Disentangling dimensions of animacy in human brain and behaviour

Kamila M Jozwik[1]*, Elias Najarro[2], Jasper JF van den Bosch[3], Ian Charest[3], Nikolaus Kriegeskorte[4] and Radoslaw M Cichy[5]

[1]University of Cambridge, Department of Psychology, Cambridge, UK.
[2]IT University of Copenhagen, Digital Design Department, Copenhagen, Denmark.
[3]University of Birmingham, School of Psychology, Birmingham, UK.
[4]Columbia University, Zuckerman Mind Brain Behavior Institute, New York, USA.
[5]Freie Universität Berlin, Department of Education and Psychology, Berlin, Germany.
*corresponding author

Abstract
The perception of animate things is of great behavioural importance to humans. Despite the prominence of the distinct brain and behavioural responses to animate and inanimate things, however, it remains unclear which of several commonly entangled properties underlie these observations. Here, we investigate the importance of five dimensions of animacy: “being alive”, “looking like an animal”, “having agency”, “having mobility”, and “being unpredictable” in brain (fMRI, EEG) and behaviour (property and similarity judgments) of 19 subjects using a stimulus set of 128 images that disentangles the five dimensions (optimized by a genetic algorithm). Our results reveal a differential pattern across brain and behaviour. The living/non-living distinction (“being alive”) was prominent in subjects’ judgments, but despite its prominence in neuroscience literature, did not explain variance in brain representations. The other dimensions of animacy explained variance in both brain and behaviour. The “having agency” dimension explained more variance in higher-level visual areas, consistent with higher cognitive contributions. The “being unpredictable” dimension instead captured representations in both lower and higher-level visual cortex, possibly because unpredictable things require attention. Animacy is multidimensional and our results show that distinct dimensions are differentially represented in human brain and behaviour.

Animacy | Dimensions | Visual object recognition | EEG | fMRI | Similarity judgements

Introduction
The perception of animate things is of great behavioural and evolutionary importance to humans and other animals and is often a matter of life and death (e.g., quick recognition of a tiger in the wild). Consistent with the importance of animacy perception in classical neuropsychological literature, lesion studies established that living things are represented in dedicated regions of the cortex (Funnell & Sheridan, 1992; Ralph et al., 1998; Silveri et al., 1996). However, it is less clear by virtue of what properties animate things are behaviourally distinguished and represented in the brain. An animal differs from an inanimate object in many respects, so animacy may be a multidimensional concept, rather than one property. The dimensions of animacy that have been explored include “being alive” (Connolly et al., 2012; Gray et al., 2007; Huth et al., 2012; Leib et al., 2016; Looser et al., 2013; Rogers et al., 2005; Wheatley et al., 2011), “looking like an animal” (Bracci et al., 2019; Connolly et al., 2012; Huth et al., 2012; Rogers et al., 2005; Sha et al., 2015; Wheatley et al., 2011), “having mobility” (Beauchamp et al., 2002, 2003; Shultz & McCarthy, 2014), “having agency” (Contini et al., 2019; Gobbini et al., 2007, 2010; Lowder & Gordon, 2015; Shultz et al., 2015; Shultz &

McCarthy, 2014; Thorat et al., 2019), and “being unpredictable” (Lowder & Gordon, 2015). The emerging picture is complex. Animacy is an important representational division in nonhuman and human ventral visual cortical areas in the inferior temporal cortex as measured by functional magnetic resonance imaging (fMRI; Kriegeskorte et al., 2008). Beyond this binary distinction, there may be an animacy continuum, where objects are perceived as more animate when they are more similar to humans (e.g., images of monkeys would be perceived as more animate than insects, even though both species belong to the animal category; Connolly et al., 2012). Animate-looking (e.g., cow mug) and animate objects (e.g., cow) are dissociated in behaviour but not in the ventral visual stream measured by fMRI (Bracci et al., 2019). Chosen animacy dimensions explain variance in the ventral visual stream fMRI measurements: agency (Thorat et al., 2019), animacy (Blumenthal et al., 2018), human-likeness (Rosenthal-von der Pütten et al., 2019), and capacity for self-movement and thought rather than face presence (Proklova & Goodale, 2020). The time course with which animacy representations emerge has been investigated with magnetoencephalography (MEG), revealing the time course of agency (Contini et al., 2019). However, MEG patterns do not seem to carry information about the animate vs inanimate object category, in contrast to fMRI (Proklova et al., 2019).

In sum, these studies offer important insights on aspects of animacy, but no unified account emerges. The likely reasons for this are that the stimuli used in previous studies were mostly handpicked to investigate a chosen dimension without controlling the other co-occurring properties of animacy. This procedure has limited our progress toward a comprehensive understanding of the representation of animacy as a multidimensional concept. Thus, despite decades of research that has established the prominence of the dimensions of animacy in brain and behaviour, it remains unclear which properties cause the distinct responses.

Here, we comprehensively investigate the importance of five dimensions of animacy: “being alive”, “looking like an animal”, “having agency”, “having mobility”, and “being unpredictable”. We optimized a stimulus set of 128 images to disentangle the five dimensions of animacy using a genetic algorithm. The 128 stimuli were chosen among a set of 300 by the algorithm to decorrelate human ratings of the five dimensions. We evaluated how well dimensions of animacy can explain behaviour in two tasks (animacy ratings and similarity judgements) and brain representations, combining fMRI and EEG to assess representations at high spatial and temporal resolution.

Results

Stimuli selection procedure and stimulus set

Evaluating the contribution of individual dimensions of animacy (“being alive”, “looking like an animal”, “having mobility”, “having agency”, and “being unpredictable”) would be best performed on a stimulus set that is as decorrelated on these dimensions as possible. We created such a stimulus set by optimizing stimuli using a genetic algorithm. We selected images that were maximally decorrelated on dimensions of animacy using a four-step procedure (Figure 1, see Methods for details). First, we created an animacy dimension grid where we asked participants to fill in names of the objects fulfilling each animacy dimension combination to then find images that satisfy all combinations of dimensions of animacy. Second, we assembled object images based on object names from step one. Third, an independent set of participants rated object images (which were assembled based on object names from step one) on each of the five dimensions of animacy to generate animacy ratings.

Fourth, we used a genetic algorithm to select a subset of images with the lowest maximum correlation between dimensions of animacy (10,000 generations). The maximum correlation between dimensions in the stimulus set was 0.35. This result was better than when randomly selecting the stimuli 10,000 times without optimization (maximum correlation between dimensions = 0.59) proving that our novel stimulus selection procedure was successful. This stimulus set was used in subsequent behavioural and brain imaging experiments.

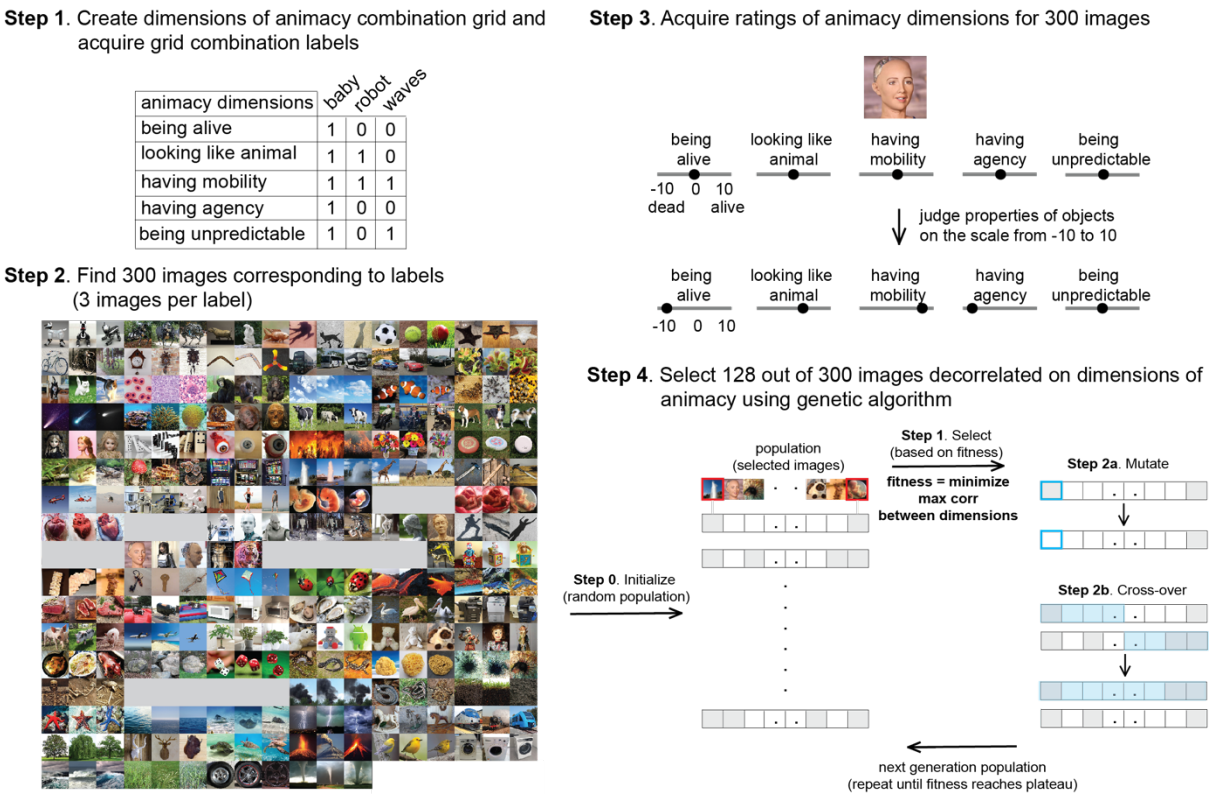

Figure 1. Stimuli selection procedure.
First, we created an animacy grid with all dimensions of animacy combinations and asked 11 participants to fill in the names of objects that fulfilled these combinations. Second, we assembled object images based on object names from step one (images containing a human face are replaced by grey squares according to bioRxiv's policy on images of individuals). Third, an independent set of 26 participants performed animacy ratings of 300 of these object images. Finally, we selected an optimal set of stimuli that had a low correlation between dimensions (as behaviourally rated) using a genetic algorithm. These stimuli were used in behavioural and brain representation experiments where a new set of participants was recruited to make sure that the stimulus generation and the actual experiments were independent.

The stimulus set (Figure 2a) consisted of 128 images spanning almost all animacy dimension combinations. A wide range of objects was present, such as humans, human fetuses, human organs, human and animal shadows, plants, corals, forces of nature, game items, toys, vehicles, and electronic equipment covering 68 categories. This stimulus set was used for two behavioural studies: animacy ratings and similarity judgements, and two brain response measurement studies: EEG (to access temporal information) and fMRI (to access spatial information). Nineteen participants performed all the studies. Importantly, participants first performed EEG and fMRI studies, followed by similarity judgements and finally animacy ratings (Figure 2b). This experimental order was to ensure that participants did not know about animacy dimensions tested until the final animacy ratings.

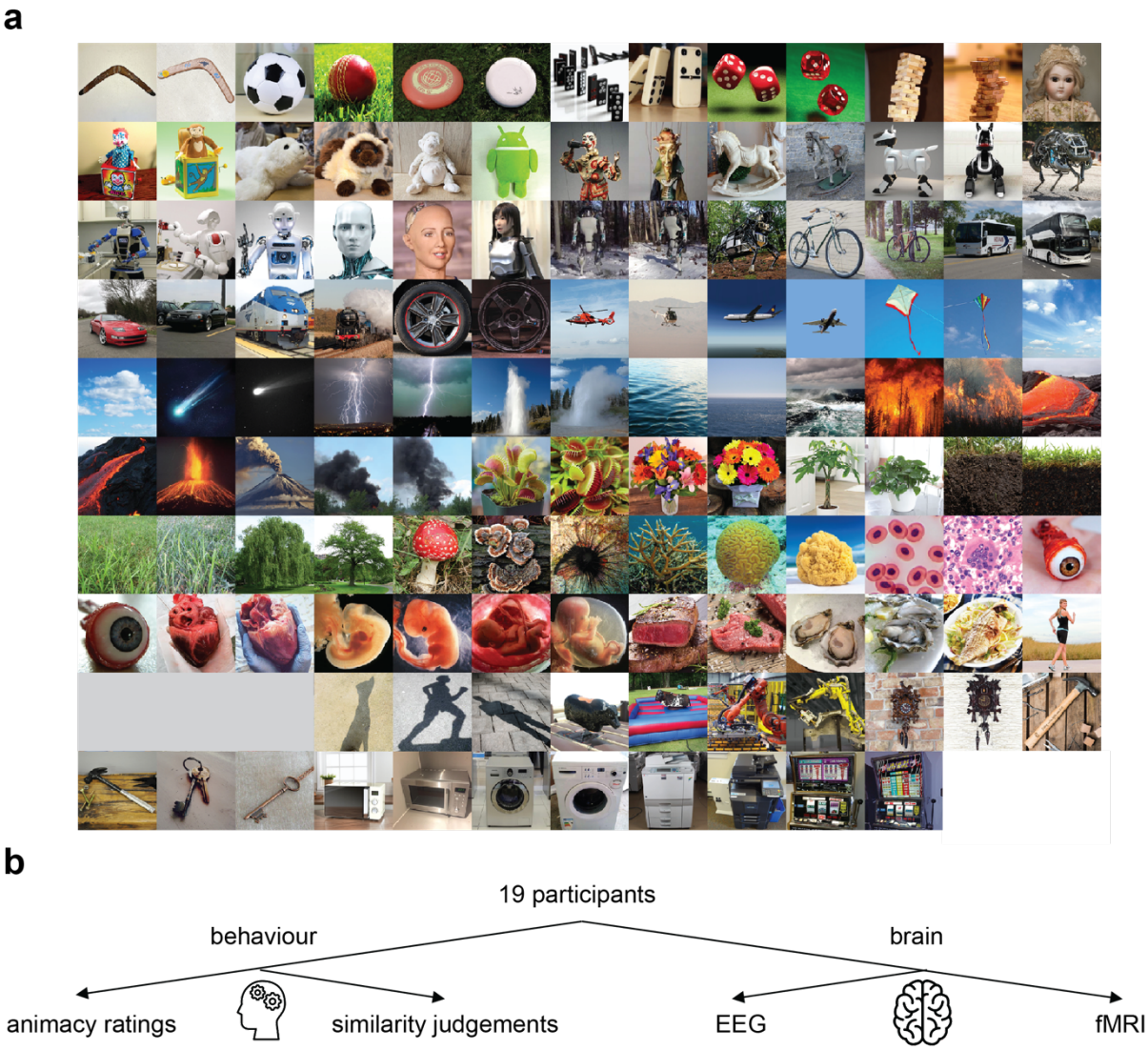

Figure 2. Stimulus set and study overview.
a. The genetic-algorithm driven stimulus set consisted of 128 images decorrelated on dimensions of animacy. The stimuli were coloured images of sport equipment, games, robots, dolls and puppets, plush toys, land vehicles, air vehicles, plants, forces of nature (water, air, fire, smoke), sea organisms, cells, organs and fetuses, humans, food, kitchen and office equipment, shadows (images containing a human face are replaced by grey squares according to bioRxiv's policy on images of individuals).
b. Study overview. All 19 participants performed two behavioural studies: animacy ratings and similarity judgements, and two brain response measurement studies: EEG (to access temporal information) and fMRI (to access spatial information). Importantly, participants first performed EEG and fMRI studies, then similarity judgements and finally animacy ratings. This experimental order was to ensure that participants did not know about animacy dimensions tested until the final animacy ratings.

Consistency in animacy ratings across participants

We first wanted to evaluate the contribution of each dimension of animacy when participants were asked to judge how animate an object image was. We first explored how consistent participants were in judging each dimension of animacy and each image. We used representational similarity analysis (RSA) to reveal which dimensions contribute most when participants judged animacy. We also examined which dimension(s) explained unique variance in the animacy ratings.

Participants judged each object image using a continuous scale from -10 to 10 for each dimension, e.g., -10 meant “inanimate” and 10 meant “animate” (Figure 3a). The same image was judged in the same way on the five investigated dimensions of animacy. The mean between-participant correlation was 0.6, which indicated that participants were generally consistent in their ratings. The mean within-participant consistency for thirty repeated stimuli was 0.89 meaning that participants consistently judged the same stimulus within a session. The raw data of animacy ratings shows that a given stimulus could be differently rated on each of the dimensions (Figure 3b).

We wanted to know how consistent participants were in judging each dimension and each stimulus. This analysis had two purposes: to be able to determine whether the variability in animacy ratings is low enough to interpret the ratings at all and to test which dimensions and stimuli were particularly controversial for participants as reflected by higher variability.

We first asked about the participants’ consistency across stimuli in all dimensions of animacy. There is variability in the consistency of ratings as some of the object images, e.g., "bike", was judged very consistently as expected, whereas other more controversial ones, e.g., “human fetus”, less consistently, with “human shadow” having the lowest consistency among the object images tested (Figure 3c).

To get more insights into animacy dimension ratings, we explored the consistency of each stimulus per each animacy dimension. For "looking like an animal”, which was judged most consistently among the dimensions, "hammer" was one of the object images that were most consistently judged, whereas an image of a "human" was not judged very consistently. The lower consistency of judging an image of a human may be related to the fact that some humans do and others do not consider themselves animals, even though from a biological point of view Homo sapiens belong to the animal category. Looking at the other side of the spectrum - "having agency" dimension was judged least consistently - we observed that an example image that had a high consistency of ratings was "eyeball", whereas an image of a "robot" was not very consistently judged (Figure 3d).

Among all the dimensions participants judged "having agency" least consistently and "looking like an animal" most consistently (Figure 3e). What about the animacy ratings (“being animate”)? Would object animacy be judged consistently across participants or given the ambiguous definition of this term, would the consistency be lower than some of the more precisely defined dimensions of animacy? We found that the latter was the case - animacy ratings had lower consistency than more precisely defined dimensions of animacy, except for the “having agency” dimension.

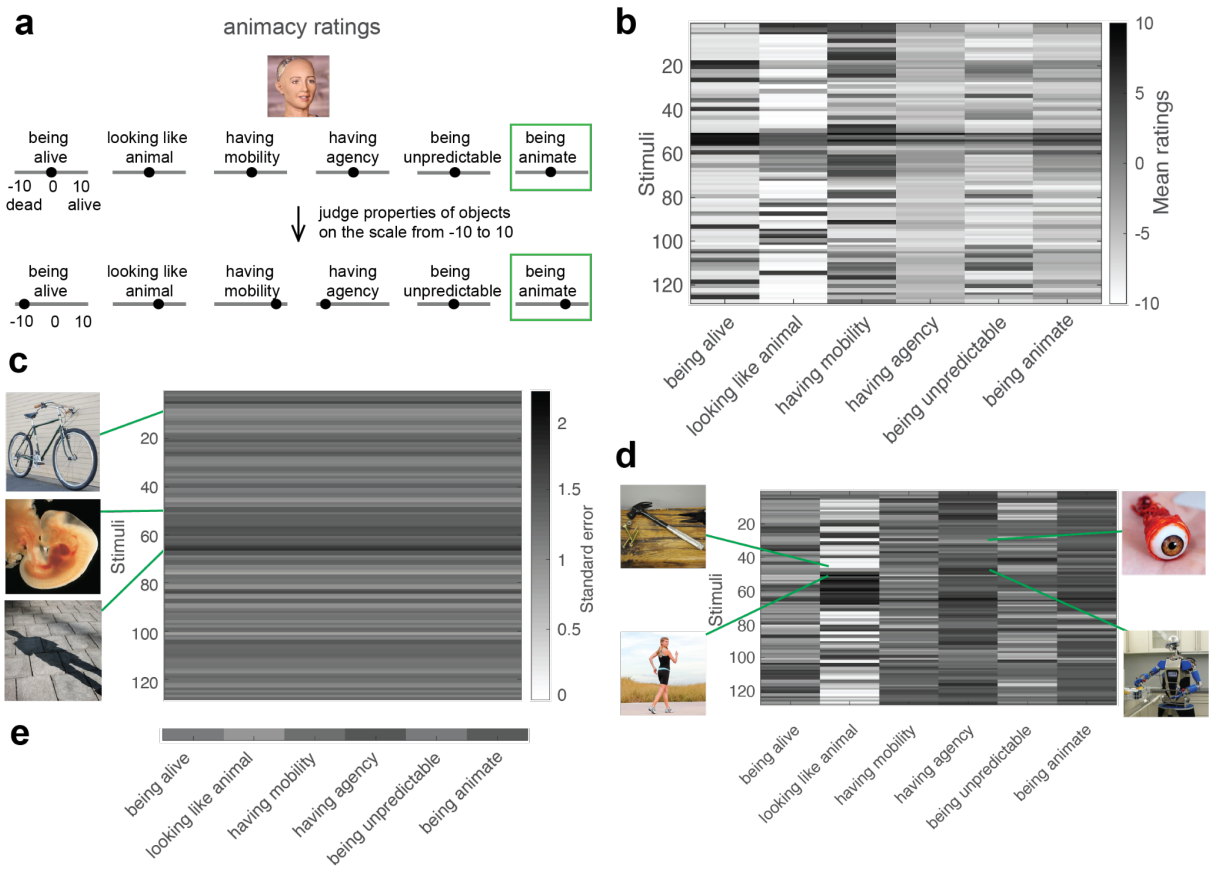

Figure 3. Animacy ratings and their consistency with examples of images judged consistently and not very consistently.
a. Illustration of animacy ratings. Participants judged each object image using a continuous scale from -10 to +10 for each animacy dimension, e.g., -10 meant “dead” and +10 meant “alive” for the “being alive” dimension. Additionally, participants performed a rating of "being animate" dimension in a similar fashion.
b. Mean ratings of each animacy dimension and stimulus across participants.
c. Consistency of each stimulus in animacy ratings across participants (standard error of the mean) with examples of stimuli with varying values of standard error.
d. Consistency of each animacy dimension and stimulus in animacy ratings across participants with examples of stimuli with varying values of standard error for the most consistently judged ("looking like an animal") and the least consistently judged ("having agency") dimensions.
e. Consistency of each animacy dimension in animacy ratings across participants (standard error of the mean).

Contribution of each animacy dimension to animacy ratings

To gain more intuition about what stimuli are considered to have the highest value on each animacy dimension, we visualized 10 object images with the overall minimum and 10 with the maximum rating on a given dimension (Figure 4a). Overall, images that had low values on animacy dimension ratings were similar among dimensions (e.g., plush toys, meat, washing machine). In contrast, images with high ratings did differ depending on a dimension tested (e.g., stimuli judged as the most unpredictable being humans and forces of nature, in contrast to humans and robots judged as having the most agency), proving that indeed these dimensions capture different aspects of animacy perception.

Among all the dimensions, "having agency" and "being alive" explained more variance than other dimensions (Figure 4b) in animacy ratings (when participants were asked to judge how animate an object image was). This result means that when asked to judge animacy humans mostly think about whether an object is alive and whether it has agency. Even though "having

agency" and "being alive" explain more variance than the other dimensions, it does not mean that they explain unique variance. To test that, we performed a unique variance analysis (see Methods for details) and observed that only one dimension, "being alive", explains significantly more unique variance than "having agency" and "being unpredictable" dimensions (Figure 4c). "Being alive" is one of the dimensions that explain the most variance in animacy ratings and also the only dimension that explains significantly more unique variance than some of the other dimensions.

In summary, when asked to rate object animacy, participants found "being alive" and "having agency" as dominant dimensions and were most consistent when judging the "looking like an animal" dimension and least consistent in judging the "having agency" dimension.

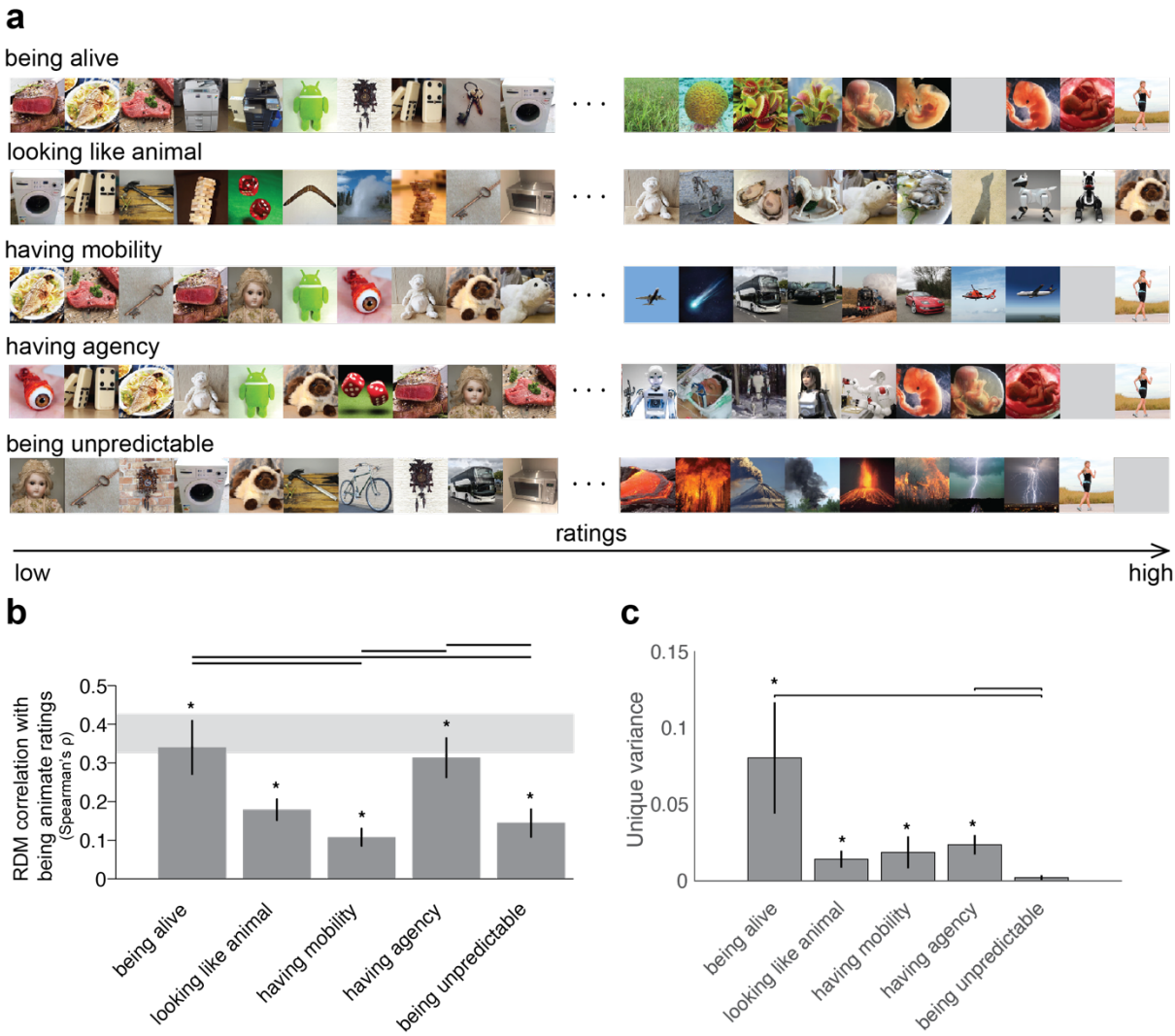

Figure 4. Animacy ratings.
a. Order of images with lowest and highest ratings on each animacy dimension. Out of 128 images, we show 10 lowest and 10 highest rated images on each animacy dimension (images containing a human face are replaced by grey squares according to bioRxiv's policy on images of individuals).
b. Animacy dimension RDMs comparisons with animacy ratings ("being animate") RDMs. Bars show the correlation between the animacy ratings RDMs and each animacy dimension RDM. A significant correlation is indicated by an asterisk (one-sided Wilcoxon signed-rank test, $p < 0.05$ corrected). Error bars show the standard error of the mean based on single-participant correlations, i.e., correlations between the single-participant animacy ratings RDMs and animacy dimension RDM. The grey bar represents the noise ceiling, which indicates the expected performance

of the true model given the noise in the data. Horizontal lines show pairwise differences between model performance ($p < 0.05$, FDR corrected across all comparisons).
c. Unique variance of each animacy dimension in explaining animacy ratings computed using a general linear model (GLM). For each animacy dimension, the unique variance is computed by subtracting the total variance explained by the reduced GLM (excluding the animacy dimension of interest) from the total variance explained by the full GLM, using non-negative least squares to find optimal weights. A significant unique variance is indicated by an asterisk (one-sided Wilcoxon signed-rank test, $p < 0.05$ corrected). The error bars show the standard error of the mean based on single-participant unique variance. Horizontal lines show pairwise differences between model performance ($p < 0.05$, FDR corrected across all comparisons).

Contribution of each animacy dimension to similarity judgements

After having established the contribution of animacy dimensions to animacy ratings, we tested whether any of the dimensions would explain similarity judgements. The similarity judgements task allows participants to evaluate objects in a more natural way concentrating on general object similarity. Please note that participants performed the similarity judgements task before competing dimensions of animacy ratings, so they were unaware of the dimensions of animacy tested. As participants were asked to arrange object images based on their similarity and not asked about animacy, dimensions of animacy may not explain these representations. Rather than the dimensions of animacy, either other categorical divisions or lower-level image features could be used for judging object similarity. Participants placed images of objects inside a circular arena according to how similar they judge them (Figure 5a). The procedure was repeated with different numbers of objects that had to be arranged indefinitely until reaching a predefined arrangement consistency (see Methods). We evaluated how well the dimensions of animacy explained the similarity judgements task using the RSA and the unique variance analysis.

Visualizing the similarity judgements as a multidimensional scaling plot helped us to determine which object images were grouped together (Figure 5b). For example, object images of most robots, fetuses, and a human on life support were grouped, with human images separated but placed close to an image of a realistic humanoid robot. This grouping is related to the agency dimension as these images received the highest rating on “having agency” dimension (Figure 4a). Animal robots formed their own cluster together with other moving objects such as boomerangs, balls, and buses. Different vehicles were grouped nearby with pictures of cars being placed near comets, clouds, and dominos. Games and toys were arranged in proximity to the “animal-like robot” group and the “human” group. Some unpredictable objects were also grouped together: geysers and game machines, or volcanos and waves. As a sanity check, images that depict the same object were grouped together, for example, two pictures of flowers or wheels.

If we assume that the similarity judgements are based only on the similarity between low-level visual features, the dimensions of animacy should not explain a large fraction of the variance. This assumption is not what we observed - all dimensions of animacy explained a significant amount of variance in the similarity judgements task (Figure 5c). None of the dimensions fully explained the similarity judgements data but the "having agency" dimension was close to explaining the total explainable variance given the noise in the data. This result means that when judging object similarity humans use all dimensions of animacy with the emphasis on agency.

Even when all dimensions of animacy explain similarity judgements, maybe one or more dimensions explain unique variance. After having performed the unique variance analysis, we observed that each dimension explained unique variance in the similarity judgements (Figure 5d). However, almost no dimension explained more unique variance than the other dimensions. This finding suggests that each dimension not only explains variance in the data but also explains a unique portion of that variance.

Overall, each animacy dimension explained a similar amount of variance in a behavioural task of similarity judgements, meaning that participants use higher-level dimensions of animacy when judging object similarity.

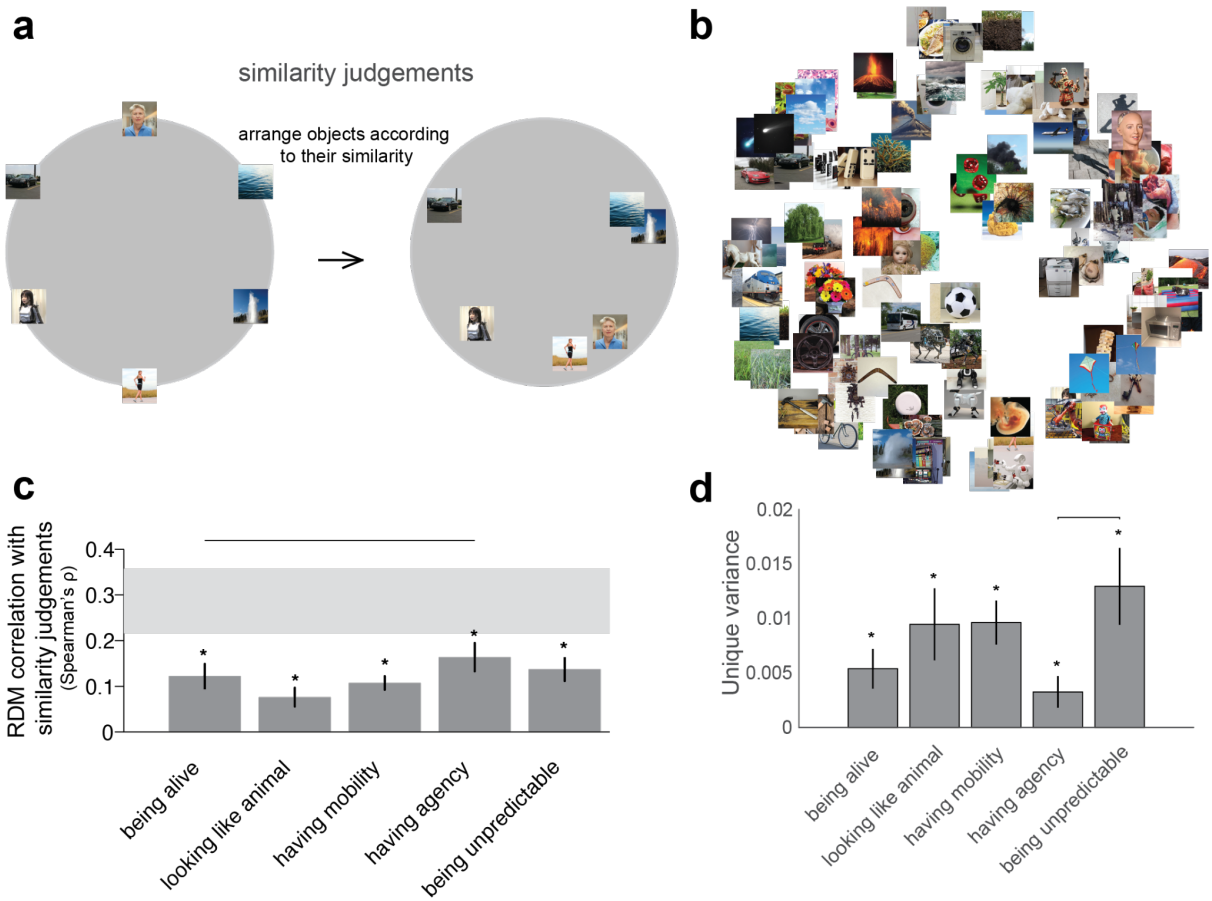

Figure 5. Dimensions of animacy and similarity judgements.
a. Similarity judgement task. During this task, object images were shown on a computer screen in a circular arena, and participants were asked to arrange the objects according to their similarity, such that similar objects were placed close together and dissimilar objects were placed further apart (see Methods for details).
b. Multidimensional scaling plot of similarity judgements (mean across participants).
c. Animacy dimension RDMs comparisons with similarity judgements RDMs. Bars show the correlation between the similarity judgements RDMs and each animacy dimension RDM using the same conventions as in Figure 4b.
d. Unique variance of each animacy dimension in explaining similarity judgements computed using the same conventions as in Figure 4c.

Having shown the contribution of each animacy dimension in explaining animacy ratings and in a behavioural task of similarity judgements, we asked whether dimensions of animacy explain the time course of object image processing in the brain using EEG. One possibility is that once an image is shown for only half a second, the brain performs only an automated image processing, and higher-level dimensions of animacy do not contribute to this process

at all. The other possibility is that beyond low-level features, higher-level dimensions of animacy do play a role in forming brain representations of images even with a short stimulus presentation. To arbitrate between those possibilities, we evaluated the amount of variance explained by each animacy dimension and tested whether any dimension(s) explains unique variance in the EEG signal.

We first performed multivariate pattern analysis (MVPA) to determine how well we could decode the pattern of activations evoked by each stimulus. We performed pair-wise decoding analysis and we could decode images in the stimulus set to a high decoding accuracy (62%) in a long time window (between 43 and 1000 ms after stimulus onset, Figure 6a). The peak decoding accuracy was at 197ms (+/- 7ms, standard error).

Once we knew that we could decode our stimuli, we asked how much variance each animacy dimension explained in EEG recordings. To answer this question, we correlated each animacy dimension with EEG representations at every time point in every participant using RSA. Do any of the dimensions explain any variance at all? We found that most animacy dimensions explained a significant amount of variance in EEG recordings (Figure 6b); however, some dimensions explained variance at slightly different times. Despite differences in the exact timing of when dimensions of animacy explained the variance, a very clear pattern that one dimension explains representations earlier than the other was not observed. However, “being unpredictable” explained significantly more variance than most dimensions in early time points: specifically more than “looking like an animal” (89-130 ms), “having mobility” (89-113 ms), and ”having agency” (79-126 ms). While “looking like an animal” explained more variance than most other dimensions in later time points: more than “being alive” (209-302 ms), ”having agency” (230-266 ms), and “being unpredictable” (146-184 ms). Finally, “having agency” explained more variance than most of the dimensions even later in time: more than “being alive” (268-301 ms), “having mobility” (261-289 ms) and “being unpredictable” (293-315 ms).

Most animacy dimensions explained variance in EEG recordings. Is it the same or unique variance? Does one dimension explain more unique variance than the others, as in the case of animacy ratings, or is there no difference between the amount of unique variance explained by each dimension, as for the similarity judgements? We found that only one dimension - "looking like an animal"- explained the unique variance in EEG recordings between 237 and 301 ms. None of the other dimensions explained any significant unique variance (Figure 6c).

Overall, most dimensions of animacy explained EEG recordings with subtle differences in timing, but only “looking like an animal” explained unique variance. Even for the rapid object recognition time course, higher-level dimensions of animacy explained a significant amount of variance in brain representations.

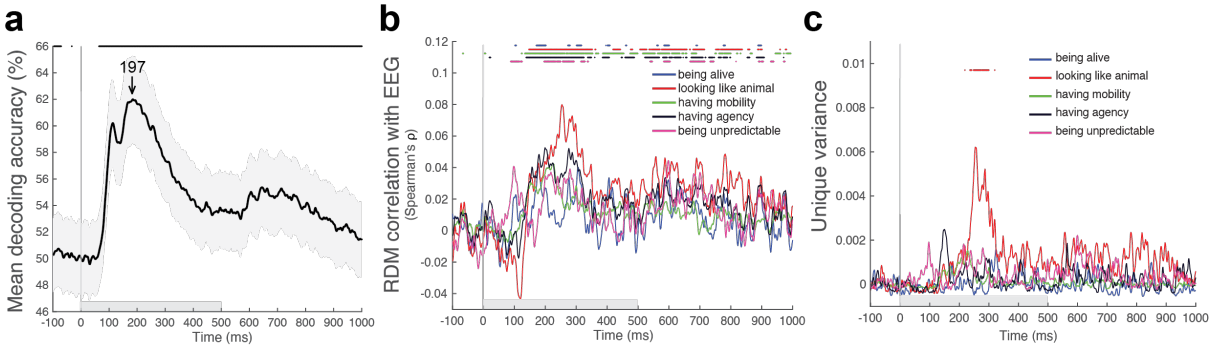

Figure 6. Dimensions of animacy and EEG time course.

a. Mean decoding curve across participants. Significant decoding is indicated by a horizontal line above the graph (one-sided Wilcoxon signed-rank test, $p < 0.05$ corrected) and starts at 43ms (+/- 2ms, standard error) with a peak latency of 197ms (+/- 7ms, standard error, indicated by an arrow). The shaded area around the lines shows the standard error of the mean based on single-participant decoding. The grey horizontal bar on the x axis indicates the stimulus duration.

b. Animacy dimension RDMs comparison with EEG RDMs across time. Lines show the correlation between the EEG RDMs and each animacy dimension RDM. A significant correlation is indicated by a horizontal line above the graph (one-sided Wilcoxon signed-rank test, $p < 0.05$ corrected). The grey horizontal bar on the x axis indicates the stimulus duration.

c. Unique variance of each animacy dimension in explaining EEG RDMs computed using a GLM. For each animacy dimension, the unique variance is computed by subtracting the total variance explained by the reduced GLM (excluding the animacy dimension of interest) from the total variance explained by the full GLM, using non-negative least squares to find optimal weights. A significant unique variance (between 237 and 301ms) is indicated by a horizontal line above the graph (one-sided Wilcoxon signed-rank test, $p < 0.05$ corrected). The grey horizontal bar on the x axis indicates the stimulus duration.

Contribution of each animacy dimension to fMRI representations

We asked where in the brain dimensions of animacy explain patterns of responses to images using fMRI. We performed both regions of interest (ROI) analysis along the ventral and dorsal visual streams and searchlight analysis. The ROI analysis was performed to evaluate the contribution of dimensions of animacy in the brain regions along the visual stream where we know object images are represented. The searchlight analysis complemented the ROI analysis testing whether other regions in the visual stream exist where dimensions of animacy explain variance that we may have missed when preselecting ROIs.

We first evaluated the contribution of each animacy dimension in ROIs across the ventral: V1v, VO2, PHC2 and dorsal: V1d, LO2, TO2 visual streams (Figure 7a). In the ventral visual stream, “being unpredictable”, “having mobility”, and to a lesser extent “having agency” explained variance in V1v, in contrast to higher-level visual areas (VO2, PHC2) where additionally “looking like an animal” explained a significant amount of variance and “having agency” explained more variance than in V1v. In the dorsal visual stream, “being unpredictable” and “having mobility” explained variance in V1d, with “having agency” explaining variance in higher-level visual areas (LO2, TO2), and “looking like an animal explaining variance only in TO2. As the dorsal stream carries information related to movement it is intuitive that “having mobility” and “being unpredictable" explain variance in dorsal regions. It was not clear, however, whether “having agency” and “looking like an animal” would explain variance in dorsal regions but it is indeed what we observe. These results suggest that the dimensions of animacy are important for both ventral and dorsal visual streams, however, a substantial variance remained unexplained.

To investigate the contribution of each animacy dimension in a spatially unbiased fashion we performed a searchlight analysis in the visual stream (Figure 7b). Consistently with the ROI analysis “being alive” did not explain any significant amount of variance in the brain. “Looking like an animal” also did not explain any significant amount of variance in searchlight analysis but it is possible that with a larger amount of data we would see this dimension explaining some variance as the ROI analysis and EEG results pointed in that direction. “Having mobility” dimension explained variance in early visual cortex only (based on the correlation strength of the searchlight analysis). In contrast, “having agency” explained variance only in higher-level visual cortex, which is consistent with the ROI analysis where “having agency” explains more variance in higher-level visual areas and further suggests that this dimension best captures higher-level representations. One dimension that explained variance in both early and higher-level visual areas was “being unpredictable” suggesting that unpredictability is important for attention and is already detected in the early visual cortex. Despite the living non-living distinction being thought to be important for brain representations “being alive” dimension did not explain any significant amount of variance in brain responses based on the ROI and searchlight analyses. This suggests that the ventral and dorsal visual streams do not represent the results of a deeper cognitive inference process that would assess whether something is alive.

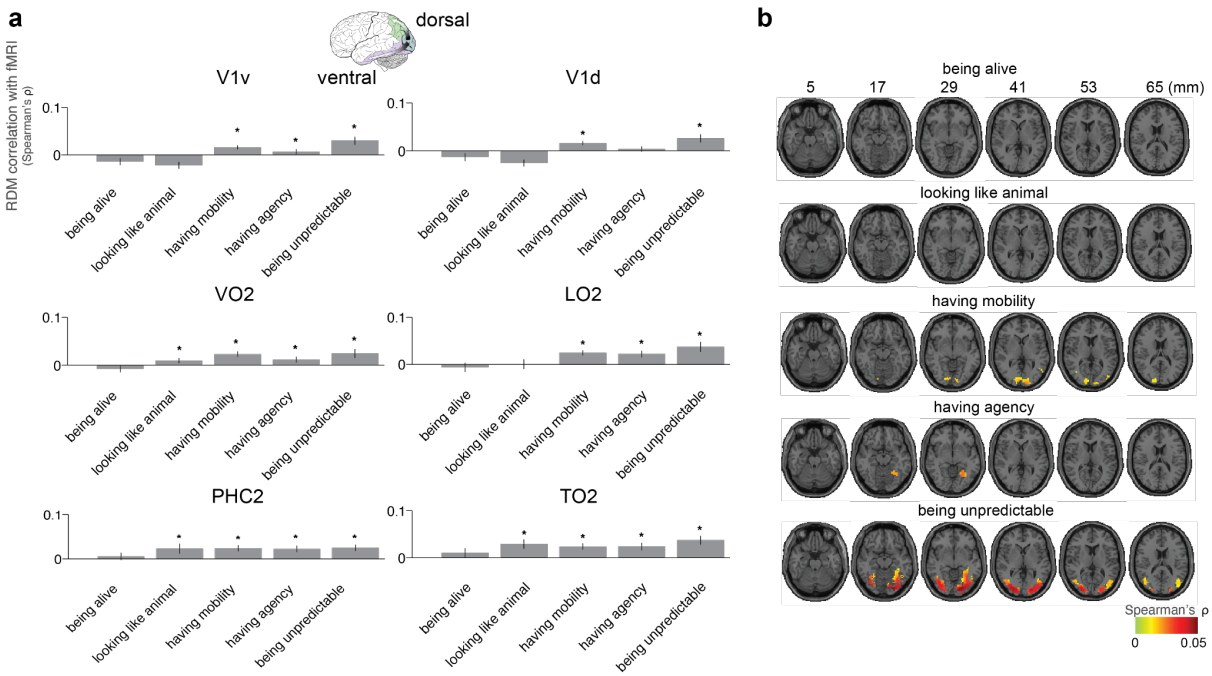

Figure 7. Dimensions of animacy and fMRI responses.
a. Animacy dimension RDMs comparisons with fMRI ROI RDMs. Bars show the correlation between each animacy dimension RDM with fMRI ROI RDMs using the same conventions as in Figure 4b. We selected ROIs across the ventral (V1v, VO2, PHC2) and dorsal (V1d, LO2, TO2) visual streams.
b. Searchlight analysis with each animacy dimension showing where in the brain animacy dimension explain image representations masked with the visual stream regions (Spearman’s ρ between animacy dimension and brain representations, one-sided Wilcoxon signed-rank test, FDR controlled at 0.05).

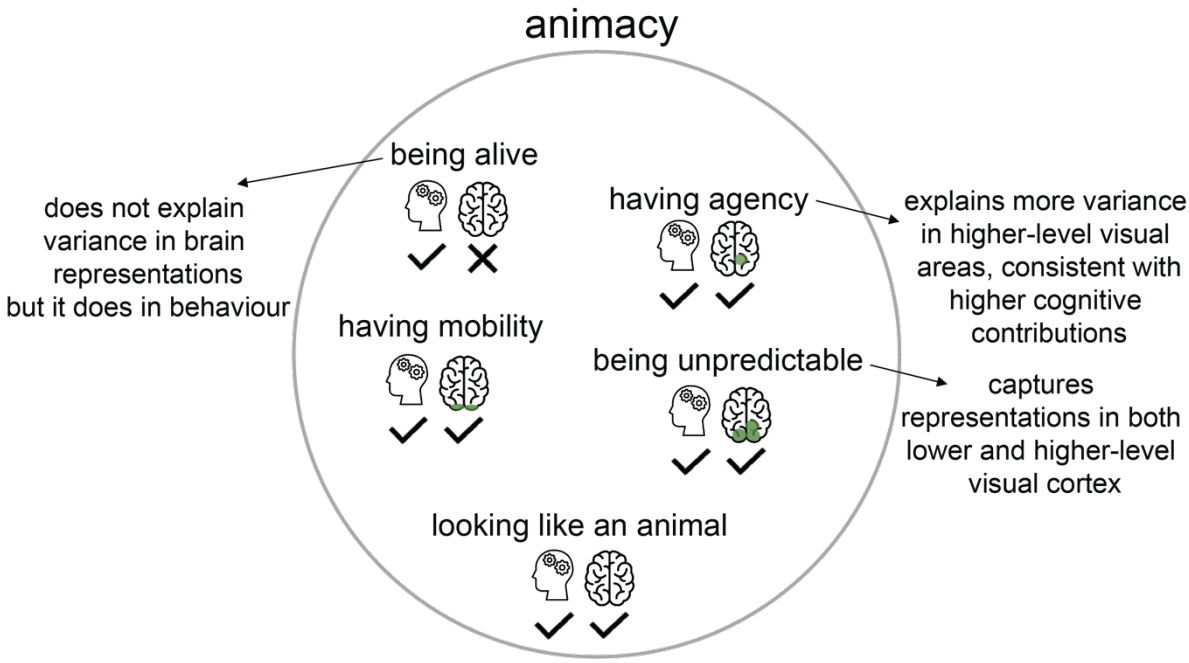

Figure 8. Summary of findings. Our results reveal a differential pattern across brain and behaviour. The living/non-living distinction ("being alive") is prominent in subjects' judgments, but despite its prominence in neuroscience literature, does not explain variance in brain representations. The other dimensions of animacy explain variance in both brain and behaviour. The "having agency" dimension explains more variance in higher-level visual areas, consistent with higher cognitive contributions. The "being unpredictable" dimension instead captures representations in both lower and higher-level visual cortex possibly because unpredictable things require attention. Animacy is multidimensional and our results show that distinct dimensions are differentially represented in human brain and behaviour.

Discussion

We investigated the representation of the multiple facets of animacy in brain and behaviour (see summary in Figure 8). Including multiple dimensions in a linear model enabled us to disentangle their roles. To increase the power of these analyses, we decorrelated the five dimensions of animacy using an optimized set of natural stimuli. We managed to reduce the maximum pairwise correlation from 0.59 (for random selection) to 0.35 (optimized). Using natural images, it may not be possible to create a stimulus set that has no correlations between dimensions at all. However, as long as the predictors of the model do not form a linearly dependent set, they can be disentangled in analysis by considering the unique variance explained by each. Our novel stimulus selection procedure using a genetic algorithm could be adopted to disentangle other multidimensional concepts beyond animacy.

It can be argued that we are not often exposed to some of the things depicted in this stimulus set, e.g., human fetuses. Despite some of the things depicted not being encountered often, all were familiar to the participants. Future studies may extend our approach to videos because some of the dimensions like "having mobility" may be better represented dynamically.

Despite its prominence in neuroscience literature, the living/non-living distinction ("being alive") did not explain variance in brain representations. This finding suggests that the ventral and dorsal visual streams do not represent the results of a deeper cognitive inference process that would assess whether something is alive. More perceptual dimensions such as "looking like an animal" explained more variance, especially in EEG time courses, and this dimension was most consistently judged by different people. For a stimulus set where dimensions of

animacy were not decorrelated, we predict that “being alive” would explain substantial variance. However, our decorrelated stimulus set revealed that other dimensions underlie the distinct responses to living things. The “being alive” dimension did explain variance in animacy ratings and similarity judgements suggesting that this dimension is present in cognition despite its lack of prevalence in the brain responses.

Consistent with higher-cognitive contributions, the “having agency” dimension explained significant variance in higher-level visual areas, explained more variance than most other animacy dimensions later in time, and was prominent in the judgments. This dimension was least consistently judged in behavioural animacy ratings suggesting that different people have different intuitions on whether something has agency; for example, our subjects were divided as to whether the robots in our stimuli had agency or not. Seeing agency (or not) in robots mirrors an ongoing debate in society and may influence how humans interact with the increasing presence of robots in their environment.

The “being unpredictable” dimension captured representations in both lower and higher-level visual cortex and earlier in time. One interpretation of this finding is that unpredictable things require attention and need to be processed early on. Humans need to know what to attend to in the visual world and if something unpredictable happens it captures attention, which enables us to stay on top of what is happening around us. This dimension of animacy has been studied mostly in the language domain (Lowder & Gordon, 2015), but we show that “being unpredictable” also explains visual representations.

In summary, we disentangled the dimensions of animacy using a novel approach for stimulus decorrelation and showed the contribution of each animacy dimension in explaining human brain representations and behavioural judgments. Future studies may expand on the representation of each of the dimensions while avoiding their entanglement and may apply this approach to other multidimensional concepts.

Methods

Stimulus set generation

Filling animacy dimension grid combinations

We created a grid with all possible combinations of dimensions of animacy (2^5 = 32). We asked participants (S = 12, mean age = 33, 6 females) to write down object category names (e.g., “humanoid robot”) for each combination in the grid to obtain a list of object categories (Figure 1 Step 1). Participants listed 100 categories, and we selected 3 images per category (total = 300 images, Figure 1 Step 2), which formed the basis for the experiment to rate the dimensions of animacy.

Ratings of dimensions of animacy to generate stimulus set

Twenty-six participants (mean age = 25, 21 females) performed animacy ratings of 300 object images through an on-line web-based interface. Participants judged each object image using a continuous scale from -10 to +10 for each dimension, e.g., -10 meant “dead” and +10 meant “alive” for the “being alive” dimension (Figure 1 Step 3). Thirty images were repeated for a within-participant consistency measure.

Stimuli subset selection using a genetic algorithm

To select a subset of 128 images for which ratings on the dimensions of animacy were maximally decorrelated we used a genetic algorithm. A genetic algorithm is an optimisation method that mimics biological evolution through natural selection. Fitness was defined as minimising the maximum correlation between dimensions of animacy (Figure 1 Step 4). We also introduced a penalty if the algorithm selected more than two stimuli from the same category (to ensure that stimuli were selected from a wide range of categories) and if the algorithm did not select at least one image of a human face and a human body (to have a reference point of object images that we know should have high ratings on the dimensions of animacy).

Animacy ratings

Stimuli

All stimuli are displayed in Figure 2a. Stimuli were 128 coloured images of real-world objects with natural backgrounds, selected from the Internet.

Participants

Nineteen participants performed an on-line animacy rating experiment (mean age = 27, 13 females). Participants had normal or corrected-to-normal vision. All of them were right-handed. Before completing the experiment, participants received information about the procedure of the experiment and gave their written informed consent. All participants received monetary reimbursement or course credits for their participation. The experiment was conducted in accordance with the Ethics Committee of the Department of Education and Psychology at Freie Universität Berlin.

Experimental design and task

Participants judged each object image using a continuous scale from -10 to +10 for each animacy dimension, e.g., -10 meant “dead” and +10 meant “alive” for the “being alive” dimension (Figure 3a). Additionally, participants performed a rating of "being animate" dimension in a similar fashion.

Similarity judgements

Stimuli

All stimuli are displayed in Figure 2a. The stimuli were the same 128 coloured object images as in the animacy rating task.

Participants

The same nineteen participants who performed the animacy rating experiment performed the web-based similarity judgements experiment. Participants first completed the EEG and fMRI experiments, then similarity judgements experiment, and finally animacy ratings experiment so that they did not know about specific dimensions of animacy tested while performing EEG, fMRI, and similarity judgements experiments.

Experimental design and task

We acquired pairwise object-similarity judgements for all 128 images by asking participants to perform an on-line multi-arrangement task using Meadows platform (www.meadows-research.com). During this task, object images were shown on a computer screen in a circular arena, and participants were asked to arrange the objects by their similarity, such that similar

objects were placed close together and dissimilar objects were placed further apart (Figure 5a). The multi-arrangement method uses an adaptive trial design, showing all object images on the first trial, and selecting subsets of objects with weak dissimilarity evidence for subsequent trials. To determine which stimuli to select for the next trial, the evidence weight of each stimulus had an evidence utility exponent (E=10) applied to it, to calculate its utility if the stimulus was picked. The similarity judgments task was completed if, among all pairs of stimuli, the pair with the lowest evidence had an evidence weight higher than 0.5. The multi-arrangement method allows the efficient acquisition of a large number of pairwise similarities. We deliberately did not specify which object properties to focus on, to avoid biasing participants' spontaneous mental representation of the similarities between objects. We aimed to obtain similarity judgements that reflect the natural representation of objects without forcing participants to rely on one given dimension. However, participants were asked after having performed the task, what dimension(s) they used in judging object similarity. All participants reported arranging the images according to categorical clusters. The reports suggest that participants used a consistent strategy throughout the experiment. The method of the object similarity judgements has been described in Mur et al., 2013, where further details can be found.

EEG

Stimuli

All stimuli are displayed in Figure 2a. The stimuli were the same 128 coloured object images as in animacy rating and similarity judgements tasks.

Participants

The same nineteen participants who performed the animacy rating and similarity judgements experiments performed the EEG experiment.

Experimental design and task

Stimuli were presented at the centre of the screen for 500 ms, while participants performed a paper clip detection task. Stimuli were overlaid with a light grey fixation cross and displayed at a width of 4° visual angle. Participants completed 15 trials. Each image was presented twice in every trial in random order with an inter-trial interval of 1–1.1 s. Participants were asked to press a button and blink their eyes in response to a paper clip image shown randomly every 3 to 5 trials (mean performance 99% (+/- 0.09, standard error)). These trials were excluded from the analysis.

Acquisition

The electroencephalogram (EEG) signals were acquired using BrainVision actiCHamp EASYCAP 64 channel system at a sampling rate of 1,000 Hz. The arrangement of the electrodes followed the standard 10-20 system.

Preprocessing

The time series were analysed with Brainstorm (http://neuroimage.usc.edu/brainstorm/). We extracted EEG patterns for each millisecond time point (from 100ms before stimulus onset to 1000ms after stimulus onset) for each trial. We filtered the responses between 0 and 50Hz.

Decoding

We performed pairwise decoding using a support vector machine (SVM) approach (Cichy et al., 2014). For each time point, EEG signals were arranged in 64-dimensional vectors (corresponding to the 64 EEG channels), yielding M = 30 pattern vectors per time point and condition. We sub-averaged the M vectors in groups of k = 5 with random assignment, obtaining L = M/k averaged pattern vectors. This procedure was performed to reduce computational load and improve the signal-to-noise ratio. Subsequently, for each pair of conditions, we assigned L-1 averaged pattern vectors to train a linear SVM using the LibSVM implementation (www.csie.ntu.edu.tw/~cjlin/libsvm). We used the trained SVM to predict the condition labels of the left-out testing data set consisting of the Lth averaged pattern vector. This process was repeated 100 times with random assignment of the M raw pattern vectors to L averaged pattern vectors. For every time point, we assigned the average decoding accuracy to a decoding accuracy matrix.

Peak latency analysis

We defined peaks of the decoding accuracy as time points with the maximum decoding accuracy.

fMRI

Stimuli

All stimuli are displayed in Figure 2a. The stimuli were the same 128 coloured object images as in EEG, animacy rating, and similarity judgements experiments.

Participants

The same nineteen participants who performed the animacy rating, similarity judgements, and EEG experiments performed the fMRI experiment.

Experimental design and task

Stimuli were presented using a rapid event-related design (stimulus duration, 500 ms) while participants performed a fixation-cross-brightness-change detection task, and their brain activity was measured with a 3T fMRI scanner. Stimuli were overlaid with a light grey fixation cross and displayed at a width of 4° visual angle. Each image was presented once per run in random order. Each run contained 32 randomly timed null trials without stimulus presentation. Participants had to report a short (100 ms) change in the luminance of the fixation cross via button press (mean performance 97% (+/- 0.14, standard error)).

Acquisition

Magnetic resonance imaging was acquired using Siemens 3T Trio with a 12-channel head coil. For structural images, we used a standard T1-weighted sequence (176 slices). The TR was 2 s and the ISI was 3 s. For fMRI, we conducted 9–13 runs in which 249 volumes were acquired for each participant. The acquisition volume covered the full brain.

Estimation of single-image activity patterns

The fMRI data were preprocessed using SPM12 (https://www.fil.ion.ucl.ac.uk/spm/). For each participant and session separately, functional data were spatially realigned, slice-time corrected, and coregistered to the participant-individual T1 structural image. We estimated the fMRI responses to the 128 image conditions with a general linear model (GLM), which included movement parameters as nuisance terms. To obtain a t-value for each voxel and condition,

GLM parameter estimates for each condition/stimulus were contrasted against a baseline. To assess the degree of the general visual stimulation we contrasted the parameter estimates for all images against the baseline.

Definition of regions of interest

For the ROI definition, we used a Probabilistic brain atlas (Wang et al., 2015). Anatomical masks were reverse-normalized from MNI-space to single-participant space. For each ROI, we extracted a multivoxel pattern of activity (t-values) for each of the 128 stimuli. We included 100 most strongly activated voxels in the ROI analysis.

Searchlight analysis

To analyse fMRI data in a spatially unbiased approach, we performed a volume-based searchlight analysis (Kriegeskorte et al., 2006) in each participant (radius of 4 voxels) with each animacy dimension RDM. We restricted the voxels included in the significance testing to visual stream areas (one-sided Wilcoxon signed-rank test).

Construction of representational dissimilarity matrix

We used the Representational Similarity Analysis toolbox for animacy dimension comparison (Nili et al., 2014). We computed response patterns (across animacy ratings, similarity judgements, EEG, and fMRI signals) for each image. We then computed response-pattern dissimilarities between images and placed these in a Representational Dissimilarity Matrix (RDM). An RDM captures which distinctions among stimuli are emphasized and which are de-emphasized.

Inferential analysis of model performance

We estimated animacy dimension performance by correlating the animacy dimension and data RDMs using Spearman's correlation coefficient. We determined whether each of the animacy dimension RDMs was significantly related to the data RDMs using a participant-as-random-effect analysis (one-sided Wilcoxon signed-rank test). We subsequently tested for differences in animacy dimension performance between each pair of dimensions of animacy using a participant-as-random-effect analysis (two-sided Wilcoxon signed-rank test). For each analysis, we accounted for multiple comparisons by controlling the FDR at 0.05.

Unique variance analysis

We used a hierarchical general linear model (GLM) to evaluate the unique variance explained by dimensions of animacy (Kietzmann et al., 2019). For each animacy dimension m, the unique variance was computed by subtracting the total variance explained by the reduced GLM (excluding the dimension of interest) from the total variance explained by the full GLM. Specifically, for dimension m, we fit GLM on X = "all dimensions but m" and Y = data, then we subtract the resulting R^2 from the total R^2 (fit GLM on X = "all dimensions" and Y = data). We performed this procedure for each participant and used non-negative least squares to find optimal weights. A constant term was included in the GLM model. We performed a one-sided Wilcoxon signed-rank test to evaluate the significance of the unique variance contributed by each dimension across participants controlling the expected false discovery rate at 0.05.

Acknowledgements

This research was supported by the Wellcome Trust [grant number 206521/Z/17/Z] awarded to KMJ; the Alexander von Humboldt Foundation postdoctoral fellowship awarded to KMJ; and

the DFG grant [CI 241-1/1, Emmy Noether Programme] to RMC. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Competing financial interests
The authors declare that they have no competing interests.

Data availability
The datasets generated during the current study are available from the corresponding author on reasonable request.

Code availability
The code generated during the current study is available from the corresponding author on reasonable request.

Author contributions
KMJ, IC, NK and RMC designed the experiments. KMJ and EN collected the data. JJFB helped to set up an on-line similarity judgements experiment. KMJ performed the analyses and wrote the paper. All authors edited the paper. NK and RMC supervised the work.

References

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron*, *34*(1), 149–159. https://doi.org/10.1016/S0896-6273(02)00642-6

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2003). FMRI responses to video and point-light displays of moving humans and manipulable objects. *Journal of Cognitive Neuroscience*, *15*(7), 991–1001. https://doi.org/10.1162/089892903770007380

Blumenthal, A., Stojanoski, B., Martin, C., Cusack, R., & Köhler, S. (2018). Animacy and real world size shape object representations in the human medial temporal lobes. *Hum Brain Mapp*, Sep;39(9):3779-3792. https://doi.org/10.1002/hbm.24212

Bracci, S., Ritchie, J. B., Kalfas, I., & Op de Beeck, H. P. (2019). The Ventral Visual Pathway Represents Animal Appearance over Animacy, Unlike Human Behavior and Deep Neural Networks. *The Journal of Neuroscience*, *39*(33), 6513–6525. https://doi.org/10.1523/JNEUROSCI.1714-18.2019

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space

and time. *Nature Neuroscience*, 17(3), 455–462. https://doi.org/10.1038/nn.3635

Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y., & Haxby, J. V. (2012). The Representation of Biological Classes in the Human Brain. *Journal of Neuroscience, 32*(8), 2608–2618. https://doi.org/10.1523/JNEUROSCI.5547-11.2012

Contini, E. W., Goddard, E., Grootswagers, T., Williams, M., & Carlson, T. (2019). A humanness dimension to visual object coding in the brain. *Neuroimage.* Nov 1;221:117139. https://doi.org/10.1016/j.neuroimage.2020.117139

Funnell, E., & Sheridan, J. (1992). Categories of knowledge? Unfamiliar aspects of living and nonliving things. *Cognitive Neuropsychology*, 9:2, 135-153, https://doi.org/10.1080/02643299208252056

Gobbini, M. I., Gentili, C., Ricciardi, E., Bellucci, C., Salvini, P., Laschi, C., Guazzelli, M., & Pietrini, P. (2010). Distinct Neural Systems Involved in Agency and Animacy Detection. *Journal of Cognitive Neuroscience*, *23*(8), 1911–1920. https://doi.org/10.1162/jocn.2010.21574

Gobbini, M. I., Koralek, A. C., Bryan, R. E., Montgomery, K. J., & Haxby, J. V. (2007). Two takes on the social brain: A comparison of theory of mind tasks. *Journal of Cognitive Neuroscience*, *19*(11), 1803–1814. https://doi.org/10.1162/jocn.2007.19.11.1803

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science.* Feb 2;315(5812):619. doi: 10.1126/science.1134475.

Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron*, *76*(6), 1210–1224. https://doi.org/10.1016/j.neuron.2012.10.014

Kietzmann, T. C., Spoerer, C. J., Sörensen, L. K. A., Cichy, R. M., Hauk, O., & Kriegeskorte, N. (2019). Recurrence is required to capture the representational dynamics of the human visual system. *Proceedings of the National Academy of Sciences*, *116*(43), 21854–21863. https://doi.org/10.1073/pnas.1905544116

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain

mapping. *Proceedings of the National Academy of Sciences*, *103*(10), 3863–3868. https://doi.org/10.1073/pnas.0600244103

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

Leib, A. Y., Kosovicheva, A., & Whitney, D. (2016). Fast ensemble representations for abstract visual impressions. *Nature Communications*, *7*, 13186. https://doi.org/10.1038/ncomms13186

Looser, C. E., Guntupalli, J. S., & Wheatley, T. (2013). Multivoxel patterns in face-sensitive temporal regions reveal an encoding schema based on detecting life in a face. *Social Cognitive and Affective Neuroscience*, *8*(7), 799–805. https://doi.org/10.1093/scan/nss078

Lowder, M. W., & Gordon, P. C. (2015). Natural forces as agents: Reconceptualizing the animate–inanimate distinction. *Cognition*, *136*, 85–90. https://doi.org/10.1016/j.cognition.2014.11.021

Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P. a, & Kriegeskorte, N. (2013). Human Object-Similarity Judgments Reflect and Transcend the Primate-IT Object Representation. *Frontiers in Psychology*, *4*(March), 128. https://doi.org/10.3389/fpsyg.2013.00128

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. *PLoS Computational Biology*, *10*(4), e1003553. https://doi.org/10.1371/journal.pcbi.1003553

Proklova, D., & Goodale, M. A. (2020). The role of animal faces in the animate-inanimate distinction in the ventral temporal cortex. *bioRxiv* https://doi.org/10.1101/2020.10.08.330639

Proklova, D., Kaiser, D., & Peelen, M. V. (2019). MEG sensor patterns reflect perceptual but not categorical similarity of animate and inanimate objects. *NeuroImage*, *193*, 167–

177. https://doi.org/10.1016/j.neuroimage.2019.03.028

Ralph, M. A. L., Howard, D., Nightingale, G., & Ellis, A. W. (1998). Are Living and Non-living Category-specific Deficits Causally Linked to Impaired Perceptual or Associative Knowledge? Evidence From a Category-specific Double Dissociation. *Neurocase*, 4:4-5, 311-338. https://doi.org/10.1080/13554799808410630

Rogers, T. T., Hocking, J., Mechelli, A., Patterson, K., & Price, C. (2005). Fusiform Activation to Animals is Driven by the Process, Not the Stimulus. *Journal of Cognitive Neuroscience* 17(3):434-45 https://doi.org/10.1162/0898929053279531

Rosenthal-von der Pütten, A. M., Krämer, N. C., Maderwald, S., Brand, M., & Grabenhorst, F. (2019). Neural Mechanisms for Accepting and Rejecting Artificial Social Partners in the Uncanny Valley. *The Journal of Neuroscience*, *39*(33), 6555–6570. https://doi.org/10.1523/JNEUROSCI.2956-18.2019

Sha, L., Haxby, J. V., Abdi, H., Guntupalli, J. S., Oosterhof, N. N., Halchenko, Y. O., & Connolly, A. C. (2015). The Animacy Continuum in the Human Ventral Vision Pathway. *Journal of Cognitive Neuroscience*, *27*(4), 665–678. https://doi.org/10.1162/jocn_a_00733

Shultz, S., & McCarthy, G. (2014). Perceived animacy influences the processing of human-like surface features in the fusiform gyrus. *Neuropsychologia*, *60*, 115–120. https://doi.org/10.1016/j.neuropsychologia.2014.05.019

Shultz, S., van den Honert, R. N., Engell, A. D., & McCarthy, G. (2015). Stimulus-induced reversal of information flow through a cortical network for animacy perception. *Social Cognitive and Affective Neuroscience*, *10*(1), 129–135. https://doi.org/10.1093/scan/nsu028

Silveri, M. C., Gainotti, G., Perani, D., Cappelletti, J. Y., Carbone, G., & Fazio, F. (1996). Naming deficit for non-living items: Neuropsychological and PET study. *Neuropsychologia*. Mar;35(3):359-67. doi: 10.1016/s0028-3932(96)00084-x

Thorat, S., Proklova, D., & Peelen, M. V. (2019). The nature of the animacy organization in human ventral temporal cortex. *ELife*, *8*, e47142. https://doi.org/10.7554/eLife.47142

Wang, L., Mruczek, R. E. B., Arcaro, M. J., & Kastner, S. (2015). Probabilistic maps of visual topography in human cortex. *Cerebral Cortex*, *25*(10), 3911–3931. https://doi.org/10.1093/cercor/bhu277

Wheatley, T., Weinberg, A., Looser, C., Moran, T., & Hajcak, G. (2011). Mind perception: Real but not artificial faces sustain neural activity beyond the N170/VPP. *PLoS ONE*, *6*(3), 1–7. https://doi.org/10.1371/journal.pone.0017960