scientific reports

Check for updates

OPEN

The Oomplet dataset toolkit as a flexible and extensible system for large-scale, multi-category image generation

John P. Kasarda[1], Angela Zhang[2], Hua Tong[2], Yuan Tan[2], Ruizi Wang[2], Timothy Verstynen[1,3]✉ & Michael J. Tarr[1,3]✉

The modern study of perceptual learning across humans, non-human animals, and artificial agents requires large-scale datasets with flexible, customizable, and controllable features for distinguishing between categories. To support this research, we developed the Oomplet Dataset Toolkit (ODT), an open-source, publicly available toolbox capable of generating 9.1 million unique visual stimuli across ten feature dimensions. Each stimulus is a cartoon-like humanoid character, termed an "Oomplet," designed to be an instance within clearly defined visual categories that are engaging and suitable for use with diverse groups, including children. Experiments show that adults can use four to five of the ten dimensions as single classification criteria in simple perceptual discrimination tasks, underscoring the toolkit's flexibility. With the ODT, researchers can dynamically generate large, novel stimulus sets to study perceptual learning across biological and artificial contexts.

The use of computer-generated stimuli in psychometric studies of behavior has a rich history. In perhaps the first study to use "computer-graphics psychophysics"[1], Shepard and Metzler used novel 3D objects "generated by digital computer and associated graphical output" to study mental simulations of physical actions (Fig. 1A)[2]. Since then, the art of visual stimulus creation has continued to be driven by advances in computer graphics. Salient examples include 3D "wire-frame" objects [3] (Fig. 1B) and 3D blob-like "amoebae" objects [4] (Fig. 1C). While these early studies relied on workstation-class computing and graphics, the advent of consumer level computer graphics tools running on desktop computers enabled a new generation of visual stimuli comprised of increasingly more complex and realistic novel objects. Helping drive this trend, beginning in the 1990's, one of the collaborators on this project developed multiple, publicly accessible [5], complex visual stimulus datasets [6,7] – two notable examples of this work being the "Greebles" [8] (Fig. 1D) and the "Fribbles" [9] (Fig. 1E), both of which have been used in 100's of psychophysical, cognitive science, cognitive neuroscience, and clinical studies. A recent and non-exhaustive list of other examples from the field includes "smoothies, spikies, and cubies" [10] (Fig. 1F), "Ziggerins" [11] (Fig. 1G), "digital embryos" [12] (Fig. 1H), the NOUN Database [13] (Fig. 1I), "Widgets" [14] (Fig. 1J), and "Sheinbugs" [15] (Fig. 1K), many of which emerged from collaborations within the Perceptual Expertise Network (PEN) [16].

Many of these visual datasets were created using a compositional approach in which individual parts from a dictionary were sampled and manually combined in different configurations to form complex objects [6,8,9,17]. Critically, these datasets were created by hand, relying on 3D modeling skills rather than explicit algorithms [7–9]. In contrast, a variety of other datasets were generated parametrically using varying values within mathematical functions to deform 3D shapes, define parts, and specify attachment points [3,4,10,12,18]. One characteristic across almost all of these datasets has been the relatively low number of available stimuli and/or stimulus categories. While datasets with a broad dictionary of discrete parts potentially allowed for thousands of distinct novel stimuli and hundreds of categories, the selection of parts, placement in different configurations, and category boundaries is laborious, making large-scale stimulus generation *ad hoc* and time-consuming [7,9]. Consequently, while the shape and configuration differences between parts enable the possibility of many different well-defined visual categories, the actual number of available categories is quite small – on the order of 10-20 at best. In contrast, visual datasets created through parametric variations allows for nearly an infinite number of different individual stimuli, but are often less suited to being organized into a large number of naturally-

[1]Department of Psychology, Carnegie Mellon University, Pittsburgh 15213, USA. [2]Entertainment Technology Center, Carnegie Mellon University, Pittsburgh 15213, USA. [3]Neuroscience Institute, Carnegie Mellon University, Pittsburgh 15213, USA. ✉email: timothyv@andrew.cmu.edu; michaeltarr@cmu.edu

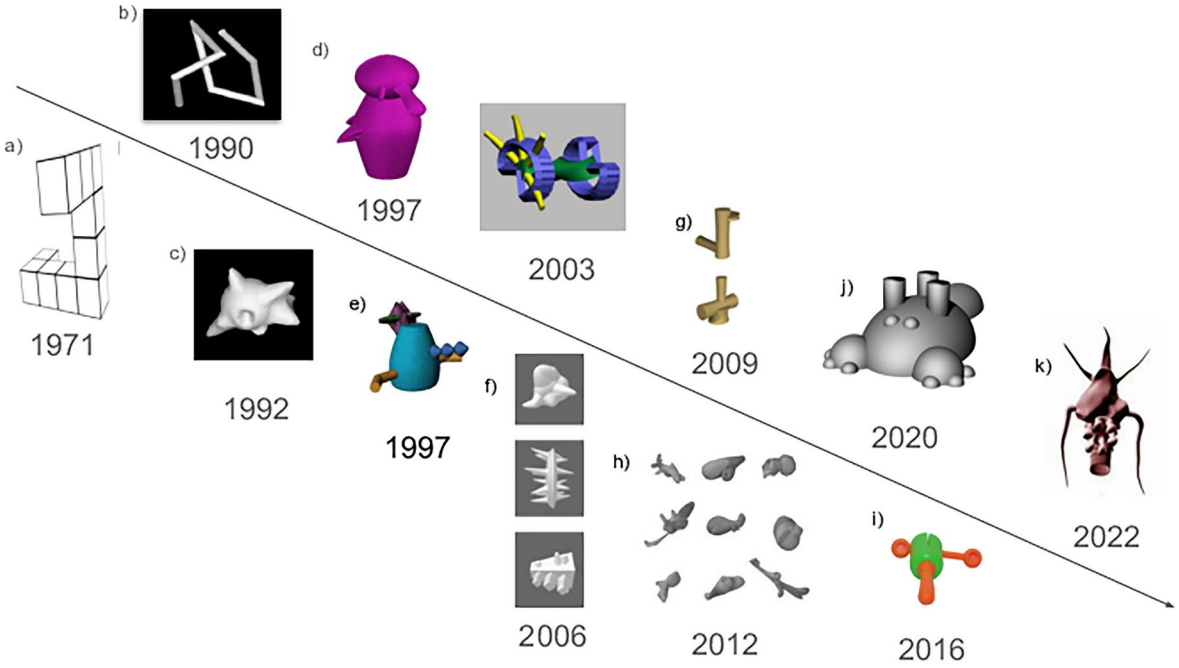

Fig. 1. A brief history of computer generated stimuli for psychometric studies. Timeline not to scale. a) Shepard, R., et al. (1971); b) Poggio, T., et al. (1990); c) Bülthoff, H., et al. (1992); d) Gauthier, I., et al. (1997); e) Williams, P. (1997); f) Op de Beeck, H. P., et al. (2006); g) Wong, A. C.-N., et al. (2009); h) Hegdé, J., et al. (2012); i) Horst, J. S., et al. (2016); j) Lebaz, S., et al. (2020); k) Jones, T., et al. (2020).

defined visual categories [3,4,10]. That is, the shape and configural variations across different parameter values are metric, meaning that visual categories, while nominally definable, are not perceptually salient or aligned with how humans typically infer categorical boundaries [19].

Until recently, limitations in dataset size and variety did not hinder most human behavioral and neuroimaging experiments, which typically have relied on using small numbers of categories and stimuli. However, with the rise of artificial intelligence and the availability of large-scale datasets [20,21], the field has shifted toward using larger, more diverse datasets [22]. This trend has enabled larger datasets [23–25] and important new findings [26–29], but these datasets often rely on internet-sourced images, mainly representing common objects and natural categories [20,21]. Such datasets are not ideal for studying perceptual categorization, as participants' prior experience with these categories is unknown and uncontrolled [8]. Furthermore, many visual processes, such as developing invariance to variations within categories, are best studied using novel stimuli to avoid memory-based strategies that can obscure the learning process [30].

To address these limitations, we developed the Oomplet Dataset Toolkit (ODT), a tool designed to generate complex and novel stimuli by precisely controlling each component's attributes and spatial arrangements. Using an extensive parts dictionary and contrastive part attributes, the ODT enables composing novel characters from distinct anthropomorphic features (e.g., arms, eyes, colors). Rather than allowing all components to occupy any position on the character, each component is restricted to a single role in each generated character image. As such, this reflects a form of compositional logic for character assembly[31]. At the same time, as implemented, compositionality in the ODT is limited to recombining which parts and attributes appear with one another. In contrast, human language (and some programming languages) often include more complex compositional structures (e.g., order or higher-order meaning). While not currently present in the ODT, relatively straightforward modifications (e.g., expanding the range of valid spatial positions for each part type) would enable a deeper compositional logic. Furthermore, the ODT quantitatively specifies spatial arrangements, allowing users to systematically alter parts within designated roles. Unlike previous compositional datasets such as Greebles [8] or Fribbles [9], which relied on manually positioned parts, the ODT's controlled assembly requires less manual adjustment for natural-looking designs.

Novelty is a key characteristic of the stimuli created by the ODT. Although observers might use general knowledge of body plans and face structures to interpret each component, the specific attributes that define individual ODT instances and categories are entirely unique. As a result, prior experience does not assist participants in learning some examples or categories faster or more accurately than others. In contrast, datasets of familiar natural objects – such as faces or everyday objects – typically have category boundaries that are intuitively recognizable or already have been learned [19]. Furthermore, individual experience with specific familiar categories impacts how they are processed and perceived [32,33]. At the same time, the generic configuration of stimuli generated by the ODT is not entirely novel in that they follow a humanoid body plan and facial configuration. As such, Oomplets are not ideal stimuli for studies of face or body processing [34].

The ODT was created as a stand-alone component of an interactive virtual environment designed to look at the dynamics of cooperative learning, where one task involves learning complex perceptual discriminations. Within this context, our objectives in creating the ODT were as follows: 1) enable the generation of a millions of individual stimuli and a large number of categories; 2) enable the use of a large dictionary of reusable parts defined by a wide variety of visual dimensions (e.g., color, shape, orientation, spacing, etc.); 3) enable visually-salient conjunctions and disjunctions of parts so as to create well-defined categories and category hierarchies; 4) build a toolkit that is user controllable to enable automatic generation of stimuli, but with fine-grained user control over parts, part attributes, and categories; 5) build a toolkit that requires only standard end-user skills (e.g., no programming knowledge or artistic skills), but that is extensible for users with those skills. The ODT is unique in realizing these objectives, providing a powerful stimulus generation toolkit that allows users to create a large number of visually-defined natural categories with potentially hundreds of thousands of hierarchically-nested, individual exemplars per category. In this way the ODT has potential applications in the psychological, neuroscientific, and artificial intelligence domains.

Methods

The ODT is a user-friendly and customizable python-based pipeline for generating large sets of unique stimuli, "Oomplets," and sorting them into hierarchically organized categories based on user-specified classification dimensions applied to the Oomplets' visual features. The pipeline consists of two python scripts (`generate.py` and `categorize.py`) and 148 component images that are combined to create 9.1 million unique visual stimuli (Fig. 2). These scripts, component images, and other relevant files are available in a publicly accessible repository (https://doi.org/10.1184/R1/25813726.v1).

Oomplet generation

Components

The components consist of images of various types of body parts or features to be used as references in creating individual Oomplets – humanoid candy stimuli. These images are stored as *.png* files in the subdirectories of the repository's "Components" directory. Each individual Oomplet stimulus image is made up of instances from seven classes of component images (Fig. 2). Because some components provide more than one attribute, a total of ten different attributes are recorded in a JSON formatted identification (ID) file associated with each generated Oomplet.

Generate

The `generate.py` script consists of Python code that creates an Oomplet by selecting one file from each of the seven component directories and compiling these components into a complete Oomplet. To accomplish this, `generate.py` employs OpenCV's[35] image processing functionality to visually parse the components and re-draw them jointly onto a common image depicting the newly created Oomplet. When invoked, `generate.py` is passed a number of required and optional arguments that allow user control of customization, computational processing, and output location. Full documentation of the script arguments and their functions is available in the repository.

Each Oomplet is defined by the user along 10 attributes nested within the 7 classes of components. To create a unique individual Oomplet, the generate process selects a value for each attribute, where there are 2-4 possible

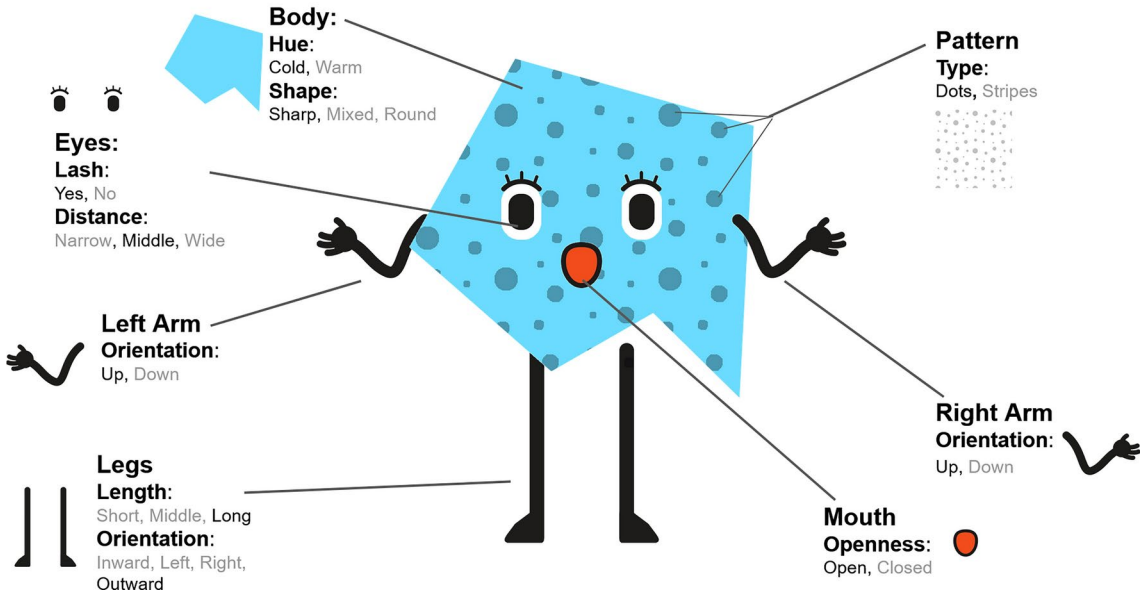

Fig. 2. An example Oomplet with each component and attribute.

values for each attribute that have been randomly ordered. As mentioned, `generate.py` captures these values and writes them into the associated JSON ID file. These ID files are what allows the pipeline to then sort the Oomplets into distinct visual categories using `categorize.py`.

Categorize

The `categorize.py` script consists of Python code that categorizes each Oomplet through a set of user-defined attribute criteria. The user specifies which attributes (a minimum of 1 and up to all 10 attributes) will be used to determine category membership. For each attribute, the user specifies the value of that attribute that helps define the category, where the complete category definition is the intersection of all 10 attribute values. As illustrated in Fig. 3, for each Oomplet that satisfies the criteria, `categorize.py` makes a copy of that Oomplet file and places it in automatically created output directories corresponding to Oomplets that match the criteria ("Match_[TIMESTAMP]") and those that do not ("NoMatch_[TIMESTAMP]"). When invoked, `categorize.py` is passed a number of required and optional arguments that allow user control of input location, categorization criteria, and other customizations.

Because categorization is based on a concatenation of values for each attribute, categorical boundaries can be along a single attribute or the intersection of many attributes. Additionally, a hierarchy may be created by running `categorize.py` multiple times (i.e., once to categorize all Oomplets with a common set of attribute values and then a second time to further sort Oomplets in one of the first sort categories based on a new set of attribute values).

Example

As a snapshot of the whole process, let us suppose that a stimulus is compiled by the `generate.py` script using the components `<mouth,open,1.png>` as reference. This stimulus would be recorded to have the attribute "open" for "mouth openness" in its ID file. The `categorize.py` script, when specified to look for images with closed mouths, would put this stimulus into the "NoMatch" sorting directory.

Usage notes

We provide detailed instructions to ensure bug-free usage of the Bit-or-Sweet pipeline. Start by completing the following steps to complete the initial setup of the stimulus generator.

Installation

1. Clone the repository locally

   % python —m venv venv

2. Set up a Python virtual environment in the root directory

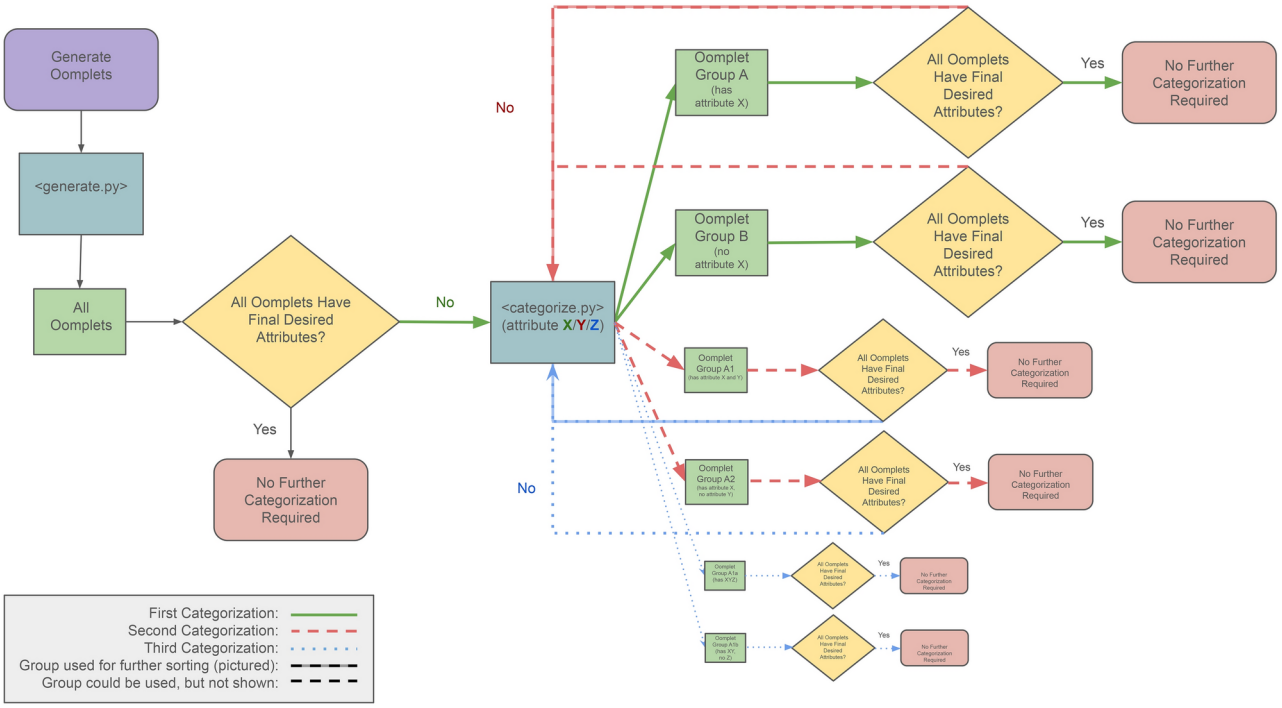

Fig. 3. The categorization pipeline. Dashed lines represent potential paths while solid lines represent executed paths for one example run of the pipeline.

MacOS or Unix:
% source venv/bin/activate
Windows:
% venv\Scripts\activate

4. Install all the requirements

% pip install —r requirements.txt

Implementation
Following installation, the `generate.py` script can be run from the command line. This is where users can specify any of the various options available to make their unique set of stimuli.

% python generate.py [—h] [—n N] [—p] [—c C] [—v] [—k] [—s S]

Options:

—h, ——help show this help message and exit
—n N number of candies to generate N (def: all combinations)
—p multiprocessing flag (def: off)
—c C max number of processes to spawn if multiprocessing (def: 4)
—v verbose (def: off)
—k keep existing files in output folder (def: off)
—s S seed value for randomly generated candies (def: 0)

The image and meta file output of this script will be located in the OompletToolkit/Output/Oomplets/ directory. Now, the `categorize.py` script may be run from the command line.

% python categorize.py [—h] [—d D] [—i I] [—k] [—a]

Options:

—h, ——help show this help message and exit
—d, ——def define your 'bitter' images (required)
—i, ——input name of the directory from which Oomplets will be sorted
—k keep existing files in output folders (def: off)
—a, ——any flags Oomplets with ANY of defining attributes as Match (def: off)

The `categorize.py` script was made to be easily customized. The `-d` option allows users to choose any number of non-contradicting attribute values to define their Match and NoMatch Oomplet groups. Attribute value specifications must be typed in the terminal exactly as shown in the list below.

'color_cool', 'color_warm',
'shape_sharp', 'shape_mixed', 'shape_round',
'lash_yes', 'lash_no',
'wide_eyes', 'middle_eyes', 'narrow_eyes',
'short_legs', 'middle_legs', 'long_legs',
'feet_left', 'feet_right', 'feet_in', 'feet_out',
'open_mouth', 'closed_mouth',
'dots_pattern', 'stripes_pattern',
'right_arm_down', 'right_arm_up', 'left_arm_down', 'left_arm_up'

Example
This section will show each step a user would take in order to generate a set of 200 images, and sort them based on their pattern and eye lashes, using a MacOS computer. First, the user needs to set up their virtual python environment.

% python —m venv venv
% source venv/bin/activate
% pip install —r requirements.txt

Next, the user must navigate to the `Scripts/ProcessingScripts` directory. From here, they will run the generation script using this command:

% python generate.py —n 200

The user has now created 200 unique images in the `Output/Oomplets` directory. Now, they must navigate to the `Scripts/AnalysisScripts` directory, where the categorization script is located. To sort the images based on their desired attributes, the user must use this command:

% python categorize.py —d stripes_pattern lash_no

Once the script has finished running, the user will now have two new directories. Each image that has a striped pattern, *and* eye lashes will be located in the `Output/Match` directory. All images that do not meet this requirement will be located in the `Output/NoMatch` directory.

Technical validation

In order to evaluate the perceptual discriminability of the different Oomplet attributes, we conducted a series of online studies using a forced choice discrimination task. We chose the eight most relevant attributes that can be used as binary classification boundaries and tested each attribute individually. In cases where attributes had more than two possible values (e.g., shape can be 'sharp', 'mixed', or 'round'), we only used the two most extreme values as the classification features (e.g., 'sharp' and 'round'). Each experiment used its own set of roughly 40,000 unique Oomplets.

All study procedures were approved by the Carnegie Mellon University Institutional Review Board and informed consent was obtained prior to each participant starting the study. All experiments were performed in accordance with relevant guidelines and regulations.

Participants

Studies were hosted on Connect[36] , CloudResearch's online crowd-sourcing platform. We recruited 50 participants for each study. Participants were excluded from the final analysis if their responses were improperly submitted to the cloud server or if they responded to fewer than 50 trials. The final sample sizes per condition were: Shape (N = 50), Pattern (N = 48), Mouth Openness (N = 49), Leg Length (N = 50), Eye Lash (N = 43), Eye Separation (N = 50), Hue (N = 47), and Arm Orientation (N = 48). Individuals who reported being colorblind were excluded from recruitment. We did not collect or restrict recruitment along any demographic category.

Task

We built the eight single-task studies using Gorilla's Experiment Builder (Task Builder 2)[37], with each task reflecting a single attribute for the classification boundary. In the task, participants were presented with 300 Oomplets (presented to participants as "candies" in this study), one at a time, and were asked whether the Oomplet is Bitter ("f" key) or Sweet ("j" key). Trials were counterbalanced, assuring that 150 images of each type were always presented. The terms "bitter" and "sweet" were chosen to avoid bias towards any of the humanoid characteristics; the bitter and sweet sets were created using Match and NoMatch criteria in the set creation with `categorize.py`.

Each trial consisted of three distinct phases (see Fig. 4). The trial started with a *fixation* phase, where the participant was presented with a centrally presented cross to bring their attention to the middle of the screen. This phase lasted 200ms, after which the cross was removed. After 100ms, the stimulus was presented (*stimulation* phase) with the words "Bitter" (left) and "Sweet" (right) presented on either side of the Oomplet, along with the keyboard response associated with each choice. Participants were given 2000ms to respond. Responses occurring after 2000ms were not recorded. Key presses were also not recorded for the first 250ms following stimulus onset in order to avoid false start responses. Finally, during the *feedback* phase, participants were informed via icons as to whether their response was correct or incorrect. Importantly, participants were not given explicit instructions as to what attributes defined the two categories and had to simply rely on this feedback to learn the relevant category boundaries.

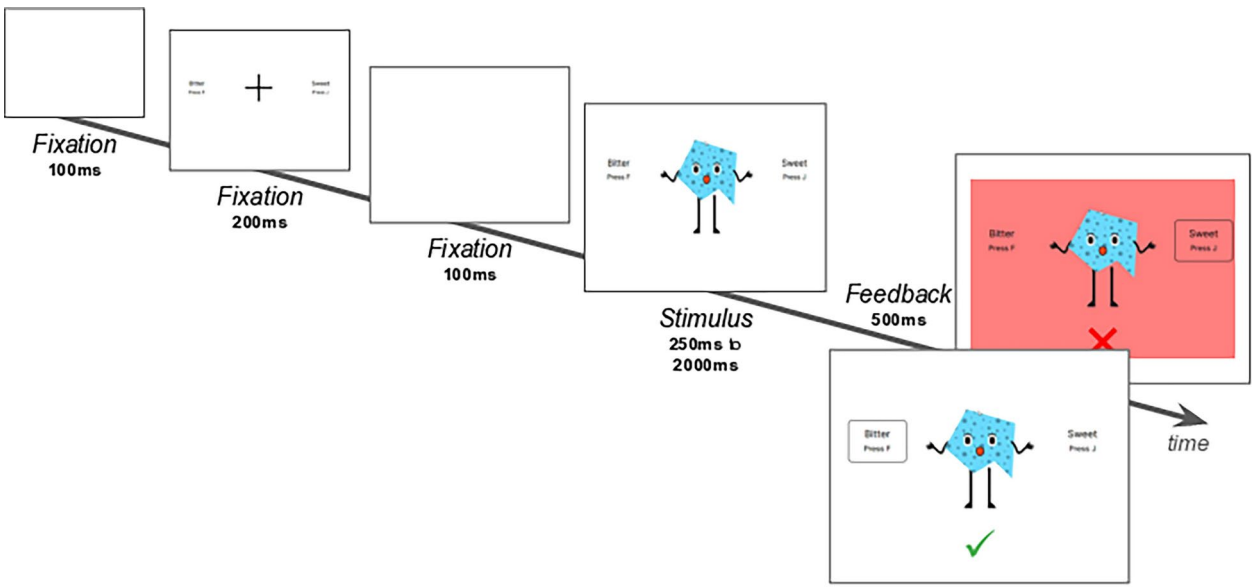

Fig. 4. The Stimulus display was shown to participants for at least 250 ms, and up to 2000ms. This display would transition early to the Feedback display when participants selected a response.

Analysis

To visualize how well each attribute could be detected as a classification dimension, we calculated two signal detection measures[38]. First, we estimated the d' for each participant as $d' = \Phi(hits) - \Phi(fa)$, where *hits* represents the true classification rate, and *fa* reflects the false alarm rate for incorrectly classifying a stimulus as sweet or bitter. The d' measure reflects the signal-to-noise ratio of the discrimination as standard deviations away from the noise distribution. The distribution of d' measures, across participants, was evaluated independently for each task. When participants had a perfect classification rate, we capped the d' value at 5.

In addition, we plotted the receiver operating characteristic (ROC) curve across tasks. This presents the joint distribution of *fa* and *hits* rates, and allows for visualizing when inter-subject responses vary along d' (reflecting consistent varying thresholds applied to the same signal-to-noise ratio) or criterion (reflecting varying signal-to-noise ratios along the same selection threshold).

In order to avoid any potential biases from stimulus characteristics (e.g., implicit assumptions on color to bitter/sweet mapping) in the resulting choices, we counterbalanced the bitter/sweet mapping across participants. Half of the participants would get one mapping and the remaining half the other. Task assignments were random without replacement, targeting 25 participants per group.

Results

Data records

We used the process described in the Methods section to generate roughly all 9 million possible unique Oomplets. The Oomplets and the code used to generate them are organized according to the TIER Protocol 4.0 directory architecture[39] (Fig. 5). Oomplet images were then stored in *PNG* format, with transparent backgrounds. Additional scripts used to help with building the validation study are included in the *Scripts* directory.

Perceptual sensitivity analysis

Figure 6 shows the distribution of d' scores, across participants, for each attribute tested. Attributes are sorted from lowest to highest average d' and errorbars reflect the 95% confidence intervals. We see that the eye distance, leg length, arm orientation, and eye lash attributes are unreliable dimensions for classification, reflected by the fact that the confidence intervals overlap with zero. The mouth openness and texture pattern show a modest discriminability, with mean d' values of 0.873 and 1.046 respectively. However, we see that this comes with a high degree of variability across participants, with a somewhat bimodal distribution of individual scores. One mode of participants sits around zero, indicating lack of discriminability. The other mode has very high d' values ranging from 1 to almost 4. Finally, body shape and color had the strongest discriminability, with mean values of 2.040 and 2.301 respectively. For both of these attributes, the spread of individual d' values was fairly broad, with some participants hovering near zero and 2 participants maxing out at d' values of 5 (reflecting perfect performance).

As an additional evaluation of participant performance, we also plotted the hit vs. false alarm rate for each participant in each task as an ROC curve. This allows for assessing the sensitivity of discrimination and response bias of each participant more clearly. We see two general patterns in these ROC plots, reflecting the general split between attributes with strong discriminability and those with weak perceptual discriminability. For attributes that had overall low d' scores, we see distributions of hit vs. false alarm rates centered near the unity diagonal, reflecting performance near chance. This suggests that those attributes have low signal-to-noise. In contrast, there is a separate, and somewhat orthogonal, cluster in the upper left portion of the plot that corresponds to attributes with high d' values. The direction of the distribution in this cluster reflects variation in the noise (and signal) distribution standard deviations, relative to a constant signal strength (i.e., d'). Thus the ROC plot for these features reveals that the primary source of between subject variance is simply variation in perceptual noise (e.g., the participant's visual acuity), not the signal-to-noise separation itself.

Our technical validation reveals a wide range of perceptual discrimination abilities for human testers, both within and across attributes. Certain attributes are easier to use as classification boundaries than others. This allows for customizing the difficulty of perceptual classification depending on the experimenter's needs.

Discussion

The Oomplet Dataset Toolkit (ODT) is a critical new tool for generating complex, novel visual stimuli through a controlled and generative process. The ODT enables the creation of large-scale datasets (millions of potential images) with engaging, compositional visual objects and categories, suitable for studying perceptual categorization in both biological (e.g., humans) and artificial agents (e.g., neural networks). Our technical validation demonstrates how different feature dimensions offer distinct levels of task complexity, guiding future experimental applications.

Flexibility is a key characteristic of the ODT. Because the toolkit allows for the generation of an extremely large number of categories, one can deploy a variety of heretofore challenging experimental paradigms. For example, examining how established mental representations of learned categories change when new categories – either more or less similar – are introduced could reveal how category structures adapt in response to new information. The generative nature of the ODT also allows the experimenter to flexibly adjust category relationships, adapting within an experiment based on how each category was initially learned or processed. This flexibility enables the ODT to fine-tune category boundaries or connections as new categories are introduced. The ODT's flexibility also allows the generated category structure to mirror the complex overlapping relationships of real-world categories, enabling experimenters to examine multiple facets of category learning in a single study. By varying factors such as participant experience, diversity within categories, and distances between categories, researchers can use controlled generation and overlap of components to simulate realistic learning conditions. This latter

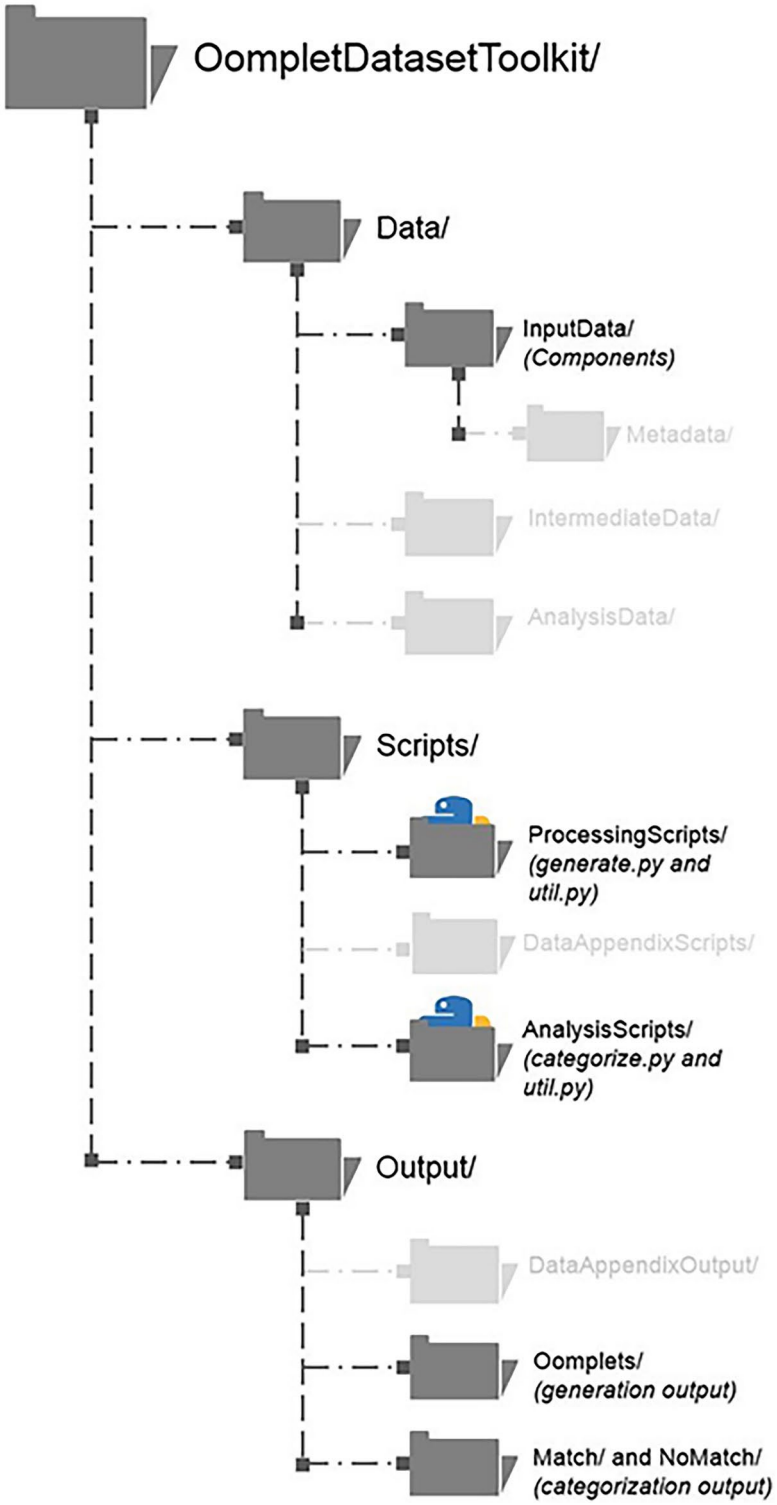

Fig. 5. Directory architecture for the ODT. Light grey folders indicate standard TIER Protocol 4.0 directories that are unused.

point is particularly important for developing AI-based studies of human-like category learning in artificial systems. Modern neural networks require millions of training examples to achieve their high performance. While humans obviously learn from much less data, there are many reasons to mirror the structure of the categories used in experiments in both natural and artificial systems [40].

Along with flexibility, the ODT is extensible. First, although the ODT currently employs a dictionary of components with discrete values, the experimenter can introduce new parts through the creation of new

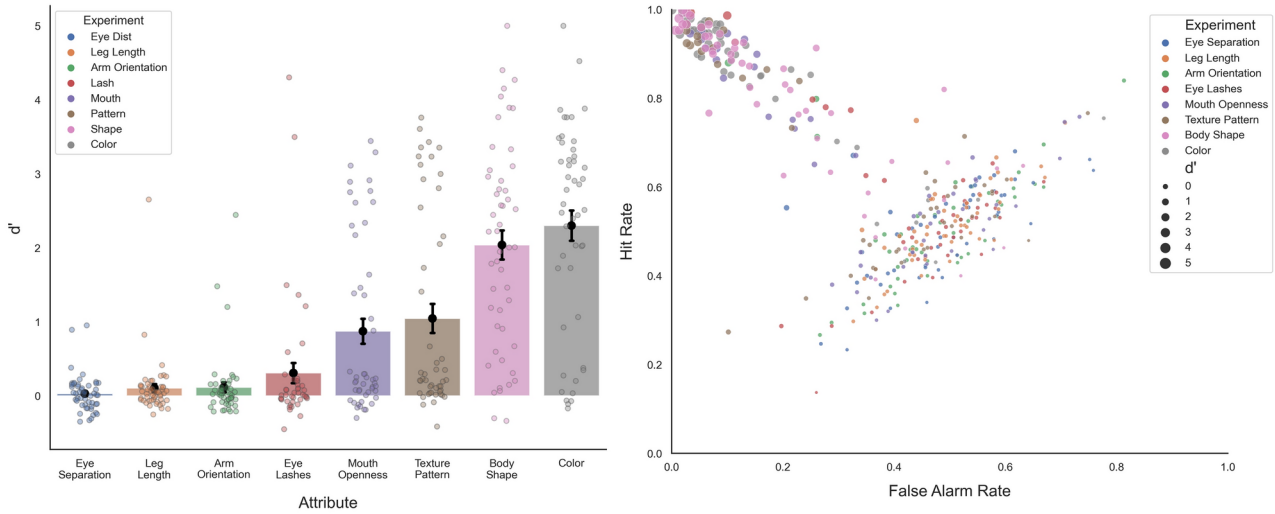

Fig. 6. Left: D-Prime scores of all participants, across all experiments. Right: ROC curve including data from all eight experiments.

component image files or increasing the number of levels per attribute. Users with some proficiency with Python can create new category boundaries by adding or altering the attribute tags associated with specific components and adapting our provided code accordingly. Likewise, users with some proficiency with art or computer graphics can create new components using existing attribute values, or, in tandem with the appropriate code modifications, can introduce new discrete or continuous attributes. For example, attributes could be varied in a continuous manner by building new components and attribute dictionaries or morphing between endpoints (e.g., continuous leg length values, body hues along a color gradient, or degree of mouth openness). These manipulations allow experimenters to identify perceptual boundaries that naturally emerge along specific feature dimensions or combinations of dimensions. By adjusting these dimensions, researchers could observe how boundaries form based on varying perceptual cues. Second, although the stimuli generated by the ODT are static, because of their humanoid appearance and compositional structure, it would relatively straightforward to animate them (e.g., using Spine software, <https://esotericsoftware.com/>). This opens up the possibility for dynamic attributes in combination with component parts or attributes.

Beyond its flexibility and extensibility, perhaps the ODT's most critical value lies in its generative design, allowing for a well-defined and manipulable distance metric between components and attributes. This makes it a versatile tool for diverse experimental and modeling applications. In contrast to the large majority of prior datasets, out-of-the-box the ODT does not require significant artistic or programming skills and is fully extensible. The ODT is user-friendly and freely available, filling a gap in the field. Instead of relying on and living with the limitations of smaller image datasets or uncurated online images, the ODT enables researchers to create well-controlled, large-scale stimulus datasets for exploring category learning in both natural and artificial systems. No comparable toolkit offers such a customizable, precise, and open-source solution.

Data availability

The code and assets used in the ODT stimulus generation pipeline are available on KiltHub: (https://doi.org/10.1184/R1/25813726.v1).

Received: 21 June 2024; Accepted: 4 March 2025
Published online: 18 March 2025

References

1. Langer, M. S. & Bülthoff, H. H. Measuring Visual Shape using Computer Graphics Psychophysics. In Péroche, B. & Rushmeier, H. (eds.) *Rendering Techniques 2000*, 1–9 (Springer Vienna, Vienna, 2000).
2. Shepard, R. N. & Metzler, J. Mental rotation of three-dimensional objects. *Science* **171**, 701–703 (1971).
3. Poggio, T. & Edelman, S. A network that learns to recognize three-dimensional objects. *Nature* **343**, 263–266 (1990).
4. Bülthoff, H. H. & Edelman, S. Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 60–64 (1992).
5. Tarr, M. J. tarrlab stimuli. https://sites.google.com/andrew.cmu.edu/tarrlab/stimuli (2024). Accessed: 01-18-2024.
6. Hayward, W. G. & Williams, P. Viewpoint dependence and object discriminability. *Psychol. Sci.* **11**, 7–12 (2000).
7. Tarr, M. J. *Visual object recognition: Can a single mechanism suffice? In Perception of Faces, Objects, and Scenes: Analytic and Holistic Processes* (Oxford University Press, Oxford, UK, 2006).
8. Gauthier, I. & Tarr, M. Becoming a 'Greeble' expert: Exploring mechanisms for face recognition. *Vision Research* **37** (1997).
9. Williams, P. *Prototypes, exemplars, and object recognition* (Yale University, United States - Connecticut, Ph.D., 1997).
10. Op de Beeck, H. P., Baker, C. I., DiCarlo, J. J. & Kanwisher, N. G. Discrimination training alters object representations in human extrastriate cortex. *J Neurosci* **26**, 13025–13036 (2006).

11. Wong, A.C.-N., Palmeri, T. J., Rogers, B. P., Gore, J. C. & Gauthier, I. Beyond shape: how you learn about objects affects how they are represented in visual cortex. *PLoS One* **4**, e8405 (2009).
12. Hegdé, J., Thompson, S., Brady, M. & Kersten, D. Object recognition in clutter: Cortical responses depend on the type of learning. *Frontiers in human neuroscience* **6**, 170 (2012).
13. Horst, J. S. & Hout, M. C. The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods* **48**, 1393–1409 (2016).
14. Lebaz, S., Sorin, A.-L., Rovira, K. & Picard, D. Widgets: A new set of parametrically defined 3D objects for use in haptic and visual categorization tasks. *European Review of Applied Psychology* **70**, 100552 (2020).
15. Jones, T. et al. Neural and behavioral effects of subordinate-level training of novel objects across manipulations of color and spatial frequency. *European Journal of Neuroscience* **52**, 4468–4479 (2020).
16. Gauthier, I., Tarr, M. & Bub, D. *Perceptual Expertise: Bridging Brain and Behavior* (Oxford University Press, Oxford, UK, 2009).
17. Tarr, M. J., Bülthoff, H. H., Zabinski, M. & Blanz, V. To What Extent Do Unique Parts Influence Recognition Across Changes in Viewpoint?. *Psychol. Sci.* **8**, 282–289 (1997).
18. Vuong, Q. C. et al. Facelikeness matters: A parametric multipart object set to understand the role of spatial configuration in visual recognition. *Visual Cognition* **24**, 406–421 (2016).
19. Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M. & Boyes-Braem, P. Basic objects in natural categories. *Cogn. Psychol.* **8**, 382–439 (1976).
20. Russakovsky, O. et al. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision* **115**, 211–252 (2015).
21. Lin, T.-Y. *et al.* Microsoft COCO: Common objects in context. In *Computer Vision - ECCV 2014*, 740–755 (Springer International Publishing, 2014).
22. Kupers, E. R., Knapen, T., Merriam, E. P. & Kay, K. N. Principles of intensive human neuroimaging. *Trends in Neurosciences* https://doi.org/10.1016/j.tins.2024.09.011 (2024).
23. Chang, N. et al. Bold5000, a public fmri dataset while viewing 5000 visual images. *Scientific Data* **6**, 1–18 (2019).
24. Allen, E. J. et al. A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nat. Neurosci.* **25**, 116–126 (2022).
25. Hebart, M. N. et al. Things-data, a multimodal collection of large-scale datasets for investigating object representations in human brain and behavior. *eLife* **12**, e82580 (2023).
26. Jain, N. et al. Selectivity for food in human ventral visual cortex. *Commun. Biol.* **6**, 175 (2023).
27. Khosla, M., Murty, Apurva Ratan N. & Kanwisher, N. A highly selective response to food in human visual cortex revealed by hypothesis-free voxel decomposition. *Current Biology* **32**, 1–13 (2022).
28. Pennock, I. M. L. et al. Color-biased regions in the ventral visual pathway are food selective. *Current Biology* **33**, 134-146.e4 (2023).
29. Wang, A. Y., Kay, K., Naselaris, T., Tarr, M. J. & Wehbe, L. Better models of human high-level visual cortex emerge from natural language supervision with a large and diverse dataset. *Nat. Mach. Intell.* **5**, 1415–1426 (2023).
30. Tarr, M. J. Rotating objects to recognize them: A case study of the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin and Review* **2**, 55–82 (1995).
31. Elmoznino, E., Jiralerspong, T., Bengio, Y. & Lajoie, G. A complexity-based theory of compositionality. arXiv preprint arXiv:2410.14817 (2024).
32. Gauthier, I., Skudlarski, P., Gore, J. C. & Anderson, A. W. Expertise for cars and birds recruits brain areas involved in face recognition. *Nat Neurosci* **3**, 191–197 (2000).
33. Tanaka, J. W., Kiefer, M. & Bukach, C. M. A holistic account of the own-race effect in face recognition: evidence from a cross-cultural study. *Cognition* **93**, B1-9 (2004).
34. Sheinberg, D. & Tarr, M. J. Objects of Expertise. In Gauthier, I., Tarr, M. J. & Bub, D. (eds.) *Perceptual Expertise: Bridging Brain and Behavior* (Oxford University Press, New York, NY, 2009). Publication Title: Perceptual Expertise: Bridging Brain and Behavior Place: New York, NY.
35. Bradski, G. The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (2000).
36. Cloudresearch. Connect by cloudresearch. https://www.cloudresearch.com (2023). Accessed: 06-05-2023 to 08-09-2023.
37. Gorilla. Gorilla task builder 2. https://www.gorilla.sc (2023). Accessed: 01-19-2023 to 08-23-2023.
38. Wickens, T. D. *Elementary signal detection theory* (Oxford university press, 2001).
39. ProjectTIER. Tier protocol 4.0. https://www.projecttier.org/tier-protocol/protocol-4-0/ (2023). Accessed: 05-05-2023.
40. Jinsi, O., Henderson, M. M. & Tarr, M. J. Early experience with low-pass filtered images facilitates visual category learning in a neural network model. *PLOS ONE* **18**, e0280145 (2023).

Acknowledgements

The authors would like to thank Michael Christel for his feedback and support. This work was sponsored, in part, by AFOSR/AFRL award FA9550-18-1-0251

Author contributions

MJT and TV conceived of the design of the project, oversaw completion of the project, and contributed to writing of this paper. AZ, HT, RW, and YT implemented the original Oomplet design and code generation. JPK contributed to designing and implementing the experimental tasks, including all data analysis, as well as leading the writing of the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to T.V. or M.J.T.