\)70537-9](https://doi.org/10.1016/S0010-9452(08)70537-9).
- Gerlach, C., 2001. Structural similarity causes different category-effects depending on task characteristics. *Neuropsychologia* 39 (9), 895–900.
- Gerlach, C., 2009. Category-specificity in visual object recognition. *Cognition* 111 (3), 281–301. <https://doi.org/10.1016/j.cognition.2009.02.005>.
- Griffin, G., Holub, A., Perona, P., 2007. Caltech-256 Object Category Dataset. Technical report, 7694. California Institute of Technology. http://www.vision.caltech.edu/Ima ge_Datasets/Caltech256/.
- Güçlü, U., van Gerven, M.A.J., 2015. Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. *The Journal of Neuroscience* 35 (27), 10005–10014. <https://doi.org/10.1523/JNEUROSCI.5023-14.2015>.
- Guest, O., Caso, A., Cooper, R.P., 2020. On simulating neural damage in connectionist networks. *Computational brain & behavior* 3 (3), 289–321. <https://doi.org/10.1007/s42113-020-00081-z>.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 770–778.
- Jones, E., Oliphant, T., Peterson, P., Others, 2001. *SciPy: Open Source Scientific Tools for Python*.
- Khaligh-Razavi, S.M., Kriegeskorte, N., 2014. Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS computational biology* 10 (11). <https://doi.org/10.1371/journal.pcbi.1003915>.
- Konkle, T., Oliva, A., 2012. A real-world size organization of object responses in occipitotemporal cortex. *Neuron* 74 (6), 1114–1124. <https://doi.org/10.1016/j.neuron.2012.04.036>.
- Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., Bandettini, P.A., 2008. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60 (6), 1126–1141. <https://doi.org/10.1016/j.neuron.2008.10.043>.
- Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., Kamali, S., Popov, S., Mallocci, M., Duerig, T., Ferrari, V., 2020. The Open Images Dataset V4: Unified Image Classification, Object Detection, and Visual Relationship Detection at Scale, 128. *International Journal of Computer Vision*, pp. 1956–1981 doi: 10.1007/s11263-020-01316-z.
- Låg, T., 2005. Category-specific effects in object identification: What is “normal”? *Cortex* 41 (6), 833–841.
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft COCO: common objects in context. *Computer Vision – ECCV 8693*, 740–755. https://doi.org/10.1007/978-3-319-10602-1_48.
- Long, B., Yu, C.-P., Konkle, T., 2018. Mid-level visual features underlie the high-level categorical organization of the ventral stream. *Proceedings of the National Academy of Sciences* 115 (38), E9015–E9024. <https://doi.org/10.1073/pnas.1719616115>.
- Mahon, B.Z., Anzellotti, S., Schwarzbach, J., Zampini, M., Caramazza, A., 2009. Category-specific organization in the human brain does not require visual experience. *Neuron* 63 (3), 397–405. <https://doi.org/10.1016/j.neuron.2009.07.012>.
- McKinney, W., Others, 2010. Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference* 445, 51–56.
- Mehta, Z., Newcombe, F., De Haan, E., 1992. Selective loss of imagery in a case of visual agnosia. *Neuropsychologia* 30 (7), 645–655. [https://doi.org/10.1016/0028-3932\(92\)90069-X](https://doi.org/10.1016/0028-3932(92)90069-X).
- Newcombe, F., Ratcliff, G., 1975. Agnosia: A disorder of object recognition. In: Michel, F., Schot, B. (Eds.), *Callosal disconnect syndromes in men*. Colloque International de, Lyon, France, Lyon, pp. 317–341.
- Nielsen, J.M., 1946. Agnosia, apraxia, aphasia: their value in cerebral localization. https://pure.mpg.de/rest/items/item_2366910/component/file_2417471/content.
- Oliphant, T.E., 2006. *A Guide to NumPy*, vol. 1. Trelgol Publishing USA.
- Oliva, A., Torralba, A., 2001. Modeling the shape of the scene: a holistic representation of the spatial envelope. *Int. J. Comput. Vis.* 42 (3), 145–175. <https://doi.org/10.1023/A:1011139631724>.
- Panis, S., Torfs, K., Gillebert, C.R., Wagemans, J., Humphreys, G.W., 2017. Neuropsychological evidence for the temporal dynamics of category-specific naming. *Vis. Cognit.* 25 (1–3), 79–99. <https://doi.org/10.1080/13506285.2017.1330790>.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, Alban, Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. Pytorch: an imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*. Curran Associates, Inc., pp. 8024–8035.
- Ratcliff, G., Newcombe, F., 1982. Object recognition: Some deductions from the clinical evidence. 171. *Academic Press*, London, UK, p. 147.
- Russell, B.C., Torralba, A., Murphy, K.P., Freeman, W.T., 2008. LabelMe: a database and web-based tool for image annotation. *Int. J. Comput. Vis.* 77 (1–3), 157–173. <https://doi.org/10.1007/s11263-007-0090-8>.
- Sartori, G., Job, R., Miozzo, M., Zago, S., Marchiori, G., 1993. Category-specific form-knowledge deficit in a patient with herpes simplex virus encephalitis. *J. Clin. Exp. Neuropsychol.* 15 (2), 280–299. <https://doi.org/10.1080/01688639308402563>.
- Scholte, H.S., 2018. Fantastic DNimals and where to find them. *Neuroimage* 180 (Part), 112–113. <https://doi.org/10.1016/j.neuroimage.2017.12.077>.
- Seabold, S., Perktold, J., 2010. Statsmodels: econometric and statistical modeling with python. *Proceedings of the 9th Python in Science Conference* 57, 61.
- Shoben, E.J., 1982. Semantic and lexical decisions. *Handbook of Research Methods in Human Memory and Cognition*, pp. 287–314. <https://doi.org/10.1016/b978-0-12-566760-9.50015-4>.
- Steeves, J.K.E., Keith Humphrey, G., Culham, J.C., Menon, R.S., David Milner, A., Goodale, M.A., 2004. Behavioral and neuroimaging evidence for a contribution of color and texture information to scene classification in a patient with visual form agnosia. *J. Cognit. Neurosci.* 16 (6), 955–965. <https://doi.org/10.1162/0898929041502715>.
- Smits, A.R., Seijdel, N., Scholte, H.S., Heywood, C.A., Kenridge, R.W., de Haan, E.H.F., 2019. Action blindsight and antipointing in a hemianopic patient. *Neuropsychologia* 128, 270–275. <https://doi.org/10.1016/j.neuropsychologia.2018.03.029>.
- Tyler, L.K., Moss, H.E., 2001. Towards a distributed account of conceptual knowledge. *Trends Cognit. Sci.* 5 (6), 244–252. [https://doi.org/10.1016/S1364-6613\(00\)01651-X](https://doi.org/10.1016/S1364-6613(00)01651-X).
- Tyler, L.K., Moss, H.E., Durrant-Peatfield, M.R., Levy, J.P., 2000. Conceptual structure and the structure of concepts: a distributed account of category-specific deficits. *Brain Lang.* 75 (2), 195–231. <https://doi.org/10.1006/brln.2000.2353>.
- Warrington, E.K., McCarthy, R.A., 1987. Categories of knowledge. Further fractionations and an attempted integration. *Brain: J. Neurol.* 110 (Pt 5), 1273–1296. <https://doi.org/10.1093/brain/110.5.1273>.
- Warrington, E.K., Shallice, T., 1984. Category specific semantic impairments. *Brain: J. Neurol.* 107 (Pt 3), 829–854. <https://doi.org/10.1093/brain/107.3.829>.
- Xiao, J., Hays, J., Ehinger, K.A., Oliva, A., Torralba, A., 2010. SUN database: large-scale scene recognition from abbey to zoo. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 3485–3492.
- Young, A.W., Newcombe, F., Hellawell, D., De Haan, E., 1989. Implicit access to semantic information. *Brain Cognit.* 11 (2), 186–209. [https://doi.org/10.1016/0278-2626\(89\)90017-1](https://doi.org/10.1016/0278-2626(89)90017-1).
- Mollon, J. D., Newcombe, F., Polden, P. G., & Ratcliff, G. (1980) On the presence of three cone mechanisms in a case of total achromatopsia. In G. Verriest (Ed.), *Colour vision deficiencies* (vol. V, pp. 130–135). Bristol, UK: Hilger.