Core dimensions of human material perception

Filipp Schmidt[1, 2, *, †] , Martin N. Hebart[1, 3 †] , Alexandra C. Schmid[4] , Roland W. Fleming[1, 2]

[1] Justus Liebig University Giessen, Germany

[2] Center for Mind, Brain and Behavior (CMBB), University of Marburg and Justus Liebig University Giessen, Germany

[3] Vision and Computational Cognition Group, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

[4] Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, USA

*Corresponding author: Filipp Schmidt

† Equal contribution

Abstract

Visually categorizing and comparing materials is crucial for our everyday behaviour. Given the dramatic variability in their visual appearance and functional significance, what organizational principles underly the internal representation of materials? To address this question, here we use a large-scale data-driven approach to uncover the core latent dimensions in our mental representation of materials. In a first step, we assembled a new image dataset (STUFF dataset) consisting of 600 photographs of 200 systematically sampled material classes. Next, we used these images to crowdsource 1.87 million triplet similarity judgments. Based on the responses, we then modelled the assumed cognitive process underlying these choices by quantifying each image as a sparse, non-negative vector in a multidimensional embedding space. The resulting embedding predicted material similarity judgments in an independent test set close to the human noise ceiling and accurately reconstructed the similarity matrix of all 600 images in the STUFF dataset. We found that representations of individual material images were captured by a combination of 36 material dimensions that were highly reproducible and interpretable, comprising perceptual (e.g., “grainy”, “blue”) as well as conceptual (e.g., “mineral”, “viscous”) dimensions. These results have broad implications for understanding material perception, its natural dimensions, and our ability to organize materials into classes.

Introduction

A large part of the human cortex is dedicated to the processing of visual signals. In recent years, there has been tremendous progress in identifying the cognitive and neural mechanisms that determine how we see. While much of this research has studied the perception of objects—the *things* in the world around us [1–4]—recent work has shown that the human visual system is also highly tuned to processing materials (e.g., [5–7])—the *stuff* that things and the world are made of [8–10]. Being able to accurately perceive material properties is crucial for interacting with our surroundings, whether we are judging if a lake is frozen, an apple is rotten, or a tool is wet and slippery.

Given the complex physical characteristics of natural materials and their enormous variability in visual appearance and functional significance (**Fig. 1**), how do we identify, compare, and categorize materials and, more generally, make sense of them so that we can interact with them in a meaningful manner? What underlying representations organize the wide range of materials and their properties to support everyday judgments and tasks?

A powerful way to characterize the mental representations of materials is in terms of material *properties*, or dimensions, paralleling previous work using objects [11]. For example, it has been shown that different classes like water and wood can be described by a particular combination of material properties [12, 13]: water is high in perceived glossiness and transparency but low in roughness and hardness, while the opposite is true for wood. These studies hint at a flexible mapping between dimensions of material perception and our ability to identify, categorise, and understand them, thus informing our everyday behaviour in relation to materials. These material dimensions could encompass information about specific visual characteristics (e.g., surface colour or texture properties), or information about conceptual characteristics (e.g., heavy or fragile).

However, while previous studies have revealed many important details of how properties and categories are inferred [13–15], they typically focused on small numbers of manually-selected properties and restricted sets of materials, which may not generalize well to the wealth of material properties in our world. Other work successfully trained supervised deep neural networks to recognize materials or material properties from images or videos [16–19]. This, however, did not allow relating patterns of network activations to real-world perceptual features. And while it is possible to come up with a sheer infinite number of candidate material dimensions, our ability to identify a core set of representational dimensions that underlie our mental representations of materials and are relevant for behavioural judgments is still surprisingly limited.

Here we sought to gain a more comprehensive and principled understanding of the properties that describe the full complexity of material representations by identifying core dimensions that determine similarity judgments between materials. Similarity judgments offer an established approach for characterizing the multidimensional space underlying mental representations [20–22] and have been central to gaining access to the mental representations of objects [3, 4], and it seems plausible that similar approaches could benefit our understanding of material perception. To provide a comprehensive characterisation of perceived material similarity, we first compiled a broad dataset of 600 natural material images, derived from 200 picturable material concepts in the English language. Next, we assessed the perceived similarity of these images in a large-scale crowdsourcing experiment comprised of 1.87 million trials, in which we asked participants which of two material images was more similar to a third reference material image. In this paradigm, the non-selected material image acts as a context, which effectively highlights the relevant dimensions shared by the other two [23, 24]. Based on a computational model of this task, we then identified a set of representational dimensions that characterizes material similarity judgements.

The dimensions of our model captured ~90% of variance in perceived similarity with only 36 dimensions. These dimensions were generally interpretable and encompassed aspects of the materials' appearance (e.g., texture, shape, and colour), mechanical properties (e.g., "viscous") as well as membership of certain material classes (e.g., metal, wood). This finding and our material embedding have broad application for studying material perception, its natural dimensions, and the representation of material features and categories in the human brain.

Results

We aimed to uncover core representational dimensions of visual material perception by taking a data-driven approach, where we (1) created a broad, systematically sampled dataset of material images, (2) collected a large-scale crowdsourced behavioural similarity dataset for these images, and (3) carried out computational modelling to identify the core dimensions underlying these similarity judgments.

STUFF: A dataset of 200 material classes in 600 images

First, we sought a systematic method for identifying material categories to mirror the full richness and complexity of material appearances in the real world. To this end, we identified an extensive list of material words, as word usage captures many of the behaviourally-relevant distinctions between different material classes [25, 26]. Specifically, we started with 8,671 concrete nouns in the American English language, which we manually distilled to 200 picturable material concepts, spanning such diverse materials as algae, brass, ebony, fleece, oil, rubber and zinc (see **Methods** and **Supplementary Materials and Methods** for details, and **Supplementary Table T1** for a complete list). The resulting list of 200 materials is much shorter than comparable lists of objects [25, 26], yet far longer than the number of material classes considered by previous studies in material perception [12, 19, 27]. This indicates that the diversity of material names may be lower than that for objects, while highlighting that broad, systematic sampling can lead to a much wider range of material classes than previously studied.

The scope of the dataset can be visualised by clustering all materials based on their semantic similarity, using correlations between 300-dimensional sense embedding vectors for each material noun [28], identified by each respective WordNet synset (cf. **Supplementary Table T1**). The resulting semantic clustering illustrates the broad semantic diversity of the materials (**Fig. 1a**). This is also supported by the fact that while our materials include WordNet hyponyms of all 10 super-classes of the influential Flickr Material Database [6] (e.g., stone, glass, water), about half of our materials (106 out of 200) extended beyond all of these super classes.

Having verified the semantic breadth of the 200 materials, we went on to collect 3 high-quality, close-up, naturalistic photographs of each material concept. Specifically, we performed an extensive web search for materials depicted in their typical aggregate state and form (e.g., liquid oil or grains of salt), including close-ups of object surfaces. The resulting 600 images of materials were cropped to square images (**Fig. 1b**). The complete STUFF dataset, and an extended version with 15 images per concept, will be available for download.

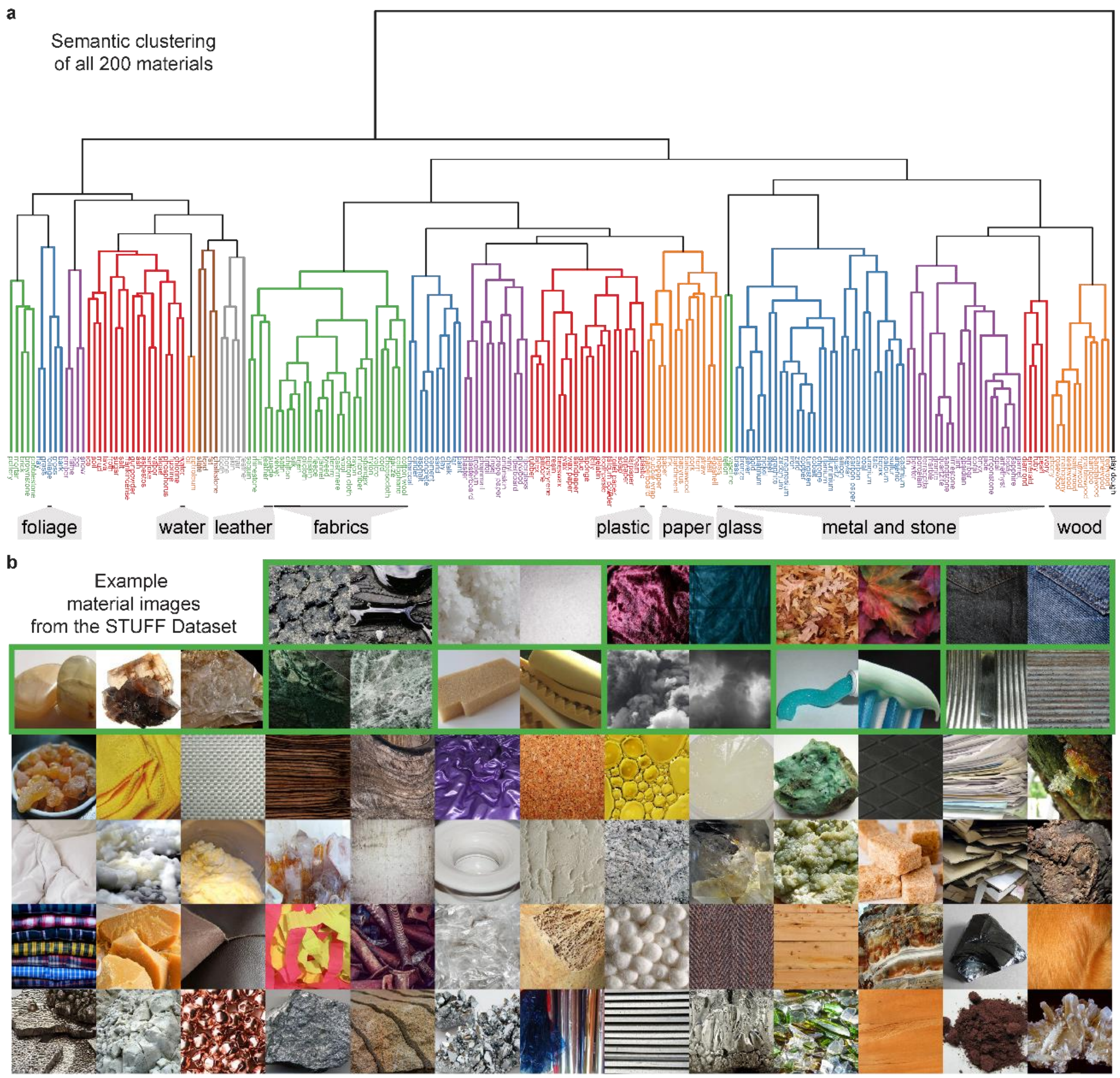


Figure 1. STUFF dataset. a. Hierarchical clustering of all 200 material classes based on an off-the-shelf semantic embedding for material nouns [28], illustrating the scope of our dataset. An approximate assignment of materials to the 10 super-classes of the Flickr Material Database [6] is shown below the dendrogram. **b.** Example images from our STUFF dataset, which contains 3 images per material class; images from the same class are grouped by green frames. Copyright information for all images is provided in **Supplementary Table T2**.

<u>A 36-dimensional representational embedding captures single-trial material similarity judgments</u>

Having assembled a broad dataset of material classes and images, we next sought to measure the perceived similarity between the 600 images in the STUFF dataset, in a triplet two-alternative forced choice (2-AFC) task. In a given trial, we presented a reference material image in a triangular arrangement with two other test materials, and we asked participants to identify which of the bottom two images was more similar to the top reference image [17]. In this task, we defined similarity as the probability of choosing two material images together, marginalized across all contexts imposed by the third image. For example, participants judged chrome as similar to moonstone when compared to paper, but as dissimilar to moonstone when compared to tin (**Fig. 2a).** Depending on the context, different properties—such as solidity, colour, or gloss—could dominate the decision. As a consequence, the triplet task highlights the relevant dimensions that form the basis of similarity judgments. By sampling very broadly across many such contexts, we can thus determine the similarity of two images according to a broad range of possible material dimensions.

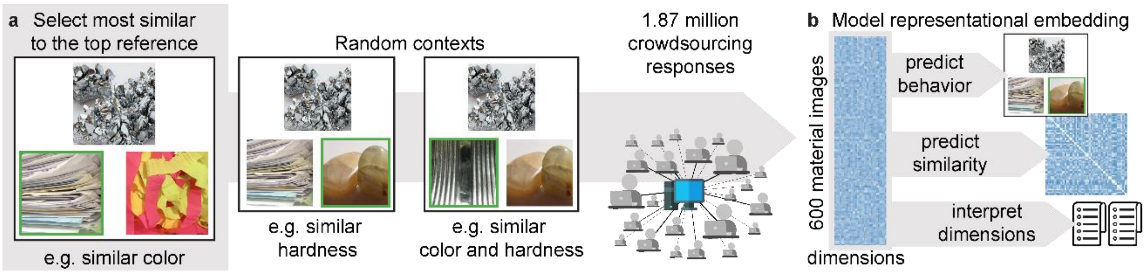


Figure 2. Experimental paradigm and modelling. a. Illustration of the triplet 2-AFC task. Material images were presented to participants in different contexts imposed by the third object in a triplet. We used online crowdsourcing to sample across a wide range of these random contexts. **b.** The goal of the modelling procedure was to learn a representational embedding [24] that (i) captures choice behaviour in the triplet 2-AFC task, (ii) predicts similarity across all pairs of materials, and (iii) provides interpretable material dimensions. Since only a subset of all possible triplets had been sampled, the model also served to complete the sparsely sampled similarity matrix. Copyright information for all images is provided in **Supplementary Table T2**.

For the triplet 2-AFC task, a full sample of the 600 × 600 similarity matrix would require a total of ~107.46 million trials. Given the excessive cost associated with acquiring a full similarity matrix, and since we anticipated a similarity matrix with a much lower rank than the number of images, we instead collected a smaller subset of 1.87 million responses from a sample of 5,038 workers on the online platform Amazon Mechanical Turk (1.74% of all possible unique triplets). We then used a computational procedure modelling the underlying cognitive process, which allowed us to fill the gaps in the similarity matrix (see below). To estimate the noise ceiling reflecting the rating consistency and thus the best possible performance any model could achieve given the data, part of this dataset was a randomly-chosen subset of 1,200 triplets which we sampled 60 times each. To test how well a model based on our sample can reproduce the full similarity matrix, we also collected a separate sample of all triplet similarity relations for a subset of 60 images, with each possible triplet repeated twice (205,320 responses).

To model the cognitive process underlying the formation of similarity judgments, we sought a computational modelling approach that could: (1) predict behavioural responses from the triplet 2-AFC task, (2) generalize to all pairs of material images in the dataset, and (3) provide interpretable material dimensions. To achieve these objectives, we used the sparse positive similarity embedding technique described in [23, 24] (code available at https://github.com/ViCCo-Group/SPoSE) adapted to a 2-AFC task and quantified material images as vectors in a multidimensional representational space (**Fig. 2b**).

Specifically, the full representational embedding constitutes a matrix in which columns correspond to material dimensions and rows to material images. Each individual row thus describes a material image as a vector in a multidimensional feature space. The representational embedding is built on three key assumptions about the material dimensions: sparsity, continuity, and positivity. The sparsity assumption reflects the fact that not all features are expressed in all materials (e.g., marble would score zero on a putative dimension of viscosity). Continuity and positivity assumptions allow us to interpret the numerical value for a given dimension as the degree to which that feature is expressed in a material (e.g., toothpaste is higher in viscosity than water) [29]. These constraints thus yield dimensions that can be combined together as behaviourally-relevant parts of images, constraints which in the past have been shown to yield interpretable dimensions [24, 30]. The model was initialized with 90 random dimensions and was trained on 90% of available trials, with the remaining 10% serving as an independent test set. To induce sparsity, the model was regularized with an L1 norm, and the regularization parameter λ which controls the trade-off between sparsity and out-of-sample model performance was determined using cross-validation on the training set ($\lambda = 0.0038$).

We iteratively adapted the weights of the 90 dimensions based on the difference between the model's predicted choice probability and the empirically measured choice. At the end of training, as a result of the sparsity constraint, 54 dimensions yielded values consistently close to 0 and were eliminated (see **Methods**). The resulting embedding thus contained 36 dimensions, which we sorted on the basis of the sum of all dimension values averaged across all materials, in descending order (**Fig. 3**). Due to the stochastic nature of the modelling procedure, fitting the model repeatedly may lead to a different embedding and a slightly different number of dimensions. Thus, to estimate the stability of the model, we ran it 50 times with different random initializations (see **Methods**). Across those models, most dimensions were highly reproducible (Pearson's $r > 0.9$ in 28/36 dimensions and $r > 0.8$ in 34/36 dimensions), indicating the stability of the embedding (see **Fig. S1a, b** for the reproducibility of all dimensions).

To evaluate the embedding's predictive performance for triplet 2-AFC judgments, we computed the human noise ceiling from the additional repeated sample of 1,200 randomly chosen triplets (see above) and estimated the consistency of choices for each triplet across participants. Averaged across all those triplets, the upper limit in fitting individual trials from the data was 73.84% (chance=50%), with the model predicting 71.86% of individual trials in the independent test data. Thus, with respect to the best possible prediction any between-participant model could achieve (human noise ceiling), the model reached a chance-corrected performance of 91.70% at the individual-trial level (**Fig. 3a**).To underscore the sensitivity of this approach to subtle changes in choice probability, we confirmed this prediction by fitting the predicted choice probability for the 1,200 test triplets to the actual choice probabilities, yielding a predictive accuracy for the noise ceiling triplets of 90.25% and a correlation coefficient of $r = 0.81$ (**Supplementary Fig. S1**).

<u>Accurate reconstruction of perceived material similarity judgments</u>

Having confirmed that the embedding could accurately predict individual trial behaviour, next we evaluated how well it could predict behaviourally measured similarity. To this end, we compared the fully-sampled similarity matrix derived from the random sample of 60 images with the predicted similarity using the 36-dimensional representational embedding (**Fig. 3b-c**). The matrices were highly correlated (Pearson's $r = 0.90$; $p < 0.001$; randomization test; 95% CI, 0.89–0.91). To reveal how well this prediction worked as a function of the noise in the data, we measured the split-half reliability of the fully-sampled similarity data of the 60 images. The split-half reliability of the similarity matrix of the 60 images was $r = 0.97$ (Spearman-Brown corrected), demonstrating that we were able to predict 86.01% of the explainable variance in similarity. These results demonstrate that despite a large variety of visual appearances in the dataset and many possible features that can contribute to material judgments, a low-dimensional representational embedding was able to accurately reproduce behaviourally measured similarity for material images.

To better understand the structure of the similarity matrix, we further examined the structure of the predicted 600×600 similarity matrix. As expected by the 2-AFC task, we observed a mean similarity of 0.50, with a wide spread of similarities (SD = 0.18, range = 0.11–0.98). Similarities within each of the 200 material classes (M = 0.84) were higher than similarities between different material classes (M = 0.50, $t(199) = 45.28$, $p < 0.001$). Specifically, similarity was highest between images within material classes of brick, straw, and lava (all 0.97) and lowest within classes of fleece, feather, and onyx (0.63, 0.61, and 0.60). Thus, even though our dataset contained multiple examples per material class, participants' responses to images within each class still exhibited a high degree of variability, likely due to differences in material appearance.

Given the semantic structure present in the STUFF dataset as revealed by the validation analysis (**Fig. 1a**), it may be expected that semantics would dominate the behavioural judgments [31]. Indeed, distributional word vector models in the past have been shown to correlate well with human word similarity judgements, explaining up to 50% of their variance (e.g., [32]). Therefore, we obtained the similarity matrix for material nouns from correlations between 300-dimensional semantic embedding vectors for material nouns [28] (cf. **Fig. 1a**) and correlated the resulting semantic similarity matrix to our behavioural similarity matrix. We found a small effect size (R^2 = 0.11), suggesting that a substantial amount of variance in behavioural similarity is explained by factors other than conceptual knowledge captured in semantic embeddings.

To test the degree to which the three examples of our representational embedding were sufficient for revealing the distinctions between material classes, we iteratively trained a linear support vector machine on 2 examples on the 36-dimensional embedding and evaluated it on the left-out third example. This yielded a pairwise accuracy of 97.21% (chance = 50%), and top-1 and top-5 accuracies of 47.00% and 74.33%, respectively (chance levels: 0.5% and 2.5%, respectively; median rank of correct answer: 2). Thus, the representational embedding is highly informative about material classes, despite being based on only 36 dimensions and using only two training examples, reinforcing the notion that perceptual representations of materials are substantially richer that purely semantic ones.

<u>Interpretable core dimensions of material representation</u>

To determine the nature of the core representational dimensions underlying material similarity judgments, we next sought to test whether the 36 dimensions in the embedding were interpretable. To this end, we first asked a separate group of observers (n = 20) to provide verbal labels for each dimension. On each trial, we visualized a given dimension by showing material images that spanned a broad range of feature values, ranging from images with high weights on these dimensions to images with low weights. Observers then entered verbal labels to describe the depicted characteristic (see **Methods**). This provided us with semantic labels for each of the 36 dimensions (**Fig. 3d**).

On average, we obtained 40.40 labels for each dimension (mean = 2.02 labels per participant; range = 1–10 labels), with good agreement considering the under-constrained task: In each category, the three most frequent labels together accounted for an average of 81% (range: 15–85%) of observers' responses. Labels were conceptual (e.g., *mineral*), or referred to optical (e.g., *metallic*), texture (e.g., *grainy*), shape (e.g., *round*) or physical properties (e.g., *viscous*) (**Fig. 3d**). All labels with > 20% agreement in the naming task are provided in **Supplementary Table T4**. By identifying dimensions by their most frequent label and sorting dimensions according to their sparsity, we can show to what extent different features and type of features explained variability in the behavioural data (**Fig. 3e**). In addition to this labelling approach, to determine the interpretability without potential rating bias, we pursued a separate, data-driven approach based on identifying semantic features that are shared between materials loading strongly on individual dimensions (see **Supplementary Materials and Methods** and **Supplementary Table T5**), which for dimensions not purely defined on image characteristics (e.g., 'mineral') closely mirrored the results found in participants.

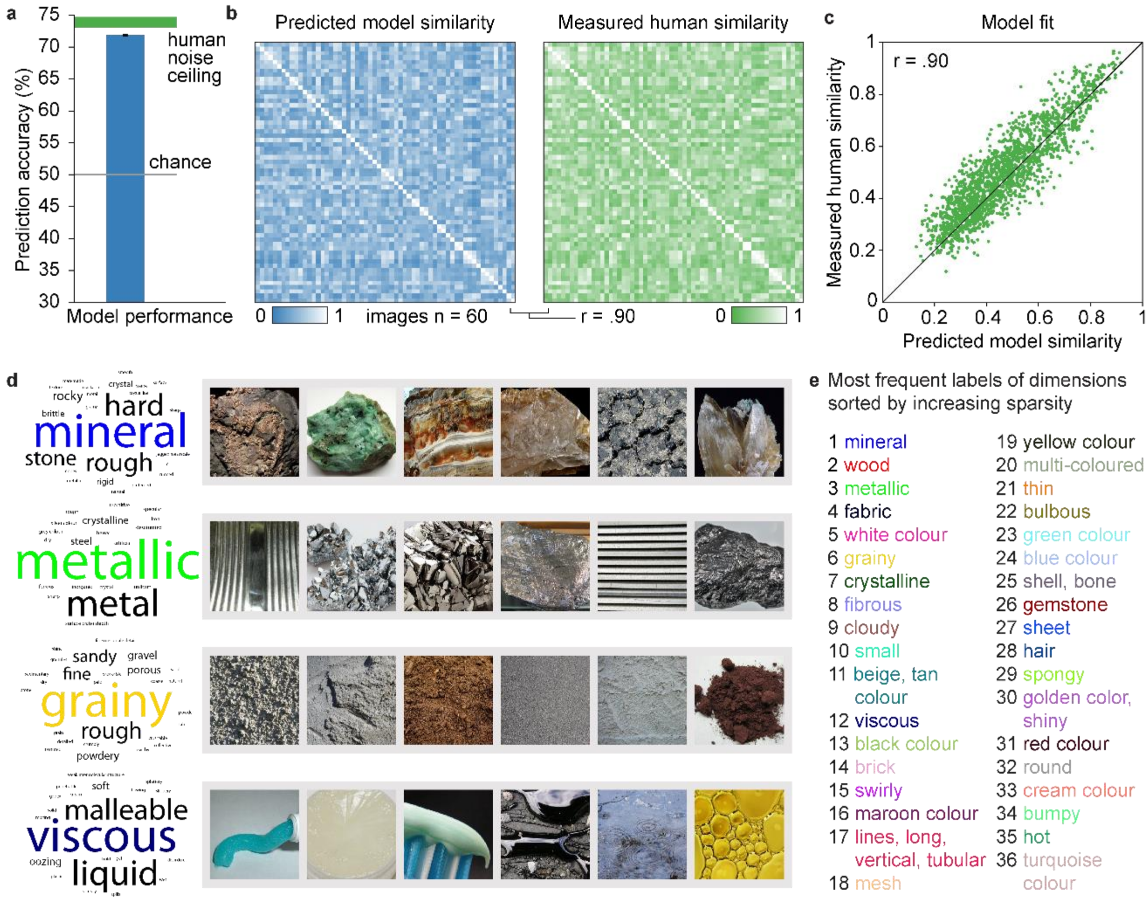


Figure 3. Modelling results and interpretability of model dimensions. a. Model prediction performance for individual trials in independent test data, relative to chance (grey) and the human noise ceiling (green). The noise ceiling denotes maximal performance given the noise in the data and is obtained by calculating the consistency in participants' responses to the same triplet. The model reached 91.7% of the noise ceiling. Error bar for prediction and width of noise ceiling denote 95% confidence intervals. **b.** To estimate how well the model predicted behavioural similarity, we compared a fully sampled behavioural similarity matrix for a subset of 60 images (blue) to the model-generated similarity matrix for these images (green). **c.** The close fit between both show that most explainable variance was captured by the model (Pearson's $r = 0.90$; $P < 0.001$; randomization test; 95% CI, 0.88–0.91). **d.** Visualization of four example dimensions and associated results of the dimension-labelling experiment. The example dimensions are visualized showing six images with large embedding weights in these dimensions. The word clouds reflect a summary of the semantic labels provided by research participants for these dimensions. Copyright information for all images is provided in **Supplementary Table T2**. **e**. The most frequent labels provided for each dimension, with dimensions 1–36 ordered according to their sparsity (i.e., mineral showed the lowest sparsity with almost all of our 600 images having nonzero values; turquoise colour showed the highest sparsity with only a few images having nonzero values).

To characterize the relationship between the 36 model dimensions, we conducted hierarchical clustering of dimensions across all material images (**Fig. 4a**). This analysis highlights what dimensions are co-expressed in material images. For example, the dimension "gemstone" formed a cluster with the dimension "round", indicating that gemstone materials tend to be round. Likewise, the dimension "hot" formed a cluster with the dimension "red colour", and the dimension "brick" with the dimension "lines, long, vertical, tubular", highlighting other important commonalities between material classes. Dimensions that covaried also extended across corresponding areas of the material similarity space, with some dimensions overlapping in materials (mineral, metallic) and others hardly ever (mineral, viscous; **Fig. 4b**). However, correlations between the 36 dimensions were generally low (maximum Pearson's r = 0.37), with only two dimension pairs showing a moderate negative correlation (mineral vs. fabric, r = -0.51; metallic vs. beige, r = -0.41). The overall distinctiveness of dimensions is also apparent in the sparsely localized maps of high features values in **Fig. 4b**. Together, these results reveal which dimensions materials tend to have in common while highlighting that dimensions reflect genuinely distinct material attributes.

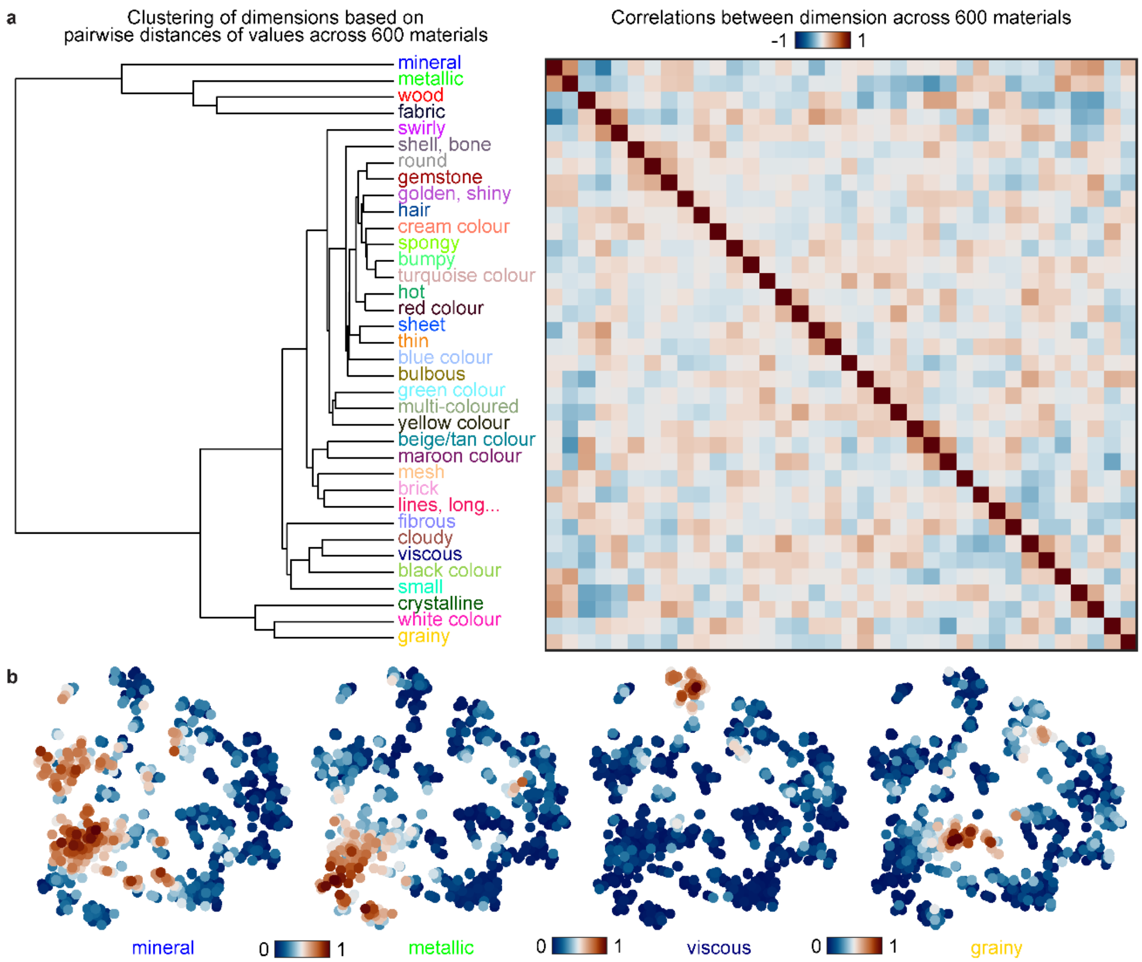


Figure 4. Similarity between model dimensions and expansion in material similarity space. a. Clustering of 36 model dimensions based on the pairwise distances between their values across all 600 materials, together with the correlation matrix, showing mostly low to moderate correlations between dimensions (mean correlation r = -0.01, standard deviation = 0.12, range = -0.51–0.37). **b.** The distribution of weights for four example dimensions across all 600 images, visualized by plotting images as points in a two-dimensional t-SNE visualization of the similarity embedding (initialized with multidimensional scaling; dual perplexity, 5 and 30; 1,000 iterations). Colour represents how strongly each image expressed the particular dimension (normalized to range 0–1: blue–red).

Visualizing and interpreting dimension profiles for individual images and global similarity

Having characterized the interpretability of the 36 dimensions across all materials, we further explored the dimensions' interpretability by visualizing dimensions for individual material images (**Fig. 5a**).

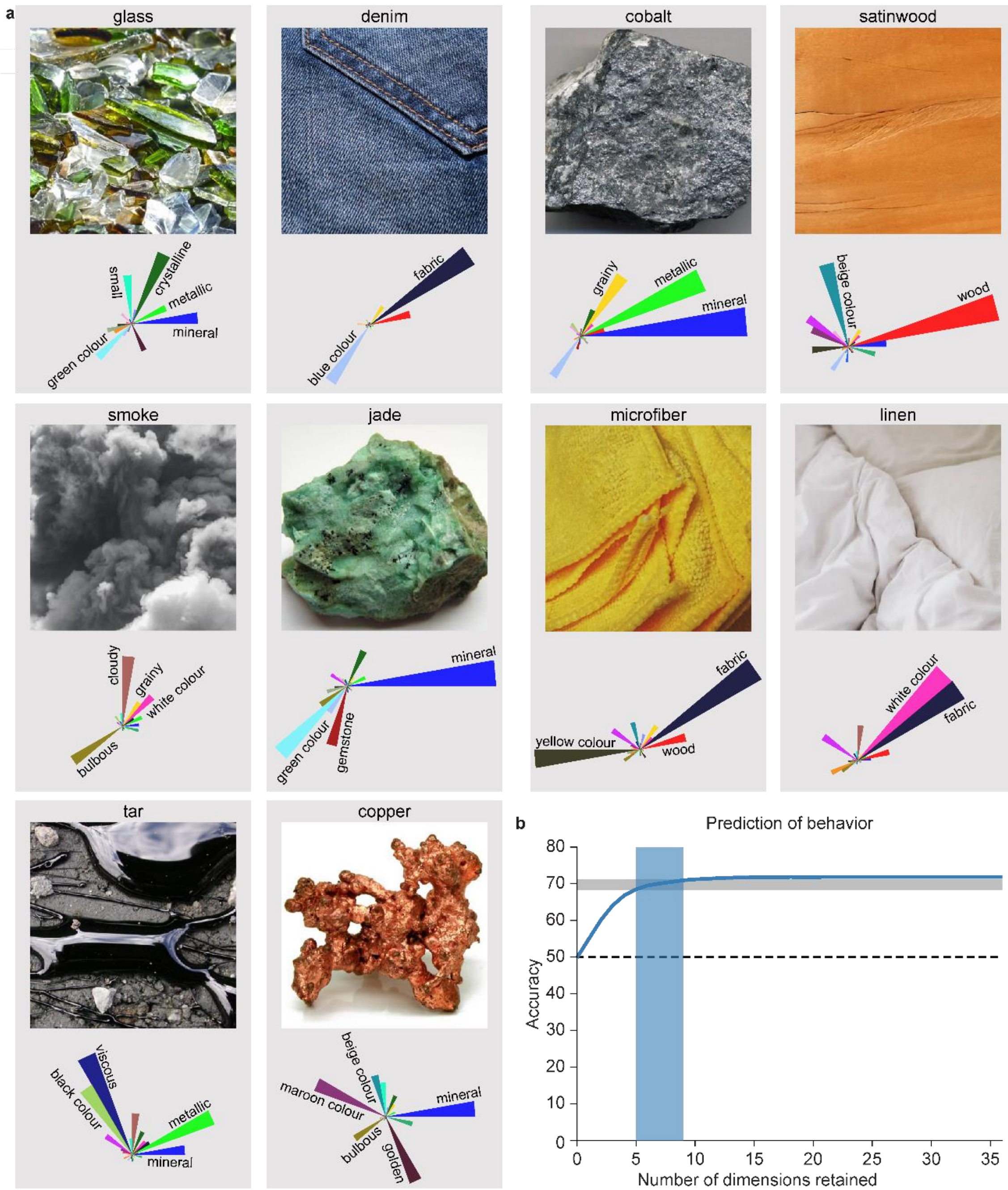


Figure 5. Behavioural judgements and similarity for individual images are well explained by 5 to 9 dimensions. a. Example material images and corresponding distributions across dimensions, using rose plots with each petal reflecting the degree a material dimension is

expressed for that image. Petal orientation and colours indicate individual dimensions and length indicates the value in that dimension. Dimension labels are only provided for weights > 0.53. Copyright information for all images is provided in **Supplementary Table T2**. **b.** For explaining 95–99% of the predictive performance in behaviour, only 5 to 9 dimensions per image are required, however these dimensions varied between images (see main text for details).

For example, the example image of smoke is characterized primarily by the dimensions of “cloudy”, “bulbous”, “white colour”, and “grainy”, while the image of jade is characterized primarily by the dimensions of “mineral”, “green colour”, and “gemstone”. The fact that each individual image can be described by a small number of dimensions suggests that not all 36 dimensions are required for all similarity judgments. To quantify this observation, we tested the predictive accuracy of the model by maintaining only the most prominent dimensions for each material. For each individual material, we set the dimension with the lowest weight to zero, predicted behaviour, recomputed the similarity matrix to measure the effect on the model’s predictive performance, and repeated this process until only one dimension was left. This analysis revealed that in order to explain 95-99% of the variance in the raw triplet-task responses, only 5 to 9 dimensions were required in any given trial (**Fig. 5b**). Accordingly, human responses to individual material images indeed appear to be driven by the expression of a small number of decisive features, while at a global level, and across individuals, observers may integrate across a larger number of these dimensions, highlighting the importance of taking diverse features into account for judging the similarity of materials [24].

Finally, to characterize the global similarity structure with respect to the 36 dimensions, we combined the visualization of dimension profiles for individual images with the t-SNE projection of the similarity embedding for all 600 material images (**Fig. 6**). The organization shows large clusters that reflect the expression of particular material properties (e.g., minerals, fabrics, woods, or viscous materials). However, the space is also highly organized within these larger clusters, with local arrangements of materials according to expressions of non-dominant properties (e.g., different colours within minerals, fabrics or viscous materials; **Supplementary Fig. 2**). At a global level, we note a gradient from images with bulbous and clumpy appearances (top right) to more linear parallel structures, like stripy wood grain or paper sheets (bottom) via grainy and fibrous materials (middle). Another notable gradient is the separation of hard materials (left and bottom) from softer materials (right), with grainy and fibrous materials in between. Hot materials (top left) appear to be relative outliers in a tight cluster separated from the rest.

This visualisation provides a more comprehensive and multifaceted view of the relationship between material properties and categories than previous studies. It is noteworthy that many of the material characteristics (e.g., grainy, fibrous, hot) are currently underrepresented in material perception research [34, 35], yet play an important role in the global organization of the perceptual representation of materials. Future research should investigate the visual cues underlying such characteristics.

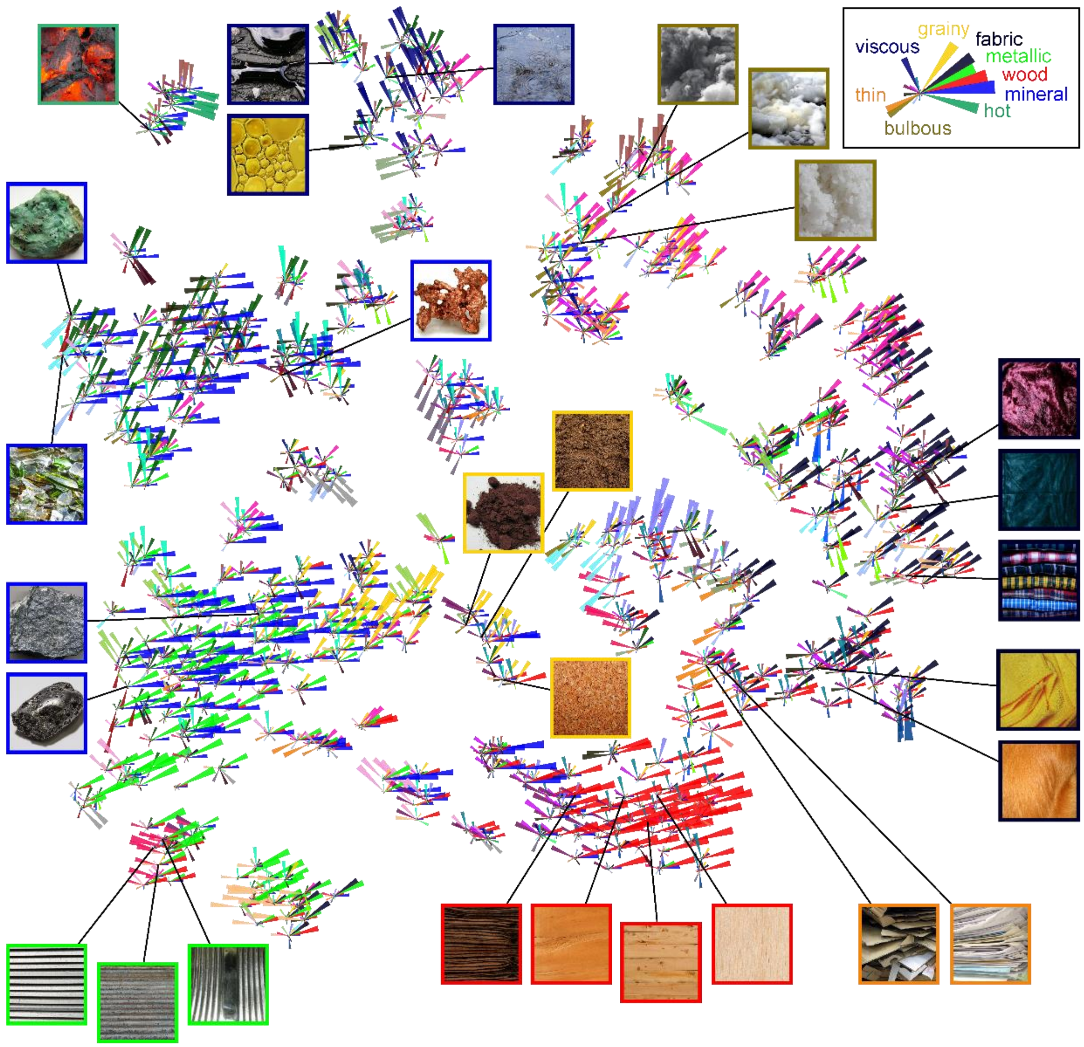


Figure 6. Two-dimensional visualization of the similarity embedding. The similarity embedding is visualized by combining rose plots for each material with t-SNE dimensionality reduction (initialized by multidimensional scaling; dual perplexity, 5 and 30; 1,000 iterations). Frames of example images are coloured according to the dominant dimension, but note that multiple other dimensions also play a role for each stimulus. Copyright information for all images is provided in **Supplementary Table T2**.

Discussion

Our world is made of stuff—and our visual system is highly attuned to processing and interpreting that stuff (e.g., [5, 6]). Material appearances vary enormously, both between and within material classes, making classifying and comparing materials computationally challenging. Our mental representation of materials must be robust and efficient, yet also sufficiently multifaceted to support diverse behaviours [8–10]. Here, we sought to characterise such representations by moving beyond small-scale experimental approaches that inevitably rely on limited stimulus sets [12, 18, 19, 27] and tasks [33, 34] restricted by the goals of the experimenter, which therefore may not capture important aspects of our perceptual experience. Instead, we used a large-scale, data-driven approach to identify core dimensions underlying similarity judgments of materials. We broadly and systematically sampled 200 material classes, assembled 600 photographs of these materials exhibiting highly diverse characteristics and appearances, collected over 1.8 million perceptual similarity judgments between triplets of these photographs, and applied computational modelling methods to derive intuitively interpretable dimensions from behavioural responses. This revealed a set of 36 dimensions along which participants appear to judge materials, capturing diverse characteristics related to surface reflectance and texture properties (e.g., *green colour*, *repetitive*), shape-related properties (e.g., *thin*, *round*), physical attributes (e.g., *malleable*, *brittle*), and category membership (e.g., *wood*, *paper*). These 36 dimensions can account for the majority of the explainable variance in the similarity judgments, and thus provide the most comprehensive and systematic mapping of the representation of materials in the human visual and cognitive system to date. These findings provide a foundation for future studies of material representations and studies that map out how these material dimensions are encoded in the human brain.

Characterising mental representations using material similarity judgments

The STUFF dataset covers 200 materials in 600 images, and we are in the process of expanding the dataset for future studies. Our findings provide a proof-of-principle demonstration that material classes—as defined by linguistic terms—are an effective way to define a stimulus set that captures information about behaviourally-relevant properties. Of course, alternative decisions about exactly which terms and images to include would have some impacts on the findings. For example, the lower sparsity of the mineral dimension likely reflects the fact that the STUFF dataset contains a lot of minerals, since there are many distinct nameable minerals (e.g., granite, jade, amethyst). Including fewer mineral samples would necessarily reduce the prominence of the 'mineral' dimension. Future studies could take into account the frequency of encounters with different material classes and their behavioural relevance when defining the dataset. Yet, despite this, our analyses suggest that the dimensions we have identified are surprisingly reproducible and robust, extending to behaviours beyond similarity judgments, such as categorisation.

One reason for the robustness of our findings is that the semantic sampling is not strongly reflected in the similarity judgments. Instead, a substantial amount of variance in the similarity data was explained by factors other than conceptual knowledge captured in semantic embeddings. This suggests a key role of appearance characteristics such as the colour, texture and shape of the materials in driving similarity judgments. Visually inspecting the ranking of stimuli along individual dimensions (e.g., 'grainy') reveals clearly appearance-based organization.

Advantages of approach for future studies

The modelling approach used here has a number of benefits. Most notable is the ability of the 36-dimensional representational embedding to accurately capture single-trial material similarity judgments (see **Fig. 3**). This low dimensionality drastically reduces the complexity of relating features to external behaviour and allows comprehensive modelling of interpretable brain signals in response to material dimensions. Another advantage of taking a data-driven approach is the possibility of discovery unconstrained by the experimenter's hypotheses about relevant dimensions. The modelling approach also allows a way to infer a multidimensional representation, while recognizing that not all dimensions are expressed in all materials. Finally, by definition, the dimensions comprise properties whose distinction is behaviourally relevant—at least in terms of determining the perceived similarity between items. As revealed by the clustering analysis, the dimensions capture distinct aspects of material appearance. The embedding of items in the multidimensional space reveals which dimensions materials tend to have in common while highlighting that dimensions reflect genuinely distinct material attributes.

Interpretability of dimensions

The dimensions that emerge from our analysis reflect a general-purpose representation that can be selectively sampled for different tasks. It is interesting to note that to explain 95-99% of the variance in the raw triplet-task responses, only 5 to 9 dimensions were typically required in any given trial. Accordingly, human responses to individual material images indeed appear to be driven by the expression of a small number of decisive features, while at a global level, and across individuals, observers may integrate across a larger number of these dimensions. Importantly, despite the relatively unconstrained task, the dimensions that emerged were highly interpretable by other participants, who tended to provide consistent labels for the dimensions: the three most frequent labels together accounted for an average of 81% of observers' responses.

Another important aspect of the dimensions is their continuous nature, which represents the degree to which a given feature is expressed in a material (e.g., toothpaste has higher viscosity than water). As the dimensions included both appearance attributes (e.g., colour and texture), and categories (e.g., metal, wood), this provides a unification of perceptual qualities and material classes [12] within a single framework. Rather than expressing category membership as a simple binary quantity, the dimensions provide continuous representations of properties that reflect the degree to which individual images exhibit particular material properties and categories. Moreover, our clustering analysis revealed the representational embedding is highly informative about material classes, despite being based on only 36 dimensions and using only two training examples per category.

Future directions

One of the most important directions for future research is linking the latent space defined by material similarity judgments to other behaviours. A key open question is how we flexibly access dimensions of this space in a variety of tasks, such as predicting the likely future behaviour of materials in response to external events, or planning and executing physical interactions with them.

Another key direction for future studies is linking the embedding of material images—and the dimensions relating them to one another—to neural representations. To date, the mapping of materials across cortical regions [35–39] is far less comprehensive than for objects, faces, or places [1, 3, 4, 40–49]. If the latent space identified here captures core, behaviourally-relevant dimensions there is hope that we may be able to establish a clearer link between perception and neural activity.

Finally, it would be interesting to ask to what extent the embedding we have identified could emerge through unsupervised (or weakly supervised) learning processes. Some form of cross-referencing across the senses, as well as active interactions with materials—and perhaps even language—presumably influence perceptual representations of materials, their properties and classes. However, there is no way for observers to acquire ground truth category labels for most physical properties, so it would be fascinating to test how much and what form of training data is required to learn a latent space resembling human perception.

Materials and Methods

Participants

A total of 6,334 workers from the online crowdsourcing platform Amazon Mechanical Turk participated in the triplet 2-AFC tasks, for the creation of the fully sampled matrix of 60 materials (1,296 workers, 1,238 after exclusion; 638 female, 596 male, 4 other; age was not assessed) and for training and evaluating the computational model (5,038 workers, 4,865 after exclusion; 2,798 female, 2,045 male, 22 other; age was not assessed). Workers were excluded if they exhibited overly fast responses (for participants with at least 60 trials, <600ms response time in >25% of trials or <900ms response time in >50% of trials) or overly deterministic responses (for participants with at least 160 trials, >60% of responses in one of the 2-AFC positions; expected value, 50%). This removed 18,860 trials (9.19% of all 205,320 trials) and 61,600 trials (3.3% of all 1,870,700 trials), respectively. All workers were located in the United States. All workers provided informed consent and were compensated financially for their time (~6.65 USD/hour based on the median response time). In addition, 20 native speaking English participants (12 female, 8 male; mean age = 34.35, standard deviation = 9.22, range = 25–59) took part in the dimension labelling experiment without compensation. All participants provided informed consent. Experimental protocols were approved by the local ethics committee of the Department of Psychology and Sports Sciences of the Justus-Liebig University Giessen (LEK-FB06; application number: 2017-0046) as well as the Ethics Committee of the Medical Faculty of Leipzig (application number: 054/20-ek) and adhered to the declaration of Helsinki.

STUFF dataset

For the selection and identification of material classes in the STUFF dataset, we followed a similar procedure as outlined in [25] for object concepts. Note that the final list is not intended to be a complete and definite set of all picturable material concepts in the English language (see discussion in [25]). However, the selection procedure yielded a highly systematic and extensive set of picturable material concepts.

Specifically, we based our selection procedure on a list of ~40,000 American English words and two-word expressions, choosing all of these that were tagged as nouns (using part-of-speech tags extended by using the British Lexicon Project; [50, 51]) and achieved a minimum concreteness rating of 4 (level at which the word could be experienced through one of the five basic senses from 1: abstract, to 5: concrete) [52]. Next, the resulting 8,671 nouns were screened by 2 authors and 2 student research assistants for whether they reflected materials (i.e., the stuff objects are made from, e.g., “granite”), with rather liberal inclusion criteria (e.g., also including material composites such as “toothpaste”), followed by a number of exclusion criteria (see **Supplementary Materials and Methods**), leaving us with a list of 200 concrete nouns referring to materials (see **Supplementary Table T1** for a list of all 200 material concepts, together with WordNet keys and definitions). Finally, for each material concept we collected three high-quality close-up naturalistic photographs from the web, showing the material in its typical aggregate state (i.e., at room temperature, e.g., solid iron and liquid oil) and form (e.g., “salt” grains and “concrete” walls and floors) without prominently featuring objects. The resulting 600 images of materials were used to collect the similarity judgments (triplet 2-AFC tasks).

Similarity judgement (triplet 2-AFC) task

To obtain similarities between images under different contexts, we employed an online triplet 2-AFC task that was carried out in sets of 20 trials. All workers were free to choose how many sets they would like to complete. In each trial, we presented three material images in a triangular arrangement on the screen. Participants were told that the image on top was the reference image and that among the two images at the bottom they should choose which one is more similar to that reference. Participants responded with a mouse click, and the next trial started after an intertrial interval of 500 ms. The instructions stated that all images would be showing a type of material or “stuff” – and if an image would show something they would not call a material, they should base their judgement on their best guess of what the "stuff" in the image could be. Material triplets and order of presentation were random but chosen in way that each cell in the 600 × 600 similarity matrix was chosen at least once.

The 60 material images for the creation of the fully sampled matrix were chosen pseudorandomly, in a way that the probability of choosing images from the same material class was as close as possible to the true probability (resulting in 48 unique classes and 6 classes that were chosen twice).

Dimension labelling experiment

To identify the extent to which the retrieved dimensions were interpretable by human participants, we sent them a pictured survey and asked them to provide labels for 36 so-called “rating scales of material properties”. These rating scales for each dimension were created by binning material images according to their dimension values, with 6 bins for values > 0.3, and a separate bin for all values < 0.3 [24]. Each bin contained a maximum of 10 images, with fewer images for very sparse dimensions. A “rating scale” to the left of the image bins marked three positions along the scale as “high”, “low” and “not at all”. Participants were asked to come up with verbal labels for these scales, whereby labels can be descriptive words but also categories. They were also specifically instructed to try to take the full range of materials into account. They should provide at least one and up to as many verbal labels for each scale as they considered reasonable.

The results were corrected for typos and spelling, and we added the term “colour” to unambiguous colour terms (e.g., blue = blue colour) and replaced “saturation” with “colour” (e.g., red saturation = red colour). Then, we removed redundant words (such as “and”, “or”, “material”) and shortened descriptions, replacing the following word endings: “-like” (e.g., jewel-like = jewel), “-ility” (e.g., flexibility = flexible), “-idity” (e.g., fluidity = fluid), and “-en” (e.g., wooden = wood). Finally, we made all terms singular (e.g., minerals = mineral) and removed synonyms (according to WordNet [53], e.g., aqua colour = turquoise colour, cloth = fabric).

Details of computational modelling

To derive core dimensions underlying material similarity judgments, we followed a recently developed modelling described in more detail in [23, 24]. The key concept is a representational embedding, in which material images are characterized as numerical vectors, with each value reflecting a different latent dimension relevant for capturing material similarity judgments. The embedding is initialized with 90 random dimensions, and the model is trained to predict human responses on 90% of triplet responses and tested on the remaining 10%, under the constraints of sparse, continuous and positive dimensions.

The model was implemented in PyTorch 1.6 (https://github.com/ViCCo-Group/SPoSE). Each triplet was encoded using three one-hot vectors (length, 200), and each vector was linked to 90 latent dimensions, but with weights replicated across all three vectors. The 200 × 90 weights were initialized randomly (range, 0–1). The dot product was chosen as a basis for proximity for computational reasons, but previous results showed similar performance when using the Euclidean distance [24]. The loss function of the model optimization consisted of the cross-entropy, which is the logarithm of the softmax function, and a regularization term based on the L1 norm:

$$\sum^{n} log\left(\frac{\exp(x_i x_j)}{\exp(x_i x_j) + \exp(x_i x_k)}\right) + \lambda \sum^{m} \|x\|_1$$

where x corresponds to an object vector; *i*, *j* and *k* to the indices of the current triplet; *n* to the number of triplets; and *m* to the number of material images. The regularization parameter λ, which controls the trade-off between sparsity and model performance, was determined using cross-validation on the training set ($\lambda = 0.0038$). The sparsity constraint results in a penalty for the number of dimensions, so that the resulting model will have fewer dimensions than the 90 initialized dimensions.

The weights in the embedding X were enforced to be positive to support interpretability of dimensions. The minimization of the loss was carried out using stochastic gradient descent as implemented in the Adam algorithm [54] using default parameters and a minibatch size of 100 triplets. After the optimization was complete, only dimensions were kept for which at least one material image had a weight larger than 0.1, leaving us with 36 dimensions. The dimensions were sorted in descending order by the sum of their weights across materials.

Reconstructing the full similarity matrix from the computational model

We defined material similarity in the triplet 2-AFC task as the probability p(i,j) of the participants choosing material image i and j to belong together, irrespective of context imposed by image k (the third image). Therefore, to compute similarity from the learned embedding for all 600 material images, we obtained the predicted choices from the model for all possible ~107.5 million triplets and then calculated the average choice probability for each pair of material images. The same procedure was used to obtain the fully sampled similarity matrix of 60 material images; after obtaining the predicted model choice for all possible 102,660 triplets, we again calculated the average choice probability for each pair of material images.

Stability of modelling dimensions

Each time the computational model is trained, the stochasticity of the optimization algorithm might produce a different set of dimensions. To test the stability of the dimensions in our 36-dimensional model, we trained the model 50 times with different random initializations. Then, we correlated each of the 36 original dimensions with all dimensions of one of the 50 reference models and chose the best-fitting dimension across all correlations as the closest match. We averaged the correlations after Fisher z-transformation and then inverted the transformation to get an average reliability for each dimension across all 50 models (**Supplementary Fig. 3**).

Supplementary Materials and Methods

STUFF dataset criteria

The following procedures were implemented in an iterative procedure where four observers (two naïve and authors F.S. and A.S.) decided on inclusion/exclusion of each word by majority vote.

When selecting materials, we excluded all nouns referring to objects (e.g., “coin”, “bagpipes”), animals and fictitious creatures (e.g., “cat”, “dragon”), people (e.g., “pilot”, “father”), navigable places (e.g. “garden”), artwork or crafts (e.g., “collage”), as well as action nouns (e.g. “smack”), times of day (“night”), units and geometric figures (e.g. “quart,” “hexagon”), non-visual but sensory nouns (e.g. “click,” “music”) and nouns that were deemed too difficult to visualize (e.g. “equipment”).

Then, we used a number of exclusion criteria to further condense the number of nouns referring to materials: (1) plural form when singular form with the identical meaning is found in the list (e.g. exclude “ashes” when “ash” is present), (2) synonyms (e.g., exclude “chinaware” when “porcelain” is present), (3) invisible materials (e.g., “butane”, “gasoline”), with the exception of water, (4) bodily fluids (e.g. “urine”), (5) reference to multiple materials (e.g., “canvas”), (6) drugs and alcohol (e.g., “marijuana”, “Irish whiskey”), (7) very rare materials (“plutonium” and “radium”), and (8) food or beverages (e.g., “lemon”, “lemonade”). Note that in contrast to [25], we did not choose materials based on whether they were named consistently by participants. Materials are more ambiguous in their appearance compared to objects (e.g., think of “elephant” and “iron”): objects come in a more limited number of typical shapes (and often of typical materials), while materials come in a plethora of shapes. For our purposes, that is, identifying dimensions underlying material similarity judgments, we considered it less important whether participants could name the materials.

Method for automatically generating semantic feature norm scores

In a first step, we generated lists of binary semantic features for all 200 materials with the large language model GPT-3 [55], closely following an approach described in previous work for objects that was shown to rival results generated by humans [56]. This yielded a list of 11,400 binary semantic features with an occurrence probability for each material. Next, we took this list of semantic features and scaled the probabilities using the well-known term-frequency inverse document frequency (tf-idf) [57]. For later comparability between dimensions, we also scaled them to a sum of 1. To identify important features for each dimension, for a given dimension, we next multiplied the scaled occurrence probability for semantic features with the dimension vector and summed this product across all material classes, yielding a score for how important a given feature is for a given dimension across all materials. Since some features are generally more common than others, we finally took the difference between the feature scores of a given dimension and the mean of all other dimensions. The resulting feature dimension scores are plotted in **Supplementary Table T5**.

Supplementary Table 1. List of all 200 material concepts, WordNet synset, and definitions. Semantic similarity was calculated by correlating 300-dimensional sense embedding vectors for each of these nouns [28], and, when not available (three classes: play dough, teflon, vaseline), we used the lexical semantic Wu-Palmer similarity measure based on depth of nodes in the WordNet taxonomies [53, 58].

Material class	WordNet Synset	WordNet Definition
algae	algae.n.1	primitive chlorophyll-containing mainly aquatic eukaryotic organisms lacking true stems and roots and leaves
aluminium	aluminium.n.1	a silvery ductile metallic element found primarily in bauxite
amber	amber.n.2	a hard yellowish to brownish translucent fossil resin; used for jewelry
amethyst	amethyst.n.1	a transparent purple variety of quartz; used as a gemstone
arsenic	arsenic.n.2	a very poisonous metallic element that has three allotropic forms; arsenic and arsenic compounds are used as herbicides and insecticides and various alloys; found in arsenopyrite and orpiment and realgar
asbestos	asbestos.n.1	fibrous amphibole; used for making fireproof articles; inhaling fibers can cause asbestosis or lung cancer
ash	ash.n.1	the residue that remains when something is burned
asphalt	asphalt.n.1	mixed asphalt and crushed gravel or sand; used especially for paving but also for roofing
balsawood	balsa.n.1	strong lightweight wood of the balsa tree used especially for floats
bamboo	bamboo.n.1	the hard woody stems of bamboo plants; used in construction and crafts and fishing poles
bark	bark.n.1	tough protective covering of the woody stems and roots of trees and other woody plants
beeswax	beeswax.n.1	a yellow to brown wax secreted by honeybees to build honeycombs

bone	bone.n.1	rigid connective tissue that makes up the skeleton of vertebrates
borax	borax.n.1	an ore of boron consisting of hydrated sodium borate; used as a flux or cleansing agent
brass	brass.n.1	an alloy of copper and zinc
brick	brick.n.1	rectangular block of clay baked by the sun or in a kiln; used as a building or paving material
bronze	bronze.n.1	an alloy of copper and tin and sometimes other elements; also any copper-base alloy containing other elements in place of tin
brownstone	brownstone.n.1	a reddish brown sandstone; used in buildings
bubble wrap	bubble_pack.n.1	packaging in which a product is sealed between a cardboard backing and clear plastic cover
cadmium	cadmium.n.1	a soft bluish-white ductile malleable toxic bivalent metallic element; occurs in association with zinc ores)
calcium	calcium.n.1	a white metallic element that burns with a brilliant light; the fifth most abundant element in the earth's crust; an important component of most plants and animals
carbon	carbon.n.1	an abundant nonmetallic tetravalent element occurring in three allotropic forms: amorphous carbon and graphite and diamond; occurs in all organic compounds
carbon paper	carbon_paper.n.1	a thin paper coated on one side with a dark waxy substance (often containing carbon); used to transfer characters from the original to an under sheet of paper
cashmere	cashmere.n.1	a soft fabric made from the wool of the Cashmere goat
cellophane	cellophane.n.1	a transparent paperlike product that is impervious to moisture and used to wrap candy or cigarettes etc.
cement	cement.n.2	a building material that is a powder made of a mixture of calcined limestone and clay; used with water and sand or gravel to make concrete and mortar
chainmail	chain_mail.n.1	flexible armor made of interlinked metal rings
chalk	chalk.n.1	a soft whitish calcite
chalkstone	chalkstone.n.1	a deposit of urates around a joint or in the external ear; diagnostic of advanced or chronic gout
charcoal	charcoal.n.1	a carbonaceous material obtained by heating wood or other organic matter in the absence of air
cheesecloth	cheesecloth.n.1	a coarse loosely woven cotton gauze; originally used to wrap cheeses
chiffon	chiffon.n.1	a sheer fabric of silk or rayon
chlorine	chlorine.n.1	a common nonmetallic element belonging to the halogens; best known as a heavy yellow irritating toxic gas; used to purify water and as a bleaching agent and disinfectant; occurs naturally only as a salt (as in sea water)
chrome	chrome.n.1	another word for chromium when it is used in dyes or pigments (chromium: a hard brittle multivalent metallic element; resistant to corrosion and tarnishing)
cinder	cinder.n.1	a fragment of incombustible matter left after a wood or coal or charcoal fire
clay	clay.n.1	a very fine-grained soil that is plastic when moist but hard when fired
coal	coal.n.1	fossil fuel consisting of carbonized vegetable matter deposited in the Carboniferous period

cobalt	cobalt.n.1	a hard ferromagnetic silver-white bivalent or trivalent metallic element; a trace element in plant and animal nutrition
cobbleston e	cobblestone.n.1	rectangular paving stone with curved top; once used to make roads
concrete	concrete.n.1	a strong hard building material composed of sand and gravel and cement and water
copper	copper.n.1	a ductile malleable reddish-brown corrosion-resistant diamagnetic metallic element; occurs in various minerals but is the only metal that occurs abundantly in large masses; used as an electrical and thermal conductor
coral	coral.n.2	the hard stony skeleton of a Mediterranean coral that has a delicate red or pink color and is used for jewelry
cord	cord.n.4	a cut pile fabric with vertical ribs; usually made of cotton
cork	cork.n.1	outer bark of the cork oak; used for stoppers for bottles etc.
cotton cloth	cotton.n.2	fabric woven from cotton fibers
cotton wool	cotton_wool.n.1	soft silky fibers from cotton plants in their raw state
crepe paper	crepe_paper.n.1	paper with a crinkled texture; usually colored and used for decorations
denim	denim.n.2	a coarse durable twill-weave cotton fabric
diamond	diamond.n.2	very hard native crystalline carbon valued as a gem
ebony	ebony.n.2	hard dark-colored heartwood of the ebony tree; used in cabinetwork and for piano keys
eggshell	eggshell.n.1	the exterior covering of a bird's egg
ember	ember.n.1	a hot fragment of wood or coal that is left from a fire and is glowing or smoldering
emerald	emerald.n.1	a green transparent form of beryl; highly valued as a gemstone
fat	fat.n.1	a soft greasy substance occurring in organic tissue and consisting of a mixture of lipids (mostly triglycerides)
feather	feather.n.1	the light horny waterproof structure forming the external covering of birds
fiberboard	fiberboard.n.1	wallboard composed of wood chips or shavings bonded together with resin and compressed into rigid sheets
fiberglass	fiberglass.n.1	a covering material made of glass fibers in resins
flame	flame.n.1	the process of combustion of inflammable materials producing heat and light and (often) smoke
flannel	flannel.n.1	a soft light woolen fabric; used for clothing
fleece	fleece.n.3	a soft bulky fabric with deep pile; used chiefly for clothing
flint	flint.n.1	a hard kind of stone; a form of silica more opaque than chalcedony
fluorine	fluorine.n.1	a nonmetallic univalent element belonging to the halogens; usually a yellow irritating toxic flammable gas; a powerful oxidizing agent; recovered from fluorite or cryolite or fluorapatite
foam	foam.n.2	a lightweight material in cellular form; made by introducing gas bubbles during manufacture
fog	fog.n.1	droplets of water vapor suspended in the air near the ground
foliage	foliage.n.1	the main organ of photosynthesis and transpiration in higher plants
frankincens e	frankincense.n.1	an aromatic gum resin obtained from various Arabian or East African trees; formerly valued for worship and for embalming and fumigation
froth	froth.n.1	(a mass of small bubbles formed in or on a liquid)

fruitwood	fruitwood.n.1	wood of various fruit trees (as apple or cherry or pear) used especially in cabinetwork
fur	fur.n.1	the dressed hairy coat of a mammal
garnet	garnet.n.1	any of a group of hard glassy minerals (silicates of various metals) used as gemstones and as an abrasive
gauze	gauze.n.2	a net of transparent fabric with a loose open weave
gelatin	gelatin.n.1	a colorless water-soluble glutinous protein obtained from animal tissues such as bone and skin
glass	glass.n.1	a brittle transparent solid with irregular atomic structure
glue	glue.n.1	cement consisting of a sticky substance that is used as an adhesive
gold	gold.n.3	a soft yellow malleable ductile (trivalent and univalent) metallic element; occurs mainly as nuggets in rocks and alluvial deposits; does not react with most chemicals but is attacked by chlorine and aqua regia
granite	granite.n.1	plutonic igneous rock having visibly crystalline texture; generally composed of feldspar and mica and quartz
graphite	graphite.n.1	used as a lubricant and as a moderator in nuclear reactors
grass	grass.n.1	narrow-leaved green herbage: grown as lawns; used as pasture for grazing animals; cut and dried as hay
gunpowder	gunpowder.n.1	a mixture of potassium nitrate, charcoal, and sulfur in a 75:15:10 ratio which is used in gunnery, time fuses, and fireworks
hair	hair.n.1	a covering for the body (or parts of it) consisting of a dense growth of threadlike structures (as on the human head); helps to prevent heat loss)
hay	hay.n.1	grass mowed and cured for use as fodder
horn	horn.n.7	the material (mostly keratin) that covers the horns of ungulates and forms hooves and claws and nails
ice	ice.n.1	water frozen in the solid state
ink	ink.n.1	a liquid used for printing or writing or drawing
iron	iron.n.1	a heavy ductile magnetic metallic element; is silver-white in pure form but readily rusts; used in construction and tools and armament; plays a role in the transport of oxygen by the blood
ivory	ivory.n.1	a hard smooth ivory colored dentine that makes up most of the tusks of elephants and walruses
jade	jade.n.1	a semiprecious gemstone that takes a high polish; is usually green but sometimes whitish; consists of jadeite or nephrite
jelly	jelly.n.3	any substance having the consistency of jelly or gelatin
kevlar	polymer.n.1	a naturally occurring or synthetic compound consisting of large molecules made up of a linked series of repeated simple monomers
latex	latex.n.1	a milky exudate from certain plants that coagulates on exposure to air
lava	lava.n.1	rock that in its molten form (as magma) issues from volcanos; lava is what magma is called when it reaches the surface
lead	lead.n.2	a soft heavy toxic malleable metallic element; bluish white when freshly cut but tarnishes readily to dull grey
leather	leather.n.1	an animal skin made smooth and flexible by removing the hair and then tanning

limestone	limestone.n.1	a sedimentary rock consisting mainly of calcium that was deposited by the remains of marine animals
linen	linen.n.1	a fabric woven with fibers from the flax plant
linoleum	linoleum.n.1	a floor covering
magnesium	magnesium.n.1	a light silver-white ductile bivalent metallic element; in pure form it burns with brilliant white flame; occurs naturally only in combination (as in magnesite and dolomite and carnallite and spinel and olivine)
mahogany	mahogany.n.1	wood of any of various mahogany trees; much used for cabinetwork and furniture
marble	marble.n.1	a hard crystalline metamorphic rock that takes a high polish; used for sculpture and as building material
marblewood	marblewood.n.1	hard marbled wood
microfiber	polyester.n.3	any of a large class of synthetic fabrics
moonstone	moonstone.n.1	a transparent or translucent gemstone with a pearly luster; some specimens are orthoclase feldspar and others are plagioclase feldspar
mortar	mortar.n.2	used as a bond in masonry or for covering a wall
moss	moss.n.1	tiny leafy-stemmed flowerless plants
mud	mud.n.1	water soaked soil; soft wet earth
nickel	nickel.n.1	a hard malleable ductile silvery metallic element that is resistant to corrosion; used in alloys; occurs in pentlandite and smaltite and garnierite and millerite
nylon	nylon.n.1	a thermoplastic polyamide; a family of strong resilient synthetic fibers
obsidian	obsidian.n.1	acid or granitic glass formed by the rapid cooling of lava without crystallization; usually dark, but transparent in thin pieces
oil	oil.n.1	a slippery or viscous liquid or liquefiable substance not miscible with water
oilcloth	oilcloth.n.1	cloth treated on one side with a drying oil or synthetic resin
oilpaper	oilpaper.n.1	paper that has been made translucent and waterproof by soaking in oil
oilskin	oilskin.n.1	a macintosh made from cotton fabric treated with oil and pigment to make it waterproof
onionskin	onionskin.n.1	a thin strong lightweight translucent paper used especially for making carbon copies
onyx	onyx.n.1	a chalcedony with alternating black and white bands; used in making cameos
opal	opal.n.1	a translucent mineral consisting of hydrated silica of variable color; some varieties are used as gemstones
paint	paint.n.1	a substance used as a coating to protect or decorate a surface (especially a mixture of pigment suspended in a liquid); dries to form a hard coating
paper	paper.n.1	a material made of cellulose pulp derived mainly from wood or rags or certain grasses
paperboard	paperboard.n.1	a cardboard suitable for making posters
papyrus	papyrus.n.1	paper made from the papyrus plant by cutting it in strips and pressing it flat; used by ancient Egyptians and Greeks and Romans
parchment	parchment.n.1	a superior paper resembling sheepskin; skin of a sheep or goat prepared for writing on

pearl	pearl.n.1	a smooth lustrous round structure inside the shell of a clam or oyster; much valued as a jewel
petroleum	petroleum.n.1	a dark oil consisting mainly of hydrocarbons
pewter	pewter.n.1	any of various alloys of tin with small amounts of other metals (especially lead)
phosphorus	phosphorus.n.1	a multivalent nonmetallic element of the nitrogen family that occurs commonly in inorganic phosphate rocks and as organic phosphates in all living cells; is highly reactive and occurs in several allotropic forms
pinewood	pinewood, pine.n.2	straight-grained durable and often resinous white to yellowish timber of any of numerous trees of the genus Pinus
plaster	plaster.n.1	a mixture of lime or gypsum with sand and water; hardens into a smooth solid; used to cover walls and ceilings
plasterboard	plasterwork.n.1	a surface of hardened plaster (as on a wall or ceiling)
plastic	plastic.n.1	generic name for certain synthetic or semisynthetic materials that can be molded or extruded into objects or films or filaments or used for making e.g. coatings and adhesives
platinum	platinum.n.1	a heavy precious metallic element; grey-white and resistant to corroding; occurs in some nickel and copper ores and is also found native in some deposits
play dough	plasticine.n.1	a synthetic material resembling clay but remaining soft; used as a substitute for clay or wax in modeling (especially in schools)
plywood	plywood.n.1	a laminate made of thin layers of wood
polystyrene	polystyrene.n.1	a polymer of styrene; a rigid transparent thermoplastic
porcelain	porcelain.n.1	ceramic ware made of a more or less translucent ceramic
pottery	pottery.n.1	ceramic ware made from clay and baked in a kiln
quartz	quartz.n.2	a hard glossy mineral consisting of silicon dioxide in crystal form; present in most rocks (especially sandstone and granite); yellow sand is quartz with iron oxide impurities
quartzite	quartzite.n.1	hard metamorphic rock consisting essentially of interlocking quartz crystals
rayon	rayon.n.1	a synthetic silklike fabric
redwood	redwood.n.1	the soft reddish wood of either of two species of sequoia trees
resin	resin.n.1	any of a class of solid or semisolid viscous substances obtained either as exudations from certain plants or prepared by polymerization of simple molecules
rhinestone	rhinestone.n.1	an imitation diamond made from rock crystal or glass or paste
ricepaper	rice_paper.n.1	a thin delicate material resembling paper; made from the rice-paper tree
rosewood	rosewood.n.1	hard dark reddish wood of a rosewood tree having a strongly marked grain; used in cabinetwork
rubber	rubber.n.1	an elastic material obtained from the latex sap of trees (especially trees of the genera Hevea and Ficus) that can be vulcanized and finished into a variety of products
ruby	ruby.n.1	a transparent piece of ruby that has been cut and polished and is valued as a precious gem
salt	salt.n.2	white crystalline form of especially sodium chloride used to season and preserve food
sand	sand.n.1	a loose material consisting of grains of rock or coral

sandalwood	sandalwood.n.1	close-grained fragrant yellowish heartwood of the true sandalwood; has insect repelling properties and is used for carving and cabinetwork
sandpaper	sandpaper.n.1	stiff paper coated with powdered emery or sand
sandstone	sandstone.n.1	a sedimentary rock consisting of sand consolidated with some cement (clay or quartz etc.)
sapphire	sapphire.n.1	a precious transparent stone of rich blue corundum valued as a gemstone
satin	satin.n.1	a smooth fabric of silk or rayon; has a glossy face and a dull back
satinwood	satinwood.n.2	hard yellowish wood of a satinwood tree having a satiny luster; used for fine cabinetwork and tools
sequin	sequin.n.1	adornment consisting of a small piece of shiny material used to decorate clothing
shell	shell.n.3	hard outer covering or case of certain organisms such as arthropods and turtles
silicon	silicon.n.1	a tetravalent nonmetallic element; next to oxygen it is the most abundant element in the earth's crust; occurs in clay and feldspar and granite and quartz and sand; used as a semiconductor in transistors
silicone	silicone.n.1	any of a large class of siloxanes that are unusually stable over a wide range of temperatures; used in lubricants and adhesives and coatings and synthetic rubber and electrical insulation
silk	silk.n.1	a fabric made from the fine threads produced by certain insect larvae
silver	silver.n.1	a soft white precious univalent metallic element having the highest electrical and thermal conductivity of any metal; occurs in argentite and in free form; used in coins and jewelry and tableware and photography
skin	skin.n.1	a natural protective body covering and site of the sense of touch; body covering of a living animal
slate	slate.n.3	a fine-grained metamorphic rock that can be split into thin layers
smoke	smoke.n.1	a cloud of fine particles suspended in a gas
snow	snow.n.1	precipitation falling from clouds in the form of ice crystals
soap	soap.n.1	a cleansing agent made from the salts of vegetable or animal fats
soil	soil.n.3	material in the top layer of the surface of the earth in which plants can grow (especially with reference to its quality or use)
spandex	spandex.n.1	an elastic synthetic fabric
sponge	sponge.n.1	a porous mass of interlacing fibers that forms the internal skeleton of various marine animals and usable to absorb water or any porous rubber or cellulose product similarly used
steel	steel.n.1	an alloy of iron with small amounts of carbon; widely used in construction; mechanical properties can be varied over a wide range
straw	straw.n.2	material consisting of seed coverings and small pieces of stem or leaves that have been separated from the seeds
suede	suede.n.1	leather with a napped surface

sugar	sugar.n.1	a white crystalline carbohydrate used as a sweetener and preservative
sulfur	sulfur.n.1	an abundant tasteless odorless multivalent nonmetallic element; best known in yellow crystals; occurs in many sulphide and sulphate minerals and even in native form (especially in volcanic regions)
talc	talc.n.1	a fine grained mineral having a soft soapy feel and consisting of hydrated magnesium silicate; used in a variety of products including talcum powder
talcum powder	talcum_powder.n.1	a toilet powder made of purified talc and usually scented; absorbs excess moisture
tar	tar.n.1	any of various dark heavy viscid substances obtained as a residue
tarpaper	tar_paper.n.1	a heavy paper impregnated with tar and used as part of a roof for waterproofing
teakwood	teakwood.n.1	hard strong durable yellowish-brown wood of teak trees; resistant to insects and to warping; used for furniture and in shipbuilding
teflon	teflon.n.1	a material used to coat cooking utensils and in industrial applications where sticking is to be avoided
terracotta	terra_cotta.n.1	a hard unglazed brownish-red earthenware
tin	tin.n.1	a silvery malleable metallic element that resists corrosion; used in many alloys and to coat other metals to prevent corrosion; obtained chiefly from cassiterite where it occurs as tin oxide
tinfoil	tinfoil.n.1	foil made of tin or an alloy of tin and lead
tinsel	tinsel.n.2	a thread with glittering metal foil attached
titanium	titanium.n.1	a light strong grey lustrous corrosion-resistant metallic element used in strong lightweight alloys (as for airplane parts); the main sources are rutile and ilmenite
toilet paper	toilet_paper.n.1	a soft thin absorbent paper for use in toilets
tooth	tooth.n.1	hard bonelike structures in the jaws of vertebrates; used for biting and chewing or for attack and defense
toothpaste	toothpaste.n.1	a dentifrice in the form of a paste
topaz	topaz.n.2	a mineral (fluosilicate of aluminum) that occurs in crystals of various colors and is used as a gemstone
tungsten	tungsten.n.1	a heavy grey-white metallic element; the pure form is used mainly in electrical applications; it is found in several ores including wolframite and scheelite
tweed	tweed.n.1	thick woolen fabric used for clothing; originated in Scotland
uranium	uranium.n.1	a heavy toxic silvery-white radioactive metallic element; occurs in many isotopes; used for nuclear fuels and nuclear weapons
vapor	vapor.n.1	a visible suspension in the air of particles of some substance
vaseline	vaseline.n.1	a trademarked brand of petroleum jelly
velcro	velcro.n.1	nylon fabric used as a fastening
velvet	velvet.n.1	a silky densely piled fabric with a plain back
vinyl	vinyl.n.2	shiny and tough and flexible plastic; used especially for floor coverings
water	water.n.1	binary compound that occurs at room temperature as a clear colorless odorless tasteless liquid; freezes into ice below 0 degrees centigrade and boils above 100 degrees centigrade; widely used as a solvent

wax	wax.n.1	any of various substances of either mineral origin or plant or animal origin; they are solid at normal temperatures and insoluble in water
wax paper	wax_paper.n.1	paper that has been waterproofed by treatment with wax or paraffin
wool	wool.n.1	a fabric made from the hair of sheep
zinc	zinc.n.1	a bluish-white lustrous metallic element; brittle at ordinary temperatures but malleable when heated; used in a wide variety of alloys and in galvanizing iron; it occurs naturally as zinc sulphide in zinc blende
zirconium	zirconium.n.1	a lustrous grey strong metallic element resembling titanium; it is used in nuclear reactors as a neutron absorber; it occurs in baddeleyite but is obtained chiefly from zircon

Supplementary Table 2. List of copyright information for all material images reprinted in this paper.

material class	Creator	Platform	Licence
aluminium	myfreetextures.com	myfreetextures.com	MyFreeTextures License
arsenic	Parent Géry	wikimedia.org	CC0 1.0
ash	Laurentius	wikimedia.org	CC BY-SA 3.0
asphalt	Kunzinger	flickr.com	CC BY 2.0
balsawood	mauro halpern	flickr.com	CC BY 2.0
balsawood	Joseph Francis	flickr.com	CC BY 2.0
beeswax	Maja Dumat	flickr.com	CC BY 2.0
bone	Maky Orel	wikimedia.org	CC0 1.0
borax	Rock Currier	wikimedia.org	CC BY 3.0
borax	Ra'ike	wikimedia.org	CC BY-SA 3.0
brownstone	Richter.V	flickr.com	CC BY-NC-SA 2.0
cellophane	nsmithtnz	flickr.com	CC BY-NC 2.0
cellophane	Itzuvit	wikimedia.org	CC BY-SA 3.0
cement	Marcu Ioachim	flickr.com	CC0 1.0
cement	Marcu Ioachim	flickr.com	CC0 1.0
chalkstone	su-lin	flickr.com	CC BY-NC-ND 2.0
chrome	Alchemist-hp	wikimedia.org	CC BY-NC-ND 3.0
clay	pixabay	pexels.com	CC0
cobalt	James St. John	wikimedia.org	CC BY 2.0
cobalt	Alchemist-hp	wikimedia.org	CC BY-NC-ND 3.0
concrete	Philippe Ramakers	unsplash.com	Unsplash Licence, https://unsplash.com/license
copper	Jonathan Zander	wikimedia.org	CC BY-SA 2.5
copper	Joanna Kosinska	unsplash.com	Unsplash Licence, https://unsplash.com/license
cork	Grapebowl	wikimedia.org	CC BY-SA 2.5
cotton wool	Ross Elliott	flickr.com	CC BY 2.0
crepe paper	under the same moon	flickr.com	CC BY 2.0
denim	Creativity103	flickr.com	CC BY 2.0
denim	Uwe_Jelting	pixabay.com	Pixabay License; https://pixabay.com/en/service/license/
ebony	Makassar-Ebenholz	wikimedia.org	CC BY-SA 2.5
fat	Nikol Lohr	flickr.com	CC BY-NC-SA 2.0

fiberglass	Cjp24	wikimedia.org	CC BY-SA 3.0
flannel	Sean Ganann	flickr.com	CC BY-NC-SA 2.0
foam	Windell Oskay	flickr.com	CC BY-SA 2.0
foam	Дар Ветер	wikimedia.org	CC BY-SA 3.0
foliage	Anne Worner	flickr.com	CC BY-NC-SA 2.0
foliage	Timothy Clough	flickr.com	CC BY-NC-SA 2.0
frankincense	Lupus in Saxonia	wikimedia.org	CC BY-SA 4.0
froth	Stefan Knaake	flickr.com	CC BY-NC 2.0
fur	Sheila in Moonducks	flickr.com	CC BY 2.0
garnet	Rachael Moore	flickr.com	CC BY-NC 2.0
glass	makamuki0	pixabay.com	Pixabay License, https://pixabay.com/en/service/license/
graphite	James St. John	flickr.com	CC BY 2.0
graphite	Ra'ike	wikimedia.org	CC BY-SA 3.0
horn	Zachi Evenor	wikimedia.org	CC BY-SA 4.0
iron	michael schaffler	unsplash.com	Unsplash Licence, https://unsplash.com/license
jade	Ra'ike	wikimedia.org	CC BY 3.0
leather	Cinnamon Cooper	flickr.com	CC BY-NC-SA 2.0
limestone	Manishwiki15	wikimedia.org	CC BY-SA 3.0
linen	Sincerely Media	unsplash.com	Unsplash Licence, https://unsplash.com/license
magnesium	Mark Fergus	wikimedia.org	CC BY 3.0
mahogany	Das Ohr	wikimedia.org	CC BY-SA 3.0
marble	shaireproductions.com	flickr.com	CC BY 2.0
marble	Olga V	pexels.com	Public domain
microfiber	Chris Yarzab	flickr.com	CC BY-NC-SA 2.0
moonstone	Robert M. Lavinsky	wikimedia.org	CC BY-SA 3.0
moonstone	Jarno	wikimedia.org	CC BY 2.0
moonstone	Géry PARENT	wikimedia.org	CC0 1.0
obsidian	Ji-Elle	wikimedia.org	CC BY-SA 3.0
oil	Martin Cooper	flickr.com	CC BY 2.0
opal	James St. John	wikimedia.org	CC-BY-2.0
paper	Niklas Bildhauer	wikimedia.org	CC BY-SA 2.0
paperboard	boris drenec	flickr.com	CC BY-NC-SA 2.0
phosphorus	Dnn87	wikimedia.org	CC BY 3.0
pinewood	decar66	flickr.com	CC BY-NC-SA 2.0
plaster	tropicalart77 (Tammy Dial Gray)	flickr.com	CC BY 2.0
plaster	Lara604	flickr.com	CC BY 2.0
plastic	freeimages.co.uk	freeimages.co.uk	Freeimage License
platinum	Alchemist-hp	wikimedia.org	CC BY-NC-ND 3.0
quartz	Andrew Gustar	flickr.com	CC BY-ND 2.0
resin	Darkone	wikimedia.org	CC BY-SA 2.5
resin	Emmanuel Boutet	wikimedia.org	CC BY-SA 3.0
rubber	bgfons.com	bgfons.com	CC BY-NC 4.0
salt	Gerwin Sturm	flickr.com	CC BY-SA 2.0
salt	Christian Mertes	wikimedia.org	CC BY-SA 3.0
sandstone	Kasia Trapszo	flickr.com	CC BY-NC-SA 2.0
satin	Jwrusa & Kjoonlee	wikimedia.org	CC0 1.0

satinwood	Philiipp Zinger	wikimedia.org	CC BY-SA 4.0
silicon	Enricoros	wikimedia.org	CC0 1.0
silver	Roger Culos	wikimedia.org	CC BY-SA 4.0
smoke	pxhere.com	pxhere.com	CC0 1.0
smoke	Jens Johnsson	unsplash.com	Unsplash Licence, https://unsplash.com/license
soil	Elvis Ripley	flickr.com	CC BY-NC 2.0
soil	United Soybean Board	flickr.com	CC BY 2.0
sugar	Marco Verch Professional Photographer	flickr.com	CC BY 2.0
sulfur	Parent Géry	wikimedia.org	CC0 1.0
talc	Stephanie Clifford	flickr.com	CC BY 2.0
talc	John Krygier	wikimedia.org	CC0 1.0
tar	ariari	flickr.com	CC BY-NC-ND 2.0
tar	Linda, Fortuna future	flickr.com	CC BY-NC 2.0
teflon	Fructibus	wikimedia.org	CC0
tin	Schtone	wikimedia.org	CC0 1.0
tin	Attribution not necessary	pexels.com	Pexels License
titanium	Alchemist-hp	wikimedia.org	CC BY-NC-ND 3.0
toothpaste	SCEhardt	wikimedia.org	CC0 1.0
toothpaste	Photos8	wikimedia.org	CC BY 3.0
topaz	Photos8	wikimedia.org	CC BY-SA 4.0
tungsten	Alchemist-hp	wikimedia.org	CC BY-NC-ND 3.0
tweed	PKM	wikimedia.org	CC0 1.0
vaseline	Kiyok	wikimedia.org	CC BY-SA 3.0
velvet	Steve Miller	flickr.com	CC BY-NC-SA 2.0
velvet	Kristiana Pinne	unsplash.com	Unsplash Licence, https://unsplash.com/license
water	Artiom Vallat	unsplash.com	Unsplash Licence, https://unsplash.com/license
water	stux	pixabay.com	Pixabay License, https://pixabay.com/en/servic e/license/

Supplementary Table 3. List of semantic labels in the dimension labelling task with at least 20% agreement between participants.

dimension	Labels (frequencies)
1	mineral (11), hard (7), rough (6), stone (5)
2	wood (10)
3	metallic (10), metal (7)
4	fabric (14), soft (7), flexible (4),
5	white colour (12), soft (5), crumbly (4),
6	grainy (8), rough (4)
7	crystalline (8), shiny (6), crystal (5), translucent (4)
8	fibrous (6), straw (4)
9	cloudy (6)
10	small (8), round (6), quantity (5)
11	beige colour (5), tan colour (5), brown colour (4)
12	viscous (8), liquid (6), malleable (5)

13	black colour (16)
14	brick (10), blocks (4)
15	swirly (7), wavy (6), creased (5), folded (4)
16	maroon colour (6), brown colour (4)
17	lines (3), long (3), tubular (3), vertical (3)
18	mesh (5), repetitive (4)
19	yellow colour (12)
20	multi-coloured (9), colourful (8)
21	thin (8), layered (5), papery (5), paper (4)
22	bulbous (9), bumpy (5)
23	green colour (16)
24	blue colour (17)
25	bone (7), shell (7), brittle (5)
26	gemstone (8), round (6), smooth (6), polished (4)
27	sheet (10), plastic (6), thin (6)
28	hair (9), soft (6)
29	spongy (5), house-hold (4)
30	golden (7), shiny (7), gold (5), metallic (5)
31	red colour (17)
32	round (8), smooth (7), ceramic (4), curved (4)
33	cream colour (8), white colour (6), ivory (4)
34	bumpy (9), lumpy (4)
35	hot (12), fire (7), molten (5)
36	turquoise colour (12), teal colour (6), blue colour (4)

Supplementary Table 4. Three most frequent descriptive labels for all 200 material concepts obtained from GPT-3 feature norms.

Material class			
algae	is green	a plant	is slimy
aluminium	is shiny	a metal	is light
amber	is yellow	used for jewelry	a fossil
amethyst	is purple	a gemstone	is transparent
arsenic	is poisonous	a metal	is white
asbestos	a mineral	is white	found in rocks
ash	is grey	made of wood	has leaves
asphalt	is black	is sticky	used for roads
balsawood	is light	made of wood	is hard
bamboo	has leaves	a plant	used for making furniture
bark	is brown	is rough	is hard
beeswax	is yellow	made by bees	used for candles
bone	is hard	is white	has marrow
borax	is white	used for cleaning	a chemical
brass	is yellow	is shiny	a metal
brick	made of clay	used for building houses	is red
bronze	a metal	is shiny	used for statues
brownstone	used for buildings	a type of rock	made of stone
bubble wrap	made of plastic	has bubbles	used for shipping
cadmium	a metal	is toxic	used in batteries
calcium	a mineral	is white	found in milk
carbon	is black	found in coal	found in diamonds
carbon paper	used for writing	used for copying	is thin

cashmere	is soft	is expensive	is warm
cellophane	is thin	is clear	is transparent
cement	is hard	is strong	is grey
chainmail	made of metal	worn by knights	is heavy
chalk	is white	used for writing	used for writing on blackboards
chalkstone	is white	is soft	a rock
charcoal	is black	used for cooking	made from wood
cheesecloth	is white	made of cotton	used for straining
chiffon	made of silk	is transparent	is thin
chlorine	is poisonous	a gas	a chemical
chrome	is shiny	a metal	is hard
cinder	is black	a rock	is hot
clay	used for making pots	used for making pottery	used for pottery
coal	is black	used for heating	found in mines
cobalt	a metal	is blue	used in batteries
cobblestone	made of stone	is hard	used for paving roads
concrete	is strong	is hard	used for roads
copper	a metal	is shiny	conducts electricity
coral	is red	found in the ocean	is pink
cord	is strong	made of rubber	made of plastic
cork	comes from trees	floats	floats on water
cotton cloth	is white	is soft	made of cotton
cotton wool	is soft	is white	used for cleaning
crepe paper	is thin	used for wrapping gifts	made of paper
denim	is blue	a fabric	made of cotton
diamond	is hard	is expensive	is clear
ebony	is black	is hard	a wood
eggshell	is white	is hard	made of calcium
ember	is hot	is red	used for cooking
emerald	is green	is precious	a gemstone
fat	is white	used for cooking	used in cooking
feather	is light	is soft	has a quill
fiberboard	made of wood	is strong	has a smooth surface
fiberglass	is strong	used for insulation	used in boats
flame	is hot	burns	is red
flannel	is soft	made of cotton	is warm
fleece	is warm	is soft	used for clothing
flint	is hard	a rock	used to make fire
fluorine	a gas	is poisonous	used in toothpaste
foam	is white	has bubbles	is soft
fog	is white	is cold	is wet
foliage	is green	is on trees	has leaves
frankincense	comes from trees	used in religious ceremonies	used in churches
froth	is foamy	is white	has bubbles
fruitwood	is hard	used for carving	used for furniture
fur	is soft	used for clothing	is warm
garnet	is red	a gemstone	a stone
gauze	is white	is thin	used for bandages

gelatin	a food	is clear	a dessert
glass	is transparent	is brittle	is hard
glue	is sticky	is white	used for sticking things together
gold	is yellow	is shiny	a metal
granite	is hard	a rock	is grey
graphite	is soft	is black	a mineral
grass	is green	has seeds	a plant
gunpowder	is explosive	is black	used in fireworks
hair	made of keratin	is long	made of protein
hay	used for feeding animals	is brown	made of grass
horn	is hard	made of bone	made of metal
ice	is cold	is frozen	is frozen water
ink	is black	used for writing	used for printing
iron	is strong	a metal	is heavy
ivory	is white	is hard	comes from elephants
jade	is green	a gemstone	a stone
jelly	is sweet	made of sugar	a dessert
kevlar	is strong	used in bulletproof vests	used for protection
latex	is sticky	is white	used in balloons
lava	is hot	is red	flows
lead	is heavy	a metal	is soft
leather	used to make belts	used to make shoes	used to make wallets
limestone	a rock	used for building	found in caves
linen	is white	is soft	made of flax
linoleum	made of plastic	used for floors	has a pattern
magnesium	a metal	is shiny	is white
mahogany	is hard	used for furniture	a wood
marble	is white	is smooth	is shiny
marblewood	is hard	used for making furniture	used for furniture
microfiber	used for cleaning	made of plastic	is soft
moonstone	is white	a gemstone	is shiny
mortar	used for construction	made of cement	has a handle
moss	is green	grows on trees	a plant
mud	is sticky	is brown	is dirty
nickel	a metal	is shiny	is hard
nylon	is strong	a fabric	used for stockings
obsidian	is black	is shiny	is sharp
oil	used for cooking	is black	is thick
oilcloth	is waterproof	made of plastic	used for covering tables
oilpaper	used for wrapping food	is thin	made of paper
oilskin	is waterproof	used for clothing	has a hood
onionskin	is thin	is white	used for writing
onyx	is black	is shiny	a gemstone
opal	a gemstone	is precious	a stone
paint	used for painting	used for walls	has a smell
paper	is white	used for writing	is thin
paperboard	used for packaging	is white	is thin

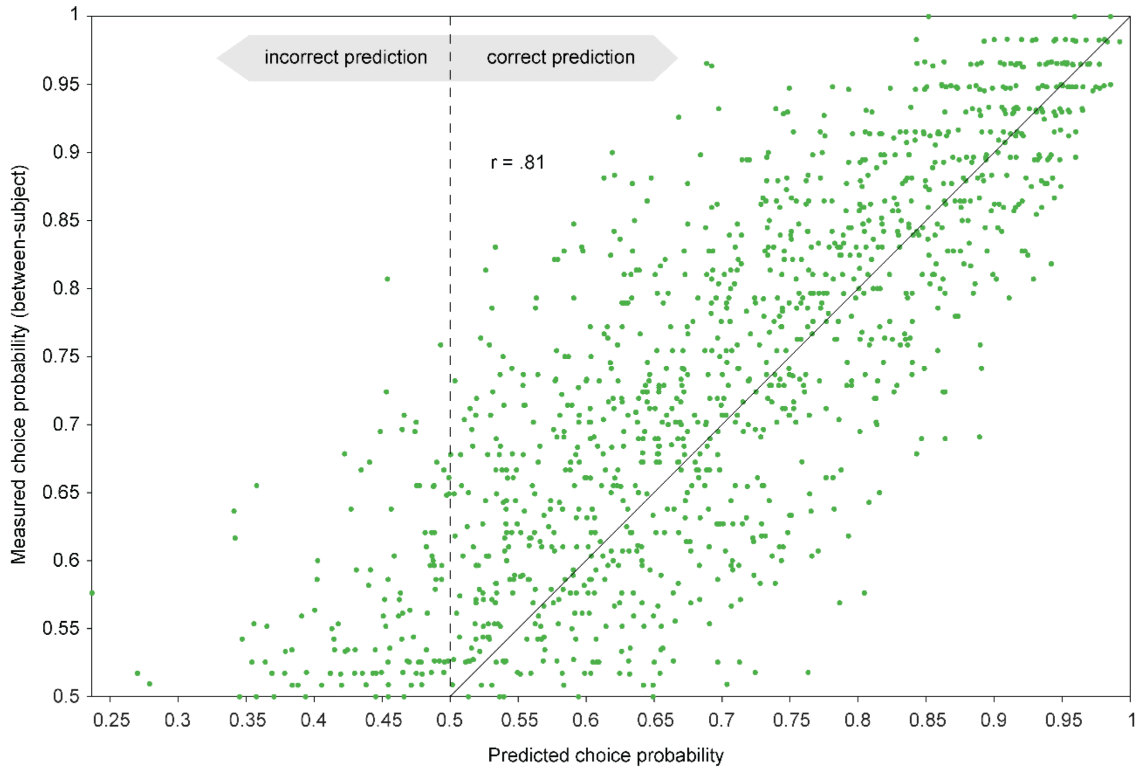
papyrus	used for writing	a plant	has leaves
parchment	used for writing	made of animal skin	is thin
pearl	is white	is shiny	is round
petroleum	a liquid	used for fuel	is black
pewter	is shiny	made of metal	a metal
phosphorus	is a solid	a chemical element	is white
pinewood	used for making furniture	used for making houses	used for building
plaster	is white	used to make casts	used for bandages
plasterboard	is white	a building material	has a smooth surface
plastic	made of oil	used for packaging	is hard
platinum	a metal	is shiny	is expensive
play dough	is soft	made of flour	used for play
plywood	made of wood	is strong	used for making furniture
polystyrene	is white	used for packaging	made of plastic
porcelain	is white	made of clay	is fragile
pottery	made of clay	has a handle	has a lid
quartz	is hard	is clear	a mineral
quartzite	is hard	a rock	used for building
rayon	made from cotton	made from wood	made from bamboo
redwood	a tree	has branches	is tall
resin	is sticky	comes from trees	is hard
rhinestone	is shiny	made of glass	used for decoration
rice paper	is thin	is white	made of rice
rosewood	is hard	used for furniture	used for making musical instruments
rubber	is elastic	is sticky	is black
ruby	is red	a stone	is precious
salt	is white	a mineral	used in cooking
sand	is white	found on beaches	found in deserts
sandalwood	a tree	used for making furniture	used in incense
sandpaper	is rough	made of paper	used for sanding
sandstone	has grains	made of sand	a rock
sapphire	is blue	a gemstone	a stone
satin	is shiny	is smooth	a fabric
satinwood	a wood	used for furniture	is hard
sequin	is shiny	used for decoration	made of metal
shell	has a shell	is small	used for protection
silicon	found in sand	found in rocks	used in computers
silicone	a liquid	is clear	used in cooking
silk	is shiny	is smooth	a fabric
silver	is shiny	a metal	is white
skin	covers the body	has hair	is thin
slate	used for writing on	a rock	is black
smoke	comes from fire	a gas	is white
snow	is white	is cold	falls from the sky
soap	is white	used for cleaning	used for washing
soil	used for growing plants	is brown	used for growing crops
spandex	is stretchy	used for making clothes	is elastic

sponge	used for cleaning	has holes	is soft
steel	is strong	is hard	is shiny
straw	used for making baskets	used for making hats	used for making brooms
suede	is soft	is brown	made of leather
sugar	is sweet	is white	has calories
sulfur	is yellow	found in rocks	found in volcanoes
talc	is soft	is white	a mineral
talcum powder	is white	used for babies	used on babies
tar	is sticky	is black	used to make roads
tarpaper	made of wood	has a rough surface	is thin
teakwood	used for furniture	is hard	used for flooring
teflon	is slippery	used for cooking	used for frying
terracotta	used for making pots	is brown	made of clay
tin	is shiny	a metal	made of metal
tinfoil	is shiny	is thin	is recycled
tinsel	is shiny	made of metal	used for decorating christmas trees
titanium	is strong	a metal	is light
toilet paper	is white	is soft	used for cleaning
tooth	is white	is hard	has a cavity
toothpaste	is white	used for cleaning teeth	has fluoride
topaz	is yellow	a gemstone	a mineral
tungsten	a metal	used in light bulbs	is hard
tweed	a fabric	is brown	made of wool
uranium	is radioactive	a metal	found in the ground
vapor	is invisible	a gas	made of water
vaseline	is white	used as a lubricant	used for lips
velcro	is sticky	used for shoes	used for clothing
velvet	is soft	a fabric	is shiny
vinyl	used for records	a type of plastic	used for making records
water	is clear	is wet	a liquid
wax	is sticky	is white	used for candles
wax paper	is thin	is white	used for wrapping food
wool	is soft	is warm	comes from sheep
zinc	a metal	is shiny	is white
zirconium	a metal	is hard	is shiny

Supplementary Table 5. Descriptive labels of all 36 model dimensions obtained from GPT-3 feature norms.

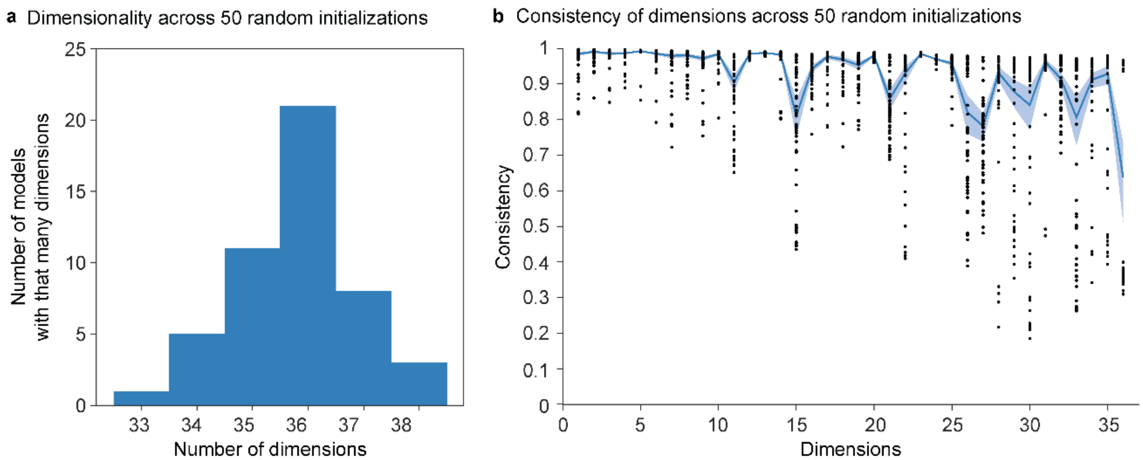
dimension (label)	Top 3 features (weights on dimensions)		
1 (mineral)	is hard (2.29)	a metal (1.98)	a mineral (1.60)
2 (wood)	used for making furniture (3.74)	a wood (2.07)	is hard (1.81)
3 (metallic)	a metal (5.00)	is shiny (4.30)	is hard (1.81)
4 (fabric)	used for making clothes (4.33)	is soft (3.46)	a fabric (2.79)
5 (white colour)	is white (6.39)	is soft (1.38)	is thin (1.16)
6 (grainy)	is black (1.05)	is grey (0.79)	used for making roads (0.79)

7 (crystalline)	a gemstone (3.94)	is shiny (3.56)	used for making jewelry (3.02)
8 (fibrous)	is green (1.73)	has leaves (1.38)	a plant (1.31)
9 (cloudy)	a gas (1.60)	is hot (1.38)	is cold (1.23)
10 (small)	is hard (1.62)	is shiny (1.54)	used for making jewelry (1.53)
11 (beige, tan colour)	is yellow (1.39)	is brown (1.06)	used for making furniture (0.88)
12 (viscous)	is sticky (2.76)	a liquid (2.20)	is clear (1.00)
13 (black colour)	is black (5.86)	is sticky (0.93)	used for writing (0.91)
14 (brick)	is hard (1.53)	made of clay (1.12)	is rectangular (1.00)
15 (swirly)	used for making clothes (2.31)	is soft (2.05)	a fabric (1.67)
16 (maroon colour)	used for making furniture (2.74)	a wood (1.83)	is brown (1.80)
17 (lines, long, vertical, tubular)	used for making furniture (2.01)	a metal (1.97)	is hard (1.61)
18 (mesh)	is strong (2.44)	is metallic (1.56)	is heavy (1.12)
19 (yellow colour)	is yellow (3.83)	is poisonous (1.02)	a plant (0.96)
20 (multi-coloured)	is colorful (2.38)	is white (1.55)	is soft (1.29)
21 (thin)	is thin (4.26)	is white (2.11)	used for writing (2.00)
22 (bulbous)	is white (2.50)	is soft (2.06)	a gas (1.33)
23 (green colour)	is green (7.50)	a plant (3.52)	has leaves (2.81)
24 (blue colour)	is blue (1.92)	is transparent (1.55)	a gemstone (1.19)
25 (shell, bone)	is white (4.64)	is hard (3.72)	made of calcium (1.30)
26 (gemstone)	used for making jewelry (3.86)	is shiny (3.73)	a gemstone (2.60)
27 (sheet)	is thin (4.13)	made of paper (2.35)	used for wrapping food (2.09)
28 (hair)	is soft (3.72)	used for making clothes (2.34)	is warm (1.46)
29 (spongy)	used for cleaning (1.91)	is white (1.54)	is thin (1.34)
30 (golden colour, shiny)	is shiny (5.87)	a metal (5.39)	is metallic (1.84)
31 (red colour)	is red (2.96)	a gemstone (1.69)	is hot (1.51)
32 (round)	made of clay (3.11)	is white (1.73)	is fragile (1.37)
33 (cream colour)	is white (3.15)	made of wood (1.01)	used for writing (0.69)
34 (bumpy)	is soft (1.20)	has bubbles (1.18)	is sweet (0.98)
35 (hot)	is hot (7.41)	is red (4.76)	is dangerous (2.71)
36 (turquoise colour)	is soft (2.00)	is clear (1.46)	is smooth (1.43)



Supplementary Figure 1. Model prediction of choice probability. Relationship between the model's predicted choice probability for the 1,200 test triplets and the measured choice probability.

Supplementary Figure 2. Two-dimensional visualization of the similarity embedding. The similarity embedding is visualized as described in Fig. 6 but each of the 600 images is represented by a square patch with its median colour in CIELAB space.



Supplementary Figure 3. Stability of model dimensions. a. Dimensionality across 50 random initializations, evaluated by plotting the frequency of resulting models, grouped by their number of dimensions. **b.** Reproducibility of dimensions across 50 random initializations in the chosen 36-dimensional embedding. Shaded areas are indicating 95% confidence intervals.

Acknowledgments

We thank Saskia Honnefeller, Christopher Landau, Judith Kanehl, and Wiebke Siedentop for help in screening and curating images for the STUFF dataset. Thanks to Hannes Hansen for applying the automatic semantic feature generation approach to our material nouns and for support with identifying the relevant feature dimensions. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 222641018 – SFB/TRR 135 TP C1, by the European Research Council (ERC) Consolidator Award SHAPE – project number ERC-CoG-2015-682859 and ERC Starting Award COREDIM–project number ERC-StG-2021-101039712, by a Max Planck Research Group Grant awarded to M.N.H. (M.TN.A.NEPF0009), and by Research Cluster “The Adaptive Mind” funded by the Hessian Ministry for Higher Education, Research, Science and the Arts.

References

1. Kietzmann TC, Spoerer CJ, Sörensen LKA et al. (2019) Recurrence is required to capture the representational dynamics of the human visual system. Proc Natl Acad Sci USA 116:21854–21863. https://doi.org/10.1073/pnas.1905544116
2. Bracci S, Ritchie JB, Beeck H op de (2017) On the partnership between neural representations of object categories and visual features in the ventral visual pathway. Neuropsychologia 105:153–164. https://doi.org/10.1016/j.neuropsychologia.2017.06.010
3. Jozwik KM, Kriegeskorte N, Mur M (2015) Visual features as stepping stones toward semantics: Explaining object similarity in IT and perception with non-negative least squares. Cold Spring Harbor Laboratory
4. Cichy RM, Kriegeskorte N, Jozwik KM et al. (2019) The spatiotemporal neural dynamics underlying perceived similarity for real-world objects. NeuroImage 194:12–24. https://doi.org/10.1016/j.neuroimage.2019.03.031
5. Wiebel CB, Valsecchi M, Gegenfurtner KR (2013) The speed and accuracy of material recognition in natural images. Atten Percept Psychophys 75:954–966. https://doi.org/10.3758/s13414-013-0436-y
6. Sharan L, Rosenholtz R, Adelson EH (2014) Accuracy and speed of material categorization in real-world images. Journal of Vision 14:12. https://doi.org/10.1167/14.9.12
7. Schmid AC, Boyaci H, Doerschner K (2021) Dynamic dot displays reveal material motion network in the human brain. NeuroImage 228:117688. https://doi.org/10.1016/j.neuroimage.2020.117688
8. Adelson EH (2001) On seeing stuff: The perception of materials by humans and machines. In: Rogowitz BE, Pappas TN (eds) Proceedings of SPIE - The International Society for Optical Engineering. SPIE, pp 1–12
9. Fleming RW (2017) Material Perception. Annu Rev Vis Sci 3:365–388. https://doi.org/10.1146/annurev-vision-102016-061429
10. Anderson BL (2011) Visual perception of materials and surfaces. Curr Biol 21:R978-83. https://doi.org/10.1016/j.cub.2011.11.022
11. DiCarlo JJ, Zoccolan D, Rust NC (2012) How does the brain solve visual object recognition? Neuron 73:415–434. https://doi.org/10.1016/j.neuron.2012.01.010
12. Fleming RW, Wiebel C, Gegenfurtner K (2013) Perceptual qualities and material classes. J Vis 13. https://doi.org/10.1167/13.8.9
13. Sharan L, Liu C, Rosenholtz R et al. (2013) Recognizing Materials using Perceptually Inspired Features. Int J Comput Vis 103:348–371. https://doi.org/10.1007/s11263-013-0609-0
14. van Assen JJR, Barla P, Fleming RW (2018) Visual Features in the Perception of Liquids. Curr Biol 28:452-458.e4. https://doi.org/10.1016/j.cub.2017.12.037
15. Liu C, Sharan L, Adelson EH et al. Exploring features in a Bayesian framework for material recognition. In: 2010 IEEE Computer Society Conference, pp 239–246
16. van Assen JJR, Nishida S, Fleming RW (2020) Visual perception of liquids: Insights from deep neural networks. PLOS Computational Biology 16:e1008018. https://doi.org/10.1371/journal.pcbi.1008018
17. Lagunas M, Malpica S, Serrano A et al. (2019) A similarity measure for material appearance. ACM Trans Graph 38:1–12. https://doi.org/10.1145/3306346.3323036
18. Schwartz G, Nishino K (2016) Material recognition from local appearance in global context (preprint). arXiv:1611.09394
19. Bell S, Upchurch P, Snavely N et al. (2015) Material recognition in the wild with the Materials in Context Database. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE
20. Edelman S (1998) Representation is representation of similarities. Behav Brain Sci 21:449-67; discussion 467-98. https://doi.org/10.1017/s0140525x98001253
21. Shepard RN (1987) Toward a universal law of generalization for psychological science. Science 237:1317–1323. https://doi.org/10.1126/science.3629243

22. Nosofsky RM (1984) Choice, similarity, and the context theory of classification. Journal of Experimental Psychology: Learning, Memory, and Cognition 10:104–114. https://doi.org/10.1037/0278-7393.10.1.104
23. Zheng CY, Pereira F, Baker CI et al. (2019) Revealing interpretable object representations from human behavior. International Conference on Learning Representations (ICLR). https://openreview.net/forum?id=ryxSrhC9KX
24. Hebart MN, Zheng CY, Pereira F et al. (2020) Revealing the multidimensional mental representations of natural objects underlying human similarity judgements. Nat Hum Behav 4:1173–1185. https://doi.org/10.1038/s41562-020-00951-3
25. Hebart MN, Dickter AH, Kidder A et al. (2019) THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. PLOS ONE 14:e0223792. https://doi.org/10.1371/journal.pone.0223792
26. Mehrer J, Spoerer CJ, Jones EC et al. (2021) An ecologically motivated image dataset for deep learning yields better models of human vision. Proc Natl Acad Sci USA 118:e2011417118. https://doi.org/10.1073/pnas.2011417118
27. Sharan L, Rosenholtz R, Adelson E (2009) Material perception: What can you see in a brief glance? J Vis 9:784. https://doi.org/10.1167/9.8.784
28. Pilehvar MT, Collier N (2016) De-Conflated Semantic Representations
29. Murphy B, Talukdar P, Mitchell T Learning effective and interpretable semantic models using non-negative sparse embedding. In: Proceedings of COLING, pp 1933–1950
30. Hoyer PO (2004) Non-negative matrix factorization with sparseness constraints. Journal of Machine Learning Research 5:1457–1469
31. Yuan W, Mo Y, Wang S et al. Active Clothing Material Perception Using Tactile Sensing and Deep Learning. In: 2018 IEEE International Conference, pp 4842–4849
32. Pereira F, Gershman S, Ritter S et al. (2016) A comparative evaluation of off-the-shelf distributed semantic representations for modelling behavioural data. Cognitive Neuropsychology 33:175–190. https://doi.org/10.1080/02643294.2016.1176907
33. Schwartz G, Nishino K (2020) Recognizing Material Properties from Images. IEEE Trans Pattern Anal Mach Intell 42:1981–1995. https://doi.org/10.1109/tpami.2019.2907850
34. Sawayama M, Dobashi Y, Okabe M et al. (2022) Visual discrimination of optical material properties: A large-scale study. Journal of Vision 22:17. https://doi.org/10.1167/jov.22.2.17
35. Schmid AC, Doerschner K (2019) Representing stuff in the human brain. Current Opinion in Behavioral Sciences 30:178–185. https://doi.org/10.1016/j.cobeha.2019.10.007
36. Komatsu H, Goda N (2018) Neural Mechanisms of Material Perception: Quest on Shitsukan. Neuroscience 392:329–347. https://doi.org/10.1016/j.neuroscience.2018.09.001
37. Cant JS, Goodale MA (2011) Scratching beneath the surface: new insights into the functional properties of the lateral occipital area and parahippocampal place area. J Neurosci 31:8248–8258. https://doi.org/10.1523/JNEUROSCI.6113-10.2011
38. Hiramatsu C, Goda N, Komatsu H (2011) Transformation from image-based to perceptual representation of materials along the human ventral visual pathway. NeuroImage 57:482–494. https://doi.org/10.1016/j.neuroimage.2011.04.056
39. Jacobs RHAH, Baumgartner E, Gegenfurtner KR (2014) The representation of material categories in the brain. Front Psychol 5:146. https://doi.org/10.3389/fpsyg.2014.00146
40. Dobs K, Isik L, Pantazis D et al. (2019) How face perception unfolds over time. Nat Commun 10:1258. https://doi.org/10.1038/s41467-019-09239-1
41. Bracci S, Beeck H op de (2016) Dissociations and Associations between Shape and Category Representations in the Two Visual Pathways. J Neurosci 36:432–444. https://doi.org/10.1523/JNEUROSCI.2314-15.2016
42. Epstein RA, Baker CI (2019) Scene Perception in the Human Brain. Annu Rev Vis Sci 5:373–397. https://doi.org/10.1146/annurev-vision-091718-014809
43. Kravitz DJ, Saleem KS, Baker CI et al. (2013) The ventral visual pathway: an expanded neural framework for the processing of object quality. Trends Cogn Sci (Regul Ed) 17:26–49. https://doi.org/10.1016/j.tics.2012.10.011

44. Kriegeskorte N, Mur M, Ruff DA et al. (2008) Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron 60:1126–1141. https://doi.org/10.1016/j.neuron.2008.10.043
45. Konkle T, Oliva A (2012) A real-world size organization of object responses in occipitotemporal cortex. Neuron 74:1114–1124. https://doi.org/10.1016/j.neuron.2012.04.036
46. Khaligh-Razavi S-M, Cichy RM, Pantazis D et al. (2018) Tracking the Spatiotemporal Neural Dynamics of Real-world Object Size and Animacy in the Human Brain. J Cogn Neurosci 30:1559–1576. https://doi.org/10.1162/jocn_a_01290
47. Martin Cichy R, Khosla A, Pantazis D et al. (2017) Dynamics of scene representations in the human brain revealed by magnetoencephalography and deep neural networks. NeuroImage 153:346–358. https://doi.org/10.1016/j.neuroimage.2016.03.063
48. Bao P, She L, McGill M et al. (2020) A map of object space in primate inferotemporal cortex. Nature 583:103–108. https://doi.org/10.1038/s41586-020-2350-5
49. Tsao DY, Livingstone MS (2008) Mechanisms of face perception. Annu Rev Neurosci 31:411–437. https://doi.org/10.1146/annurev.neuro.30.051606.094238
50. Brysbaert M, New B, Keuleers E (2012) Adding part-of-speech information to the SUBTLEX-US word frequencies. Behav Res 44:991–997. https://doi.org/10.3758/s13428-012-0190-4
51. Keuleers E, Lacey P, Rastle K et al. (2012) The British Lexicon Project: lexical decision data for 28,730 monosyllabic and disyllabic English words. Behav Res 44:287–304. https://doi.org/10.3758/s13428-011-0118-4
52. Brysbaert M, Warriner AB, Kuperman V (2014) Concreteness ratings for 40 thousand generally known English word lemmas. Behav Res 46:904–911. https://doi.org/10.3758/s13428-013-0403-5
53. Princeton University (2010) “About WordNet.” WordNet. https://wordnet.princeton.edu/
54. Kingma DP, Ba J (2014) Adam: A Method for Stochastic Optimization. arXiv
55. Brown TB, Mann B, Ryder N et al. (2020) Language Models are Few-Shot Learners
56. Hansen H, Hebart MN (2022) Semantic features of object concepts generated with GPT-3
57. Salton G, Buckley C (1988) Term-weighting approaches in automatic text retrieval. Information Processing & Management 24:513–523. https://doi.org/10.1016/0306-4573(88)90021-0
58. Wu Z, Palmer M (1994) Verb Semantics and Lexical Selection