

Supplementary Information for

Flexible Usage of Object and Global Scene Information During Human Scene Categorization

Sandro L. Wiesmann and Melissa L.-H. Vö

Department of Psychology, Johann Wolfgang Goethe-Universität, Frankfurt, Germany

Correspondence concerning this article should be addressed to Sandro L. Wiesmann,
Department of Psychology, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 6, 60323
Frankfurt am Main, Germany. Email: wiesmann@psych.uni-frankfurt.de

Note: In this document, we report results of the original experiments reported in the initial submission of the article cited above. After reviewers raised methodological concerns regarding these experiments, we decided to rerun all experiments and report only the revised experiments in the published article. The original experiments are reported here for the sake of transparency alone. We advise the readers to interpret results with caution. Figure and table identifiers refer to this document only.

This document contains information on the methods and results of the four original experiments reported in the initial submission of the manuscript entitled *Flexible Usage of Object and Global Scene Information During Human Scene Categorization* (Wiesmann & Vö, 2025). The first round of reviews revealed substantial methodological concerns by the reviewers regarding this initial set of experiments, especially the use of pink noise masks and the question whether they equally mask object and texture stimuli. Additionally, reviewers pointed out the unequal forward and backward masking of stimuli in Experiment 3 and potential pop-out effects of object stimuli. We therefore decided to address these criticisms by rerunning all experiments with different masks, forward masks in Experiment 3 and more rigorous controls of stimulus luminance. The published article therefore contains only the revised experiments (Experiment 1 in the published manuscript jointly addresses the questions of original Experiments 1 and 4). The original experiments are reported here for the sake of transparency alone. We advise the readers to interpret their results with caution. The theoretical background of this study is outlined in the published manuscript. Figure and table identifiers refer to this document only (supplementary tables and figures are reported at the end of the document).

Experiment 1

In Experiment 1, we compared the utility of object and global scene information for human scene categorization and assessed whether a combination of both types of information is sufficient to explain the human ability to effortlessly categorize real-world scenes. We created six different stimulus conditions: two different global scene information textures with varying spatial resolutions based on an algorithm introduced by Oliva and Torralba (2006; see also Brady et al., 2017; Wiesmann & Vö, 2022), versions of the original stimuli that were reduced to single objects (see Wiesmann & Vö, 2023), combinations of the two scene textures with a single object from the same scene, and the original scenes. By directly comparing the scene categorization accuracy for these conditions within subjects, we were able to draw

conclusions about the relative utility of object and global scene information and their combination for human scene categorization.

Method

Participants

A simulation-based power analysis based on preliminary data using the *mixedpower* package for R (Kumle et al., 2020) showed that a sample size of 30 participants yielded sufficient power ($> .80$) to test the differences between conditions we were interested in. We recruited a total of 32 German-speaking undergraduate students to participate in Experiment 1 for course credit. Data were collected in 2021. We discarded data from two participants who had not performed the task conscientiously (as indicated by less than 50% correct trials in the original scene control condition). Data from the remaining 30 participants were used for analysis (27 female, 3 male; 18–38 years old, $M = 23.3$ years, $SD = 4.7$). All participants reported normal or corrected-to-normal vision, were naïve with regard to the experimental hypotheses and the stimulus material, and provided informed consent before the study. The study was conducted in accordance with guidelines provided by the Goethe University's ethics committee (ethics approval ID# 2014-106R1).

Stimuli

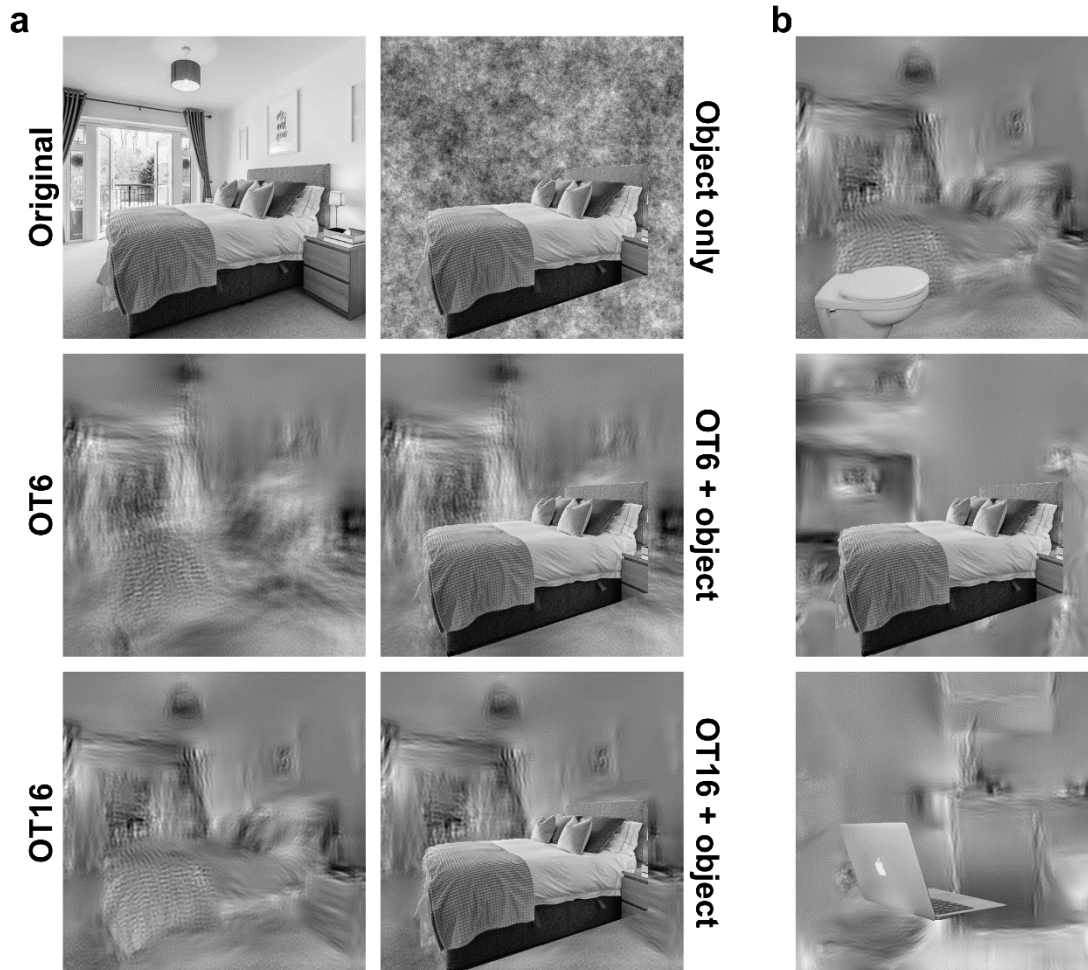
We used the same base stimulus set described in previous work (Wiesmann & Vö, 2022, 2023). Stimuli consisted of 240 black and white photographs of indoor and outdoor scenes belonging to 16 different basic-level categories (indoor: bathroom, bedroom, kitchen, living room, classroom, office, restaurant, shop; outdoor: beach, desert, forest, mountain, city, highway, playground, train station). All stimuli were 512 x 512 pixels in size. We created six different versions of each scene image (see Figure 1a): (1) A scene texture based on the algorithm by Oliva and Torralba (2006) preserving global information of the scene with a grid size of 6×6 windows (hereafter referred to as *OT6*); (2) analogously, a texture with a grid size of 16×16 windows (*OT16*); (3) a version showing only a single object cut-out from the

original image; (4) a combination of the single object with the OT6 texture in the background; (5) a combination of the single object with the OT16 texture in the background; and (6) the original image.

The OT6 and OT16 textures were created by dividing each input image into a grid of 6×6 or 16×16 windows, respectively, and then calculating the image energy at four spatial scales and six orientations in each window (for details see Brady et al., 2017; Wiesmann & Vö, 2022). A random white noise image was then iteratively matched to the parameters

Figure 1

Examples of Stimuli Used in (a) Experiments 1, 3 and 4 and (b) Experiment 2.



Note. In Experiment 1, all six stimulus categories shown in (a) were used. Note that object and texture combinations always contained information from the same scene. In Experiments 3 and 4, we only used Object only and OT16 stimuli. In Experiment 2, inconsistent object-texture combinations were presented (b). Object information from one scene was always combined with an OT16 texture background of a scene from another category in that experiment. See text for details. Original images in this figure taken from pixabay.com.

extracted within each window. The resulting textures thus preserve varying degrees of spatial resolution (the OT16 stimuli having a higher resolution due to the smaller grid windows) and therefore presumably different degrees of spatial layout information. The objects used in conditions (3-5) were hand-labelled by the first author using the LabelMe online tool (Russell et al., 2008). Afterwards, the objects were either pasted on a texture background or, in the object-only condition, on a pink-noise background to create comparable visual clutter around the objects. When pasting the cut-out objects on the different backgrounds, they were always reproduced in their original size and position.

Apparatus

The experiment was carried out online using jsPsych (version 6.1.0, de Leeuw, 2015). Participants were required to use a computer or laptop with a screen size of at least 14” and a physical keyboard. The size of all displayed elements was calibrated using a built-in calibration feature of jsPsych. The absolute size of all stimuli was thus independent of the physical screen size and resolution and the same for all participants (approximately 14 cm x 14 cm). Note, however, that we did not control viewing distance.

Procedure

All instructions were presented on screen in writing. Participants performed a speeded 16AFC scene categorization task on stimuli showing different combinations of object and global scene information. Each trial consisted of the same elements: a fixation cross (jittered presentation time between 1000 and 1500 ms), a randomly selected stimulus in one of the six stimulus conditions (50 ms), a blank screen (50 ms), and a pink noise mask (100 ms). Afterwards, participants first categorized the stimulus they saw at the superordinate level (indoor vs. outdoor) and then at the basic level (e.g., bathroom, kitchen, etc.) using the number keys on their keyboard. Last, participants indicated how confident they were that their categorization was correct on a scale from 1 (*not confident at all*) to 6 (*very confident*). The categorization task and confidence rating were not temporally constrained. Before the

experimental trials, participants completed 16 practice trials in which unprocessed scene stimuli were presented and another 16 trials with scene stimuli in one of the six stimulus conditions described above. During the practice trials, participants received visual feedback on the accuracy of their categorization. The main experiment comprised 240 trials and took approximately 30 minutes to complete.

Analysis

We discarded single trials with response times more than three standard deviations from the participant's individual mean (3.6%), leaving data from 6941 trials for analysis. We then set up generalized linear mixed effect models (GLMMs) to test the effect of the stimulus condition on superordinate- and basic-level categorization accuracy. Following Barr et al. (2013), we set the random effects structure of all models maximal, including a random intercept for participants (30), basic-level scene categories (16) and stimuli (240). Random slopes for the stimulus condition were specified within participants and basic-level scene categories. To test whether superordinate-level and basic-level categorization accuracy were significantly different from chance, we set up zero-intercept models and compared the estimated accuracies for each condition against the theoretical chance level of 50% and 6.25%, respectively, using Wald z tests. To test for differences between the individual stimulus conditions, we calculated pairwise comparisons between accuracy scores using the Holm-Bonferroni correction. We set up an analogous cumulative link mixed model (CLMM) using the ordinal package (Christensen, 2022) to assess the effect of the stimulus condition on confidence ratings. To test the effect of the covariates object size and eccentricity on categorization accuracy, we set up three additional models using only data from trials in which an object was presented (i.e., the object only, OT6 + object, OT16 + object conditions). In model M1, we only entered stimulus condition as a predictor. In model M2, we added the covariates object size and eccentricity and in M3 also the interactions between condition, object size and eccentricity.

Results

Observed superordinate- and basic-level categorization accuracies as well as confidence ratings are summarized in Figure 2 (for detailed modelling results see Supplementary Tables S1 and S2). As expected, the texture-only conditions yielded the lowest categorization accuracies, and the OT6 textures were the only condition with categorization accuracies that did not significantly differ from chance (both at the superordinate and the basic categorization level) whereas OT16 textures were significantly more useful than OT6 textures. Single objects allowed for significantly higher categorization accuracy than OT6 and OT16 textures, but the latter only at the basic categorization level. Apparently, when participants only had to decide whether a scene is indoors or outdoors (superordinate-level categorization), the utility of a single object was not significantly higher compared to the rough layout information contained in the OT16 textures.

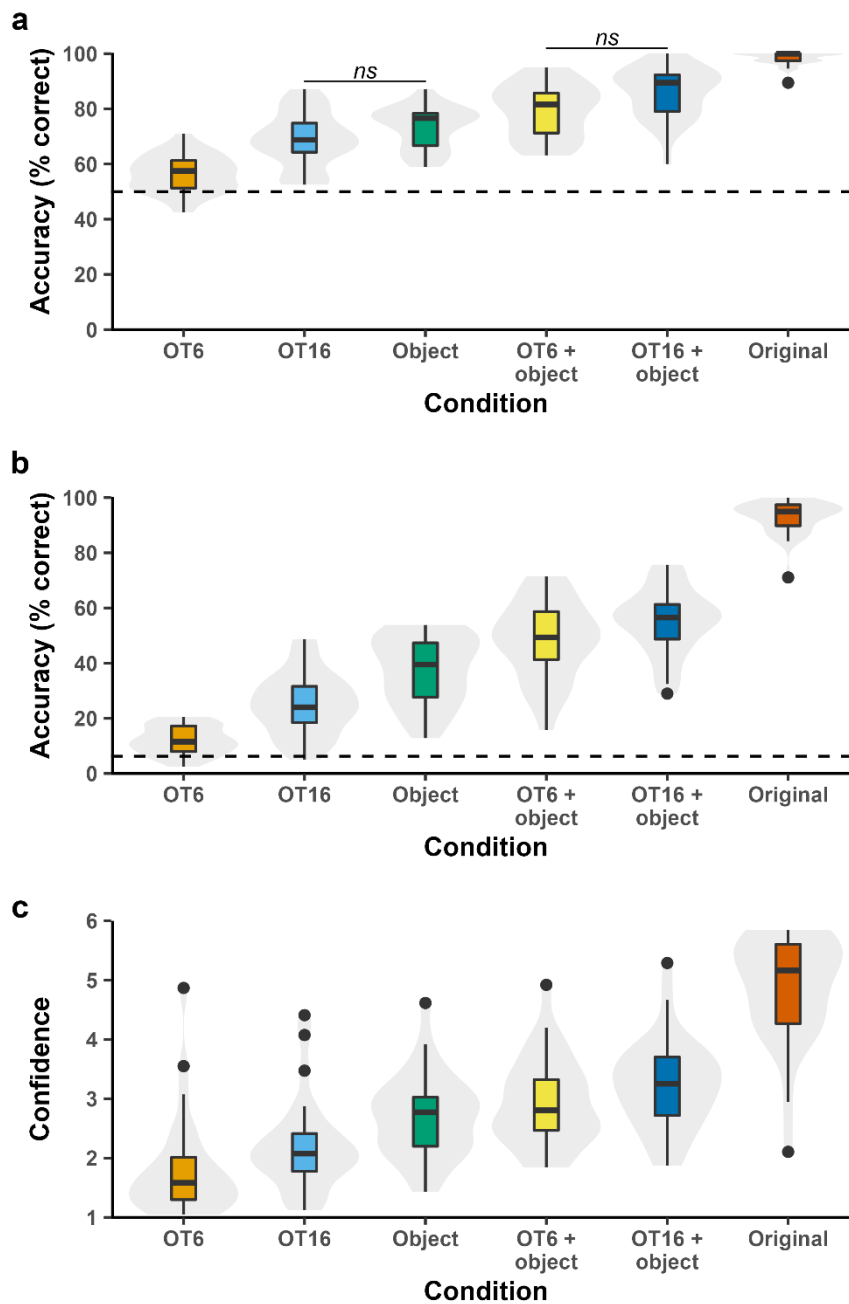
Performance further increased significantly in the conditions in which we combined a single object in full resolution with a texture background, both in comparison to the texture-only conditions and the single-object condition. However, because of the baseline differences between the object-only and the texture-only conditions, the increase was numerically smaller relative to the object-only condition as compared to the texture-only conditions. Within the texture + object conditions, the stimuli with OT16 textures in the background yielded higher accuracies than those with OT6 backgrounds, but this difference was only significant at the basic categorization level. And, importantly, in both the OT6 + object and the OT16 + object conditions (mean basic-level categorization accuracy 48.4% and 55.2%, respectively), performance was still significantly worse than in the original scenes (92.7%). This indicates that combining information of an object in full resolution with a background containing primarily spatial layout information of the scene is not sufficient to explain fast scene categorization. In line with the above results of basic-level categorization accuracy, confidence ratings also differed significantly between the conditions (see Figure 2c). Object

images yielded higher confidence ratings than texture images, and confidence increased even further when object and texture information was combined, but it did not reach the confidence level of the original scenes.

Lastly, we were interested in exploring the effects of the covariates object size and eccentricity on basic-level categorization accuracy in those conditions in which an object was presented (see Supplementary Figure S1). Not surprisingly, predicted accuracies for stimuli containing large objects presented close to the screen centre were close to 100%, and for these stimuli adding a texture background improved performance only marginally. For larger eccentricities and smaller object sizes, the effect of adding the texture background was more pronounced (i.e., there was a stronger increase of categorization performance from the object only to OT6 + object and OT16 + object conditions). Interestingly, predicted accuracy was always numerically above chance except for the smallest object size and the largest eccentricity in our dataset, implying that even smaller objects or objects presented in the periphery are useful for scene categorization.

Figure 2

Observed (a) Superordinate-Level Categorization Accuracy, (b) Basic-Level Categorization Accuracy and (c) Confidence Ratings in Experiment 1 by Stimulus Condition.



Note. All contrasts between conditions were significant except where noted otherwise ($\alpha = .05$, ns = non-significant comparisons). The dashed lines represent the respective chance level for each task (superordinate-level categorization: 50%; basic-level categorization: 6.25%).

Experiment 2

Experiment 1 implies that object information is more useful for participants in a scene categorization task than global scene information. In Experiment 2, we specifically tested this ‘object advantage’ hypothesis. We created stimuli containing object and global scene information with conflicting semantic meaning (e.g., a fridge from a kitchen scene in full resolution on top of the texture of a train station). By enforcing a choice between the two presented categories (here *kitchen* vs. *train station*) in a 2AFC task, we were able to test whether participants preferred object over global scene information in a basic-level scene categorization task. Secondly, we were interested in studying the time course of this preference. Specifically, we assessed whether the preference for object relative to global scene information changed when we manipulated the stimulus-onset asynchrony (SOA) between stimulus and mask. According to global-to-local theories of scene perception, global scene information is used before local information in the scene categorization process, implying that global scene information should be available and used even at short SOAs. We would then expect that participants base their categorizations more on global scene information at short SOAs and increasingly use object information given longer SOAs. In contrast, if information usage were to follow a local-to-global pattern, thus implying that object information is available with shorter SOAs compared to global scene information, we would expect decreasing usage of object information across SOAs. We assessed these alternative hypotheses by presenting the semantically inconsistent object-texture combinations described above with varying SOAs between stimulus and mask, thus testing the object-preference hypothesis derived from Experiment 1 and assessing the time course of this preference.

Method

The method of Experiment 2 was identical to that of Experiment 1 except for the deviations noted below.

Participants

A simulation-based power analysis using preliminary data showed that a sample size of 16 participants already yielded sufficient power ($> .99$) to test the effect of SOA on the tendency to choose the object or texture category. However, since this experiment was run 2021 as part of a psychology undergraduate class in which students participated for course credit, we recruited a total of 51 German-speaking volunteers to participate in Experiment 2. We discarded data from seven participants due to technical issues (inaccurate presentation times on more than 10% of trials) and two participants who had not performed the task conscientiously (as indicated by more than 10% of trials with response times under 100 ms). Data from the remaining 42 participants were used for analyses (33 female, 9 male; 18–30 years old, $M = 21.2$ years, $SD = 3.0$). None of the participants of Experiment 1 participated in Experiment 2.

Stimuli

We used the same base stimulus set of 240 black and white scene images as in Experiment 1. We created inconsistent combinations of single objects and OT16 scene textures by pairing a single cut-out object of one scene with the texture of another scene from a different basic-level scene category (e.g., a stove from a kitchen scene with a texture background from a desert scene; see Figure 1b). For each basic-level category, the 15 object stimuli of that category were always paired with one texture stimulus of each of the remaining 15 scene categories. We thus ensured that there was no systematic bias in the stimuli due to texture-object combinations from specific categories. We further systematically counterbalanced the specific object-texture pairings across participants to avoid biases created by single objects that fit specific backgrounds very well. Lastly, we also counterbalanced the specific object-texture pairings and SOAs across participants.

Apparatus

The experiment was carried out online using jsPsych (version 6.3.0, de Leeuw, 2015) and the jspsych-psychophysics plugin (Kuroki, 2021) to ensure accurate presentation times and SOAs. Before the experiment, participants were additionally required to check and confirm that their screen was set to a refresh rate of 60 Hz. The actual refresh rate was calculated by the jspsych-psychophysics plugin and verified before data analysis (see below).

Procedure

Participants performed a speeded 2AFC scene categorization task on stimuli showing combinations of inconsistent object and global scene information with varying SOAs between stimulus and mask. Note that the SOA reflects the time participants have to sample information from the retina (i.e., stimulus presentation time plus inter-stimulus-interval between stimulus and mask). As VanRullen (2011) points out, a manipulation of this time is not an index of the timing of underlying brain processes and only reflects differences in the amount of information needed to perform a task. Each trial comprised a fixation cross (jittered presentation time between 1000 and 1500 ms), a randomly selected stimulus showing a single object of one scene on top of a texture background of a scene from another scene category (33 ms), a blank screen (0, 33, 67 or 100 ms), and a pink noise mask (67 ms). Afterwards, the names of two basic-level scene categories were presented on the left and right of the screen centre, respectively, and participants were instructed to indicate as quickly as possible which category the stimulus they had seen belonged to by pressing the X or M keys on the keyboard. Crucially, the two displayed category labels were always the category of the scene the object was cut-out from (e.g., ‘kitchen’ for an oven) and the category of the scene the texture was created from (e.g., ‘desert’). We will refer to the choice of either category as an ‘object response’ or ‘texture response’ throughout this article. Last, participants indicated how confident they were that their response was correct on a scale from 1 (*not confident at all*) to 6 (*very confident*).

Before the experimental trials, participants completed two blocks of 16 practice trials each. In the first block, we presented unprocessed scene stimuli, and there was always one correct category label and one randomly selected distractor category. Here, participants received visual feedback on the accuracy of their categorization. In the second block, mismatching object and texture information was presented as in the rest of the experiment. Since both response options were ‘correct’ in this block (one referred to the scene category of the presented object, the other to that of the presented texture), no feedback was presented to avoid biasing participants towards the object or texture category. After the main experiment, participants saw all stimuli again and rated how well the presented objects fit the background on a scale from 1 (*not fitting at all*) to 6 (*very fitting*). When they were not able to perceive an object within the stimulus, participants were instructed to press the 0 key (this was sometimes true for smaller objects that ‘blended in’ with the textures). Completion of the experiment and rating took approximately 40 minutes in total.

Analysis

We excluded trials with imprecise presentation times (average single frame duration longer than 18 ms or presentation times of single screens more than 16.5 ms longer than specified in the experimental protocol; 0.9%), trials with response times more than three standard deviations from the participants’ individual mean (1.4%) and trials in which participants reported not to be able to perceive an object (1.2%; see consistency rating described above), leaving data from 9728 trials for analysis. We used GLMMs for analyses of the proportion of object responses and CLMMs for analyses of confidence ratings. To test the object-preference hypothesis, we set up a zero-intercept model testing the proportion of object responses at each SOA against the theoretical chance level of 50% using Wald z tests. To test the effect of the SOA between stimulus and mask on category choice (scene category of the object vs. the texture), we set up three different models. In model M1, we only entered SOA as a predictor. In M2, we added the covariates object size and eccentricity and in M3 also the

interactions between SOA, object size and eccentricity. To test differences in confidence ratings between object and texture responses, we set up a CLMM estimating contrasts between both responses at each SOA. In all models, we included separate random effects for participants (42), objects (240) and textures (240). Random slopes for all predictors and their interactions were specified within all random grouping variables. The exact model specifications are available in the Online Supplement.

Results

Results of Experiment 2 are summarized in Figure 3 (for detailed modelling results see Supplementary Tables S3–S7). Note that in our 2AFC categorization task participants always had to choose between the category of the scene from which we created the texture and the category of the scene from which we cut out the single object. Throughout this article, we will refer to the choice of either category as a ‘texture response’ or ‘object response’. Since the proportions of both responses are complementary, we will only explicitly report the proportion of object responses in this section.

In line with the object-preference hypothesis we derived from Experiment 1, there was a significant preference for the object category across SOA conditions. However, contrary to the marked difference between the OT16 and the object-only condition in Experiment 1 (24.1% vs. 35.2% basic-level categorization accuracy, respectively), the preference for the object condition in Experiment 2 was rather small (55.8% object responses across SOAs vs. 50% expected by chance). Even more surprisingly, and in contrast to predictions made by global-to-local theories of scene categorization, the preference for the object category decreased significantly over time. Importantly, this effect was significant even when we controlled for object size and eccentricity (M2), but when we included interactions between SOA, object size and eccentricity (M3), we found that the effect of SOA depended on the object size (see model predictions in Figure 3c). Larger objects showed high rates of object responses with slight increases or no changes across SOAs whereas small objects showed

decreasing rates of object responses across SOAs. Note that the proportion of object responses in Figure 3c appears overall higher compared to the observed data because object size and eccentricity were not evenly distributed in the dataset (see Figure 3b). More than 75% of the objects covered less than 10% of the image pixels, and more than 80% of objects had their centroid more than 100 pixels from the fixation point (roughly 2.7 cm on screen). As a result, the average data we observe are much closer to the chance line (compare to the orange and grey lines of predicted data in the middle panel of Figure 3c).

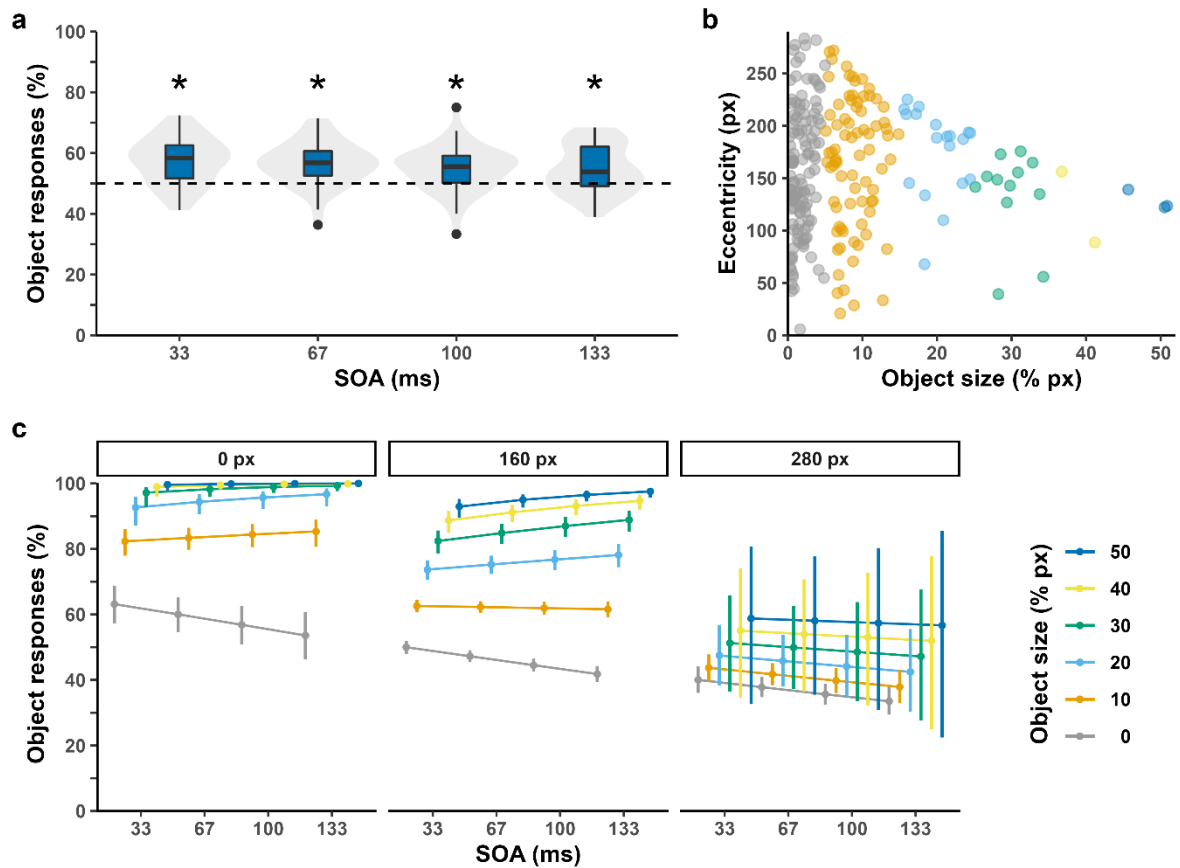
Since the overall difficulty and the choice between only two categories in the 2AFC task may have distorted our results and lead to adoption of a guessing strategy, we repeated our analyses with a subset of our data discarding all trials in which participants had responded with a confidence rating of 1, indicating that they were ‘not confident at all’ in the accuracy of their response (26.7% of data, see Supplementary Table S6). Note that this simple confidence rating of course does not imply that these data are not meaningful. Some visual information probably still entered the participants’ visual systems, and participants may simply be unsure how to integrate conflicting global scene properties and object information. However, since the proportion of object/texture responses was at chance across all SOAs for these trials, we decided to compare our results to an analysis of only trials in which participants did not feel like they were only guessing (i.e., with confidence ratings > 1). Indeed, the proportion of object responses in this analysis was slightly higher, especially for the lower SOAs, and the decline of object responses across SOAs was stronger (compare Supplementary Tables S4 and S5). However, the variance was also higher in this analysis due to irregular dropout of data from single participants in single SOAs. We therefore focus on the findings based on the complete data presented above but maintain that our results are not purely the result of random guessing and that the negative trend across SOAs is reliable.

We further compared the confidence ratings for object and texture responses across SOAs (see Supplementary Figure S2 and Supplementary Table S6) and found that object

responses were related to higher confidence ratings compared to texture responses at SOAs between 33-100 ms. However, at the 133 ms SOA, the longest SOA in our experiment, the confidence ratings did not differ significantly between object and texture responses anymore. This observation implies that participants were able to reach confident categorization

Figure 3

Results of Experiment 2. (a) Observed Proportion of Trials in Which Participants Chose the Scene Category of the Presented Object as a Function of SOA. (b) Distribution of Object Size and Eccentricity in the Stimulus Set. (c) Predicted Proportion of Trials in Which Participants Chose the Scene Category of the Presented Object as a Function of SOA, Object Size (Colours) and Eccentricity (Panels).



Note. Since participants always had the choice between the scene category of the presented object and the presented texture in a 2AFC task, the proportion of texture responses is complementary to the proportion of object responses.

(a) The dashed line represents the expected chance level in the 2AFC task (50%). Asterisks indicate proportions significantly different from chance ($\alpha = .05$). (c) The three selected levels of eccentricity (0, 160 and 280 px) represent the approximate minimum, mean and maximum in the stimulus set, respectively. The object sizes 0% and 50% represent the rounded minimums and maximums in the stimulus set, respectively. Error bars represent standard errors of the mean. Note that error bars for some combinations of object size and eccentricity are larger due to the irregular distribution of these features in our stimulus set (see b).

decisions with shorter presentation times when basing their choice on the object as compared to global scene information in the image. Interestingly, the consistency ratings also turned out to be a significant predictor in our models (see Supplementary Table S7). When participants perceived the object-texture combinations as consistent (i.e., the object matched the texturized scene background), they were more likely to choose the object category. Importantly, we still observed decreasing usage of object information across SOAs after controlling for the effect of consistency.

Experiment 3

In Experiment 2, we assessed the time course of object and global scene information usage by manipulating the SOA between stimuli containing category-inconsistent object and global scene information and a mask. Surprisingly, and in contrast to the notion that information usage during scene categorization proceeds in a global-to-local manner, we found that participants rather made use of object information at low SOAs and increasingly used global scene information at longer SOAs. In Experiment 3, we therefore directly compared scene categorization performance when stimuli were presented in a local-to-global vs. a global-to-local sequence (i.e., either object information before global scene information or vice versa). If information usage during scene categorization indeed follows a local-to-global pattern, we would expect better performance following stimuli presented in an object-texture sequence.

Method

The method of Experiment 3 was identical to that of Experiment 1 except for the deviations noted below.

Participants

A simulation-based power analysis using preliminary data showed that a sample size of 45 participants yielded sufficient power to test the effect of the stimulus sequence on response time and categorization accuracy. We recruited a total of 48 German-speaking

volunteers among undergraduate students and via Prolific to participate in Experiment 3 for course credit or a financial compensation of £6. Data were collected in 2022. We discarded data from three participants who had not performed the task conscientiously (as indicated by mean categorization accuracies of 55% and below compared to a theoretical chance level of 50%). Data from the remaining 45 participants were used for analysis (33 female, 12 male; 18–40 years old, $M = 24.1$ years, $SD = 6.6$). None of the participants of Experiment 1 and 2 participated in Experiment 3.

Stimuli

In Experiment 3, we only used the OT16 texture and Object only stimuli (see Experiment 1 and Figure 1a).

Apparatus

The experiment was carried out online using jsPsych (version 6.3.0, de Leeuw, 2015) and the jspsych-psychophysics plugin (Kuroki, 2021) to ensure accurate presentation times and SOAs. Before the experiment, participants were additionally required to check and confirm that their screen was set to a refresh rate of 60 Hz. The actual refresh rate was calculated by the jspsych-psychophysics plugin and verified before the experiment.

Procedure

Participants performed a speeded 2AFC scene categorization task on scene stimuli that were either presented in an object-texture sequence (e.g., a bed cropped from a bedroom scene first, followed by the texture representing the global scene information of the same bedroom image) or a texture-object sequence (e.g., the texture of the bedroom first, followed by the image of the bed; see Figure 4). After the fixation cross (jittered presentation time between 1000 and 1600 ms), a randomly selected stimulus sequence was presented (object-texture or texture-object) consisting of the first stimulus of the sequence, a blank screen, the second stimulus, another blank screen (33 ms each) and finally a pink noise mask (67 ms). Afterwards, the names of two basic-level scene categories were presented on the left and right

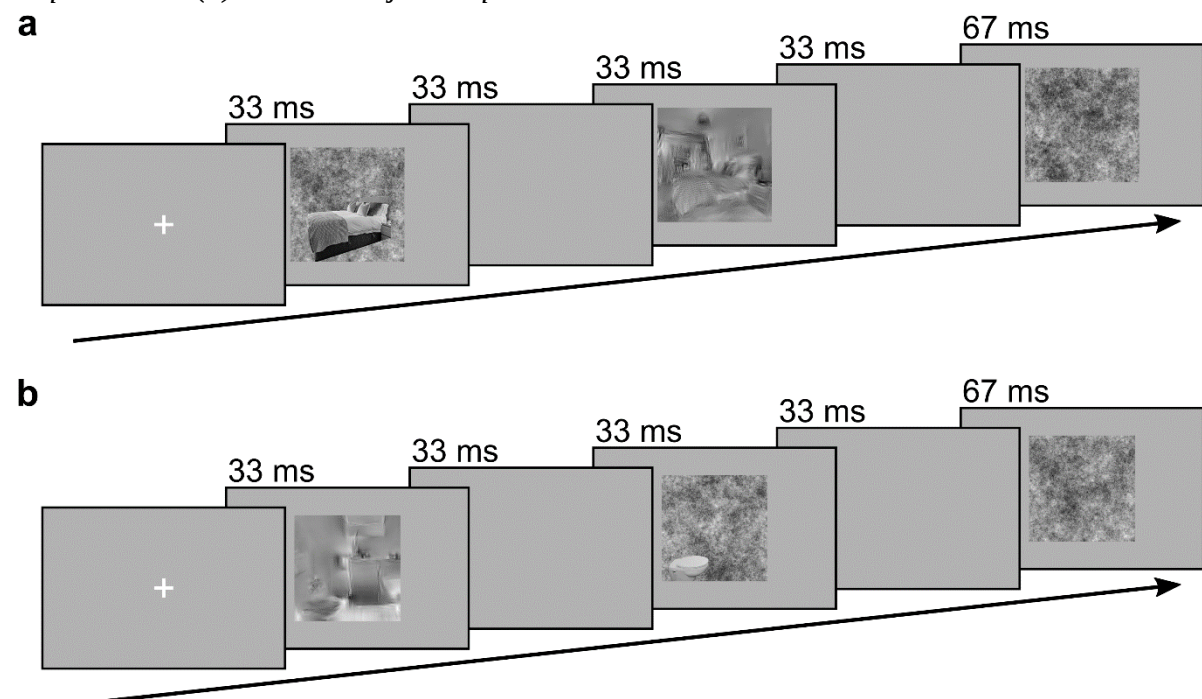
of the screen centre, respectively, and participants were instructed to indicate as quickly as possible which category the stimulus they had seen belonged to by pressing the X or M key on the keyboard. A randomly selected distractor category was presented on each trial. The distractor categories for each individual image and the allocation of images to the two different presentation sequences were counterbalanced across participants. Last, participants indicated how confident they were that their response was correct on a scale from 1 (*not confident at all*) to 6 (*very confident*). As in Experiment 2, participants completed two blocks of 16 practice trials each before the experimental trials, one consisting of single, unprocessed scene stimuli, and one consisting of stimuli presented in object-texture and texture-object sequences. Completion of the experiment took approximately 40 minutes.

Analysis

We discarded trials in which the average single frame duration was longer than 18 ms ($< 0.1\%$ of data) and trials with response times more than three standard deviations from the

Figure 4

Two Different Stimulus Presentation Sequences Used in Experiment 3: (a) Object-Texture Sequence and (b) Texture-Object Sequence.



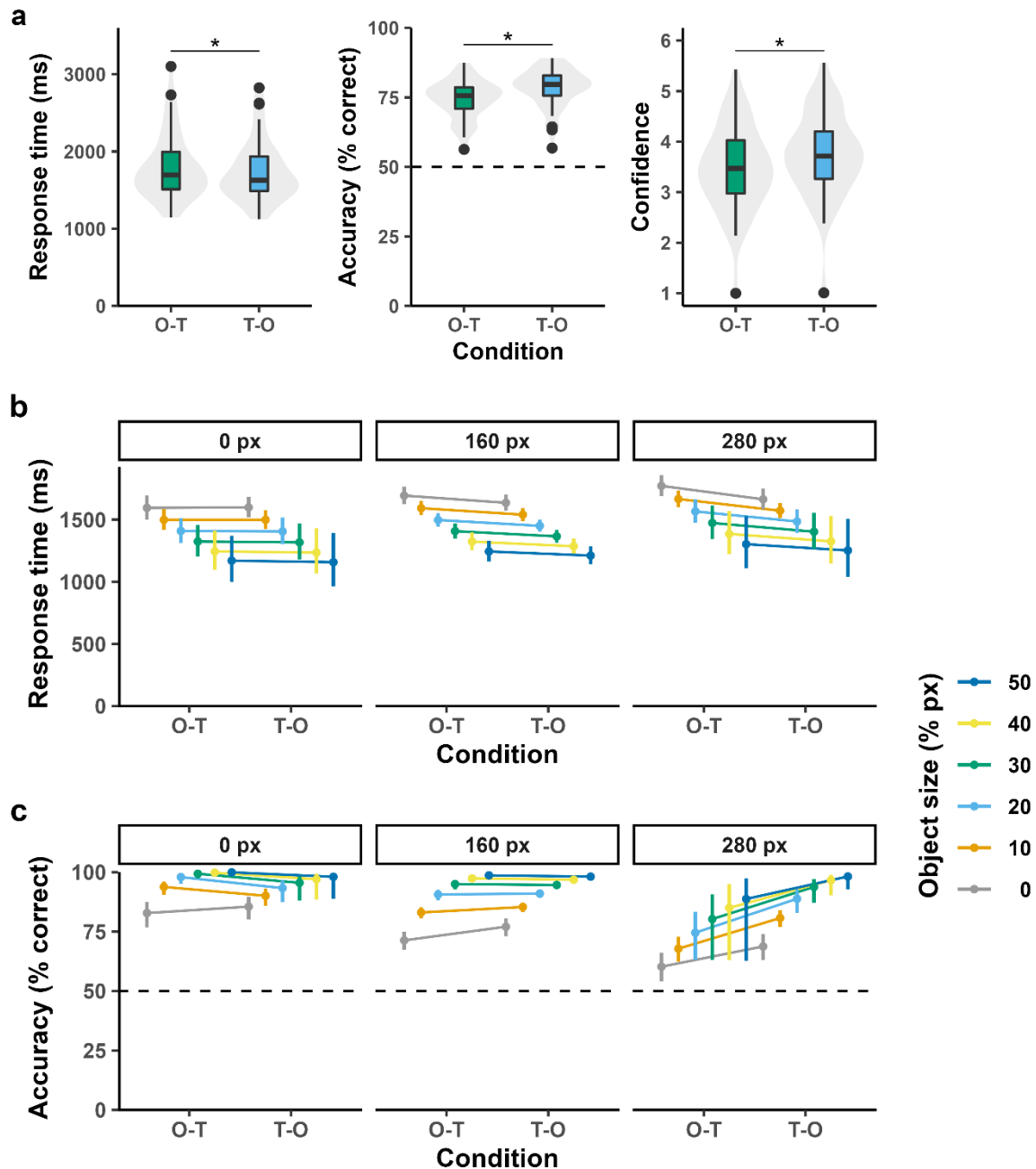
participants' individual mean (1.4%), leaving data from 10649 trials for analysis. We assessed the effect of the presentation sequence (object-texture vs. texture-object) on log response times of correct trials, categorization accuracy and confidence ratings using linear mixed models (LMMs), GLMMs and CLMMs, respectively. As in Experiment 2, we initially set up models only containing the predictor presentation sequence (M1), and then added the covariates object size and eccentricity (M2) and their interactions (M3). We included separate random effects for participants (45), basic-level scene categories (16) and stimuli (240). We specified random slopes for all predictors in a model within participants and basic-level categories (see Online Supplement for final model specifications). Note that there is currently no consensus on how to calculate p values for LMMs. However, given the size of our dataset, we treat effects with $|t| > 2$ as significant (see also Kliegl et al., 2010).

Results

The observed response times, scene categorization accuracies and confidence ratings as a function of the presentation sequence are summarized in Figure 5a (for detailed modelling results see Supplementary Tables S8–S10). In line with the notion of a default global-to-local pattern of information usage during scene categorization, and in contrast to our findings from Experiment 2, we observed significantly faster, more accurate and more confident responses following texture-object sequences as compared to object-texture responses. Crucially, however, we observed a significant interaction between the presentation sequence and eccentricity in the accuracy model (see Figure 5c). While the model predicted an overall pattern of better performance for texture-object compared to object-texture sequences when objects were presented in the periphery, this pattern was less clear when objects were presented close to the screen centre. A similar trend could be observed in the response time data (see Figure 5b), but it did not reach significance.

Figure 5

Results of Experiment 3. (a) Observed Response Time, Basic-Level Categorization Accuracy and Confidence Ratings for Object-Texture and Texture-Object Stimulus Sequences. Predicted (b) Response Time and (c) Basic-Level Categorization Accuracy for Object-Texture and Texture-Object Stimulus Sequences as a Function of Object Size and Eccentricity.



Note. (a) Asterisks indicate significant differences between conditions ($\alpha = .05$). (b, c) The three selected levels of eccentricity (0, 160 and 280 px) represent the approximate minimum, mean and maximum in the stimulus set, respectively. The object sizes 0% and 50% represent the rounded minimums and maximums in the stimulus set, respectively. Error bars represent standard errors of the mean. Note that error bars for some combinations of object size and eccentricity are larger due to the irregular distribution of these features in our stimulus set (see Figure 3b). The dashed line in (c) represents the chance level for the basic-level categorization task (50%). O-T = object-texture stimulus sequence, T-O = texture-object stimulus sequence.

Experiment 4

In Experiment 2 and 3, we found variable usage of object and global scene information across SOAs. Intriguingly, our results suggest that global scene information is available even at short SOAs (otherwise it could not facilitate subsequent object processing in Experiment 3), but in looking at results from Experiment 2, it appears that the global information extracted at short SOAs is itself not sufficient to allow for access to the scene category (as indicated by the persisting object bias in that experiment). When conflicting object and texture information was presented in Experiment 2, participants may initially have relied on object information more because it was more useful (as indicated by a significant object bias across SOAs) and only started to use texture information increasingly when presentation times were longer and the participants' confidence in their texture responses increased (as indicated by significantly lower confidence ratings following texture responses at shorter SOAs).

In Experiment 4, we sought to reconcile the findings of Experiments 2 and 3 by assessing the ability to extract scene category information from object and global scene information at different presentation times. Specifically, we manipulated the SOA between presentation of an object or texture stimulus and a visual mask. Based on Experiments 2 and 3 and our discussion of the results, we expected that (1) object information should lead to higher categorization accuracies than global scene information at lower SOAs (thus explaining the early object bias in Experiment 2 and supporting the hypothesis that global scene information rather has a facilitating effect on object information usage in Experiment 3), (2) global scene information should show a marked increase of categorization accuracy at higher SOAs (thus explaining increasing usage of global scene information over time in Experiment 2), and (3) object information should lead to higher confidence ratings than global scene information at lower SOAs (thus explaining the higher importance of object

information in Experiments 2 and 3 and supporting the hypothesis that global scene information rather has a facilitating effect on object information usage in Experiment 3).

Method

The method of Experiment 4 was identical to that of Experiment 3 except for the deviations noted below.

Participants

A simulation-based power analysis using preliminary data showed that a sample size of 40 participants yielded sufficient power to test the effect of the stimulus condition within each level of the factor SOA on categorization accuracy. We recruited a total of 46 German-speaking volunteers among undergraduate students and via Prolific to participate in Experiment 4 for course credit or a financial compensation of £9. Data were collected in 2023. We discarded data from four participants due to technical issues (inaccurate presentation times on more than 10% of trials) and two participants who had not performed the task conscientiously (as indicated by mean basic-level categorization accuracies $< 7\%$ at a theoretical chance level of 6.25%). Data from the remaining 40 participants were used for analysis (17 female, 22 male, 1 non-binary; 19–36 years old, $M = 26.0$ years, $SD = 4.9$). None of the participants of Experiments 1–3 participated in Experiment 4.

Stimuli

We used the same OT16 texture and Object only stimuli as in Experiment 3 (see Figure 1a). Note, however, that we did not present sequences of stimuli in Experiment 4.

Procedure

Participants performed a speeded 16AFC scene categorization task on OT16 texture and object-only stimuli that were presented with varying SOAs between stimulus and mask (33, 50, 67, 83, 100, 117, 133 and 250 ms). We selected SOAs in the range of Experiments 2 and 3 (i.e., 33–133 ms) and additionally a longer SOA (250 ms) to test whether scene categorization performance reaches a plateau after this time. Each trial comprised a fixation

cross (jittered presentation time between 1000 and 1600 ms), a randomly selected stimulus presented either as a texture or object only (33 ms), followed by a blank screen (0-217 ms) and a pink noise mask (67 ms). Afterwards, participants categorized the scene stimulus at the superordinate and basic level and rated their confidence (see above). Each participant saw all 240 scenes both as a texture and an object stimulus for a total of 480 trials. The texture and object trials were randomly interleaved, but the allocation of the images to the eight different SOAs and their presentation order (i.e., whether an individual scene was first presented as a texture or object stimulus within the experiment) were counterbalanced across participants. As in previous experiments, participants completed two blocks of 16 practice trials each before the experimental trials, one consisting of single, unprocessed scene stimuli, and one consisting of scenes reduced to object and texture stimuli. Completion of the experiment took approximately 60 minutes.

Analysis

We discarded trials in which the average single frame duration was longer than 18 ms (1.6% of data) and trials with response times more than three standard deviations from the participants' individual mean (1.7%), leaving data from 18564 trials for analysis. We used GLMMs to analyse accuracy data and CLMMs to analyse confidence ratings. To test whether basic-level categorization accuracy was significantly different from chance, we set up a zero-intercept model and compared the estimated accuracies against the theoretical chance level of 6.25% using Wald z tests. To test the effect of the stimulus type (object vs. texture) on categorization accuracy across SOAs, we first set up a model in which we entered SOA (continuous), the stimulus condition and their interaction. To specifically assess accuracy and confidence differences between object and texture stimuli at each SOA, we set up additional models in which we entered SOA as a factor and had the model directly estimate contrasts between object and texture stimuli at each SOA. We included separate random effects for participants (40), basic-level scene categories (16) and stimuli (240) in all models and always

specified random slopes for all predictors within participants and basic-level categories. Note that, due to the high computational demands of the complex models we tested here, we suppressed correlations between random effects in these models (see Online Supplement for final model specifications).

Results

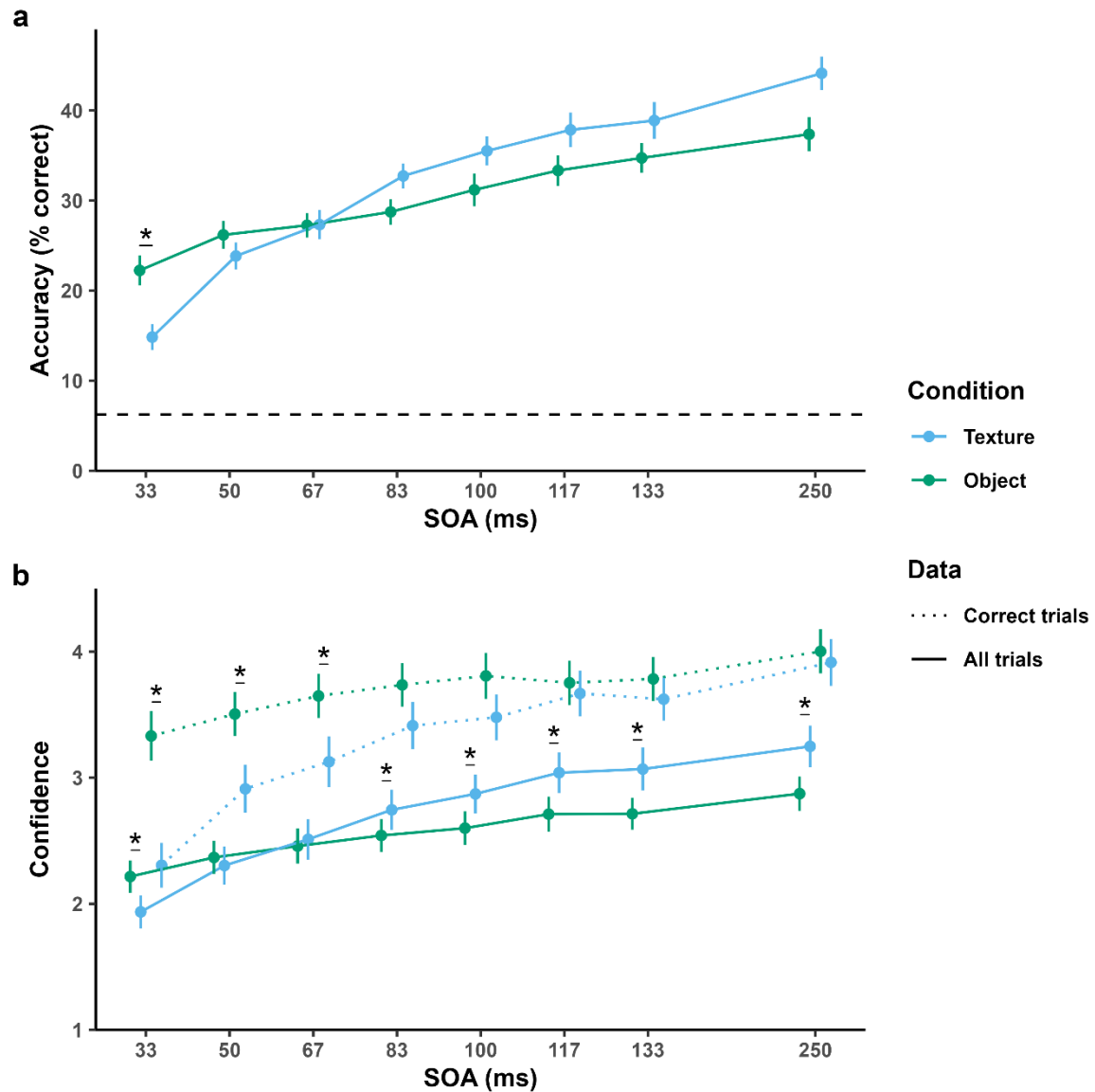
Observed scene categorization accuracies and confidence ratings as a function of stimulus condition and SOA are summarized in Figure 6 (for detailed modelling responses see Supplementary Tables S11 and S12). We observed above-chance basic-level categorization accuracy at all SOAs except for texture stimuli at an SOA of 33 ms. We further found a significant effect of SOA on basic-level categorization accuracy but no significant difference between stimulus conditions (at the mean of the scaled SOA predictor). However, there was a significant positive interaction between both predictors, indicating that there was a stronger increase in categorization performance across SOAs for texture compared to object stimuli (see Figure 6a). In looking at differences between objects and textures at each SOA, we found that objects yielded significantly higher basic-level categorization accuracy than textures at an SOA of 33 ms but not at higher SOAs. We further found significantly lower confidence ratings for textures compared to objects at an SOA of 33 ms but also significantly higher confidence ratings for textures compared to objects at SOAs of 83 ms and above. However, when we only considered confidence ratings of correct trials, we found significantly lower confidence ratings for textures compared to objects at SOAs between 33 and 67 ms but no significant differences at higher SOAs (see Figure 6b).

Discussion

In Experiment 4, we tested the participants' ability to categorize object and texture stimuli given different presentation times. In line with the notion of local-to-global information usage, we observed higher scene categorization accuracy and confidence ratings for object information compared to global scene information at low SOAs and a stronger

Figure 6

Observed (a) Basic-Level Categorization Accuracy and (b) Confidence Ratings in Experiment 4 by SOA and Stimulus Condition.



Note. Asterisks indicate significant differences between conditions within an SOA ($\alpha = .05$). The dashed line in (a) represents the chance level in the basic-level categorization task (6.25%). Confidence ratings in (b) are shown for all trials (solid lines) and correct trials only (dotted lines). Error bars represent standard errors of the mean. Note that we truncated y axes for better readability.

increase of global scene information utility compared to object information across SOAs. To our knowledge, our study is the first to provide causal evidence for lower information sampling demands for basic-level scene categorization based on object compared to global scene information (i.e., the extraction of object information for the purpose of scene categorization). Previous studies either assessed object recognition instead of scene categorization as the dependent variable (for overviews see Fabre-Thorpe, 2011; Joubert et al., 2007; Mack & Palmeri, 2011), correlated object and global scene information with behavioural performance and neural responses (e.g., Greene & Hansen, 2020) or examined scene-incongruity effects which may not be based on the same information as scene categorization (Gagne & MacEvoy, 2014). However, while the findings of Experiment 4 are consistent with local-to-global theories of scene perception, they do not provide evidence for the assumption that processing of object information precedes processing of global scene information at the neural level. They only suggest that shorter presentation times are sufficient to sample diagnostic object information compared to global scene information in our stimuli (Mack & Palmeri, 2011; VanRullen, 2011).

Our results further support our interpretation of Experiment 2 by showing that object information, compared to global scene information, leads to scene category access with shorter presentation times (which is reflected in the bias towards using object information at low SOAs in Experiment 2) and that participants need longer presentation times to reach the same accuracy and confidence levels using global scene information (which is reflected in the increasing usage of global scene information across SOAs in Experiment 2). The low categorization accuracy and confidence ratings for texture stimuli at an SOA of 33 ms further support our interpretation of Experiment 3 in which we related better performance for texture-object sequences to facilitating effects of global scene information rather than scene category access based on global scene information.

References

- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Brady, T. F., Shafer-Skelton, A., & Alvarez, G. A. (2017). Global ensemble texture representations are critical to rapid scene perception. *Journal of Experimental Psychology: Human Perception and Performance*, 43(6), 1160–1176. <https://doi.org/10.1037/xhp0000399>
- Christensen, R. H. B. (2022). *ordinal—Regression Models for Ordinal Data* (Version 2022.11-16) [Computer software]. <https://CRAN.R-project.org/package=ordinal>
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47(1), 1–12. <https://doi.org/10.3758/s13428-014-0458-y>
- Fabre-Thorpe, M. (2011). The characteristics and limits of rapid visual categorization. *Frontiers in Psychology*, 2, 243. <https://doi.org/10.3389/fpsyg.2011.00243>
- Gagne, C. R., & MacEvoy, S. P. (2014). Do Simultaneously Viewed Objects Influence Scene Recognition Individually or as Groups? Two Perceptual Studies. *PLoS ONE*, 9(8), e102819. <https://doi.org/10.1371/journal.pone.0102819>
- Greene, M. R., & Hansen, B. C. (2020). Disentangling the Independent Contributions of Visual and Conceptual Features to the Spatiotemporal Dynamics of Scene Categorization. *The Journal of Neuroscience*, 40(27), 5283–5299. <https://doi.org/10.1523/JNEUROSCI.2088-19.2020>
- Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research*, 47(26), 3286–3297. <https://doi.org/10.1016/j.visres.2007.09.013>

- Kliegl, R., Masson, M. E. J., & Richter, E. M. (2010). A linear mixed model analysis of masked repetition priming. *Visual Cognition*, 18(5), 655–681.
<https://doi.org/10.1080/13506280902986058>
- Kumle, L., Võ, M. L.-H., & Draschkow, D. (2020). *Mixedpower: A library for estimating simulation-based power for mixed models in R* (Version Version 0.1.0) [Computer software]. <https://doi.org/10.5281/zenodo.3733023>
- Kuroki, D. (2021). A new jsPsych plugin for psychophysics, providing accurate display duration and stimulus onset asynchrony. *Behavior Research Methods*, 53(1), 301–310.
<https://doi.org/10.3758/s13428-020-01445-w>
- Mack, M. L., & Palmeri, T. J. (2011). The Timing of Visual Object Categorization. *Frontiers in Psychology*, 2, 165. <https://doi.org/10.3389/fpsyg.2011.00165>
- Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research*, 155, 23–36.
[https://doi.org/10.1016/S0079-6123\(06\)55002-2](https://doi.org/10.1016/S0079-6123(06)55002-2)
- Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). LabelMe: A database and web-based tool for image annotation. *International Journal of Computer Vision*, 77(1–3), 157–173. <https://doi.org/10.1007/s11263-007-0090-8>
- VanRullen, R. (2011). Four Common Conceptual Fallacies in Mapping the Time Course of Recognition. *Frontiers in Psychology*, 2, 365.
<https://doi.org/10.3389/fpsyg.2011.00365>
- Wiesmann, S. L., & Võ, M. L.-H. (2022). What makes a scene? Fast scene categorization as a function of global scene information at different resolutions. *Journal of Experimental Psychology: Human Perception and Performance*, 48(8), 871–888.
<https://doi.org/10.1037/xhp0001020>

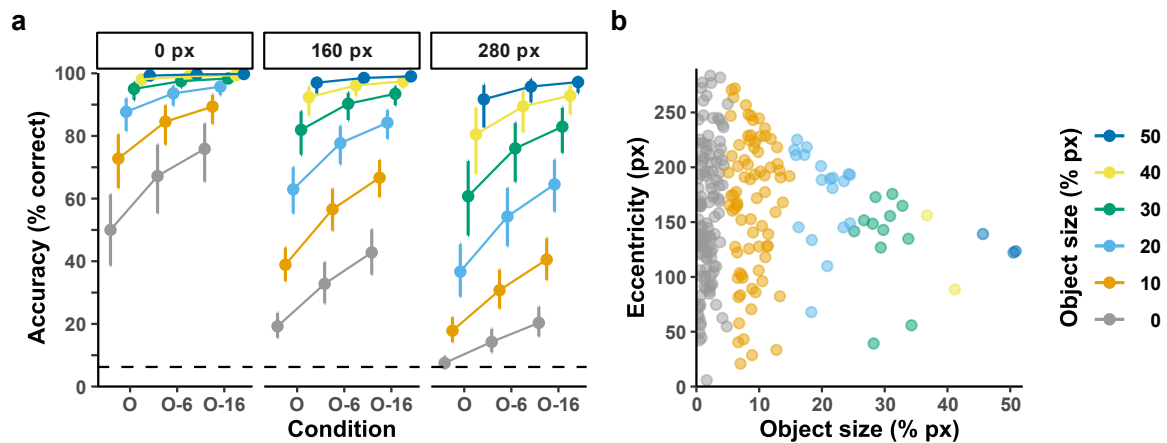
Wiesmann, S. L., & Võ, M. L.-H. (2023). Disentangling diagnostic object properties for human scene categorization. *Scientific Reports*, 13(1), 5912.

<https://doi.org/10.1038/s41598-023-32385-y>

Supplementary Figures and Tables for the Analysis of the Original Experiments

Supplementary Figure S1

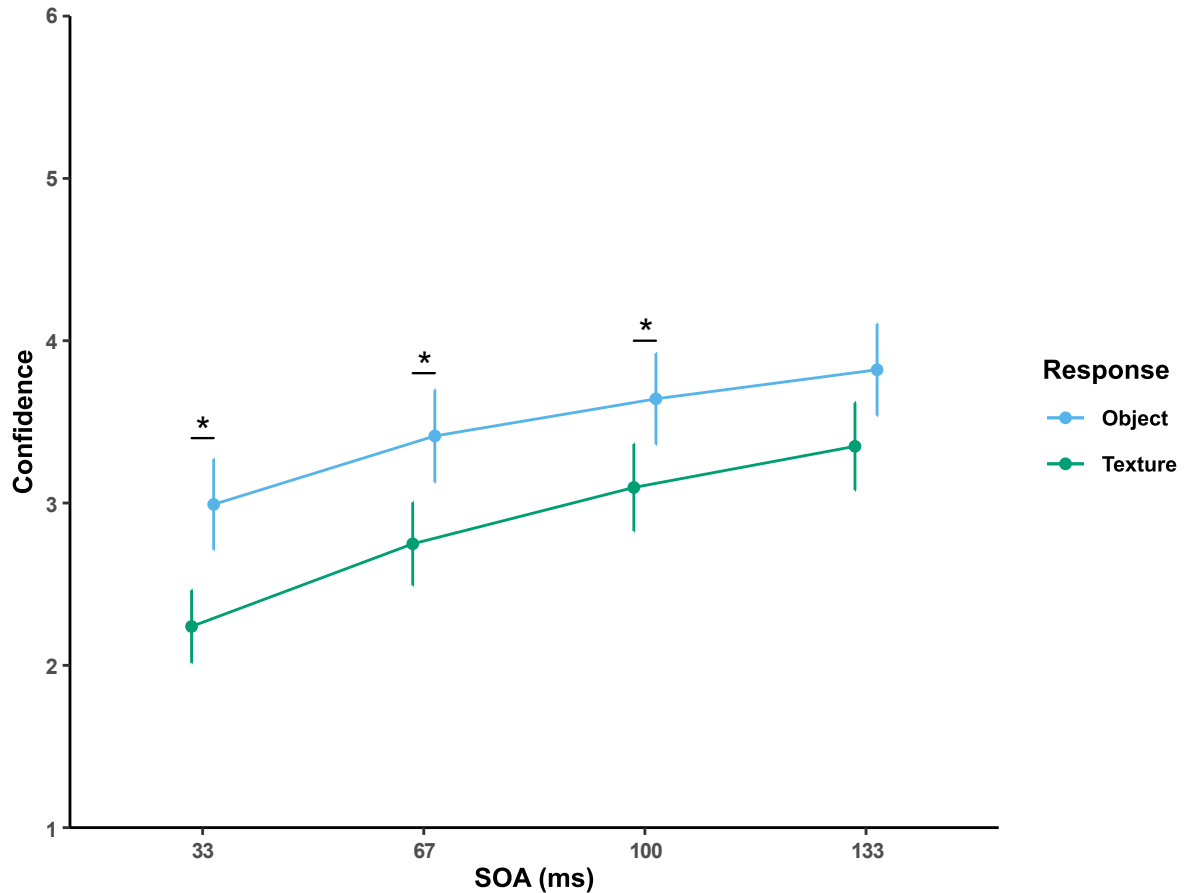
Effect of object size and eccentricity in Experiment 1. (a) Predicted scene categorization accuracy as a function of stimulus condition, object size and eccentricity. (b) Distribution of object size and eccentricity in the stimulus set.



Note. The three selected levels of eccentricity (0, 160 and 280 px) represent the approximate minimum, mean and maximum in the stimulus set, respectively. The object sizes 0% and 50% represent the rounded minimums and maximums in the stimulus set, respectively. Error bars represent standard errors of the mean. O = Object only, O-6 = OT6 + Object, O-16 = OT16 + Object.

Supplementary Figure S2

Mean confidence ratings following object and texture responses at different SOAs in Experiment 2.



Note. Asterisks indicate significant differences between object and texture responses within an SOA ($\alpha = .05$). Texture and object response refer to the scene category selected in the 2AFC scene categorization task on stimuli containing object information from one scene (e.g., a car from a city scene) and an inconsistent global scene information texture from a scene from another category (e.g., a restaurant). In this case, categorizing the stimulus as a city would be considered an object response and categorizing it as a restaurant would be considered a texture response.

Supplementary Table S1

Test of estimated superordinate-level and basic-level categorization accuracy against chance in Experiment 1.

Condition	Superordinate level				Basic level			
	Accuracy	SE	<i>z</i>	<i>p</i>	Accuracy	SE	<i>z</i>	<i>p</i>
OT6	.593	0.358	1.05	.295	.049	0.387	-0.68	.751
OT16	.735	0.270	3.78	<.001	.164	0.334	3.24	<.001
Object	.772	0.226	5.42	<.001	.335	0.222	9.12	<.001
OT6 + Object	.864	0.309	5.97	<.001	.480	0.258	10.18	<.001
OT16 + Object	.909	0.273	8.44	<.001	.567	0.245	12.16	<.001
Original	.992	0.476	10.16	<.001	.962	0.240	24.76	<.001

Note. The theoretical chance levels at the superordinate and basic categorization level were 50% and 6.25%, respectively.

Supplementary Table S2

Pairwise comparisons of categorization accuracies and confidence ratings in the different stimulus conditions in Experiment 1.

Contrast	Superordinate-level cat. acc.				Basic-level categorization acc.				Confidence			
	Estimate	SE	<i>z</i>	<i>p</i>	Estimate	SE	<i>z</i>	<i>p</i>	Estimate	SE	<i>z</i>	<i>p</i>
OT6 – OT16	-0.647	0.204	-3.18	.004	-1.345	0.289	-4.66	<.001	-0.773	0.130	-5.96	<.001
OT6 – Object	-0.850	0.202	-4.20	<.001	-2.285	0.383	-5.96	<.001	-1.408	0.189	-7.45	<.001
OT6 – OT6.Object	-1.473	0.195	-7.57	<.001	-2.890	0.363	-7.95	<.001	-1.752	0.208	-8.44	<.001
OT6 – OT16.Object	-1.931	0.309	-6.25	<.001	-3.240	0.345	-9.40	<.001	-2.095	0.192	-10.93	<.001
OT6 – Original	-4.456	0.503	-8.87	<.001	-6.212	0.518	-11.99	<.001	-4.286	0.383	-11.19	<.001
OT16 – Object	-0.203	0.137	-1.48	.138	-0.940	0.303	-3.11	.004	-0.636	0.118	-5.38	<.001
OT16 – OT6.Object	-0.826	0.151	-5.48	<.001	-1.546	0.267	-5.79	<.001	-0.979	0.140	-7.00	<.001
OT16 – OT16.Object	-1.284	0.188	-6.82	<.001	-1.896	0.234	-8.11	<.001	-1.323	0.134	-9.85	<.001
OT16 – Original	-3.809	0.463	-8.23	<.001	-4.867	0.402	-12.12	<.001	-3.513	0.318	-11.04	<.001
Object – OT6.Object	-0.623	0.167	-3.72	<.001	-0.605	0.130	-4.65	<.001	-0.343	0.093	-3.69	<.001
Object – OT16.Object	-1.080	0.188	-5.74	<.001	-0.955	0.143	-6.67	<.001	-0.687	0.112	-6.13	<.001
Object – Original	-3.606	0.471	-7.65	<.001	-3.927	0.291	-13.50	<.001	-2.878	0.250	-11.49	<.001
OT6.Object – OT16.Object	-0.457	0.232	-1.97	.098	-0.350	0.115	-3.05	.004	-0.344	0.098	-3.52	<.001
OT6.Object – Original	-2.982	0.493	-6.05	<.001	-3.322	0.317	-10.47	<.001	-2.534	0.231	-10.95	<.001
OT16.Object – Original	-2.525	0.490	-5.16	<.001	-2.972	0.314	-9.46	<.001	-2.191	0.215	-10.21	<.001

Note. All *p* values are corrected using the Holm-Bonferroni correction to keep the family-wise error rate at $\alpha = .05$.

Supplementary Table S3

Test of the estimated proportion of object responses against chance across SOAs in Experiment 2.

SOA	Proportion of object responses	SE	<i>z</i>	<i>p</i>
33 ms	.588	0.073	4.91	<.001
67 ms	.579	0.079	4.01	<.001
100 ms	.565	0.095	2.76	.006
133 ms	.554	0.086	2.53	.011

Note. The theoretical chance level was 50%.

Supplementary Table S4

Models assessing the effect of SOA and the covariates object size and eccentricity on the proportion of object responses in Experiment 2.

Predictors	M1				M2				M3			
	Logit	CI	<i>z</i>	<i>p</i>	Logit	CI	<i>z</i>	<i>p</i>	Logit	CI	<i>z</i>	<i>p</i>
(Intercept)	0.29	0.15 – 0.42	4.05	<.001	0.31	0.18 – 0.44	4.80	<.001	0.36	0.22 – 0.49	5.22	<.001
SOA	-0.05	-0.10 – -0.00	-1.97	.048	-0.06	-0.12 – -0.01	-2.29	.022	-0.04	-0.10 – 0.02	-1.28	.199
Object size					0.49	0.35 – 0.64	6.63	<.001	0.62	0.45 – 0.80	6.91	<.001
Eccentricity					-0.30	-0.38 – -0.21	-6.80	<.001	-0.38	-0.49 – -0.27	-6.56	<.001
Object size * Eccentricity									-0.23	-0.44 – -0.02	-2.10	.036
Object size * SOA									0.10	0.02 – 0.18	2.51	.012
Eccentricity * SOA									-0.03	-0.10 – 0.05	-0.74	.462
Object size * Eccentricity * SOA									-0.04	-0.17 – 0.08	-0.66	.507

Note. All predictors in these analyses were standardized. The nonsignificant effect of SOA in M3 is therefore the effect of SOA for an object with an average size and eccentricity. As can be seen in Figure 4b and the interaction between object size and SOA, SOA still showed a negative effect for smaller objects that made up the majority of our stimuli (see Figure 4c).

Supplementary Table S5

Models assessing the effect of SOA and the covariates object size and eccentricity on the proportion of object responses only for trials with confidence ratings > 1 in Experiment 2.

Predictors	M1				M2				M3			
	Logit	CI	<i>z</i>	<i>p</i>	Logit	CI	<i>z</i>	<i>p</i>	Logit	CI	<i>z</i>	<i>p</i>
(Intercept)	0.40	0.21 – 0.59	4.20	<.001	0.46	0.28 – 0.64	5.12	<.001	0.53	0.34 – 0.71	5.53	<.001
SOA	-0.11	-0.18 – -0.04	-3.09	.002	-0.12	-0.20 – -0.05	-3.33	.001	-0.10	-0.18 – -0.02	-2.44	.015
Object size					0.73	0.53 – 0.92	7.30	<.001	0.87	0.65 – 1.09	7.64	<.001
Eccentricity					-0.39	-0.50 – -0.27	-6.76	<.001	-0.49	-0.63 – -0.34	-6.43	<.001
Object size * Eccentricity									-0.26	-0.53 – 0.01	-1.87	.062
Object size * SOA									0.06	-0.05 – 0.17	1.03	.303
Eccentricity * SOA									-0.03	-0.12 – 0.07	-0.57	.568
Object size * Eccentricity * SOA									-0.09	-0.26 – 0.07	-1.07	.283

Note. Trials in which participants rated their confidence as “not confident at all” made up 26.7% of data. All predictors in these analyses were standardized.

Supplementary Table S6

Model assessing confidence rating differences between object and texture responses at each SOA in Experiment 2.

SOA	Logit	CI	<i>z</i>	<i>p</i>
33	-0.56	-0.84 – -0.28	-3.92	<.001
67	-0.44	-0.68 – -0.20	-3.64	<.001
100	-0.21	-0.41 – -0.01	-2.11	.035
133	-0.15	-0.37 – 0.08	-1.25	.211

Supplementary Table S7

Models assessing the effect of consistency ratings, SOA, object size and eccentricity on the proportion of object responses in Experiment 2.

Predictors	M1				M2				M3			
	Logit	CI	<i>z</i>	<i>p</i>	Logit	CI	<i>z</i>	<i>p</i>	Logit	CI	<i>z</i>	<i>p</i>
(Intercept)	0.29	0.15 – 0.43	4.11	<.001	0.34	0.21 – 0.46	5.25	<.001	0.35	0.22 – 0.48	5.40	<.001
SOA	-0.06	-0.11 – -0.01	-2.54	.011	-0.07	-0.12 – -0.02	-2.58	.010	-0.07	-0.12 – -0.01	-2.43	.015
Consistency	0.06	0.01 – 0.11	2.37	.018	0.05	0.00 – 0.10	2.10	.036	0.06	0.01 – 0.10	2.20	.028
Object size					0.53	0.39 – 0.67	7.26	<.001	0.53	0.38 – 0.68	7.05	<.001
Eccentricity					-0.30	-0.39 – -0.21	-6.56	<.001	-0.37	-0.49 – -0.26	-6.56	<.001
Object size * Eccentricity									-0.22	-0.40 – -0.04	-2.38	.017
Object size * SOA									0.04	-0.01 – 0.10	1.59	.112
Eccentricity * SOA									-0.02	-0.07 – 0.03	-0.71	.475
Object size * Consistency									0.03	-0.02 – 0.08	1.12	.261
Eccentricity * Consistency									0.01	-0.04 – 0.06	0.52	.605
SOA * Consistency									0.01	-0.04 – 0.07	0.41	.683
Object size * Eccentricity * SOA									-0.04	-0.11 – 0.04	-0.91	.363
Object size * Eccentricity * Con.									0.01	-0.07 – 0.08	0.18	.855
Object size * SOA * Consistency									0.06	0.01 – 0.11	2.25	.024
Eccentricity * SOA * Consistency									-0.00	-0.06 – 0.05	-0.10	.919
Object size * Ecc. * SOA * Con.									-0.03	-0.10 – 0.04	-0.86	.390

Note. All predictors in these analyses were standardized. Con. = Consistency, Ecc. = Eccentricity.

Supplementary Table S8

Models assessing the effect of the presentation sequence condition and the covariates object size and eccentricity on log response times in Experiment 3.

Predictors	M1			M2			M3		
	log(RT)	CI	<i>t</i>	log(RT)	CI	<i>t</i>	log(RT)	CI	<i>t</i>
(Intercept)	7.39	7.31 – 7.46	192.60	7.39	7.31 – 7.46	201.14	7.39	7.31 – 7.46	200.35
Condition [T-O]	-0.03	-0.05 – -0.01	-3.59	-0.03	-0.05 – -0.01	-3.58	-0.03	-0.05 – -0.01	-3.43
Object size				-0.06	-0.08 – -0.04	-5.00	-0.06	-0.09 – -0.03	-3.86
Eccentricity				0.02	-0.00 – 0.04	1.62	0.02	-0.00 – 0.05	1.80
Object size * Eccentricity							0.00	-0.03 – 0.03	0.01
Object size * Condition							0.00	-0.02 – 0.03	0.11
Eccentricity * Condition							-0.01	-0.03 – 0.01	-1.36
Object size * Eccentricity * Condition							0.00	-0.03 – 0.03	0.09

Note. The covariates object size and eccentricity were standardized. The condition predictor was dummy coded with the object-texture sequence as the baseline. T-O = texture-object sequence.

Supplementary Table S9

Models assessing the effect of the presentation sequence condition and the covariates object size and eccentricity on basic-level categorization accuracy in Experiment 3.

Predictors	M1				M2				M3			
	Logit	CI	z	p	Logit	CI	z	p	Logit	CI	z	p
(Intercept)	1.30	0.96 – 1.63	7.59	<.001	1.38	1.08 – 1.68	9.11	<.001	1.45	1.13 – 1.76	9.01	<.001
Condition [T-O]	0.29	0.16 – 0.41	4.38	<.001	0.29	0.15 – 0.43	4.04	<.001	0.20	0.03 – 0.36	2.35	.019
Object size					0.53	0.31 – 0.75	4.73	<.001	0.64	0.36 – 0.93	4.41	<.001
Eccentricity					-0.26	-0.39 – -0.12	-3.70	<.001	-0.38	-0.59 – -0.17	-3.60	<.001
Object size * Eccentricity									-0.16	-0.48 – 0.16	-1.01	.314
Object size * Condition									-0.12	-0.33 – 0.09	-1.15	.252
Eccentricity * Condition									0.21	0.01 – 0.40	2.09	.037
Object size * Eccentricity * Condition									0.21	-0.05 – 0.47	1.55	.121

Note. The covariates object size and eccentricity were standardized. The condition predictor was dummy coded with the object-texture sequence as the baseline. T-O = texture-object sequence.

Supplementary Table S10

Models assessing the effect of the presentation sequence condition and the covariates object size and eccentricity on confidence ratings in Experiment 3.

Predictors	M1				M2			
	Logit	CI	<i>z</i>	<i>p</i>	Logit	CI	<i>z</i>	<i>p</i>
Condition [T-O]	0.37	0.23 – 0.51	5.21	<.001	0.38	0.24 – 0.52	5.29	<.001
Object size					0.68	0.44 – 0.92	5.59	<.001
Eccentricity					-0.37	-0.54 – -0.20	-4.18	<.001

Note. The covariates object size and eccentricity were standardized. The Condition predictor was dummy coded with the object-texture sequence as the baseline. We attempted to fit a model containing interactions between condition, object size and eccentricity (i.e., M3) for the confidence ratings, but the model structure was not supported by the data and the model failed to converge. T-O = texture-object sequence.

Supplementary Table S11

Test of the estimated basic-level categorization accuracies against chance for object and texture stimuli across SOAs in Experiment 4.

SOA	Object stimuli				Texture stimuli			
	Accuracy	SE	<i>z</i>	<i>p</i>	Accuracy	SE	<i>z</i>	<i>p</i>
33	.161	0.141	7.49	<.001	.081	0.264	1.04	.148
50	.206	0.127	10.71	<.001	.157	0.246	4.19	<.001
67	.215	0.141	10.05	<.001	.203	0.243	5.53	<.001
83	.241	0.117	13.38	<.001	.268	0.200	8.51	<.001
100	.263	0.130	12.86	<.001	.295	0.275	6.68	<.001
117	.288	0.125	14.41	<.001	.325	0.245	8.07	<.001
133	.305	0.121	15.59	<.001	.331	0.301	6.65	<.001
250	.339	0.126	16.16	<.001	.421	0.224	10.68	<.001

Note. The theoretical chance level was 6.25%.

Supplementary Table S12

Models assessing the contrasts between texture and object stimuli at each SOA for basic-level categorization accuracy, confidence ratings (using all trials) and confidence ratings (using only correct trials) in Experiment 4.

SOA	Categorization accuracy				Confidence (all trials)				Confidence (correct trials)			
	Logit	CI	<i>z</i>	<i>p</i>	Logit	CI	<i>z</i>	<i>p</i>	Logit	CI	<i>z</i>	<i>p</i>
33	-0.78	-1.28 – -0.29	-3.09	.002	-0.50	-0.88 – -0.13	-2.62	.009	-1.43	-2.14 – -0.72	-3.96	<.001
50	-0.30	-0.76 – 0.15	-1.30	.193	-0.07	-0.42 – 0.29	-0.36	.716	-0.73	-1.13 – -0.32	-3.48	.001
67	-0.06	-0.52 – 0.40	-0.25	.800	0.11	-0.26 – 0.48	0.58	.562	-0.47	-0.91 – -0.03	-2.09	.037
83	0.14	-0.26 – 0.54	0.67	.500	0.34	0.01 – 0.67	2.00	.045	-0.14	-0.53 – 0.25	-0.72	.473
100	0.17	-0.35 – 0.69	0.63	.531	0.42	0.08 – 0.76	2.40	.016	-0.23	-0.60 – 0.14	-1.20	.232
117	0.19	-0.26 – 0.65	0.84	.402	0.46	0.12 – 0.80	2.63	.008	-0.05	-0.46 – 0.36	-0.25	.804
133	0.14	-0.42 – 0.70	0.50	.620	0.47	0.10 – 0.85	2.48	.013	-0.11	-0.51 – 0.29	-0.55	.580
250	0.36	-0.07 – 0.79	1.64	.102	0.48	0.14 – 0.82	2.74	.006	-0.02	-0.37 – 0.34	-0.09	.928