

ARTICLE

Labelling and iconicity facilitate visual categorisation and discrimination

James Scott¹ , Robert Foley²  and Mirjana Bozic¹ 

¹Department of Psychology, University of Cambridge, Cambridge, UK and ²Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, Cambridge, UK

Corresponding author: James Scott; Email: jhs74@cam.ac.uk

(Received 26 July 2024; Revised 25 June 2025; Accepted 01 July 2025)

Abstract

We investigated how the presence of linguistic labels, their iconicity and mode of presentation (cued vs not cued) affect non-linguistic cognitive processing, focussing on the learning and visual discrimination of new categories. Novel species of aliens that mimicked natural categories were paired with iconic labels, non-iconic labels or no labels across two tasks. In the Training task participants learnt to categorise the aliens, with results showing that both labels and iconicity improved categorisation. We then used a Match to Sample task to test how these variables affect rapid visual discrimination. Results showed that the presence of labels, their iconicity and label cueing all lead to more rapid and accurate visual discrimination of newly acquired categories. We argue that this is due to iconicity exaggerating sensory expectations provided by linguistic labels, made more readily accessible by cueing. We also examine the possible implications of our results for the discussion about language evolution.

Keywords: iconicity; language evolution; labels; non-linguistic cognition

1. Introduction

Words are not passive tags. Instead, it is now recognised that they dynamically shape not only linguistic processing (Perea & Rosa, 2002) but also other (non-linguistic) aspects of cognitive functioning too (Lupyan & Spivey, 2008). Research has also uncovered non-arbitrary mapping between word forms and their meanings, examples of which include sound symbolism and iconicity (e.g., Dingemanse et al., 2015). The current study investigates whether the presence of labels and their iconicity affect non-linguistic cognitive processing, focusing on learning and visual discrimination of novel categories. We also consider how any such effects may contribute to the discussion about language evolution.



1.1. Labels

Describing a linguistic label is not straightforward (Haspelmath, 2023), but in line with others, we define it as any sound comprising phonemes which is linked to a referent (Lupyan & Thompson-Schill, 2012). Labels have been shown to improve non-linguistic cognitive processes, including categorisation and visual discrimination (e.g., Gauthier et al., 2003; Lupyan et al., 2007; Winawer et al., 2007). The clearest evidence that labels facilitate categorisation comes from studies employing novel categories, as this rules out prior knowledge as an explanatory factor. For example, several experiments tasked participants with dividing novel ‘alien’ stimuli into different categories based on their visual features. Pairing aliens with novel pseudoword labels caused participants to learn these artificial categories more quickly and accurately (Lupyan et al., 2007). Other work indicates that real labels (i.e., existing words) can also facilitate visual discrimination. For example, Russian distinguishes between light and dark blue with the labels ‘goluboy’ and ‘siniy’. Compared to English, Russian speakers exhibit faster discrimination of shades of blue over this labelled category border (Winawer et al., 2007).

One way in which labels are proposed to influence non-linguistic cognition is engagement with predictive processing (Lupyan & Clark, 2015). Such accounts characterise cognitive functions such as categorisation and discrimination as ‘best guesses’ derived from an interplay between prior knowledge and the available sensory data (Clark, 2013; Fletcher & Frith, 2009; Friston & Kiebel, 2009; Teufel et al., 2018). By this view, labels act as priors that provide categorical and abstracted sensory expectations (Lupyan et al., 2020; Lupyan & Clark, 2015). Label-induced sensory predictions are categorical because labels denote categories, and thus over time become associated with category diagnostic features (Edmiston & Lupyan, 2015; Lupyan & Bergen, 2016). For example, while members of the category DOG vary hugely in their perceptual properties (e.g., chihuahuas vs dalmatians), the same label (e.g., ‘dog’) is always used. Thus, ‘dog’ denotes features most typical of dogs, and abstracts over incidental variation. Label-induced predictions also interact with sensory processing to warp representations to appear more typical, by emphasising category-typical item features and minimising incidental variation (Lupyan, 2008). For example, labels might more strongly predict and therefore emphasise typical DOG features such as a tail and snout, while underweighting a given dog’s particular coat colour. This effect would be induced through label knowledge alone and emphasised by hearing or using the label (see below). By increasing typicality, labels increase within-category similarity and between-category dissimilarity (Lupyan, 2012), effectively inducing categorical perception (Goldstone, 1994; Goldstone & Hendrickson, 2010). Neuroimaging data provide evidence for this hypothesis: for example, an electroencephalogram (EEG) study by Samaha et al. (2018) showed that the presence of labels improved both the recognition of and discrimination between ambiguous distorted images of categorical objects, which was associated with early occipital-parietal activation.

A corollary of the proposed mechanism is the ‘perceptual magnet effect’. This term originally referred to the poorer discrimination of more typical phonemes (Kuhl, 1991, 1994), with related typicality effects including the shift-to-prototype effect (Huttenlocher et al., 1991), and broader theory of representational shift (Lupyan, 2008). Here, we use the perceptual magnet effect to refer to the greater representational warping of typical compared to atypical items. The ‘magnet’ metaphor

highlights how the strength of ‘attraction’ (i.e., warping) decreases with distance (i.e., as items become less typical), comparable to the inverse square law (Solov’ev et al., 2023). Labels are hypothesised to exhibit the perceptual magnet effect because their categorical sensory predictions will match and therefore emphasise category-typical item features (Lupyan, 2008). More typical items will have more features which concur with these predictions and thus experience greater warping than less typical objects. This, in turn, means that label effects on categorisation will be more pronounced for more typical items (Edmiston & Lupyan, 2015; Lupyan & Thompson-Schill, 2012). Since the labels-as-priors account predicts the perceptual magnet effect, experimental evidence for the perceptual magnet effect in categorisation and discrimination tasks can be taken as further support that labels indeed provide categorical sensory predictions to these cognitive processes.

1.2. Iconicity

The non-arbitrariness of linguistic labels has long been a topic of scientific interest (Nielsen & Rendall, 2011), and numerous studies have identified correspondences between linguistic form and meaning, many of which are robust cross-culturally (Blasi et al., 2016). The recent surge in research has yielded both a wealth of insights and a proliferation of terminology, with classifications and definitions frequently overlapping or even contradictory (Barker & Bozic, 2024; Winter et al., 2023). We adopt here the classification provided by Dingemanse et al. (2015), and distinguish between arbitrariness (no mapping between the label and its meaning) and non-arbitrariness, examples of which include systematicity and iconicity. Systematicity is the statistical relationship between the sounds in large numbers of words and their abstract categories (e.g., the prosody of nouns vs verbs in English), and although this is indicative of links between labels and their meanings, it only captures broad patterns across the lexicon. Iconicity, meanwhile, is defined as when ‘aspects of the form and meaning of words are related by means of perceptuomotor analogies’ (Dingemanse et al., 2015). For example, larger objects are more readily associated with labels that contain vowels with lower resonant frequencies, or require a larger mouth opening to produce, such as ‘mal’ as opposed to ‘mil’ (Sapir, 1929). Iconic labels thus possess perceptual similarities to their referents. Sound symbolism is a specific form of iconicity at the level of phonetic features – e.g., compared to voiceless stops, voiced stops are associated with more rounded shapes (D’Onofrio, 2014; Shen et al., 2022). Iconicity is not a minor phenomenon. Since the early investigations of the takete-maluma/bouba-kiki effects (Köhler, 1929; Ramachandran & Hubbard, 2001; Spence & Parise, 2012) that linked label form with visual shape, iconic associations have been shown between labels and size (Ahlner & Zlatev, 2010; Nielsen & Rendall, 2011), colour (Johansson et al., 2020), brightness (Hirata et al., 2011); and less obvious properties including spatial deixis (Johansson & Zlatev, 2013), weight (Davis et al., 2019), intensity (Dingemanse et al., 2015), speed (Cuskley, 2013), taste (Gallace et al., 2011) and even social dominance (Auracher, 2017). Evidently, iconicity is alive and kiki-ng.

Iconicity is thought to be underpinned by associations between features presented to different modalities (Sidhu & Pexman, 2018). These cross-modal associations constitute priors that interact with sensory processing (Ernst, 2007): for example, the sound ‘wee’ is associated with the wide mouth shape required to produce it, and

auditory perception of this sound warps visual representations to increase their perceived horizontal elongation (Sweeny et al., 2012). Other findings which may support this position include priming evidence that iconic labels can influence shape perception (Sidhu & Pexman, 2017). The processes that give rise to iconic cross-modal associations thus arguably bear a striking resemblance to the mechanism by which labels at large influence non-linguistic cognition – they both provide sensory priors that interact with sensory processing. Iconic predictions highlight category-typical features of the categories they denote, which is precisely what makes such labels iconic: for example, if an iconic label that predicts roundedness (e.g., ‘bouba’) is applied to a category of rounded objects (e.g., MELONS), then the label ‘bouba’ will elicit sensory predictions which are highly category diagnostic of melons.

By this view, labels at large provide category-diagnostic predictions through cross-modal association; while iconicity does so through form. Thus, iconic labels may benefit from both sources to provide especially strong category diagnostic predictions. We therefore suggest that the sensory priors provided by *iconic* labels might be particularly strong and exert powerful effects on various aspects of cognitive processing. Furthermore, if iconicity exaggerates general label mechanisms in this manner, iconic labels would also be expected to more strongly display the ‘perceptual magnet effect’ discussed above (i.e., the greater representational warping of typical compared to atypical items) compared to non-iconic labels.

Existing evidence indeed confirms that iconic labels significantly facilitate linguistic cognition. As covered in a recent review (Nielsen & Dingemanse, 2021), iconicity facilitates label learning, and this iconic advantage is present in both young children (Imai et al., 2008; Perry et al., 2015) and adults (Lockwood et al., 2016; Nygaard et al., 2009). Iconic facilitation also extends beyond word learning: Kovic et al. (2010) trained participants to associate novel iconic labels with categories of visual stimuli. Participants were faster to verify that iconic labels matched congruent visual images, and EEG results indicated that participants were sensitive to iconicity within 200 ms of visual stimulus presentation, suggesting that behavioural effects were underpinned by top-down predictions provided by iconic labels. Iconic labels were also found to be processed more quickly in visual and auditory lexical decision tasks (Sidhu et al., 2020), and even in aphasic patients (Meteyard et al., 2015), suggesting that iconic labels might enjoy more direct links to semantics.

While iconic labels have been reliably implicated in linguistic cognition, their effects on non-linguistic cognition remain largely unexplored. For example, Maglio et al. (2014) showed that iconicity affects the precision of visual and conceptual representations, which implies that iconicity could affect other aspects of non-linguistic cognition. However, to our knowledge, only two studies have investigated the role of iconicity in categorisation, and none have examined visual discrimination. One previously mentioned study paired novel iconic labels with nine novel objects and manipulated congruency (Kovic et al., 2010). While the subsequent test phase found that iconic labels were easier to process, they had no effect on categorisation. A second study asked participants to categorise a limited set of novel ‘aliens’ in a between-subjects design (Lupyan & Casasanto, 2015). Depending on group, aliens were paired with congruent iconic labels (‘crelch’ for pointy aliens), incongruent labels (‘crelch’ for smooth aliens), no label or real words. Congruent novel iconic labels facilitated categorisation performance as much as using real labels (e.g., ‘pointy’). This advantage was not present when alien categories were paired with incongruent pseudowords or were not labelled at all. These results hint that iconic

labels may augment the non-linguistic cognitive task of categorisation. However, this is hard to verify because the study only contrasted congruent vs incongruent iconic labels and thus did not employ a direct non-iconic control condition. It therefore remains an open question as to whether the hypothesised stronger sensory predictions of iconic labels shape non-linguistic cognitive processes too, for example facilitating categorisation and visual discrimination.

1.3. *Mode of presentation*

We have proposed that the presence of labels provides priors that can facilitate aspects of non-linguistic cognition, and that iconicity might amplify this mechanism through the provision of stronger sensory predictions. If this is the case, iconic labels should facilitate non-linguistic cognitive processes such as categorisation and visual discrimination more strongly than their non-iconic counterparts. An additional variable that might test the hypothesis about their comparable ‘sensory priors’ mechanism is the mode in which labels are presented. As noted earlier, the existence of a label has been shown to affect various aspects of cognitive processing (Lupyan & Clark, 2015). While this can emerge due to its mere presence in the mental lexicon, cueing a label (for example, by hearing it spoken) activates the label more reliably, triggering stronger sensory expectations (Lupyan et al., 2020). Inspired by Lupyan et al. (2020), we refer to these two types of presentation as ‘offline’ (not cued) and ‘online’ (cued), respectively. Online presentation (or cueing) of a label is thus expected to enable label-induced priors to play even more active roles in modulating non-linguistic cognition, which is in line with the available evidence: cueing labels improves visual search time and efficiency (Lupyan, 2007), exaggerates representational warping of colour (Forder & Lupyan, 2019) and even extends to self-directed speech (Hebert et al., 2021; Lupyan & Swingley, 2010). However, no studies so far have tested whether the same upregulation of label effects through cueing also holds for iconicity. If iconicity influences non-linguistic cognition via the same mechanism as labels at large, iconic labels are expected to be liable to cueing too.

1.4. *Present study*

The present study explores how novel labels, their iconicity, and mode of presentation affect non-linguistic cognitive processing, focussing on the categorisation and visual discrimination of novel categories. Consistent with existing findings (Lupyan et al., 2007) we first hypothesised that the presence of labels would facilitate the process of learning and categorisation of novel objects. We predicted that the presence of labels would also impact the subsequent task of rapid visual discrimination of members of these learnt categories, extending the existing evidence on discrimination advantages for real labels (Samaha et al., 2018) to novel labels. In line with the proposed mechanism of labels as priors, we predicted that labels would trigger the perceptual magnet effect in both categorisation and visual discrimination, interacting with item typicality (with labelled items more sensitive to item typicality). We further predicted that iconic labels would lead to a greater enhancement of these processes. Finally, we anticipated that cueing the labels would increase the strength of their effects on non-linguistic cognition, resulting in faster and more accurate responses.

Our study builds on the well-established paradigms (Lupyan et al., 2007; Lupyan & Casasanto, 2015) but uses an expanded design and more stringent conditions to test these hypotheses. Participants learnt to categorise and discriminate novel species of aliens across two tasks, Training and Match to Sample (MTS). In the Training task, participants learnt two novel categories of aliens that were paired with iconic or non-iconic labels, or no label. In contrast to the existing work (Lupyan et al., 2007; Lupyan & Casasanto, 2015), individual aliens in our study were novel on every trial, rendering recognition and memory of individual exemplars impossible and forcing abstraction over category diagnostic features. Furthermore, aliens systematically varied in typicality, allowing us to examine the perceptual magnet effect.

The MTS task tested whether the effects of labels and iconicity on learning new categories also extend to their rapid visual discrimination. Here, participants were required to quickly decide which of two competing aliens belonged to the same category as the sample alien. The target stimuli in MTS were again different on every trial (and to those presented in Training), precluding reliance on recognition of individual exemplars. The use of MTS is a novel addition to the existing literature, enabling us to interrogate the effects of labels and iconicity on on-the-fly visual discrimination. Our alien typicality manipulation once again allowed us to test for perceptual magnet effects. Finally, the MTS task allowed us to test whether cueing strengthened the effects of labels and iconicity, and to distinguish the effects of 'online' label cueing from learning advantages carried over from Training.

2. Methods

2.1. Participants

A total of 159 participants were recruited using Prolific (Prolific, 2021), SONA (SONA-Systems, n.d.) and social media advertising. Five participants were excluded due to software error. 34 datasets obtained from the SONA platform were excluded because they could not be verified as coming from a genuine participant. Given online data collection, we ensured participant effort and attention by screening for careless responding (Stosic et al., 2024), resulting in the exclusion of a further 20 participants. This was achieved by using the R package 'careless' (version 1.2.2) (Yentes & Wilhelm, 2023) to generate long string, average string and reaction time based on intraindividual response variability variables (Hong et al., 2020; Ward & Meade, 2023), as well as screening for inattentive and imbalanced keystrokes. This left 100 participants in the analyses (48 female; age 18–40, $M = 24.9$, $SD = 5.6$). Participants had good command of English and no neurological, language or uncorrected auditory or visual impairments. Ethical approval was granted by the Department of Psychology Ethics Committee. Participants provided informed consent and were compensated for taking part, equally across all recruitment platforms.

2.2. Tasks and conditions

The experiment comprised two tasks: Training and Match to Sample (MTS). In the Training task, participants had to learn two novel categories (aliens) in one of three randomly assigned between-subjects conditions: Iconic Label, Non-Iconic Label and No Label (Table 1). In the Iconic Label condition, the categories were paired with spoken iconic labels. In the Non-Iconic Label condition, the categories were paired

Table 1. Experimental conditions

Training conditions	MTS conditions
Iconic label	Iconic label <u>Online</u> (spoken label presentation) Iconic label <u>Offline</u> (no label presentation)
Non-iconic label	Non-iconic label <u>Online</u> (spoken label presentation) Non-iconic label <u>Offline</u> (no label presentation)
No label	No label

with spoken non-iconic labels. In the No Label condition, the categories were not paired with labels.

The Match to Sample (MTS) task was built upon Training to assess the effects of labels and iconicity on the visual discrimination of learnt alien categories. The same three conditions as Training were used. In addition, to distinguish from Training advantages and investigate the effects of cueing, the Iconic and Non-Iconic Label conditions were each split into Online and Offline groups. Online groups were trained with labels and heard these same labels during MTS. Offline groups were also trained with labels but did not hear a spoken label during MTS. The Online vs Offline distinction thus examined active facilitation (cueing) vs the ‘passive’ effects carried over from learning in Training alone, respectively.

2.3. Stimuli

2.3.1. Aliens

Two new categories were created, consisting of aliens which varied on four visual dimensions. They were created by generating a four-dimensional tensor, with five steps on each dimension. These dimensions were: number of spokes (Spoke Number); how sharp the point on each spoke was (Spikiness); the size of the body of the alien in proportion to its spokes (Fatness); and colour from dark green to bright yellow (Darkness). To make aliens appear more naturalistic, a textured ‘skin’ and eyes were added. Alien stimuli were generated using Inkscape (Inkscape-Project, 2020). See Figure 1A for examples of aliens.

A 4D tensor with five steps on each dimension generates 625 unique combinations of features, and hence 625 unique aliens. Importantly, these aliens varied systematically in appearance according to the unique position of each along the four dimensions. This tensor was then bisected to create two distinct ‘species’ of aliens. First, two orthogonally opposite category exemplars were selected from the 16 possible combinations of dimensional extremes. The first category exemplar had the lowest Spoke Number; lowest Spikiness; greatest Fatness; and greatest Darkness. The orthogonally opposite one formed the exemplar for the other category, characterised by the highest Spoke Number; highest Spikiness; least Fatness; and least Darkness. These two categories of exemplars are marked by blue stars in the top left and bottom right corners of Figure 1A, which is a simplified illustration of the tensor used to generate the alien stimuli.

The aliens were coded according to proximity to the exemplars by collapsing their values across the four dimensions into one dimension (combined dimensional value; grey boxes in Figure 1A). For example, the exemplar of category 1 (Figure 1A, top left) has the lowest value on each of the four dimensions (1:1:1:1), giving a value of 4. Conversely, the exemplar of the other category (Figure 1A, bottom right) has a

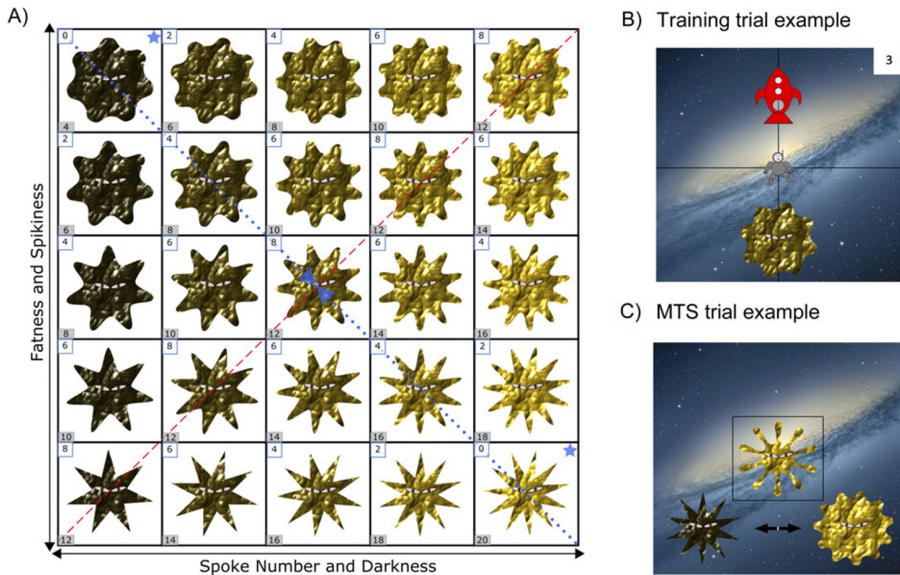


Figure 1. Stimuli and tasks. (A) Simplified illustration of the tensor containing 25 aliens only. Glulge and skysk exemplars presented in the top left and bottom right corners, respectively (marked by blue stars). Dashed red line indicates category boundary; dashed blue lines indicate Target Distance from exemplar. Numbers in blue boxes are Target Distance values; grey boxes indicate combined dimensional values for each alien. (B) Screenshot of a single Training trial. (C) Screenshot of a single Match to Sample trial.

value of 20 (5:5:5:5). Aliens were then segregated into three categories. Those with values closer to either exemplar were placed into the two respective categories. Aliens equidistant between exemplars were designated borderline. Due to the combinatorial nature of the tensor, frequencies of aliens at each distance from an exemplar varied systematically.

2.3.2. Labels

One iconic and one non-iconic pseudoword label was chosen per alien stimulus category. They were short, in accordance with English phonotactics and with minimal phonological neighbours. Iconic labels were constructed using iconic mappings found cross-linguistically and cross-culturally (Auracher, 2017; Blasi et al., 2016). To create the labels, candidate pseudowords were taken from the ARC Non-word Database (Rastle et al., 2002) and pre-tested in a sample of 34 participants who did not take part in the main experiment. To select iconic labels, participants were presented with category exemplars and asked to choose potential labels that most resembled each category. The same procedure was then repeated on a separate list of potential non-iconic labels, where participants selected two labels that they felt bore the least resemblance to either category. Labels with the highest mean ratings across participants were used to select iconic and non-iconic labels, respectively. This resulted in 'glulge' (/glʌldʒ/) and 'skysk' (/ski:sk/) chosen as iconic labels for the dark and round, and light and spiky categories, respectively. The non-iconic labels were 'stoise' (/stɔɪz/) and 'phrav' (/fræv/), with 'stoise' always applied to the dark and round alien category, and 'phrav' to the light and spiky category. Labels were

recorded in Audacity (Audacity-Team, 2021). Further information regarding label selection can be found in the [Appendix](#).

2.4. Procedure

For each trial in Training ($n = 180$), participants were presented with an astronaut avatar in the middle of the screen; an alien in one of four positions (above, below, left, or right of the astronaut); and a spaceship on the opposite side of the alien. Participants were instructed to move their avatar to approach the aliens of the 'friendly species' (category) and retreat to the spaceship if presented with the 'unfriendly' category of alien. Whether the glurge or skysk was the friendly category (i.e., was to be approached) was counterbalanced and consistent throughout for each participant. Following the response, participants in label conditions heard the relevant label. Category learning was initially trial and error, and participants were given accuracy feedback after each trial. A screenshot from Training is presented in [Figure 1B](#).

For each trial in the MTS task ($n = 200$), participants were presented with three aliens. In the middle of the screen was the sample alien, and participants were asked to select which of the two competing aliens presented below belonged to the same category as the sample alien. Importantly, at least one of the two aliens below was from a different category to the sample. For most trials, the target alien was of the same category as the sample; though for some, the target was substituted for a Borderline alien. In these trials, the correct response is to select the Borderline alien, as it is still closer to the sample than the distractor. This 'same or different' judgement is central to other visual discrimination tasks (Samaha et al., 2018; Winawer et al., 2007), although here matching stimuli were never physically identical. Hence, successful discrimination depended on the integration of both perceptual and categorical information. Participants in the two Online conditions heard the label for the sample alien at the beginning of each trial. A screenshot from MTS is presented in [Figure 1C](#). Every target alien in MTS appeared only once, while some distractors were repeated once. The intention was to preclude reliance upon recognition of individual exemplars, instead forcing participants to abstract over individual features and perceptual dimensions as per real categories (Lupyan & Bergen, 2016). The experiment was coded in PsyToolkit (Stoet, 2010, 2017) and conducted online.

2.5. Analyses

In both the Training and MTS tasks, we recorded participants' reaction times (RTs) and error rates. As an initial quality check, all RTs under 200 ms in both tasks were eliminated. The RTs for correct trials were then log transformed to eliminate skew, and datapoints more than 1.5 interquartile ranges above the upper quartile or below the lower quartile were removed. This resulted in exclusion of 221/12150 datapoints (1.8%) in Training; and 28/12895 (0.2% of the total) in MTS. RTs of correct target responses were compared across conditions using linear mixed-effect models as implemented in the lme4 R package (Bates et al., 2015a) (R version 4.3.3, package version 1.1.35.4). Accuracy data were modelled using generalised mixed-effects models (GLMM) with a binomial error distribution and a logit link function, fitted using the glmer function from the lme4 package in R.

Accuracy and correct RT responses from the Training session were fitted with models that included the fixed effects of Condition, Target Distance, Trial Number and interactions between Condition and Target Distance and Condition and Trial Number. Since the main questions driving this research concerned the effects of labelling and iconicity across different conditions (in addition to the effects of cueing in MTS), these effects were coded via orthogonal planned contrasts of Label vs No Label conditions (testing for Labelling effects) and Iconic vs Non-Iconic conditions (testing for Iconicity effects). This allowed us to both reduce reliance on multiple explanatory post-hoc comparisons and to interrogate any possible interactions; for instance, to establish how the change in accuracy or RTs across Target Distance might vary between the Iconic and Non-Iconic Label conditions. Target Distance defines how atypical an alien is and was used to detect perceptual magnet effects (i.e., if labelling or iconicity effects were more pronounced for more typical aliens). Target Distance was calculated by counting the number of orthogonal steps to get to its exemplar. For example, an alien with a value of 16 would be 4 orthogonal steps from the exemplar (score of 20), giving it a Target Distance of 4, and borderline aliens had the maximum value of 8 (see [Figure 1A](#)). The inclusion of Trial Number allowed us to test for learning effects over the course of Training. The RT and accuracy data from the MTS task were fitted with models including the fixed effects of Condition (including orthogonal planned contrasts examining the effects of Labelling, Iconicity and Cueing), Target Distance and their interaction. In MTS, Target Distance referred to the atypicality of the target alien. Nested random intercepts for Item and Participant were included in all analyses (Bates et al., 2015b), with Item nested within Target Distance and Participant nested within Condition. Trial Number and Target Distance were centred. All models and outputs are presented in the [Appendix](#). The Satterthwaite approximation was used for degrees of freedom, and significant p -values are reported at $p < .05$.

3. Results

3.1. Training

Accuracy: The Accuracy model included fixed effects of Condition (with planned contrasts coding for the effects of Labelling and Iconicity), Target Distance, Trial Number and the interactions between Condition and Target Distance, and Condition and Trial Number. Results are shown in [Figure 2A–C](#) and [Table 2\(a\)](#). There was no significant effect of Condition on accuracy ($X^2(2,100) = 1.29$, $p = .525$). However, both Trial and Target Distance were significant predictors ($X^2(1,100) = 35.22$, $p < .001$ and $X^2(1,100) = 165.68$, $p < .001$, respectively), with accuracy increasing over trials and decreasing with Target Distance (i.e., as aliens became less typical). Condition interacted with both Trial ($X^2(2,100) = 17.20$, $p < .001$) and Target Distance ($X^2(2,100) = 6.33$, $p = .042$). Planned contrasts revealed that this was driven by interactions with Labelling, with increases in Trial and Target Distance differently affecting Label conditions compared to the No Label condition. Trial additionally interacted with Iconicity, affecting the Iconic Label condition more strongly than the Non-Iconic Label condition (see [Table A1](#) in the [Appendix](#) for full details of planned contrast results).

Reaction Times: Correct RTs were modelled using the same fixed effects of Condition, Target Distance, Trial Number and the interaction between Condition

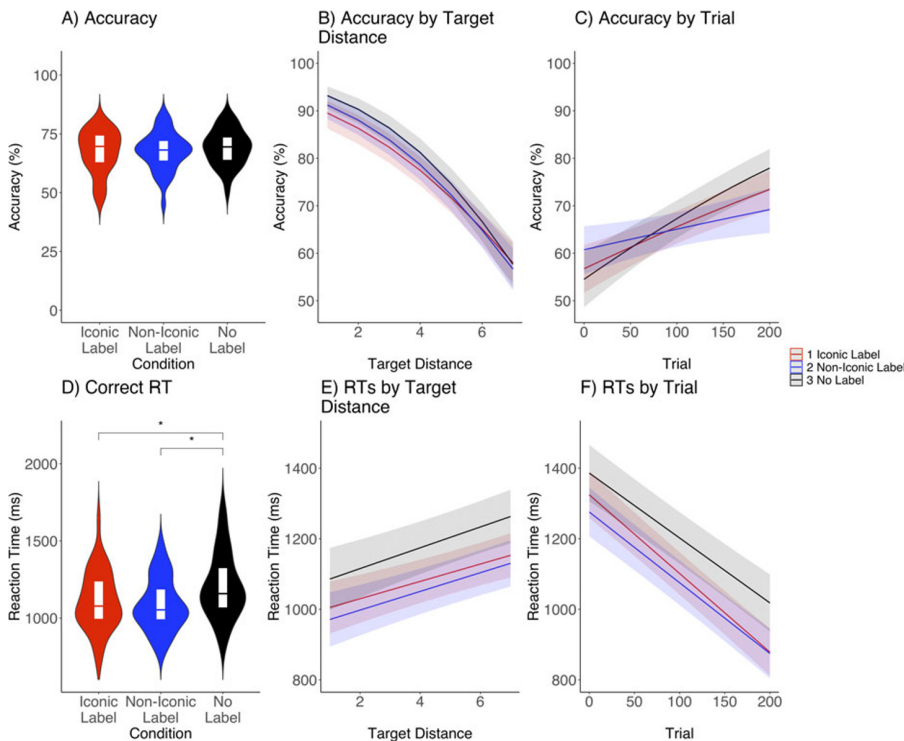


Figure 2. Accuracy and RT results in training. (A) Distribution of accuracy rates across participants in the three training conditions. (B) Average accuracy per condition at different target distances. (C) Average accuracy per condition over trials. (D) Distribution of correct RTs across participants in the three training conditions. (E) Average RTs per condition at different target distances. (F) Average RTs per condition over trials.

Table 2. Results for (a) accuracy and (b) correct RT models in the Training task

Effect	(a) Accuracy			(b) RT						
	Chisq	df	Pr (>Chisq)	Sum Sq	Mean Sq	NumDF	DenDF	F	Pr(>F)	η^2
(Intercept)	236.98	1	<.001***							
Condition	1.29	2	0.52	0.16	0.08	2	99.9	3.67	0.02*	0.07
Target distance	165.68	1	<.001***	1.02	1.02	1	173.3	46.27	<.001***	0.21
Trial	35.22	1	<.001***	9.01	9.01	1	299.5	408.54	<.001***	0.58
Condition:Target Distance	6.33	2	0.04 *	0.01	0.00	2	11713.9	0.13	0.88	0.00
Condition:Trial	17.20	2	<.001 ***	0.12	0.06	2	11722.7	2.81	0.06	0.00

*p<.05; **p<.01; ***p<.001.

and Target Distance, and Condition and Trial Number. Results are illustrated in Figure 2D–F and Table 2(b). Condition affected correct RTs ($F(2,99.9) = 3.67$, $p = .029$; $\eta^2 = .07$), with Label conditions faster than the No Label condition but no difference between the two. Target Distance and Trial were again significant predictors ($F(1,173.3) = 46.27$, $p < .001$; $\eta^2 = .21$; and $F(1,299.5) = 408.54$, $p < .001$;

$\eta^2 = .58$, respectively), with RTs decreasing over trials and increasing with target distance. There were no significant interactions between Condition and either Target Distance or Trial (see Table A2 in the Appendix).

Together, the Training data showed that participants were able to successfully learn the new categories in a short period of time. As expected, both accuracy and RTs improved over the course of Training, with responses becoming more accurate and faster as the session progressed. The presence of labels led to faster correct RTs. Alien typicality (Target Distance) significantly influenced the process, with more typical aliens proving easier to learn. Finally, Labelling interacted with both Target Distance and Trial number, suggesting a potentially subtly different learning mechanism in the Label conditions compared to the No Label condition. The only hint of Iconicity effects came from the accuracy analysis, where Iconicity interacted with Trial, such that iconic labels significantly improved participants' accuracy over time compared to non-iconic labels.

3.2. Match to sample

Accuracy: The Accuracy model in MTS included the fixed factors of Condition (with planned contrasts coding for the effects of Labelling, Iconicity and Cueing), Target Distance and their interaction. Results are illustrated in Figure 3A,B and Table 3(a). Condition robustly affected accuracy ($X^2(4,100) = 35.05$, $p < .001$), driven by significant effects of Labelling ($b = .04$, $p < .05$; with Label conditions more accurate than the No Label condition); and Cueing ($b = .21$, $p < .001$; with Online conditions more accurate than Offline conditions). Target Distance was also a significant predictor ($X^2(1,100) = 7.74$, $p = .005$), with accuracy decreasing with Target Distance (as aliens became less typical) across all conditions. There was also a significant Condition by Target Distance interaction ($X^2(4,100) = 28.74$, $p < .001$), driven by interactions between Target Distance and Labelling ($b = -.009$, $p < .05$; with effects stronger in Label conditions compared to the No Label Condition) and Target Distance and Cueing ($b = -.05$, $p < .001$; with effects stronger in Online conditions compared to Offline conditions). See Table A3 in the Appendix for full details of planned contrast results.

Reaction Times: The RT model on correct responses in MTS included the same fixed effects of Condition (with planned contrast for the effects of Labelling, Iconicity and Cueing), Target Distance and their interaction. Results are illustrated in Figure 3C,D and Table 3(b). Correct RTs in MTS were affected by Condition ($F(4,99.9) = 2.54$, $p = .045$; $\eta^2 = .09$), reflecting significant effects of both Iconicity (Iconic faster than Non-Iconic conditions, $b = -.03$, $p = .034$) and Cueing (Online faster than Offline conditions, $b = -.03$, $p = .024$). Target Distance was also a significant predictor ($F(1,200.2) = 42.43$, $p < .001$; $\eta^2 = .17$), with the RTs increasing with Target Distance across all conditions. Results also showed a significant Condition by Target Distance interaction ($F(4,12613.3) = 4.41$, $p = .001$; $\eta^2 < .001$), with Target Distance interacting with both Labelling ($b = .001$, $p = .02$, with effects more prominent in Label conditions compared to the No Label condition) and Cueing ($b = .003$, $p < .001$, with effects more prominent in Online conditions compared to Offline conditions). See Table A4 in the Appendix.

In sum, MTS data revealed that, following the Training session, participants' visual discrimination of the newly learnt categories was more accurate if these

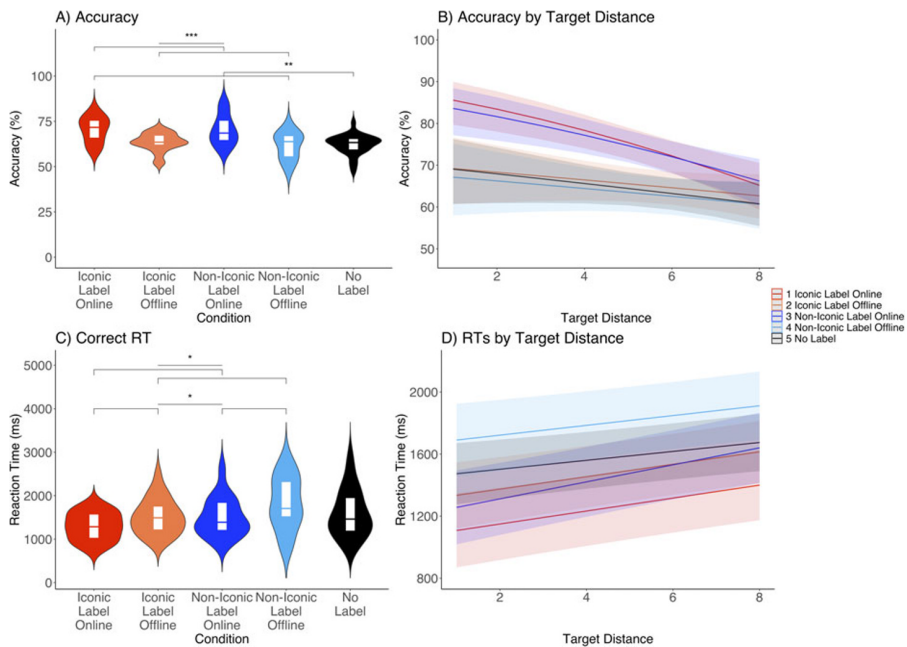


Figure 3. MTS accuracy and correct RT results. (A) Distribution of accuracy rates across participants in the five MTS conditions. (B) Average accuracy per condition at different target distances. (C) Distribution of correct RTs across participants in the five MTS conditions. (D) Average RTs per condition at different target distances.

Table 3. Results for (a) accuracy and (b) correct RT models in the MTS task

Effect	(a) Accuracy			(b) RT						
	Chisq	df	Pr (>Chisq)	Sum Sq	Mean Sq	NumDF	DenDF	F	Pr(>F)	η^2
(Intercept)	136.31	1	<.001 ***							
Condition	35.05	4	<.001 ***	0.26	0.06	4	99.9	2.54	0.04 *	0.09
Target distance	7.74	1	0.01 **	1.08	1.08	1	200.2	42.44	<.001 ***	0.17
Condition: Target distance	28.74	4	<.001 ***	0.45	0.11	4	12613.3	4.41	0.001 **	>0.01

*p<.05; **p<.01; ***p<.001.

categories were labelled and directly cued (i.e., presented online). Correct responses were faster if labels were iconic and cued. Target Distance affected RTs and accuracy as expected, with participants’ responses becoming slower and less accurate as aliens got less typical across all conditions, though most prominently in the Online and Label conditions. We discuss the implications of these results below.

4. Discussion

The present study investigated the effects of labelling and iconicity on non-linguistic cognitive processing, focussing on the learning and visual discrimination of novel categories. To do this, we compared the performance of participants across Training

and Match to Sample tasks, contrasting label conditions with the no label condition, iconic with non-iconic conditions and online and offline label presentation. As summarised above, the Training data showed that participants were able to successfully learn the new categories, with the presence of labels and item typicality (Target Distance) both significantly influencing their learning. The subsequent MTS task revealed that visual discrimination of the newly learned categories was indeed helped by the presence of labels, their iconicity and cueing.

4.1. Training

The purpose of Training was for participants to learn the alien categories and their accompanying labels. The results demonstrated that the task functioned as intended, with all three conditions improving in both speed and accuracy across trials, and accuracy reaching well above chance across all three conditions by the end of Training. Consistently greater reaction times and lower response accuracies at greater Target Distances confirmed that stimulus typicality translated into difficulty as intended. Crucially, we also saw evidence that Label conditions produced faster RTs on average compared to the No Label condition. This indicates that labels facilitate categorisation, consistent with the literature (Lupyan et al., 2007). The only hint of the effects of iconicity in Training came from the interaction between Iconicity and Trial in the accuracy analysis, with a steeper learning rate in the Iconic condition compared to the Non-Iconic condition – thus providing only weak evidence for the hypothesis that iconic labels confer advantage to the learning of new categories (Lupyan & Casasanto, 2015).

4.2. Match to sample

The MTS task assessed visual discrimination of these newly learned categories. The results provided evidence in support of nearly all our hypotheses. As laid out in the Introduction, we expected that the presence of labels would result in faster and more accurate discrimination of novel alien category members. We further predicted that iconic labels would lead to a greater enhancement of these processes, as would cueing these labels. Less typical aliens were expected to be more difficult to discriminate. Next, we expected that labelled, iconic and cued conditions would be more sensitive to alien typicality, resulting in significant interactions with Target Distance. Such results would be indicative of the perceptual magnet effect (Lupyan, 2008) and support the hypothesis that the effects were due to labels providing sensory expectations to predictive cognitive processes. MTS results support these hypotheses: Label conditions were more accurate than the No Label condition; and correct responses were faster in Iconic conditions than in the Non-Iconic conditions. Online conditions outperformed Offline conditions in both speed and accuracy. Target Distance consistently affected performance as predicted, and the interactions between Target Distance and Condition confirmed that labels and online conditions were exhibiting the perceptual magnet effect.

4.2.1. Label vs no label

The Label vs No Label distinction is a necessary starting point, as showing an effect of labelling on non-linguistic cognitive processing underpins examination of iconicity

and mode of label presentation. As reviewed in the Introduction, labels have been shown to facilitate both linguistic and non-linguistic cognition, including visual discrimination (Winawer et al., 2007). It has been argued that labels achieve this by providing categorical sensory predictions, which interact with sensory processing to warp item representations to appear more typical (Lupyan, 2008; Lupyan & Clark, 2015), thereby easing between-category discrimination. Our results are consistent with this account. Target aliens used in MTS were always novel, meaning that no previous encounters were available to rely upon. Furthermore, correct execution of a trial could not have resulted from simple comparisons of singular dimensions between aliens, for example colour, since the four-dimensional nature of the tensor meant that two aliens of the same category could look vastly different along multiple dimensions (Figure 1). Hence, correct execution of a trial necessitated abstraction of perceptual features over all four dimensions. Furthermore, the fact that each dimension extended continuously and linearly into both categories meant there was substantial overlap in perceptual features between categories. Participants therefore had to extrapolate distributions of perceptual features for each category to identify category membership for any given alien. As seen in the data, this became progressively more difficult at higher Target Distances, consistent with both equal weighting of perceptual features and greater category uncertainty at atypical extremes, due to greater overlap of predicted distributions of perceptual features.

While there is existing evidence that real labels (i.e., existing words) facilitate visual discrimination via this mechanism (Samaha et al., 2018), our results show that the same effect holds for novel labels. Label conditions were more accurate than the No Label condition, extending previously documented label advantages (Lupyan & Casasanto, 2015) to a novel task context. As noted in the Introduction, these results arguably arose as a result of labels interfacing with the predictive processing (Lupyan & Clark, 2015) necessary for successful MTS trial execution. They provided an additional level of prior information, superior in abstraction and flexibility, that was unavailable to participants in the No Label condition. As a prior uniting a category, a label necessarily deals with generalisations rather than particulars (Lupyan & Bergen, 2016). Thus, labels exaggerate category-diagnostic features while minimising idiosyncratic ones. Applied to MTS, this would increase the weighting of perceptual features which matched sensory predictions, highlighting similarities between aliens of the same category (Boutonnet & Lupyan, 2015). This interpretation is also consistent with the evidence for the perceptual magnet effect in Label conditions in MTS, which implies that participants in Label conditions perceived aliens as more typical of their category than they were, easing category comparisons. In other words, alien posteriors were warped towards their priors, suggesting that label-induced priors guided interpretation to form a percept, enhancing MTS performance. Overall, therefore, our results provide new evidence that novel labels facilitate visual discrimination and suggest that this occurs through a predictive mechanism.

4.2.2. Iconicity

The second key finding from MTS was the speed advantage of iconic over non-iconic labels on correct trials. This provides evidence that iconicity facilitates visual discrimination, which we argue results from iconic exaggeration of label-induced sensory predictions. Iconic words have been shown to elicit cross-modal sensory

predictions, most obviously in the Bouba-Kiki effect (Ramachandran & Hubbard, 2001). The iconic labels in this experiment were designed to reflect the visual properties of the aliens. For example, the 'g' and 'l' in 'glulge' are associated with roundness (Nielsen & Rendall, 2011); and iconic correspondence was confirmed through pretesting. Hence, iconic labels arguably triggered sensory representations which matched category-typical features. This would strengthen label predictions in a category-specific fashion, facilitating MTS performance.

A lack of interactions between Iconicity and Target Distance in MTS meant that no perceptual magnet effects were observed for Iconic conditions over and above other contrasts, contrary to our predictions. One possibility is that iconic perceptual magnet effects were overshadowed by those of Label and Online conditions – iconic effects may be relatively weak in an additive scenario. Another is that iconicity did not interact directly with the abstract label prior, which would be required for such effects. Instead, iconicity might have provided a second pathway to sensory prediction characterised by lower-level, more concrete representations. Iconic-induced sensory associations are generic and not characterised by the abstraction afforded by labels. Indeed, iconicity is a property of the labels separate from the categories they denote. This would still result in an additive advantage, leaving the Iconic Label Online condition outperforming all others. This interpretation is clearly speculative and contrary to our predictions. Further, it cannot explain Online modulation of Iconic Labels as discussed below, and more research is therefore required to elucidate the mechanisms of iconicity in this context. Regardless, iconic labels still displayed perceptual magnet effects within the Label–No Label contrast (Lupyan, 2008), and thus behaved like real words. Given this, it seems likely that iconicity triggered sensory representations which served to exaggerate label-induced sensory predictions and help participants identify and categorise aliens in MTS.

While the iconic speed advantage observed in the MTS task builds upon previous findings in the literature, previous work also evidences an iconic accuracy advantage during category learning (Lupyan & Casasanto, 2015). Our Training task mimicked this experimental structure, but as discussed above, only weakly replicated this result in the accuracy analysis. One possibility is that this is due to our more stringent experimental design with novel stimuli on every trial. Another is that the harsher test of iconic vs non-iconic labels in our study produced a subtler effect than the reversed iconicity of the Lupyan & Casasanto's 2015 study. However, this does not explain the stronger iconic results in MTS, which also used a direct iconic contrast and novel stimuli on every trial. It is possible that the same factors weakened the iconic *accuracy* advantage in MTS, but did not affect *speed*. The latter may have been less important in Training given the slower pace of the task. This interpretation is speculative, however, and requires further investigation.

4.2.3. *Online vs offline*

Finally, MTS data showed that Online conditions outperformed the Offline conditions. In Online conditions, labels were heard just before aliens were seen. It is likely that this cueing activated labels more reliably, forcing stronger and more consistent sensory predictions. Evidence of stronger predictions is also borne by Target Distance results. Significantly steeper interactions between Online conditions and Target Distance compared to Offline conditions for both speed and accuracy point to stronger perceptual magnet effects in the Online conditions. This demonstrates that

in Online conditions, more typical alien features were more heavily weighted as typicality increased. Thus, their prior predictions must have been stronger. Online label presentation, therefore, elicited stronger label-induced sensory predictions, resulting in MTS advantages.

Results from Online conditions may provide further mechanistic clarification. Iconic labels were liable to cueing modulation, with the Iconic Label Online condition exhibiting the highest accuracy and fastest responses of all conditions. This provides further evidence that iconic labels operate by the same general predictive mechanism as non-iconic labels. Online results may also suggest that label and iconic MTS advantages do not stem from superior Training performance. In Training, labels generally exhibited faster categorisation; and iconic labels showed a faster learning rate. A possible objection is that this better category learning might have translated into MTS performance advantages. However, the fact that cueing improved label performance indicates that advantages stem instead from label-derived predictions. Confirmation of this interpretation requires further testing.

In sum, Match to Sample results support our hypotheses of label, iconic and cueing advantages. Target Distance and Online results suggest that this enhancement might result from sensory predictions provided by label priors. Cueing improved both speed and accuracy, leaving the Iconic Label Online condition as the numerically best-performing condition. We propose that these effects arise due to iconicity exaggerating sensory expectations provided by labels, which are made readily accessible by their cueing.

4.3. *Potential evolutionary implications*

Labels constitute the core of language (Barham & Everett, 2021), and characterising their emergence is therefore a key question for language evolution research. There are clearly limitations in extrapolating data derived from modern humans to interpret putative early hominin capabilities, but experimentation provides an important avenue of research which complements other approaches. Hence, while full evolutionary discussion is beyond the scope of this paper, we believe our findings complement other suggestions that early labels were iconic.

Previous work suggests that the earliest labels were iconic. Labels may have emerged from the standardisation of non-human primate communicative calls (Tallerman, 2014). However, such standardisation presents a tension between the advantages this entails and the simultaneously greater difficulty in grounding meaning. Many propose that iconicity bridges this gap by providing referential insight (Imai & Kita, 2014; Perniss & Vigliocco, 2014). This suggestion is supported by experimental evidence using novel vocalisations (Ćwiek et al., 2021; Perlman et al., 2015; Perlman & Lupyan, 2018) and more broadly concurs with mimetic proposals for the origin of labels (Knight & Lewis, 2017). Iconic labels are also easier to learn (Lockwood et al., 2016) and may therefore have been promoted by the cultural evolutionary forces which influence language evolution (Chater & Christiansen, 2010; Laland, 2017). Further, experimental work with panins implies that crossmodal correspondences were possible in the ancestral state (Ludwig et al., 2011), although this is contested (Margiotoudi et al., 2022).

To these existing lines of evidence which support early iconic labels, we add our present findings as a possible complement. While communication is clearly central to

label evolution (Levinson, 2022), this does not exclude other proximate advantages. As noted elsewhere, labels' ability to augment non-linguistic cognition may provide one such advantage (Lupyan & Bergen, 2016). Since our findings indicate that iconicity exaggerates this advantage for categorisation and visual discrimination, we suggest that iconic labels may have been particularly powerful ways of enhancing non-linguistic cognition. Such an advantage may have promoted iconic label form through cultural evolutionary processes. This suggestion complements previous work by providing another reason to suspect that the earliest labels were iconic. Clearly, however, further research is required to consolidate this claim.

5. Conclusion

In conclusion, this study investigated the effects of labelling and iconicity on non-linguistic cognition across two tasks. The Training task allowed participants to learn labels and categories, and demonstrated a label advantage in the process. The MTS task showed that the presence of labels, their iconicity and cueing enhanced the non-linguistic cognitive processes of categorisation and visual discrimination. We argue that these effects arise due to iconicity exaggerating sensory expectations provided by labels, which were made more readily accessible by their cueing. Finally, we reviewed several reasons to suggest that labels in the earliest stages of evolution were iconic.

Data availability statement. The data and code that support the findings of this study are openly available in OSF at https://osf.io/v4u38/?view_only=3534f5d83e084a8cb680117b457bcfd7.

Funding statement. This research was supported by funds from the Economic and Social Research Council Doctoral Training Partnership awarded to JHS (award reference: ES/J500033/1).

Competing interests. The authors declare none.

Ethics approval. Ethical approval for this study was granted by the relevant Psychology Ethics Committee. Informed consent was obtained from all participants.

References

- Ahlner, F., & Zlatev, J. (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies*, 38(1/4), 298–348. <https://doi.org/10.12697/sss.2010.38.1-4.11>.
- Audacity-Team. (2021). *Audacity(R)* (3.0.0). <https://audacityteam.org/>. [Computer Program]
- Auracher, J. (2017). Sound iconicity of abstract concepts: Place of articulation is implicitly associated with abstract concepts of size and social dominance. *PLoS One*, 12(11), 1–25. <https://doi.org/10.1371/journal.pone.0187196>.
- Barham, L., & Everett, D. (2021). Semiotics and the origin of language in the lower Palaeolithic. *Journal of Archaeological Method and Theory*, 28(2), 535–579. <https://doi.org/10.1007/s10816-020-09480-9>.
- Barker, H., & Bozic, M. (2024). *The forms, mechanisms, and roles of iconicity: A review*. Authorea Preprints.
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Bolker, M. B. (2015a). Package 'lme4'. *Convergence*, 12(1), 1–130. [Computer Software].
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015b). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound-meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences of the United States of America*, 113(39), 10818–10823. <https://doi.org/10.1073/pnas.1605782113>.
- Boutonnet, B., & Lupyan, G. (2015). Words jump-start vision: A label advantage in object recognition ords jump-start vision: A label advantage in object recognition. *Journal of Neuroscience*, 35(25), 9329–9335.
- Chater, N., & Christiansen, M. H. (2010). Language acquisition meets language evolution. *Cognitive Science*, 34(7), 1131–1157. <https://doi.org/10.1111/j.1551-6709.2009.01049.x>.

- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–204. <https://doi.org/10.1017/S0140525X12000477>.
- Cuskley, C. (2013). Mappings between linguistic sound and motion. *Public Journal of Semiotics*, 5(1), 39–62. <https://doi.org/10.37693/pjos.2013.5.9651>.
- Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., Kawahara, S., Koutalidis, S., Krifka, M., Lippus, P., Lupyan, G., Oh, G. E., Paul, J., Petrone, C., Ridouane, R., Reiter, S., Schümchen, N., Szalontai, Á., Únal-Logacev, Ö., ... Perlman, M. (2021). Novel vocalizations are understood across cultures. *Scientific Reports*, 11(1), 1–12. <https://doi.org/10.1038/s41598-021-89445-4>.
- D'Onofrio, A. D. (2014). Phonetic detail and dimensionality in sound-shape correspondences: Refining the Bouba-Kiki paradigm. *Language and Speech*, 57(3), 367–393. <https://doi.org/10.1177/0023830913507694>.
- Davis, C. P., Morrow, H. M., & Lupyan, G. (2019). What does a Horgous look like? Nonsense words elicit meaningful drawings. *Cognitive Science*, 43(10), 1–18. <https://doi.org/10.1111/cogs.12791>.
- Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, 19(10), 603–615. <https://doi.org/10.1016/j.tics.2015.07.013>.
- Edmiston, P., & Lupyan, G. (2015). What makes words special? Words as unmotivated cues. *Cognition*, 143, 93–100. <https://doi.org/10.1016/j.cognition.2015.06.008>.
- Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, 7(5), 1–14. <https://doi.org/10.1167/7.5.7>.
- Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: A Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, 10(1), 48–58. <https://doi.org/10.1038/nrn2536>.
- Forder, L., & Lupyan, G. (2019). Hearing words changes color perception: Facilitation of color discrimination by verbal and visual cues. *Journal of Experimental Psychology: General*, 148(7), 1105–1123. <https://doi.org/10.1037/xge0000560.supp>.
- Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521), 1211–1221. <https://doi.org/10.1098/RSTB.2008.0300>.
- Gallace, A., Boschini, E., & Spence, C. (2011). On the taste of “Bouba” and “Kiki”: An exploration of word-food associations in neurologically normal participants. *Cognitive Neuroscience*, 2(1), 34–46. <https://doi.org/10.1080/17588928.2010.516820>.
- Gauthier, I., James, T. W., Curby, K. M., & Tarr, M. J. (2003). The influence of conceptual knowledge on visual discrimination. *Cognitive Neuropsychology*, 20(3–6), 507–523. <https://doi.org/10.1080/02643290244000275>.
- Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, 123(2), 178–200. <https://doi.org/10.1037/0096-3445.123.2.178>.
- Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(1), 69–78. <https://doi.org/10.1002/WCS.26>.
- Haspelmath, M. (2023). Defining the word. *Word*, 69(3), 283–297. <https://doi.org/10.1080/00437956.2023.2237272>.
- Hebert, K. P., Goldinger, S. D., & Walenchok, S. C. (2021). Eye movements and the label feedback effect: Speaking modulates visual search via template integrity. *Cognition*, 210, 104587. <https://doi.org/10.1016/j.cognition.2021.104587>.
- Hirata, S., Ukita, J., & Kita, S. (2011). Implicit phonetic symbolism in voicing of consonants and visual lightness using Garner’s speeded classification task. *Perceptual and Motor Skills*, 113(3), 929–940. <https://doi.org/10.2466/15.21.28.PMS.113.6.929-940>.
- Hong, M., Steedle, J. T., & Cheng, Y. (s2020). Methods of detecting insufficient effort responding: Comparisons and practical recommendations. *Educational and Psychological Measurement*, 2020(2), 312–345. <https://doi.org/10.1177/0013164419865316>.
- Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review*, 98(3), 352–376. <https://doi.org/10.1037/0033-295x.98.3.352>.
- Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), 1–13. <https://doi.org/10.1098/rstb.2013.0298>.
- Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, 109(1), 54–65. <https://doi.org/10.1016/j.cognition.2008.07.015>.
- Inkscape-Project. (2020). *Inkscape*. <https://inkscape.org/>. [Computer Program]

- Johansson, N. E., Anikin, A., & Aseyev, N. (2020). Color sound symbolism in natural languages. *Language and Cognition*, 12, 56–83. <https://doi.org/10.1017/langcog.2019.35>.
- Johansson, N. E., & Zlatev, J. (2013). Motivations for sound symbolism in spatial deixis: A typological study of 101 languages iconicity view project diachronic atlas of comparative linguistics-DiACL view project motivations for sound symbolism in spatial Deixis: A typological study of 101 Lan. *PJOS*, 5(1), 1–20. <https://www.researchgate.net/publication/313436697>
- Knight, C., & Lewis, J. (2017). Wild voices: Mimicry, reversal, metaphor, and the emergence of language. *Current Anthropology*, 58(4), 435–453. <https://doi.org/10.1086/692905>.
- Köhler, W. (1929). *Gestalt psychology*. Liveright. <https://archive.org/details/gestaltpsycholog0000kohl/page/n5/mode/2up>
- Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition*, 114(1), 19–28. <https://doi.org/10.1016/j.cognition.2009.08.016>.
- Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50(2), 93–107. <https://doi.org/10.3758/BF03212211>.
- Kuhl, P. K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology*, 4(6), 812–822. [https://doi.org/10.1016/0959-4388\(94\)90128-7](https://doi.org/10.1016/0959-4388(94)90128-7).
- Laland, K. N. (2017). The origins of language in teaching. *Psychonomic Bulletin and Review*, 24(1), 225–231. <https://doi.org/10.3758/s13423-016-1077-7>.
- Levinson, S. C. (2022). The interaction engine: Cuteness selection and the evolution of the interactional base for language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1859), 1–7. <https://doi.org/10.1098/rstb.2021.0108>.
- Lockwood, G., Dingemanse, M., & Hagoort, P. (2016). Sound-symbolism boosts novel word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9763–9768. <https://doi.org/10.1037/xlm0000235.supp>.
- Ludwig, V. U., Adachi, I., & Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (pan troglodytes) and humans. *Proceedings of the National Academy of Sciences of the United States of America*, 108(51), 20661–20665. <https://doi.org/10.1073/pnas.1112605108>.
- Lupyan, G. (2007). Reuniting categories, language, and perception. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 29, 1247–1252.
- Lupyan, G. (2008). From chair to “chair”: A representational shift account of object Labeling effects on memory. *Journal of Experimental Psychology: General*, 137(2), 348–369. <https://doi.org/10.1037/0096-3445.137.2.348>.
- Lupyan, G. (2012). What do words do? Toward a theory of language-augmented thought. *Psychology of Learning and Motivation – Advances in Research and Theory*, 57, 255–297. <https://doi.org/10.1016/B978-0-12-394293-7.00007-8>.
- Lupyan, G., Abdel Rahman, R., Boroditsky, L., & Clark, A. (2020). Effects of language on visual perception. *Trends in Cognitive Sciences*, 24(11), 930–944. <https://doi.org/10.1016/j.tics.2020.08.005>.
- Lupyan, G., & Bergen, B. K. (2016). How language programs the mind. *Topics in Cognitive Science*, 8(2), 408–424. <https://doi.org/10.1111/tops.12155>.
- Lupyan, G., & Casasanto, D. (2015). Meaningless words promote meaningful categorization. *Language and Cognition*, 7(2), 167–193. <https://doi.org/10.1017/langcog.2014.21>.
- Lupyan, G., & Clark, A. (2015). Words and the world: Predictive coding and the language-perception-cognition Interface. *Current Directions in Psychological Science*, 24(4), 279–284. <https://doi.org/10.1177/0963721415570732>.
- Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science*, 18(12), 1077–1083. <https://doi.org/10.1111/j.1467-9280.2007.02028.x>.
- Lupyan, G., & Spivey, M. J. (2008). Now you see it, now you Don’t: Verbal but not visual cues facilitate visual object detection. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 30, 963–968.
- Lupyan, G., & Swingle, D. (2010). Self-directed speech alters visual processing. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 32(32), 1210–1215. <https://escholarship.org/uc/item/1wx3983m>
- Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General*, 170–186. <https://doi.org/10.1037/a0024904.supp>.

- Margiotoudi, K., Bohn, M., Schwob, N., Tagliatalata, J., Pulvermüller, F., Epping, A., Schweller, K., & Allritz, M. (2022). Bo-NO-bouba-kiki: Picture-word mapping but no spontaneous sound symbolic speech-shape mapping in a language trained bonobo. *Proceedings of the Royal Society B*, 289(1968), 1–9. <https://doi.org/10.1098/RSPB.2021.1717>.
- Meteyard, L., Stoppard, E., Snudden, D., Cappa, S. F., & Vigliocco, G. (2015). When semantics aids phonology: A processing advantage for iconic word forms in aphasia. *Neuropsychologia*, 76, 264–275. <https://doi.org/10.1016/j.neuropsychologia.2015.01.042>.
- Nielsen, A. K. S., & Dingemanse, M. (2021). Iconicity in word learning and beyond: A critical review. *Language and Speech*, 64(1), 52–72. <https://doi.org/10.1177/0023830920914339>.
- Nielsen, A. K. S., & Rendall, D. (2011). The sound of round: Evaluating the sound-symbolic role of consonants in the classic takete-maluma phenomenon. *Canadian Journal of Experimental Psychology*, 65(2), 115–124. <https://doi.org/10.1037/a0022268>.
- Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition*, 112(1), 181–186. <https://doi.org/10.1016/j.cognition.2009.04.001>.
- Perea, M., & Rosa, E. (2002). The effects of associative and semantic priming in the lexical decision task. *Psychological Research*, 66(3), 180–194. <https://doi.org/10.1007/s00426-002-0086-5>.
- Perlman, M., Dale, R., & Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. *Royal Society Open Science*, 2(8), 1–16. <https://doi.org/10.1098/rsos.150152>.
- Perlman, M., & Lupyan, G. (2018). People can create iconic vocalizations to communicate various meanings to naïve listeners. *Scientific Reports*, 8(1), 1–14. <https://doi.org/10.1038/s41598-018-20961-6>.
- Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: From a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), 1–14. <https://doi.org/10.1098/RSTB.2013.0300>.
- Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in english and Spanish and its relation to lexical category and age of acquisition. *PLoS One*, 10(9), 1–17. <https://doi.org/10.1371/journal.pone.0137147>.
- Prolific. (2021). <https://www.prolific.com>. [Computer Software]
- Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia—a window into perception, thought and language. *Journal of Consciousness Studies*, 8(12), 3–34.
- Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 55(4), 1339–1362. <https://doi.org/10.1080/02724980244000099>.
- Samaha, J., Boutonnet, B., Postle, B. R., & Lupyan, G. (2018). Effects of meaningfulness on perception: Alpha-band oscillations carry perceptual expectations and influence early visual responses. *Scientific Reports*, 8(1), 1–14. <https://doi.org/10.1038/s41598-018-25093-5>.
- Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology*, 12(3), 225–239. <https://doi.org/10.1037/h0070931>.
- Shen, Y.-C., Chen, Y.-C., & Huang, P.-C. (2022). Standard article i-perception. *I-Perception*, 13(2), 1–11. <https://doi.org/10.1177/20416695221084724>.
- Sidhu, D. M., & Pexman, P. M. (2017). A prime example of the Maluma/Takete effect? Testing for sound symbolic priming. *Cognitive Science*, 41(7), 1958–1987. <https://doi.org/10.1111/cogs.12438>.
- Sidhu, D. M., & Pexman, P. M. (2018). Five mechanisms of sound symbolic association. *Psychonomic Bulletin and Review*, 25(5), 1619–1643. <https://doi.org/10.3758/s13423-017-1361-1>.
- Sidhu, D. M., Vigliocco, G., & Pexman, P. M. (2020). Effects of iconicity in lexical decision. *Language and Cognition*, 12(1), 164–181. <https://doi.org/10.1017/langcog.2019.36>.
- Solov'ev, A. A., Guseva, S. A., & Shramko, A. D. (2023). The magnetic field of a coronal hole in the Heliosphere: The Inverse Square law. *Geomagnetism and Aeronomy*, 63(8), 1238–1247. <https://doi.org/10.1134/S0016793223080212>.
- SONA-Systems. (n.d.). *SONA Systems Participant Management Software*. <https://cam.sona-systems.com>. [Computer Software]
- Spence, C., & Parise, C. V. (2012). The cognitive neuroscience of crossmodal correspondences. *I-Perception*, 3(7), 410–412. <https://doi.org/10.1068/i0540ic>.
- Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, 42(4), 1096–1104. <https://doi.org/10.3758/BRM.42.4.1096>.
- Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1), 24–31. <https://doi.org/10.1177/0098628316677643>.

Stosic, M. D., Murphy, B. A., Duong, F., Fultz, A. A., Harvey, S. E., & Bernieri, F. (2024). Careless responding: Why many findings are spurious or spuriously inflated. *Advances in Methods and Practices in Psychological Science*, 7(1), 1–19. <https://doi.org/10.1177/25152459241231581>.

Sweeny, T. D., Guzman-Martinez, E., Ortega, L., Grabowecy, M., & Suzuki, S. (2012). Sounds exaggerate visual shape. *Cognition*, 124(2), 194–200. <https://doi.org/10.1016/j.cognition.2012.04.009>.

Tallerman, M. (2014). No syntax saltation in language evolution. *Language Sciences*, 46, 207–219. <https://doi.org/10.1016/j.LANGSCI.2014.08.002>.

Teufel, C., Dakin, S. C., & Fletcher, P. C. (2018). Prior object-knowledge sharpens properties of early visual feature-detectors. *Scientific Reports*, 8(1), 1–12. <https://doi.org/10.1038/s41598-018-28845-5>.

Ward, M. K., & Meade, A. W. (2023). Dealing with careless responding in survey data: Prevention, Identification, and recommended best practices. *Annual Review of Psychology*, 74, 2022. <https://doi.org/10.1146/annurev-psych-040422>.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of the United States of America*, 104(19), 7780–7785. <https://doi.org/10.1073/pnas.0701644104>.

Winter, B., Woodin, G., & Perlman, M. (2023). Defining iconicity for the cognitive sciences. In *The Oxford handbook of iconicity in language* (pp. 1–38). Oxford University Press. <https://doi.org/10.31219/osf.io/5e3rc>

Yentes, R., & Wilhelm, F. (2023). *careless: Procedures for computing indices of careless responding* (1.2.2). [Computer Software]

Appendix

Training task

Table A1. Accuracy ~ condition×target distance + condition×trial + (1|target distance:item) + (1|condition: participant)

Accuracy	Estimate	Std. error	z-value	Pr(> z)	Significance	OR
(Intercept)	0.8501047	0.055222	15.394	<2.00E–16	***	2.34
Contrast1 (label vs no label)	–0.0364923	0.0324978	–1.123	2.61E–01		0.96
Contrast2 (iconic vs non-iconic)	–0.0067126	0.0495441	–0.135	0.892226		0.99
Target distance	–0.3447893	0.0267865	–12.872	<2.00E16	***	0.71
Trial	0.0036652	0.0006176	5.935	2.94E–09	***	1.00
Contrast1 (label vs no label):target distance	0.0198488	0.0098117	2.023	0.043076	*	1.02
Contrast2 (iconic vs non-iconic):target distance	0.0206277	0.0144955	1.423	0.154723		1.02
Contrast1 (label vs no label):trial	–0.0008674	0.0002543	–3.412	0.000646	***	1.00
Contrast2 (iconic vs non-iconic):trial	0.000931	0.0003807	2.445	0.014466	*	1.00

Condition was coded with the orthogonal contrasts 1 (iconic and non-iconic label conditions vs no label condition) and 2 (iconic label vs non-iconic label conditions).

Table A2. RT ~ condition×target distance + condition×trial + (1|target distance:item) + (1|condition: participant)

RT data	Estimate	Std. error	df	t value	Pr(> t)	Significance
(Intercept)	3.02E+00	7.53E−03	1.09E+02	401.378	<2.00E−16	***
Contrast1 (label vs no label)	−1.49E−02	5.52E−03	9.99E+01	−2.696	0.00823	**
Contrast2 (iconic vs non- iconic)	2.83E−03	8.46E−03	1.00E+02	0.335	0.73849	
Target distance	9.55E−03	1.40E−03	1.73E+02	6.802	1.60E−10	***
Trial	−7.64E−04	3.78E−05	3.00E+02	−20.212	<1.00E−16	***
Contrast1 (label vs no label): target distance	2.41E−04	6.87E−04	1.17E+04	0.35	0.72603	
Contrast2 (iconic vs non- iconic):target distance	−4.00E−04	1.06E−03	1.17E+04	−0.377	0.70602	
Contrast1 (label vs no label): trial	−4.73E−05	2.01E−05	1.17E+04	−2.354	0.01857	*
Contrast2 (iconic vs non- iconic):trial	−6.69E−06	3.08E−05	1.17E+04	−0.217	0.82788	

Condition was coded with the orthogonal contrasts 1 (iconic and non-iconic label conditions vs no label condition) and 2 (iconic label vs non-iconic label conditions).

MTS task

Table A3. Accuracy ~ condition×target distance + (1|target distance:item) + (1|condition:participant)

Effect	Estimate	Std. error	z value	Pr(> z)	Significance	OR
(Intercept)	0.727218	0.062287	11.675	<2.00E−16	***	2.07
Contrast1 (label vs no label)	0.043965	0.015208	2.891	0.00384	**	1.04
Contrast2 (iconic vs non- iconic)	0.026095	0.039201	0.666	0.50562		1.03
Contrast3 (online vs offline)	0.206378	0.039212	5.263	1.42E−07	***	1.23
Target distance	−0.087047	0.031279	−2.783	0.00539	**	0.92
Contrast1 (Label vs no label):target distance	−0.008723	0.004102	−2.127	0.03344	*	0.99
Contrast2 (iconic vs non- iconic):target distance	−0.007388	0.010888	−0.678	0.49746		0.99
Contrast3 (online vs offline):target distance	−0.054824	0.010914	−5.023	5.08E−07	***	0.95

Condition was coded with the orthogonal contrasts 1 (Iconic and Non-Iconic Label conditions vs No Label condition), 2 (Iconic Label vs Non-Iconic Label conditions) and 3 (Online vs Offline conditions).

Table A4. RT ~ condition×target distance + (1|target distance:item) + (1|condition:participant)

Effect	Estimate	Std. error	df	t value	Pr(> t)	Significance
(Intercept)	3.14E+00	1.36E−02	1.09E+02	230.358	<2.00E−16	***
Contrast1 (label vs no label)	−4.38E−03	6.01E−03	1.00E+02	−0.729	0.467581	
Contrast2 (iconic vs non- iconic)	−3.31E−02	1.54E−02	9.99E+01	−2.15	0.033955	*
Contrast3 (online vs offline)	−3.53E−02	1.54E−02	9.99E+01	−2.296	0.023777	*
Target distance	1.15E−02	1.77E−03	2.00E+02	6.514	5.78E−10	***
Contrast1 (label vs no label): target distance	8.33E−04	3.73E−04	1.26E+04	2.234	0.025518	*
Contrast2 (iconic vs non- iconic):target distance	1.25E−04	9.28E−04	1.26E+04	0.134	0.893247	
Contrast3 (online vs offline): target distance	3.11E−03	9.29E−04	1.26E+04	3.351	0.000807	***

Condition was coded with the orthogonal contrasts 1 (iconic and non-iconic label conditions vs no label condition), 2 (iconic label vs non-iconic label conditions) and 3 (online vs offline conditions).

Iconic label generation

The pretest survey ran as follows. Participants were first presented with an image of the exemplar alien from the round and dark category. They were then visually presented with a list of possible labels for this alien category. All participants were presented with the same list of labels, but the order was randomised for each participant. Participants were asked to think about how each label would sound and then to rank the labels from most to least fitting for this alien. The phrasing of the instructions was deliberately kept neutral in an attempt to avoid introducing bias and allow for immediate associations. This process was then repeated for the light and spiky alien exemplar, with a separate list of possible labels. Finally, participants were presented with a list of possible non-iconic labels and asked to select two labels that fit neither alien more than the other.

The pretesting proceeded in two steps. First, an initial sample of 10 participants completed the pretest survey. This contained 14 possible round and dark category iconic labels, and 15 possible spiky and light category iconic labels. We used this pretest to reduce the number of possible label options to give more meaningful results. Hence, we retained the top five label choices from each category for the next stage.

We then reran the survey with 9 of the original 10 participants, plus 22 new participants. This survey was identical to the first but included 5 possible round and dark category iconic labels, 5 possible spiky and light category iconic labels and 31 potential non-iconic labels. We selected final iconic labels with the highest average rank and non-iconic labels that were selected most commonly, while avoiding any phonemic overlap between labels within a condition. For example, the potential non-iconic labels ‘flep’ and ‘stoise’ were equally commonly selected after ‘phrav’. Given the phonemic similarities between ‘flep’ and ‘phrav’, we opted to include only ‘phrav’ and to select ‘stoise’ as our second non-iconic label.