Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Research Report

Vision matters for shape representation: Evidence from sculpturing and drawing in the blind

Shuang Tian [a,1], Lingjuan Chen [a,1], Xiaoying Wang [a], Guochao Li [a], Ze Fu [a], Yufeng Ji [d,e], Jiahui Lu [a], Xiaosha Wang [a], Shiguang Shan [d,e] and Yanchao Bi [a,b,c,*]

[a] *State Key Laboratory of Cognitive Neuroscience and Learning & IDG, McGovern Institute for Brain Research, Beijing Normal University, Beijing, China*

[b] *Beijing Key Laboratory of Brain Imaging and Connectomics, Beijing Normal University, Beijing, China*

[c] *Chinese Institute for Brain Research, Beijing, China*

[d] *Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China*

[e] *University of Chinese Academy of Sciences, Beijing, China*

ARTICLE INFO

Article history:
Received 1 November 2023
Reviewed 3 January 2024
Revised 23 January 2024
Accepted 27 February 2024
Action editor Ryo Kitada
Published online 19 March 2024

Keywords:
Shape
Multimodal
Blind
Object modelling
Drawing

ABSTRACT

Shape is a property that could be perceived by vision and touch, and is classically considered to be supramodal. While there is mounting evidence for the shared cognitive and neural representation space between visual and tactile shape, previous research tended to rely on dissimilarity structures between objects and had not examined the detailed properties of shape representation in the absence of vision. To address this gap, we conducted three explicit object shape knowledge production experiments with congenitally blind and sighted participants, who were asked to produce verbal features, 3D clay models, and 2D drawings of familiar objects with varying levels of tactile exposure, including tools, large nonmanipulable objects, and animals. We found that the absence of visual experience (i.e., in the blind group) led to stronger differences in animals than in tools and large objects, suggesting that direct tactile experience of objects is essential for shape representation when vision is unavailable. For tools with rich tactile/manipulation experiences, the blind produced overall good shapes comparable to the sighted, yet also showed intriguing differences. The blind group had more variations and a systematic bias in the geometric property of tools (making them stubbier than the sighted), indicating that visual experience contributes to aligning internal representations and calibrating overall object configurations, at least for tools. Taken together, the object shape representation reflects the intricate orchestration of vision, touch and language.

* *Corresponding author*. State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, 100875, China.

E-mail address: ybi@bnu.edu.cn (Y. Bi).

[1] Co-first authors.

https://doi.org/10.1016/j.cortex.2024.02.016

1. Introduction

Object recognition is a multisensory process that involves different types of neural representation, with modality-specific sensory signals merging into a coherent object representation. For example, we can recognize a cup by seeing it, touching it, hearing someone tap on it, or listening to someone describe it with words. How these different sensory inputs merge into a supramodal object representation, and what the nature of such potentially supramodal representation is, remain to be fully understood.

Shape is a classic example of a presumed supramodal representation. It could be acquired and accessed by vision, touch, and language description, and there is rich evidence in the literature suggesting at least some shared cognitive and neural representations across these input modalities. For people with typical sight development who can perceive object shape with vision and touch simultaneously, the perceptual similarity space input from these two modalities was found highly correlated behaviorally, indicating a good alignment between them (Erdogan et al., 2015; Lee Masson et al., 2016). Neuroimaging studies supported this alignment by showing that both visual and haptic input of objects activated a common brain area in the lateral occipital cortex (LOC), the neural representation similarity space of vision and touch was correlated with the behavioral similarity space (Amedi et al., 2001, 2002; Erdogan et al., 2016; Lee Masson et al., 2016; Stilla & Sathian, 2008), and the neural representation could decode cross-modally (Erdogan et al., 2016; Tian et al., 2023). While such visual-tactile overlap observed in sighted individuals could be attributed to imagery, compelling evidence has also been obtained with congenitally blind individuals. Behavioral and neuroimaging studies showed that, without visual experience, blind people's behavioral patterns in terms of shape similarity ratings and sorting tasks were highly similar to those of the sighted people (Kim et al., 2019; Peelen et al., 2014). Such behavioral similarity was also supported by a similar neural basis. Visual ventral cortex shows similar activation profiles in sighted and congenitally blind individuals (e.g., Mahon et al., 2009; Pietrini et al., 2004). The LOC could be activated by touching or hearing visual-to-auditory sensory substitution soundscapes in both sighted and blind subjects (Amedi et al., 2007, 2010), the representation space in this area in the blind subjects showed a correlated structure with that of the pictorial form in the sighted subjects (Handjaras et al., 2016; Handjaras et al., 2017), and the neural representational similarity pattern was significantly correlated with the behavioral shape similarity ratings in both groups (Peelen et al., 2014; Xu et al., 2023). Finally, language experience also contributes to object property learning (Bi, 2021; Wang et al., 2020), and shape knowledge could be constructed at least partly based on language descriptions (see discussions in Kim et al., 2019).

Although these lines of evidence converge on a supramodal shape representation across visual, tactile, and language experiences, the evidence is predominantly based on (dis)similarity structures (e.g., a paintbrush is more similar in shape to a razor than to a pair of scissors), which may well result from different individual shape representations. Even similar representations that allow for cross-modal decoding might have different tuning properties (Breedlove et al., 2020; Favila et al., 2022). We do not know from the existing evidence whether there are any differences in how the congenitally blind, without vision, represent the shape of a cup compared to sighted people. In fact, past research has implicated a complex interaction among object properties (object identity *vs* shape), domain (animate *vs* artifacts), and information modality (visual *vs* nonvisual; sighted *vs* blind). Modality independence tended to be observed when shape information was accessed explicitly (input or output), for items where shape properties match with nonvisual properties more directly (e.g., when object sound conveys shape information in cases of sensory substitutional mapping or emotional expressions with systematic facial shape correspondences), relative to when object identity was accessed and/or when object shape properties do not match with other properties transparently (e.g., identity sound or vocal sound without transparent mapping with shape; Amedi et al., 2002, 2007; Bola et al., 2022; Mattioni et al., 2020; Wang et al., 2015; see discussions about interaction with object domains in Bi et al., 2016). Thus, an explicit assessment of object shape representations across different types of objects and tasks is warranted.

There are different possibilities about how multi-modality inputs work together to derive shape representations with different predictions. One is that there is a same supramodal representation that can be generated independently from visual or tactile inputs, i.e., either modality is sufficient, which predicts that the blind and the sighted have the same shape representation for objects with which they have full tactile experience. For example, touching or seeing a cup yields the same representation in the blind and the sighted. For objects that they have not touched (e.g., a rocket or an elephant), there are two possibilities. One possibility is that shape representations are thus impoverished in the blind; the other is that they can nonetheless learn the shape compositions from language descriptions and combine them with their tactile experiences of certain shape elements despite the lack of experience with the actual object. For instance, a person may learn from language that an elephant is a big mammal with a nose like a long tube and big ears like fans, forming a composite representation with the shape of tube and fans. This latter possibility accommodates the same shape representation for the blind and the sighted people even for those objects without direct tactile experiences. Finally, there is also a possibility that visual and tactile modalities have intrinsic differences in shape derivation. For instance, the visual modality tends to access object shape in parallel and is less constrained by size, while the tactile modality accesses shape more sequentially for objects larger than palm and is more tightly related to the manipulation and functional object properties. Having them available together in the sighted leads to convergent shape representations, but in the absence of one modality, the shape representation constructed from the remaining modality might differ. According to this hypothesis, even for objects with which they have rich tactile experience, the congenitally blind might have different shape representations from the sighted.

To assess how visual, tactile, and language inputs affect shape representations, we compared congenitally blind and sighted participants in explicit object shape production experiments. We chose three domains of objects with which both groups were highly familiar but had different degrees of tactile experiences: tools (high), large nonmanipulable objects (medium), and animals (low). Three shape knowledge production behavioral experiments were conducted on these objects: object feature verbal generation (a language task), clay modelling with Play-Doh (3D shape representation), and drawing (2D shape representation). The clay modelling experiment (Exp 2) is the most transparent one and would be considered as the main experiment, with the verbal feature task replicating and extending the literature and drawing task exploring the 2D transformation of the object shape in the two populations. We tested potential group differences in two aspects: the general quality of their shape knowledge production (How good is each response?) and the inter-subject consistency (How variable is each response with that of the other participants?). Specific measures for these two aspects are chosen based on the feasibility and optimality for each experiment.

2. Experiment 1: Verbal feature generation

2.1. *Methods*

2.1.1. *Participants*

Thirteen early blind people (age: mean ± SD = 38 ± 12; years of education: mean ± SD = 11 ± 3; four females) and 15 sighted people (age: mean ± SD = 44 ± 10; years of education: mean ± SD = 12 ± 3; seven females) who were matched to each other in age [$t_{(26)} = 1.55$, $p = .13$] and years of education [$t_{(26)} = .93$; $p = .36$] participated in the verbal feature generation task. The sample size was determined based on the availability of early blind individuals. All blind participants in the three experiments reported that they were born blind (see Table S1 for detailed characteristics), although we could not verify the exact time of their blindness onset due to the lack of their medical records. None reported any memory of being able to see and identify shapes visually. All sighted participants had normal or corrected-to-normal vision. All participants were native Mandarin Chinese speakers. None had any history of psychiatric or neurological disorders or suffered from head injuries. All participants provided informed consent and received monetary compensation for their participation.

2.1.2. *Materials*

The stimuli included 3 domains of 10 objects each, including tools (such as fan/hammer/key), large nonmanipulable objects (such as bed/couch/bus), and animals (such as cat/lizard/hairtail). The familiarity of the stimuli was rated on a 7-point scale by the same group of participants (except two sighted participants) and was equally high to the blind and sighted groups in every domain (tools: means ± SD = 6.55 ± .31 and 6.45 ± .45, $p = 1$; large objects: means ± SD = 6.24 ± .62 and 6.49 ± .36, $p = .84$; animals: means ± SD = 4.88 ± 1.06 and 5.15 ± 1.00, $p = 1$), although the animals were generally less familiar than the tools and large objects ($ps < 2.02 \times 10^{-7}$). Note that in the actual experiment we included additional categories (e.g., celebrity, cloth, face part, flower, food, fruit and vegetable, musical instrument, vehicle) for another purpose and are not described here.

2.1.3. *Procedure*

The experiment was conducted in the laboratory or in the subjects' homes according to their preference. During each trial, the subjects listened to the name of an item and were asked to say the first three features they associated with this item. These features could be what the item looks like, what sound it makes, how it feels to touch, what smell or taste it has, how to use it, what it is used for, what connections it has with other things, or other unique features of this item. All responses were recorded and input to a sheet.

2.1.4. *Analysis*

We adopted two approaches to quantify the generated features data (see Fig. 1A): (1) Manually coding. We categorized the generated features according to the degree to which they are related to modality-specific properties (see Fig. S1). We further grouped them into three broad feature types for analyses given the interest of the visual absence effect: "visual-only" (color and surface texture), "multimodal" (i.e., possibly acquired by either vision or nonvisual modalities – shape, size, and motion), and "others" to which the two subject groups have similar manners of access (properties related to nonvisual sensory modalities, such as gustatory, olfactory, auditory; encyclopedic knowledge). For example, for the features generated for crow, we coded "their fur is black" as visual-only, "a kind of bird" as encyclopedic, and "it flies" as multimodal (See Table S5 for more examples; https://osf.io/59wkf/ for the complete list). We counted the number of different types of features, converted the numbers to proportions, and performed a three-way (domain × group × feature type) repeated-measures ANOVA to test whether the absence of visual experience affected the predominance of retrieval across different feature types. (2) Implicit shape representation extraction using text analysis. We identified the nearest 50 words to the word "shape" in Chinese by calculating the cosine distance of the word vectors extracted by a pre-trained fastText embedding model (Joulin et al., 2016; Word vectors for 157 languages · fastText). We manually deleted the nonsense words and adjusted the wrongly segmented words, resulting in 25 shape-related words as key shape dimensions. Next, we used a pre-trained Bert model for Chinese (Devlin et al., 2018; bert-base-chinese · Hugging Face) to transform the generated features into sentence-level embedding vectors and calculated the cosine distance between each feature and each shape-related word. To test whether the features generated by the two groups implicated different shape association patterns and whether the difference varied between different domains, we performed group × domain repeated-measures ANOVA on the distances. To test the agreement within and across the two groups, we also averaged the distances along the shape dimensions for each item each subject, and calculated pairwise inter-subject (1 – correlation). One sighted participant's mean distance from the others in the large object condition exceeded the group mean distance by 3 standard deviations, their data were considered outliers and were excluded from the analyses. Repeated-measures ANOVA and pairwise t-tests

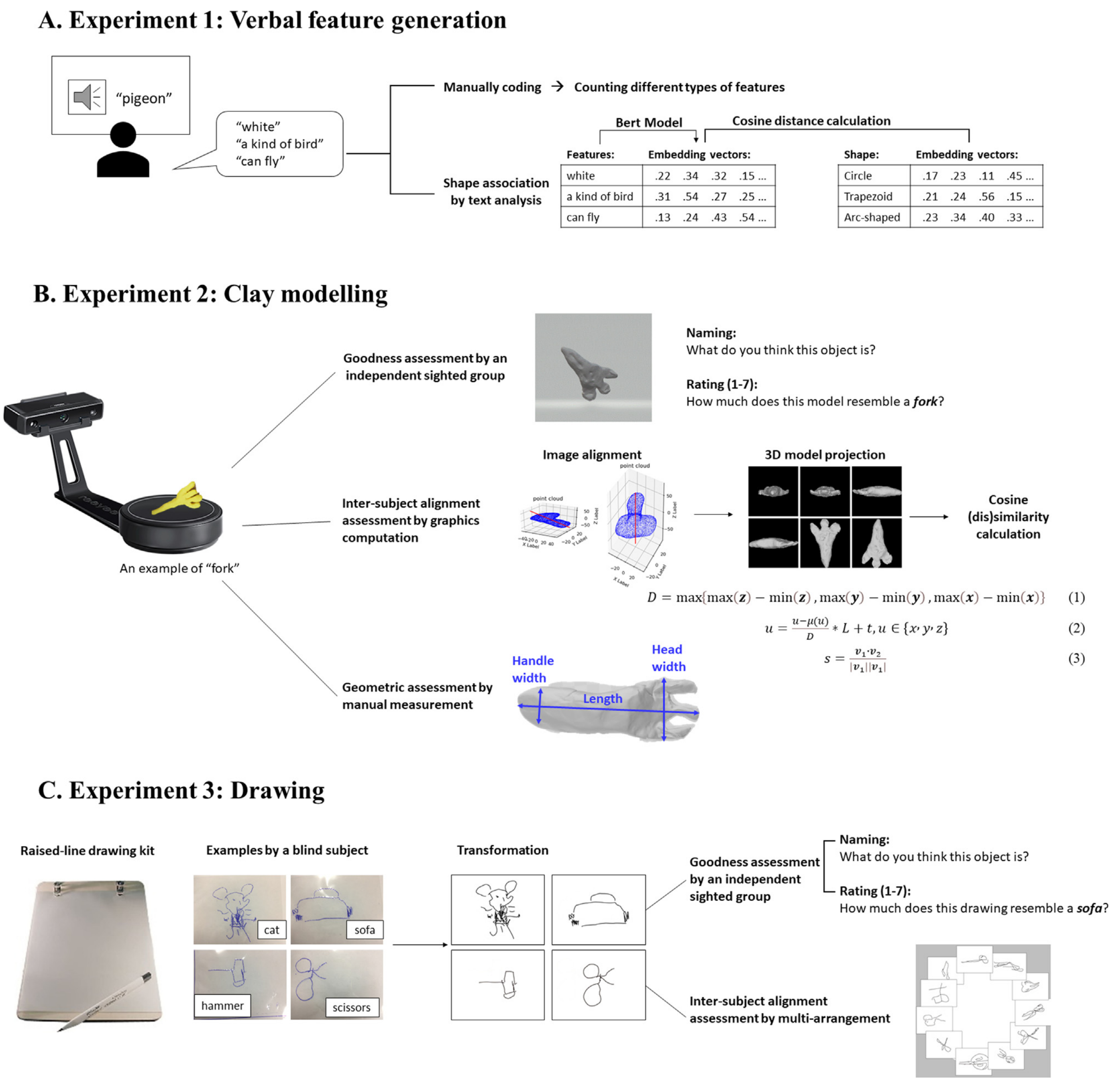

Fig. 1 – Schematic diagrams of the experimental procedures and measurements. (A) Verbal feature generation experiment (Experiment 1): The subjects were instructed to report 3 critical features for each orally presented item from 3 domains (tools, large nonmanipulable objects, and animals). The generated features were coded as "visual-only", "multimodal", and "others" types and analyzed by group × feature type × domain repeated-measures ANOVA. The features were also converted to embedding vectors and represented by the patterns of their distances to 25 shape-related words. Inter-subject alignment analyses were done on these representations. (B) Clay modelling experiment (Experiment 2): The subjects (blindfolded if sighted) shaped a 448-gram Hasbro Play-Doh into the object named by the experimenter. The models were then scanned by a 3D scanner. An independent group of sighted raters judged the goodness of the models by naming and resemblance rating. A graphics computation method calculated inter-subject similarities of the models. The geometric parameters of tool models were manually measured. Formulas 1 and 2 were used for scale alignment. Formula 3 was used to calculate cosine similarity of histogram of oriented gradients of the projected images (Dalal et al., 2005). (C) Drawing experiment (Experiment 3): The subjects (blindfolded if sighted) drew the object using a raised-line drawing kit. The drawings were photographed, resized, background-removed, and centralized. An independent group of sighted raters evaluated the goodness of the drawings by naming and resemblance rating. Another group of participants performed a multi-arrangement task to measure inter-subject alignment of the drawings.

were performed to compare the within-group and between-group disagreement. All multiple comparisons tests in the analyses were adjusted with the Bonferroni method, unless stated explicitly. All statistical analyses were conducted using rstatix package (Kassambara, 2021) in R (R Core Team, 2019). Considering that the data did not strictly follow the normal distribution, we also performed permutation-based (5000 permutations) ANOVA and t-tests using permuco (Frossard & Renaud, 2021) and RVAideMemoire packages (Hervé, 2020) respectively, to validate the parametric statistical results. We further visualized the shape representation space using the inter-subject × inter-item dissimilarity data which were averaged across all within-group and between-group subject pairs to obtain a 20 × 20 group-item dissimilarity matrix for each domain. We performed a multi-dimensional scaling analysis (MDS) using a 2-dimensional ratio MDS model with the SMACOF package (de Leeuw, 2016) which uses a stress majorization algorithm, and we visualized the results using the factoextra package (Kassambara, 2016) implemented in R.

2.2. *Results*

We adopted a widely used verbal feature generation experiment to examine the effects of visual absence on verbal feature production of objects. We first evaluated the distribution of feature types by manually labeling features into different sensory types: visual-only, multimodal, and others (see Fig. S1 for the detailed distributions of feature types). Note that we performed both parametric and nonparametric (permutation-based) testing for the ANOVA and t-test below given that not all data distributions were normal. The results were highly convergent. We present the parametric statistical figures in the main text and nonparametric ones in the Supplementary materials.

A group × feature type × domain repeated-measures ANOVA (see Table S2A for both the parametric and permutation-based results) showed a trend of significant interaction between group and feature type [$F_{(2,\ 52)} = 2.74$, $p = .07$] in that the blind group generated more multimodal features ($p = .03$) than the sighted group, while they did not differ on the other two types (*ps* > .09), or on the more fine-grained feature categories (*ps* > .42; Fig. S1). The interaction was not modulated by different domains [$F_{(4,\ 104)} = 1.25$, $p = .30$]. The interaction of domain and feature type was significant [$F_{(4,\ 104)} = 4.11$, $p = .004$]. Simple main effect analyses showed that significant domain effect only existed in the visual-only features ($p = 9.45 \times 10^{-7}$) with animals having more visual-only features than large objects and tools (*ps* < .002), and tools having slightly more visual-only features than large objects ($p = .05$). This result agrees with the object domain × modality interaction pattern that is well recognized and discussed in the literature (e.g., Bi et al., 2016; Mahon & Caramazza, 2009, 2011).

The primary focus is object shape representation. We did not ask the subjects to generate shape features explicitly given the poverty of shape terms in general. Instead, we analyzed the implicit shape representations potentially embedded in the features they generated. Given the widely associative nature of language descriptions (Grand et al., 2022), we used a large text analysis approach to explore the patterns in which the verbal features associate with shape features. We achieved 25 words that are most related to shape from a pre-trained fastText embedding model. We computed the cosine distance between each created feature and each of the 25 shape words as the shape-dimensional-vector representation (see Methods). A two-way repeated-measures ANOVA (see Fig. 2A and Table S2B) showed a main effect of domain [$F_{(2,\ 52)} = 31.50$, $p = 1.09 \times 10^{-9}$] with the features of tools being nearer to shape features than those of animals and large objects (*ps* < 3.00×10^{-182}) which did not differ with each other ($p = .44$). This is in agreement with the findings and discussions about the close relation between tool use and shape property (Chen et al., 2017; Fabbri et al., 2016; Wang et al., 2018). There was also a main effect of blindness [$F_{(1,\ 26)} = 6.53$, $p = .02$], with the features generated by the blind people being less strongly linguistically associated with shape features than those by the sighted group. The interaction between domain and group was not significant [$F_{(2,\ 52)} = 1.13$, $p = .33$]. These findings suggest the sensitivity of such text-analysis-based shape representation extraction, although the nature of the representation is not transparent.

To test whether the absence of visual experience affects individual shape knowledge idiosyncrasy, we calculated the inter-subject dissimilarity (i.e., 1 − correlation) of the mean feature patterns along the 25 shape dimensions for each item (see Fig. 2B and Table S2C). The data from one sighted subject were excluded because their mean distance from the other subjects was more than the group mean by 3 standard deviations in the large object condition. The main effect of domain was significant [$F_{(2,\ 696)} = 102.65$, $p = 8.61 \times 10^{-40}$], the main effect of group was not [$F_{(2,\ 348)} = 1.65$, $p = .19$]. There was a significant interaction between group and domain [$F_{(4,\ 696)} = 6.69$, $p = 2.76 \times 10^{-5}$]. Simple main effect analysis revealed that for tools and large objects the effects of group were significant (*ps* < .02). For tools, the within-group agreement in the blind was higher than both in the sighted group and the between-group (*ps* < .0006). For large objects, the within-sighted group agreement was the highest (*ps* < .02).

Finally, we visualized the shape-dimensional-vector representation space by performing a multi-dimensional scaling analysis (MDS) on the inter-subject × inter-item dissimilarity data (a 20 × 20 group-item mean dissimilarity matrix) for each domain. As shown in Fig. 2C, although it was not transparent what the two dimensions reflected, it was visible that the blind and sighted groups tended to cluster into different spaces with the blind tending to have a systematic shift.

3. Experiment 2: Clay modelling

3.1. *Methods*

3.1.1. *Participants*

Twelve blind people (age: mean ± SD = 45 ± 13; years of education: mean ± SD = 11 ± 4; four females) and 15 sighted people (age: mean ± SD = 51 ± 7; years of education: mean ± SD = 9 ± 2; nine females) participated in the clay modelling experiment. They were matched in age and years of education (*ps* > .1). Ten blind subjects and five sighted ones from Experiment 1 also participated in this experiment.

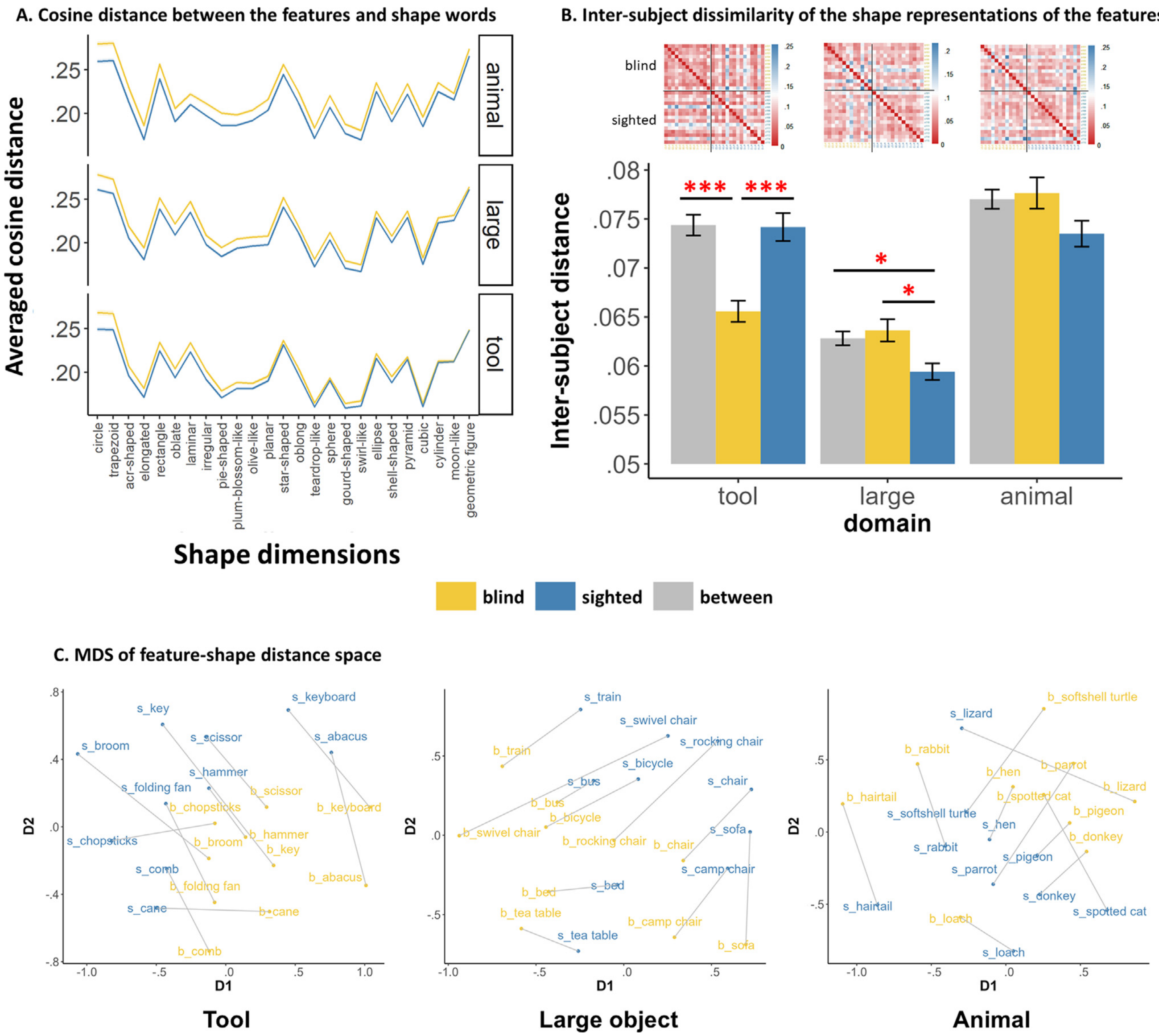

Fig. 2 – Results of the verbal feature generation experiment. (A) Average cosine distance of verbal features to shape-related words. (B) Inter-subject alignment calculated by pairwise (1 − correlation) of mean shape-dimensional-vector representations. Upper panel: Mean heatmaps of inter-subject dissimilarity within and between blind and sighted groups. Red: more similar; blue: less similar. Lower panel: Bar plot of inter-subject dissimilarity by group. Error bars: standard error of the mean (SEM) across subject pairs. *$p < .05$, **$p < .01$, ***$p < .001$, Bonferroni-corrected. (C) Multi-dimensional scaling of the mean shape spaces of the shape-dimensional-vector representations.

3.1.2. *Materials*

We selected the stimuli from the three domains: tools, large nonmanipulable objects, and animals. Each domain contained 8 items (see Fig. S2 for the items). They were equally familiar to the blind and sighted groups, as indicated by their 1–7 ratings (tools: means ± SD = 6.36 ± .30 and 6.36 ± .44, $p = 1$; large objects: means ± SD = 6.19 ± .72 and 6.19 ± .88, $p = 1$; animals: means ± SD = 5.20 ± .57 and 5.70 ± .28, $p = .15$). Consistent with Experiment 1, animals were less familiar to the subjects than large objects and tools ($ps < .002$).

3.1.3. *Procedure*

The experiment took place in the laboratory for all subjects. The experimenter orally presented each item to the subjects, who then tried to shape it with a 448-gram Hasbro Play-Doh, adjusting the amount of material as needed. The subjects could take as much time as they needed until they thought the model was recognizable or decided to quit. The sighted group were blind-folded during the experiment. After the subjects finished shaping, each model was scanned with a 3D scanner (Wiiboox Reeyee; see Fig. 1B).

3.1.4. Analysis

We measured the goodness of the models generated by the subjects using a subjective evaluation method. We recruited an independent sighted group of 31 raters (age: mean ± SD = 22 ± 3; 29 females) and separated them into three subgroups to evaluate three subsets of the models respectively. Within each subset the models made by the blind and the sighted were interleaved, with two common subjects' models appearing among every subset to calculate the consistency between raters. The models were transformed to GIF images, where they kept rotating at a constant gentle speed (see examples in Fig. S2). They were presented randomly one by one on the screen. One evaluation was to see if the model was intelligible by asking the raters to type the name of the object or “unknown” if they could not recognize it. Another evaluation was to rate how similar the model was to the target object on a 7-point Likert scale. The intraclass correlation (ICC) between the three averaged group ratings was high (ICC = .93, 95% CI: .83–.96), indicating excellent reliability (Koo & Li, 2016), thus the rating scores for each item were averaged across raters for further statistical analyses.

To measure how similar the models generated by the blind were to those by the sighted, we used a computer graphics approach to measure the (dis)agreement within and across subject groups for the models (The computation code is available at https://osf.io/59wkf/). We performed images alignment for the original scanned 3D images (i.e., stl files). In the step of pose alignment, we achieved the main axis (v) using principal component analysis (Chaouch & Verroust-Blondet, 2009; Liu et al., 2010; Paquet et al., 2000; Vranić et al., 2001), and rotate v to z-axis, then aligned x or y in the same way. The step of scale alignment is to normalize by the maximum value of the difference of the x, y, z coordinate components of all points (see Fig. 1B, formula 1 and 2). Next, we projected the models from six viewpoints: up, down, left, right, front, and back, and calculated the similarity of the projected images at each angle, which was achieved by calculating the cosine similarity (Fig. 1B, formula 3) between the features of histogram of oriented gradient of the images (Dalal et al., 2005). Considering the possibility that two models may be oriented differently in a certain angle (for example, one model's up may correspond to another model's down), we calculated the pairwise similarity of two models for each pair of opposite angles (up–down, left–right, front–back). We then took the maximum similarity value for each pair of angles, and averaged these three values to obtain the overall similarity of two models, which was transformed to dissimilarity by subtracting it from 1. To evaluate the validity of this computational measure, we compared such results with sampled rating results: 1) We sampled pairs with different ranges of computed similarity (10 pairs with high similarity (.98–.97); 10 with medium similarity (.5); and 10 with low similarity (.13–.06)). We then recruited 10 sighted raters to subjectively judge the shape similarity for each model pair on a 1–7 scale. 2) We evenly sampled 10 model pairs from each domain and recruited another group of 11 raters for subjective similarity judgment. For both kinds of pair samples, the computed similarity and the subjectively judged similarity were highly correlated (rs > .70; 5000 permutations, $p < .0005$).

Then we used this approach and tested the within-group and between-group dissimilarity differences for each domain. Given that the blind group failed to made 15 (16%) items among animal and large object models, resulting in missing values, we used linear mixed-effect modelling (comparison pair and item as random effects) with the lme4 package (Bates et al., 2015) and the multcomp package (Hothorn et al., 2008) implemented in R to test the domain and group effects (see Supplementary material for the details about the models). We adjusted all multiple comparisons tests in the analyses with the Bonferroni method, unless stated otherwise. We further visualized the shape representation space for each domain using the same method of MDS as in Experiment 1 (see Methods in Experiment 1). Motivated by the MDS result (Fig. 3D), we manually measured the length, handle-width, and head-width of the tool models (Fig. 1B) with 3D Builder in Windows 11, and conducted t-tests to compare group differences for the three parameters separately.

3.2. Results

In this clay modelling experiment, we collected 633 clay models in total (273 by the blind, 360 by the sighted; see Table S3A). Regarding to the goodness of the models, we recruited an independent group of 31 sighted raters (3 subsets) to name and score the clay models. The naming accuracy reflects the intelligibility of the models, and the rating score (1–7) reflects the degree to which the models were similar to the target objects. A domain × group repeated-measures ANOVA of naming accuracy (Fig. 3A and Table S3B for both the parametric and permutation-based results) revealed a significant main effect of group [$F_{(1, 25)} = 6.61$, $p = .02$; blind < sighted], and a significant main effect of domain [$F_{(2, 50)} = 107.17$, $p = 8.32 \times 10^{-19}$; tools > animals, tools > large objects, $ps < 1.22 \times 10^{-5}$]. The interaction between domain and group was also significant [$F_{(2, 50)} = 6.10$, $p = .004$, with the accuracy differed between the two groups only for animals, $p = .02$, not for tools and large objects, $ps > .07$]. The ratings yielded similar results (Fig. 3B and Table S3C): The main effects of group [$F_{(1, 25)} = 6.46$, $p = .02$; blind < sighted] and domain [$F_{(2, 50)} = 100.59$, $p = 2.98 \times 10^{-18}$; tools > animals, tools > large objects, $ps < 4.93 \times 10^{-5}$] were significant. The interaction was also significant [$F_{(2, 50)} = 5.92$, $p = .005$] in that the group difference was only significant in the animal models ($p = .02$) and not in the tool or large object models ($ps > .07$).

To compare the agreement of object shape representations within and across groups, we used a computer 3D graphic approach (see Methods) to calculate the cosine (dis)similarity between each pair of the models for each item (Fig. 3C). We used linear mixed-effect modelling (random effects for comparison pair and item) to test the domain and group effects. The likelihood-ratio test indicated that the group effect was significant [$\chi^2_{(2)} = 13.56$, $p = .001$], with dissimilarity within the sighted being smaller (within-sighted < within-blind, $p = .0002$, within-sighted < between-groups, $p = .001$), and the within-blind group dissimilarity not differing from the between-group dissimilarity ($p = .50$). The main effect of domain was also significant [$\chi^2_{(2)} = 40.32$, $p = 1.76 \times 10^{-9}$], with the tool models being the most consistent among individuals

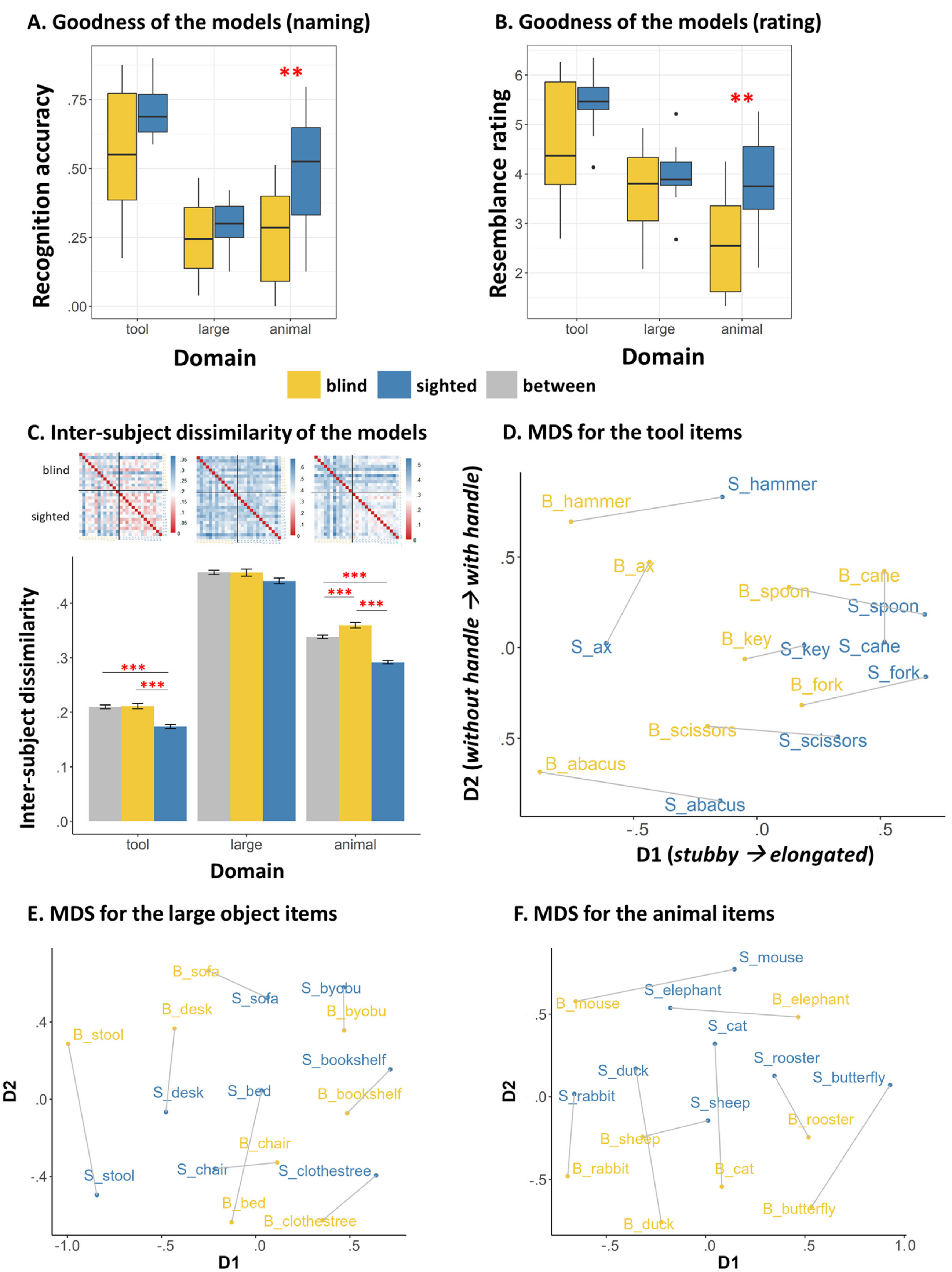

Fig. 3 – Results of the clay modelling experiment. (A) Goodness evaluated by the naming task. An independent group of sighted raters named the models. The accuracy was the proportion of correct recognition. (B) Goodness evaluated by the resemblance rating task. The same raters scored the models' similarity to the target objects on a 7-point Likert scale. (C)

($ps < 1.32 \times 10^{-6}$) and the large object models the most variable ($ps < 2.62 \times 10^{-5}$). Importantly, the interaction of domain and group was also significant [$\chi^2_{(4)} = 9.58$, $p < .05$]. Simple effect analyses showed significant group difference in terms of idiosyncrasy for tool models [$\chi^2_{(2)} = 33.28$, $p = 1.78 \times 10^{-7}$] and animal models [$\chi^2_{(2)} = 15.29$, $p < .002$]. For tools, the dissimilarity within the sighted was smaller compared to within blind and between group differences ($ps < .0003$), and the latter two were comparable ($p = 1$); for animals, the within-sighted dissimilarity was smaller than the between-group ($p = .01$), which was also smaller than the within-blind ($p = .003$).

In summary, the absence of visual experience led to poorer abilities to make 3D models of animals and did not affect the ability to generate recognizable, well-formed 3D structures for tools and large objects on the group average level (although the sighted subjects did not perform well on large objects, either). However, for tools it did lead to greater idiosyncrasies – without visual experience, the blind subjects had greater variations in their 3D modelling for tools relative to the sighted, which was not a general effect of less experience with Play-Doh as such effects were not apparent for large objects.

We used MDS to further explore the shape representation spaces of the two groups (see Methods and Fig. 3D–F). From the space of tools (Fig. 3D), it is again confirmed that the blind and sighted do share item-based consistency in terms of 3D shapes, although an overall left-ward shift of the blind group relative to the sighted was also revealed (except for “ax” and “cane”). Viewing items positions across Dimension 1 and 2 suggests that D1 roughly corresponded to a continuum of elongated versus stubby shapes, and D2 to a rough distinction of having or not a handle. We manually measured the length of the whole tool and the width of the head and handle parts separately (see Fig. 1B), excluding scissors because they varied in their open/closed state and were difficult to measure consistently, and compared them between the two groups. Welch t-tests revealed that the two groups differed significantly in length [$t_{(184)} = 2.20$, $p = .03$, uncorrected; blind made shorter/stubbier models], but not in the width of either the handle or the head ($ts < .51$; $ps > .61$). As for the MDS spaces for the large object and animal items (Fig. 3E–F), we did not observe a transparent interpretation for the dimensions and systematic differences between the groups as for the tools, which probably resulted from the overall higher inter-subject variances.

4. Experiment 3: Drawing

Drawing is a common cognitive tool to tap into mental shape representations (Bainbridge et al., 2021; Fan et al., 2023; Heller et al., 2002; Kennedy, 2003; Kennedy & Juricevic, 2003). We conducted the drawing experiment to explore how the absence of visual experience would impact participants' production of 2D shapes for objects, bearing in mind that blindness may additionally affect the process of mapping 3D representations onto 2D surface.

4.1. *Methods*

4.1.1. *Participants*

The subjects in the drawing experiment were the same as in the modelling experiment, with the addition of two blind subjects (14 blind, age: mean ± SD = 44 ± 14; years of education: mean ± SD = 11 ± 4; four females; 15 sighted, age: mean ± SD = 51 ± 7; years of education: mean ± SD = 9 ± 2; nine females). They were matched in both age and education ($ps > .05$). The drawings by one blind subject were excluded because he drew every item as messy circles without discriminative information.

4.1.2. *Materials*

The drawings were made using a Swedish raised-line drawing kit (Kennedy, 2003, see Fig. 1C), with which the lines drawn on the plastic paper were raised and thus could be detected by touching – i.e., tactile feedback of the traces was available. The size of a sheet of plastic paper was an A4 size. Each sheet was split into 4 equal areas with a cross line, and each object was drawn within the area of one quadrant.

The stimuli included six tools, six large nonmanipulable objects, and six animals (see Fig. S3 for the items) with the familiarity matched across the blind and sighted groups (tools: means ± SD = 6.38 ± .35 and 6.49 ± .33, $p = 1$; large objects: means ± SD = 6.64 ± .41 and 6.71 ± .19, $p = 1$; animals: means ± SD = 5.19 ± .67 and 5.78 ± .20, $p = .25$). Again, the animals were the least familiar domain ($ps < 6.35 \times 10^{-6}$). We chose this smaller item set because some subjects complained about the time consumption for drawing.

4.1.3. *Procedure*

For the blind subjects, the experiment was conducted either in the laboratory or in the subjects' home, according to their preference. For the sighted subjects, the experiment was conducted in the laboratory. The subjects were taught to manipulate the drawing kit, and were asked to freely draw something with the kit to familiarize themselves before the experiment started. In the experiment, each object name was orally presented by the experimenter. The subjects were instructed to draw the object within each quadrant and stop when they thought the drawing was recognizable or they gave up. No time limit was set. The sighted subjects were blindfolded during the experiment. We photographed all the generated drawings, resized them to 1024 × 768 pixels, and performed centering and background removal (Fig. 1C).

4.1.4. *Analysis*

To quantify the goodness of the drawings, we recruited an independent sighted group of raters (15 raters; age:

Inter-subject alignment calculated by pairwise dissimilarity with graphics computation method. Upper panel: Mean heatmaps of inter-subject dissimilarity within and between blind and sighted groups. Red: more similar; blue: less similar. Lower panel: Bar plot of inter-subject dissimilarity by group. Error bars: standard error of the mean (SEM) across subject pairs. *$p < .05$, **$p < .01$, ***$p < .001$, Bonferroni-corrected. (D–F) Multi-dimensional scaling of the models made by the blind and sighted subjects.

mean ± SD = 23 ± 1; 11 females) in the college to name and score the drawings. We separated the drawings and raters into three subsets, within which the drawings by the blind and the sighted were interleaved. We used two common subjects' drawings in every subset to calculate the intraclass correlation between raters. The drawings were presented randomly one by one on the screen. The raters typed the name of the object or skipped if they could not recognize it; and then the true name of the object was presented, the raters were asked to rate how similar the drawing was to the actual target object on a 7-point Likert scale. For the naming task, we did a domain × group repeated-measures ANOVA of the accuracy, i.e., the percentage of correct recognition for each item, each subject. For the rating task, the ICC between the three averaged group ratings was high (ICC = .86, 95% CI: .73–.93, based on a mean-rating, consistency, 2-way random effects model), indicating good reliability (Koo & Li, 2016). Therefore, we averaged the scores for each item across raters and did statistical tests.

In the inter-subject alignment measurement, as the technical limitations of objectively calculating pairwise similarity for open-contour line drawings exist, we recruited another independent sighted group of raters (120 raters, age: mean ± SD = 22 ± 2; 62 females) to perform a multi-arrangement task (Kriegeskorte & Mur, 2012) to subjectively evaluate inter-subject (dis)similarities for each item. During the task (Fig. 1C), the drawings by all subjects for one item were presented at the same time on the screen, and the raters were instructed to organize the drawings according to their shape similarity by dragging and dropping items with mouse. The items were separated into six subsets with each subset being arranged by 20 raters. Each rater only needed to finish one trial. Dissimilarity was measured by the on-screen Euclidean distance, which was scaled to have a root mean square of 1. We performed linear mixed-effect model analysis (random effects for comparison pair and item) to test the main effects and interaction effect (see the Supplementary material for the details about the models). We adjusted all multiple comparisons tests in the analyses with the Bonferroni method, unless we stated otherwise. Since there was no inter-item dissimilarity data, we could not visualize the representation spaces as in Experiment 1 and Experiment 2.

4.2. *Results*

All blind subjects and blindfolded sighted subjects were asked to draw objects using a raised-line drawing kit (Kennedy, 1997) that allows for tactile feedback of the drawing trace. We collected 484 drawings in total (215 from the blind, 269 from the sighted; see Table S4A). Similar to Experiment 2, an independent group of 15 sighted raters named and rated the drawings on a 7-point Likert scale. For the naming accuracy (Fig. 4A and Table S4B), there was a significant main effect of group, with the drawings by the sighted being significantly more intelligible than those by the blind [$F_{(1,26)} = 5.36$, $p = .03$], and a main effect of domain [$F_{(2, 52)} = 37.85$, $p = 7.15 \times 10^{-11}$], with tool drawings being the easiest to recognize (*ps* < .002) and large object drawings the hardest (*ps* < .007). Notably, the group × domain interaction was significant [$F_{(2, 52)} = 5.70$, $p = .006$]. Simple effect tests showed that the two groups only differed in drawing animals [$t_{(20.9)} = 3.02$, $p = .02$] and not tools or large objects (*ps* ≥ .54). The result of rating was consistent with the naming task (Fig. 4B and Table S4C), with significant group main effect [$F_{(1,26)} = 7.95$, $p = .009$, blind < sighted], domain effect [$F_{(2, 52)} = 22.32$, $p = 1 \times 10^{-7}$, tools > animals and large objects, *ps* < .0002], and interaction [$F_{(2, 52)} = 8.67$, $p = .0006$]. Again the sighted group was better than blind group only on the animal drawings [$t_{(16.2)} = 3.48$, $p = .009$] and not tools or large objects (*ps* > .21).

To see whether blindness affects the inter-subject alignment in object representation, we recruited another independent group of 120 sighted subjects (six subsets) using the multi-arrangement task (Kriegeskorte & Mur, 2012), where they were asked to organize the drawings according to their shape similarity for every item (Fig. 4C). We performed a linear mixed-effect model analysis and the likelihood-ratio test revealed that the group effect was significant [$\chi^2_{(2)} = 17.86$, $p = .0001$], with the between-group distance being larger than the within-group distance (*ps* < .006), and the within-sighted not significantly differing with the within-blind ($p = .13$). Domain effect was also significant [$\chi^2_{(2)} = 28.34$, $p = 7.03 \times 10^{-7}$; dissimilarity in animals > tools and large objects, $ps < 2.41 \times 10^{-5}$; tools *vs* large objects, $p = 1$]. The interaction was not significant though [$\chi^2_{(4)} = 7.68$, $p = .10$].

In summary, the lack of visual experience had the strongest impact on how well they drew animals, while it did not affect their drawings of tools or large objects significantly (although the sighted participants did not perform well on large objects, either). However, there was a consistent profile in that the blind and sighted participants showed similar within-group variations but greater deviations from each other across the three domains.

As in Experiment 2, we analyzed the geometry of tool drawings, but we used length/width ratio instead of absolute size, since the drawing size was not fixed as in the clay experiment (where we gave the same amount of Play-Doh). We found no significant differences ($p = .76$), possibly because drawing involves 2D conversion and perspective taking, which may affect the length/width measures.

5. Discussion

We used three shape knowledge production experiments in three modalities with sighted and congenitally blind subjects to elucidate the role of vision, touch, and language in shape representation. Two key findings were obtained. First, while the absence of visual experience (i.e., in blind group) led to production of language feature with an overall greater distance with shape across domains, explicit (nonverbal) shape productions revealed stronger difference in animals than in tools and large objects, suggesting that tactile experience of actual objects matters when vision is absent. Second, for tools with rich tactile/manipulation experiences, the blind group performed with similar quality as the sighted group, but with greater variation and systematically different geometric property (stubbier than the sighted), indicating that vision helps to align and proportion the shapes at least of tools.

Consistent with the literature that the representation of animals relies more on visual experience (Farah & McClelland, 1991), our study found that for animals more visual-only

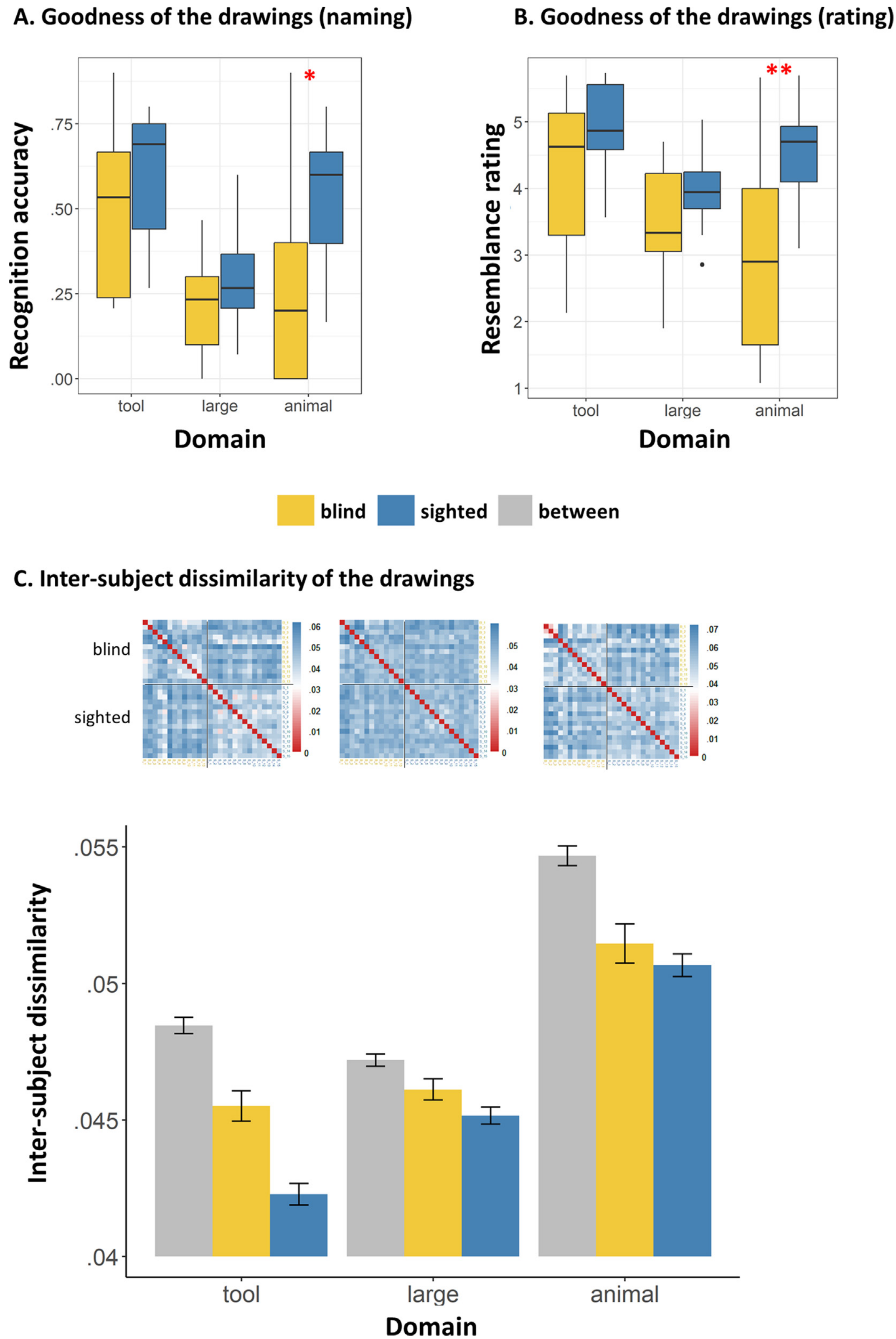

Fig. 4 – **Results of the drawing experiment. (A) Goodness evaluated by the naming task. An independent group of sighted raters named the drawings. The accuracy was the proportion of correct recognition. (B) Goodness evaluated by the resemblance rating task. The same raters scored the drawings' similarity to the target objects on a 7-point Likert scale. (C) Inter-subject alignment acquired through the multi-arrangement task performed by another independent group of sighted participants. Upper panel: Mean heatmaps of inter-subject dissimilarity within and between blind and sighted groups. Red: more similar; blue: less similar. Lower panel: Bar plot of inter-subject dissimilarity by group. Error bars: standard error of the mean (SEM) across subject pairs. *p < .05, **p < .01, ***p < .001, Bonferroni-corrected.**

features were generated in the verbal feature generation experiment, with no significant difference between the blind and sighted groups. While the features generated by the blind group showed overall greater distance with shape words in the embedding space, animal items did not exhibit stronger difference compared to artifacts in either the mean space or the individual variations. In contrast, in the two nonverbal explicit (3D and 2D) shape production tasks, animal shape representations were affected the most by the absence of visual experience relative to tools and large objects. As reasoned in the Introduction, although actual tactile experience with real animals is rare (except for pets), blind people may still construct animal shapes by learning from toys or by extracting learned tactile-based shape elements and combining them in accordance with language descriptions. The results showed that this is not fully the case. Although congenitally blind individuals could produce 3D and 2D forms for animals that they rarely have actual tactile experience of (e.g., butterfly, elephant), their produced 3D and 2D animal forms were not as good or as recognizable as those produced by the typically developed people. That is, language and primary tactile experience are not sufficient to generalize to form well-structured animal shapes. One might argue that the difference in animal representations between the two groups resulted from the complexity in modelling or drawing animals because animals have more irregular contours. This is unlikely because their performance was actually better in drawing animals than drawing large objects, which tend to have more simple geometric shape elements and were even harder or equally hard to be recognized in drawings and models compared to animals. Another argument could be that the subjective goodness assessment was done by sighted people, who might have a bias to the works that conform to the visual appearance. However, this is unlikely, either, because the tool drawings were evaluated equally good between the two groups by the same group of sighted raters (see discussion below).

By contrast, tool shapes were less affected by the absence of visual experience. The 3D clay models and 2D drawings produced by blind individuals were made with overall similar degree of recognizability and with high fidelity, suggesting that actual tactile experience of tools alone generates shape representations that are largely convergent with visual-tactile experience (e.g., see Reiner, 2008). However, there were also intriguing differences. Lack of visual experience led to lower agreement among model shapes, consistent with the recent findings that sensory deprivations increase idiosyncrasy (e.g., Sen et al., 2022). How does vision contribute to aligning shape representations? At least for objects larger than palm, tactile exploration derives their shape by obtaining parts in sequence and compiling them into whole "gestalt" representations, while vision tends to process whole objects in parallel. Such complementary information processes may enhance representation consistency and accuracy. Indeed, we observed a geometric difference between the tool models made by the blind and sighted groups, with the blind tended to underestimate the tool length. We speculate that this underestimation tendency without vision may stem from how vision calibrates the length estimation for irregular-shaped objects whose localization of center of mass misaligns with shape center (see Chan, 1995; Solomon & Turvey, 1988; Withagen & Michaels, 2004, 2005).

The results for the large nonmanipulable objects were puzzling. While little difference between the blind and sighted groups was observed for explicit shape production of these objects, which could be predicted by the similarly rich tactile interaction with these objects across the two groups, both groups performed rather poorly on these objects. One possibility is that the large size of these objects makes them difficult to reproduce as much smaller models for both groups. For the current purpose, we did not see large size act differently on people having or not sight, aligning with the previous brain imaging findings (see Bi et al., 2016 for reviews), but the conclusion requires caution of floor effects. The potential effects of size in shape representation remain to be further studied (see Konkle & Caramazza, 2013).

We reasoned that our production tasks are more explicit in reflecting shape representation than those implicit judgment ones used in the literature. However, it should be noted that each production task manifests the internal mental representation in one explicit manner and may be subject to properties of this particular output channel. The three experiments shared a similar input (object names) and varied critically in the output formats and contents (verbal semantic features, 3D shape, and 2D shape). As presented in the Introduction, past neuroimaging research has implicated complex interactions among object property, domain, and information modality in terms of neural activity patterns. The three behavioral experiments here yielded overall convergent pattern with the neuroimaging results in that 1) object properties mattered: Experiment 2 and 3 were more aligned with each other (both shape tasks) and less so with Experiment 1 (a semantic feature task), and 2) object domain mattered: in Experiment 2 and 3 the overall behavioral profiles for the blind and sighted groups were more similar for manmade items (with comparable goodness) than for animal items (with reduced goodness in the blind), although with behavioral experiments it is difficult to draw conclusions about the representational format of such similar productions. Note that the subtle yet reliable differences in the shape representation for tool items across groups promote further studies to reveal the underlying neural structures.

A few issues warrant further discussion. First, for Experiment 1, we measured implicit shape representation from the generated verbal features by calculating their embedding distance with shape terms. If the blind subjects produced and received language contents that are systematically different from the sighted, analyses using their own embedding space may result in different relational profiles with the shape features. We had the assumption that the blind and sighted have very similar language experiences (Bi, 2021; Kim et al., 2019), but this assumption requires scrutiny in the future investigations. Indeed, we do not have a good explanation for the group and domain interaction in terms of idiosyncrasy (significant difference in tools and large objects and less so in animals) in this particular experiment. Second, we have mainly focused on the results of the 3D modelling (Experiment 2) as it is the most natural form to render shape representation. For the drawing experiment, we did not want to overinterpret results, because converting 3D to 2D introduces more variance, such as perspective taking, and the blind group had little drawing experience. Nonetheless, we did observe an interesting pattern in the drawings of the two groups. The blind people tended to draw more often with matchstick-figure-like style (see examples in Fig. S3). An ad hoc analysis of chi-square test of independence confirmed that the proportion of these matchstick style drawings in the blind group was significantly higher than that in the sighted group [$\chi^2_{(1)} = 32.7$, $p = 1.10 \times 10^{-8}$]. The implications of such drawing pattern require discussions beyond the scope of the current study. Finally, our results have broader implications for neurorehabilitation of people with sensory impairments. We found that tactile experiences are crucial for forming internal shape representations when vision is absent, and that visual experiences help to align and calibrate object shapes even for those objects that are familiar to touch. These observations raise awareness of the potential group differences in the most salient object property, promote neuroimaging studies that investigate the neural representations of objects in this special population in finer details, encourage ways to enrich tactile inputs in experimental settings (e.g., with toy versions) for objects that are difficult to touch in real life, and to design products or equipment adapting to the shape representation properties in the blind populations.

To conclude, by comparing object shape production properties in congenitally blind and sighted individuals, we showed that even for such a classic supramodal property, its representation reflects the intricate orchestration of vision, touch and language. Without vision, tactile experience (with the help of language) may derive shape representation even for things that people have never touched for real (the case of many animals), but the representation is significantly impoverished. When tactile experience is fully available (the case for familiar tools), the shape representation is fully well-formed, yet vision deprivation led to greater idiosyncrasies and subtle geometry biases. These findings invite further understanding for the exact neural and computational mechanisms across different modalities in generating a coherent representation.

Funding

This work was supported by the National Science and Technology Innovation 2030 Major Program (2021ZD0204104 to Y.B.), National Natural Science Foundation of China (31925020, 82021004 to Y.B., 32071050 to X.Y.W., 32171052 to X.S.W.), Changjiang Scholar Professorship Award (T2016031 to Y.B.), the Fundamental Research Funds for the Central Universities.

CRediT authorship contribution statement

Shuang Tian: Data curation, Formal analysis, Investigation, Methodology, Project administration, Validation, Writing – original draft, Writing – review & editing. **Lingjuan Chen:** Data curation, Formal analysis, Investigation, Project administration, Validation, Writing – review & editing. **Xiaoying Wang:** Funding acquisition, Investigation, Methodology, Project administration, Supervision. **Guochao Li:** Investigation. **Ze Fu:** Investigation, Methodology. **Yufeng Ji:** Methodology, Software. **Jiahui Lu:** Investigation. **Xiaosha Wang:** Funding acquisition, Investigation, Methodology. **Shiguang Shan:** Methodology. **Yanchao Bi:** Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Open practices

The study in this article has earned Open Data and Open Material Badges for transparent practices. The data and materials used in this study are available at: https://osf.io/59wkf/.

Declaration of competing interest

The authors declare no competing interests.

Acknowledgments

We are grateful to Guangyao Zhang and Han-Wu-Shuang Bao for their help with statistical analyses.

No part of the study procedures or analyses was pre-registered prior to the research being conducted. We reported how we determined our sample size, all data exclusions (if any), all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cortex.2024.02.016.

The experimental data reported in this paper, including the complete list of verbal features with their coded types, the 3D clay models in gif format, and 2D drawings, as well as the 3D model similarity calculation code are available on Open Science Framework (https://osf.io/59wkf/).

REFERENCES

Amedi, A., Jacobson, G., Hendler, T., Malach, R., & Zohary, E. (2002). Convergence of visual and tactile shape processing in the human lateral occipital complex. *Cerebral Cortex, 12*(11), 1202–1212. https://doi.org/10.1093/cercor/12.11.1202

Amedi, A., Malach, R., Hendler, T., Peled, S., & Zohary, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. *Nature Neuroscience, 4*(3), 324–330. https://doi.org/10.1038/85201

Amedi, A., Raz, N., Azulay, H., Malach, R., & Zohary, E. (2010). Cortical activity during tactile exploration of objects in blind and sighted humans. *Restorative Neurology and Neuroscience, 28*(2), 143–156. https://doi.org/10.3233/RNN-2010-0503

Amedi, A., Stern, W. M., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., Hemond, C., Meijer, P., & Pascual-

Leone, A. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nature Neuroscience, 10*(6), 687–689. https://doi.org/10.1038/nn1912

Bainbridge, W. A., Pounder, Z., Eardley, A. F., & Baker, C. I. (2021). Quantifying aphantasia through drawing: Those without visual imagery show deficits in object but not spatial memory. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior, 135*, 159–172. https://doi.org/10.1016/j.cortex.2020.11.014

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1). https://doi.org/10.18637/jss.v067.i01

Bi, Y. (2021). Dual coding of knowledge in the human brain. *Trends in Cognitive Sciences, 25*(10), 883–895. https://doi.org/10.1016/j.tics.2021.07.006

Bi, Y., Wang, X., & Caramazza, A. (2016). Object domain and modality in the ventral visual pathway. *Trends in Cognitive Sciences, 20*(4), 282–290. https://doi.org/10.1016/j.tics.2016.02.002

Bola, Ł., Yang, H., Caramazza, A., & Bi, Y. (2022). Preference for animate domain sounds in the fusiform gyrus of blind individuals is modulated by shape–action mapping. *Cerebral Cortex*, 1–21. https://doi.org/10.1093/cercor/bhab524

Breedlove, J. L., St-Yves, G., Olman, C. A., & Naselaris, T. (2020). Generative feedback explains distinct brain activity codes for seen and mental images. *Current Biology, 30*(12), 2211–2224.e6. https://doi.org/10.1016/j.cub.2020.04.014

Chan, T.-C. (1995). The effect of density and diameter on haptic perception of rod length. *Perception & Psychophysics, 57*(6), 778–786. https://doi.org/10.3758/BF03206793

Chaouch, M., & Verroust-Blondet, A. (2009). Alignment of 3D models. *Graphical Models, 71*(2), 63–76. https://doi.org/10.1016/j.gmod.2008.12.006

Chen, J., Snow, J. C., Culham, J. C., & Goodale, M. A. (2017). What role does “elongation” play in “tool-specific” activation and connectivity in the dorsal and ventral visual streams? *Cerebral Cortex, 28*(4), 1117–1131. https://doi.org/10.1093/cercor/bhx017

Dalal, N., Triggs, B., Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection to cite this version: Histograms of oriented gradients for human detection. In *IEEE computer society conference on computer vision and pattern recognition* (pp. 886–893). http://lear.inrialpes.fr.

de Leeuw, J. (2016). Multidimensional scaling in R: SMACOF. *Vignettes, 2007*, 1–5. https://cran.r-project.org/web/packages/smacof/.

Devlin, J., Chang, M.-W., Lee, K., Google, K. T., & Language, A. I. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Naacl-Hlt 2019, Mlm* (pp. 4171–4186).

Erdogan, G., Chen, Q., Garcea, F. E., Mahon, B. Z., & Jacobs, R. A. (2016). Multisensory part-based representations of objects in human lateral occipital cortex. *Journal of Cognitive Neuroscience, 28*(6), 869–881. https://doi.org/10.1162/jocn_a_00937

Erdogan, G., Yildirim, I., & Jacobs, R. A. (2015). From sensory signals to modality-independent conceptual representations: A probabilistic language of thought approach. *PLoS Computational Biology, 11*(11), 1–32. https://doi.org/10.1371/journal.pcbi.1004610

Fabbri, S., Stubbs, K. M., Cusack, R., & Culham, J. C. (2016). Disentangling representations of object and grasp properties in the human brain. *Journal of Neuroscience, 36*(29), 7648–7662. https://doi.org/10.1523/JNEUROSCI.0313-16.2016

Fan, J. E., Bainbridge, W. A., Chamberlain, R., & Wammes, J. D. (2023). Drawing as a versatile cognitive tool. *Nature Reviews Psychology*. https://doi.org/10.1038/s44159-023-00212-w

Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General, 120*(4), 339–357. https://doi.org/10.1037/0096-3445.120.4.339

Favila, S. E., Kuhl, B. A., & Winawer, J. (2022). Perception and memory have distinct spatial tuning properties in human visual cortex. *Nature Communications, 13*(1), 5864. https://doi.org/10.1038/s41467-022-33161-8

Frossard, J., & Renaud, O. (2021). Permutation tests for regression, anova, and comparison of signals: The permuco package. *Journal of Statistical Software, 99*(15). https://doi.org/10.18637/JSS.V099.I15

Grand, G., Blank, I. A., Pereira, F., & Fedorenko, E. (2022). Semantic projection recovers rich human knowledge of multiple object features from word embeddings. *Nature Human Behaviour, 6*(7), 975–987. https://doi.org/10.1038/s41562-022-01316-8

Handjaras, G., Ricciardi, E., Leo, A., Lenci, A., Cecchetti, L., Cosottini, M., Marotta, G., & Pietrini, P. (2016). How concepts are encoded in the human brain: A modality independent, category-based cortical organization of semantic knowledge. *NeuroImage, 135*(May), 232–242. https://doi.org/10.1016/j.neuroimage.2016.04.063

Handjaras, G., Leo, A., Cecchetti, L., Papale, P., Lenci, A., Marotta, G., … Ricciardi, E. (2017). Modality-independent encoding of individual concepts in the left parietal cortex. *Neuropsychologia, 105*, 39–49. https://doi.org/10.1016/j.neuropsychologia.2017.05.001

Heller, M. A., Brackett, D. D., Scroggs, E., Steffen, H., Heatherly, K., & Salik, S. (2002). Tangible pictures: Viewpoint effects and linear perspective in visually impaired people. *Perception, 31*(6), 747–769. https://doi.org/10.1068/p3253

Hervé, M. (2020). *Package ‘RVAideMemoire’*. See https://CRAN.R-Project.Org/Package=RVAideMemoire,0–9.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. Technical Report Number 019 *Biometrical Journal, 50*(3), 346–363 http://www.stat.uni-muenchen.de.

Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., & Mikolov, T. (2016). *FastText.zip: Compressing text classification models* (pp. 1–13). http://arxiv.org/abs/1612.03651.

Kassambara, A. (2016). *Factoextra: Extract and visualize the results of multivariate data analyses.* R Package Version, 1.

Kassambara, A. (2021). *rstatix: Pipe-friendly framework for basic statistical tests.* R package version 0.7. 0.

Kennedy, K. M. (1997). How the blind draw. *Scientific American, 276*(1), 76–81. https://doi.org/10.1038/scientificamerican0197-76

Kennedy, J. M. (2003). Drawings from Gaia, a blind girl. *Perception, 32*(3), 321–340. https://doi.org/10.1068/p3436

Kennedy, J. M., & Juricevic, I. (2003). Haptics and projection: Drawings by Tracy, a blind adult. *Perception, 32*(9), 1059–1071. https://doi.org/10.1068/p3425

Kim, J. S., Elli, G. V., & Bedny, M. (2019). Knowledge of animal appearance among sighted and blind adults. *Proceedings of the National Academy of Sciences.* , Article 201900952. https://doi.org/10.1073/pnas.1900952116

Konkle, T., & Caramazza, A. (2013). Tripartite organization of the ventral stream by animacy and object size. *Journal of Neuroscience, 33*(25), 10235–10242. https://doi.org/10.1523/JNEUROSCI.0983-13.2013

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kriegeskorte, N., & Mur, M. (2012). Inverse MDS: Inferring dissimilarity structure from multiple item arrangements.

Frontiers in Psychology, 3(Jul), 1–13. https://doi.org/10.3389/fpsyg.2012.00245

Lee Masson, H., Bulthé, J., Op De Beeck, H. P., & Wallraven, C. (2016). Visual and haptic shape processing in the human brain: Unisensory processing, multisensory convergence, and top-down influences. *Cerebral Cortex, 26*(8), 3402–3412. https://doi.org/10.1093/cercor/bhv170

Liu, Z., Wang, Z., Ma, C., Zhang, C., Mitani, J., & Fukui, Y. (2010). Shape alignment and shape orientation analysis-based 3D shape retrieval system. *Multimedia Systems, 16*(4–5), 319–333. https://doi.org/10.1007/s00530-010-0193-x

Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M., & Caramazza, A. (2009). Category-specific organization in the human brain does not require visual experience. *Neuron, 63*(3), 397–405. https://doi.org/10.1016/j.neuron.2009.07.012

Mahon, B. Z., & Caramazza, A. (2009). Concepts and categories: A cognitive neuropsychological perspective. *Annual Review of Psychology, 60*(1), 27–51. https://doi.org/10.1146/annurev.psych.60.110707.163532

Mahon, B. Z., & Caramazza, A. (2011). What drives the organization of object knowledge in the brain? *Trends in Cognitive Sciences, 15*(3), 97–103. https://doi.org/10.1016/j.tics.2011.01.004

Mattioni, S., Rezk, M., Battal, C., Bottini, R., Cuculiza Mendoza, K. E., Oosterhof, N. N., & Collignon, O. (2020). Categorical representation from sound and sight in the ventral occipito-temporal cortex of sighted and blind. *ELife, 9*. https://doi.org/10.7554/eLife.50732

Paquet, E., Rioux, M., Murching, A., Naveen, T., & Tabatabai, A. (2000). Description of shape information for 2-D and 3-D objects. *Signal Processing: Image Communication, 16*(1), 103–122. https://doi.org/10.1016/S0923-5965(00)00020-5

Peelen, M. V., He, C., Han, Z., Caramazza, A., & Bi, Y. (2014). Nonvisual and visual object shape representations in occipitotemporal cortex: Evidence from congenitally blind and sighted adults. *Journal of Neuroscience, 34*(1), 163–170. https://doi.org/10.1523/JNEUROSCI.1114-13.2014

Pietrini, P., Furey, M. L., Ricciardi, E., Gobbini, M. I., Wu, W.-H. C., Cohen, L., Guazzelli, M., & Haxby, J. V. (2004). Beyond sensory images: Object-based representation in the human ventral pathway. *Proceedings of the National Academy of Sciences, 101*(15), 5658–5663. https://doi.org/10.1073/pnas.0400707101

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 3 https://www.r-project.org/.

Reiner, M. (2008). Seeing through touch: The role of haptic information in visualization. In *Visualization: Theory and practice in science education* (pp. 73–84). Springer.

Sen, S., Khalsa, N. N., Tong, N., Ovadia-Caro, S., Wang, X., Bi, Y., & Striem-Amit, E. (2022). The role of visual experience in individual differences of brain connectivity. *Journal of Neuroscience, 42*(25), 5070–5084. https://doi.org/10.1523/JNEUROSCI.1700-21.2022

Solomon, H. Y., & Turvey, M. T. (1988). Haptically perceiving the distances reachable with hand-held objects. *Journal of Experimental Psychology: Human Perception and Performance, 14*(3), 404–427. https://doi.org/10.1037/0096-1523.14.3.404

Stilla, R., & Sathian, K. (2008). Selective visuo-haptic processing of shape and texture. *Human Brain Mapping, 29*(10), 1123–1138. https://doi.org/10.1002/hbm.20456

Tian, S., Chen, Y., Fu, Z., Wang, X., & Bi, Y. (2023). Simple shape feature computation across modalities: Convergence and divergence between the ventral and dorsal visual streams. *Cerebral Cortex, 33*(15), 9280–9290. https://doi.org/10.1093/cercor/bhad200

Vranić, D. V., Saupe, D., & Richter, J. (2001). Tools for 3D-object retrieval: Karhunen-Loeve transform and spherical harmonics. In *2001 IEEE Fourth Workshop on Multimedia Signal Processing* (pp. 293–298). https://doi.org/10.1109/mmsp.2001.962749

Wang, X., Men, W., Gao, J., Caramazza, A., & Bi, Y. (2020). Two forms of knowledge representations in the human brain. *Neuron, 107*(2), 1–11. https://doi.org/10.1016/j.neuron.2020.04.010

Wang, X., Peelen, M. V. V., Han, Z., He, C., Caramazza, A., & Bi, Y. (2015). How visual is the visual cortex? Comparing connectional and functional fingerprints between congenitally blind and sighted individuals. *Journal of Neuroscience, 35*(36), 12545–12559. https://doi.org/10.1523/JNEUROSCI.3914-14.2015

Wang, X., Zhuang, T., Shen, J., & Bi, Y. (2018). Disentangling representations of shape and action components in the tool network. *Neuropsychologia, 117*(May), 199–210. https://doi.org/10.1016/j.neuropsychologia.2018.05.026

Withagen, R., & Michaels, C. F. (2004). Transfer of calibration in length perception by dynamic touch. *Perception & Psychophysics, 66*(8), 1282–1292. https://doi.org/10.3758/BF03194998

Withagen, R., & Michaels, C. F. (2005). The role of feedback information for calibration and attunement in perceiving length by dynamic touch. *Journal of Experimental Psychology: Human Perception and Performance, 31*(6), 1379–1390. https://doi.org/10.1037/0096-1523.31.6.1379

Xu, Y., Vignali, L., Sigismondi, F., Crepaldi, D., Bottini, R., & Collignon, O. (2023). Similar object shape representation encoded in the inferolateral occipitotemporal cortex of sighted and early blind people. *PLoS Biology, 21*(7), Article e3001930. https://doi.org/10.1371/journal.pbio.3001930