Psychological Research
https://doi.org/10.1007/s00426-019-01247-6

ORIGINAL ARTICLE

Screen size matches of familiar images are biased by canonical size, rather than showing a memory size effect

Matteo Valsecchi[1,2]

Received: 4 April 2019 / Accepted: 9 September 2019

Abstract

Being confronted with the depiction of a familiar object activates a number of properties of the object that are stored in memory. Memory properties such as color and size have been shown to interfere with the processing of the color and of the size of the depiction, so that that reaction times are longer when the color or size of the depiction are incongruent with the stored knowledge about the object. In the case of color, it is known that the memorized information also affects the appearance of the depiction, for example when a gray banana appears slightly yellow, a phenomenon known as memory color effect. Here, I tested whether a memory size effect also occurs. To this aim, I conducted one experiment where observers matched either the screen size or the real-world size of pairs of animals or vehicles. The results indicate that the screen matches are biased in the same direction as the real-world size matches, opposite of what would be predicted by a memory color effect. This result was replicated in a second experiment using a different and larger set of animal images. Overall, I confirm that observers cannot ignore the real-world size information when they attempt to match the screen size of two items, although this results in a bias towards the canonical size of the items, rather than in a memory size effect.

Introduction

Among the various aspects that compose our knowledge of objects in the world is their usual size. Through experience, we learn that cats are generally smaller than cows, and that bikes are generally smaller than cars. At the same time, given a large enough difference, we are obviously able to judge which of two pictures of different objects is smaller or larger. It has long been known that we can use the familiar size of an object for perceptual judgments, for instance we can use it as a cue to its distance when contextual cues are limited (Epstein, 1963; Paivio, 1975; Smith, 1952; Tyer, Allen, & Pasnak, 1983). Although world size information can inform perceptual judgments, it appears that human observers have a certain degree of flexibility in dissociating the size-constant, real-world size of an object and its retinal size if instructed accordingly (Baird, 1963, 1965; Gilinsky, 1955; Higashiyama, 1984).

✉ Matteo Valsecchi
matteo.valsecchi@unibo.it

1 Dipartimento di Psicologia, Universitá di Bologna, Viale Berti Pichat, 5, 40127 Bologna, Italy

2 Abteilung Allgemeine Psychologie, Justus-Liebig Universität Giessen, Giessen, Germany

Does this mean that we can use our stored knowledge about object size when needed, and ignore it completely when it is not needed? A very well-developed line of research involving speeded judgments suggests that this is not the case. When observers are asked to judge the relative screen size of two images, they are slower to respond to incongruent trials where relative screen and real-world size mismatch (Konkle & Oliva, 2012), akin to a Stroop interference. Such interference effects can occur even without full-object recognition, based on the association between size and mid-level attributes of the image (Long & Konkle, 2017). The degree of interference in incongruent trials seems to be similar both for real world and screen depiction size judgments (Gliksman, Itamar, Leibovich, Melman, & Henik, 2016). In a similar vein, Fisher and Sperandio (2018) showed that single objects can be detected quicker if they are presented in a size closer to their real-world size. Similar congruency costs can be observed even when observers compare the size of words or symbols that represent smaller or larger objects, indicating that the activation of the memory size attribute can occur and interfere with size judgments even without a pictorial presentation (Rubinsten & Henik, 2002; Shen et al., 2016).

Additional phenomena indicate that the stored representation of object size can be triggered automatically by a pictorial representation and influence visual processing. One is attentional scaling (Collegio, Nah, Scotti, & Shomstein, 2019), i.e., the finding that attention is oriented less efficiently within depictions of objects whose real-world size is larger. Another example is the finding that real-world size, albeit irrelevant for the task, can interfere with judgments of numerosity (Reynvoet, Vos, & Henik, 2018).

The fact that real-world size has such a clear effect on the processing speed of image size suggests the question of whether the stored knowledge can even modify the appearance of the depiction itself. Such a claim might seem far-fetched, but it has been shown to be the case for color. It is well known that the memory color, i.e., the color that is associated with a familiar object, beside creating Stroop-like interference when the observer is asked to name the color of the depiction (Naor-Raz, Tarr, & Kersten, 2003), influences the appearance of the color as it is depicted. This phenomenon is known as a memory color effect (see Witzel & Gegenfurtner, 2018 for a recent review). When observers are asked to set the color of an object depiction to gray, their settings will be biased in a chromatic direction approximately opposite to the one of the chromaticity that they associate to the object itself. For instance they will adjust a banana to a slightly bluish color, so as to compensate for the memory color effect which makes it appear slightly yellow (Hansen, Olkkonen, Walter, & Gegenfurtner, 2006; Olkkonen, Hansen, & Gegenfurtner, 2008; Witzel, Olkkonen, & Gegenfurtner, 2016). The memory color effect can even be shown to alter the neural representation of images in visual cortices as early as V1 (Bannert & Bartels, 2013; Scholte, Meuwese, Lamme, Vandenbroucke, & Fahrenfort, 2014), as evidenced by the fact that the memory color can be decoded based on the pattern of activation produced by a gray image. Notice that the memory size of an image can be decoded from early visual areas representations as well (Coutanche & Koch, 2018; Coutanche & Thompson-Schill, 2019).

In the present study I address the question of whether something akin to a memory size effect occurs. The motivation comes from the observation that size, similar to color, is a semantic property readily activated by the presentation of the image of an object. To answer this question I conducted two experiments where observers were asked to match the relative size of two stimuli in terms of their screen size and in terms of their real-world memorized size. In the presence of a memory size effect, the match values should be anticorrelated. If the information stored in memory makes the depiction of a cow be perceived to be larger than the one of a cat, as a compensatory mechanism, observers should produce a slightly smaller cow when matching it to the image of a cat in terms of screen size. If observers are able to completely ignore the real-world size of the items they are matching, they will perform the matches in an unbiased way. It is however also conceivable that the observers will be biased in their adjustments in the direction of the size relationship that exists in the world. This result would still be indexing an effect of semantic knowledge on screen size adjustments, although its effect would rather be mediated by a preference for a given pictorial representation, rather than by a perceptual illusion that needs to be counteracted. Indeed, there is evidence that the canonical size of a known item can determine the relative size at which observers prefer to view it or generate it (Konkle & Oliva, 2011; Linsen, Leyssen, Sammartino, & Palmer, 2011).

The results of both experiments supported the conclusion that the memorized size of an object influences its adjusted size through the second mechanism. When matching two images in terms of screen size, observers produced matches biased towards the real-world ratio of the size of the objects. When asked to match the screen size of two depictions, observers are rather biased towards the canonical size associated with the pictorial representation, as opposed to showing a memory size effect.

Experiment 1

In Experiment 1, I investigate whether the familiar real-world size of an object has an effect on size adjustments when observers are asked to match the screen size of the image that depicts it. To this aim, I had observers match pairs of images of animals or vehicles in terms of their screen size and in terms of their real size. I evaluated whether the first were dependent on the second.

Materials and methods

Participants

21 observers participated in Experiment 1 (11 female, mean age 22.1 years).

Stimuli

Stimuli were two sets of six images belonging to the category “animal” and “vehicle”, respectively. The images were downloaded from the internet based on the criteria of belonging to animals or means of transport of different size, depicted as much as possible perpendicular to the viewing direction and in high definition. The images were converted to grayscale and manually segmented from the background and centered in a square image.

Testing both animals and vehicles is essential, since animacy, as well as real-world size, determines the cortical representation of images of meaningful items (Konkle & Caramazza, 2013). Finding an effect of memory size on size adjustments for animal images would not per se warrant that the effects generalize to images of inanimate objects. I chose, however, not to test size perception across image categories, so as to avoid having to come up with stimulus sets balanced for image structure across categories. While it is relatively easy to find examples of small and large items within one category that have a similar structure (animals with four legs and one head at one extremity, vehicles with anterior and posterior wheels, etc.), I do not think it would be possible to find inanimate and animate objects that are built on the same principle.

While a single descriptor of size might be sufficient when dealing with items that have exactly the same structure or when dealing with extremely large size differences, I decided to use multiple descriptors when analyzing the data in this study. This was motivated by the need to account for the fact that small size differences such as the ones that I expected to observe when participants attempted to match the object sizes, in the presence of small structural differences between items might produce different results for different descriptors of size. For instance, if the observer managed to exactly match the height of the two objects, the area might still have differed. Moreover, I did not instruct observers to use one specific descriptor when performing the matches, so I could not choose one specific descriptor a priori. For each image, I constructed look-up tables for different descriptors of size at different image screen sizes (image square reproduced between 5 and 1201 pixel wide in 2 pixel increments). The first step to do this was to show the image on a pc screen at multiple sizes using the "DrawTexture" procedure implemented in the Psych Toolbox (Kleiner et al., 2007). The resulting image was captured from the screen, convolved with a 2D Gaussian filter with 2 pixel standard deviation and thresholded at 1 8-bit RGB value. Finally, I applied the "regionprops" algorithm from Matlab (The MathWorks, Inc., Natick, MA, USA) to extract the Major Axis Length, Minor Axis Length, Area, Width, Height, Equivalent Diameter and Perimeter properties.

To assess whether the estimated real-world size of the images was correlated with structural aspects of the images I computed for each image the ratios of Major Axis Length to Minor Axis Length, as an indicator of image elongation, and the ratio of Convex Radius (computed from the Convex Area) to Perimeter, as an indicator of contour complexity. For all images the ratios were computed using a texture screen size of 301 pixels. The real-world size of the images is computed as the average of the minimum and maximum ranges across all observers. For each image, I also computed the rectilinearity index described by Nasr et al. (2014). This represents a measure of rectilinearity computed across scales, angles and curvature. Values are normalized across images for each scale and angle, and then averaged across scales and angles. For this analysis, I used four scales with spatial frequencies of 204.8, 113.7, 68.3 und 37.9 cycles per image. Rectilinearity is normally associated with real-world size due to physical constraints related to gravity (Long, Konkle, Cohen, & Alvarez, 2016), so some degree of association with size is expected if the sample is drawn randomly from a sample of images. The values are plotted in Fig. 1. Generally speaking, there does not seem to be a clear association between structural properties and estimated size within a category with the exception of rectilinearity in vehicles.

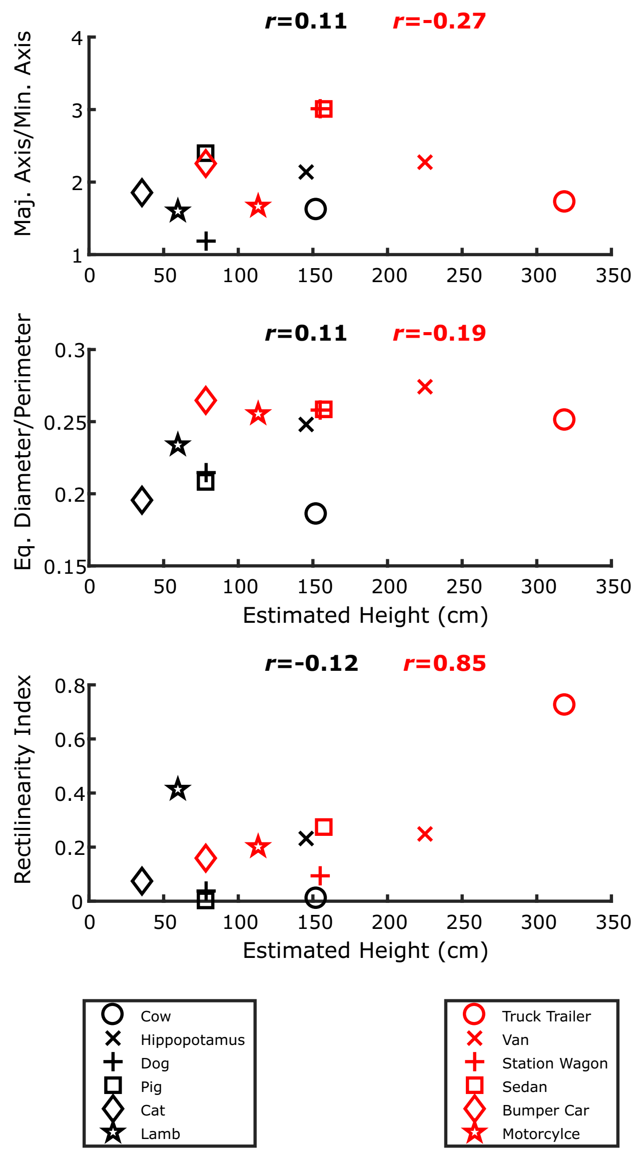


Fig. 1 Ratio of major and minor axis, ratio of equivalent diameter to perimeter and rectilinearity index for each image used in Experiment 1, as a function of estimated real-world object or animal height. Most descriptors have weak correlation with object real-world size, with the exception of the rectilinearity index and possibly of the axes ratio in vehicles. In both cases, the correlation seems to be driven mostly by one specific item (truck trailer). Notice also that animal and vehicle items differ substantially in the ratio of equivalent diameter to perimeter and to a certain degree also in the axes ratio, i.e., the animal items are less compact and for the most part more elongated

All correlations are weak, and even the correlation of size and rectilinearity in vehicles seems mostly driven by the largest stimulus (truck trailer). Incidentally, the truck trailer item seems also to be the one driving the (much weaker) negative correlation between size and axes ratio in vehicles. Notice that a strong correlation of real-world size and image structure would be problematic because the different image structures could be associated with differences in perceived size of the picture itself (e.g., Warren & Pinneau, 1955), and thus confound a possible correlation between real-world size and perceived screen size. Notice also that vehicles were for the most part more elongated and less compact than animals.

Procedure

Observers sat in a dimly illuminated room, in front of a Display++ 32 in. LCD monitor (Cambridge Research Systems Ltd, Rochester, UK). Viewing distance of 48 cm was ensured by means of a chin rest. The screen resolution was 1920 × 1080 pixels, with 2.743 pixels per mm and 22.976 pixels per degree of visual angle. Stimuli were displayed using Matlab and the PsychToolbox (Kleiner et al., 2007). At the beginning of each trial, two images were presented, either top-left and bottom-right of the center of the screen, or bottom-left and top-right (Fig. 2), with a center-to-center distance of 368 pixels. This was done to make it less tempting for observers to simply match the height or the width of the images by aligning their boundary in the "Screen Size" condition. The initial size of the two images was set starting from a 241 pixels (square texture side length). An equal pixel value (randomly drawn in steps of two between 0 and 120) was subtracted from the size of one image (randomly assigned) and added to the size of the other image. Observers could modify the relative size of the images by pressing the up or down arrow key on the PC keyboard. Pressing one arrow increased the size of one image (randomly assigned) and decreased the size of the other, whereas the other arrow had the opposite effect. The size changes took place in steps of two pixels and I ensured that the responsivity of the keys was such that through a brief key press observers were able to produce a two pixel change.

Observers underwent two subsequent sessions with identical stimuli. In the first session (Screen Size), their task was to modify the size of the two images until they appeared to have the same size on the screen. In the second session (Real Size), the task was instead to modify the size of the images until the two animals or objects appeared to have the same relative size that they would have had when seen next to each other. When observers were satisfied with the adjustment, they had to confirm it by pressing the enter key, which would initiate the next trial. Observers had no time limit to complete the task, but the overall duration of the experiment was between 30 and 60 min.

Each combination of the 6 images in each category was tested twice in the Screen Size session, except for the combinations of 2 identical pictures, which were only tested once, yielding a total of 36 trials per category. The two categories were tested in two separate trial blocks, whose order was balanced across observers. In the Real Size condition I only tested each combination of pictures once, yielding a total of 21 trials per category, again conducted in separate, counterbalanced blocks. Notice that the choice of testing the Real

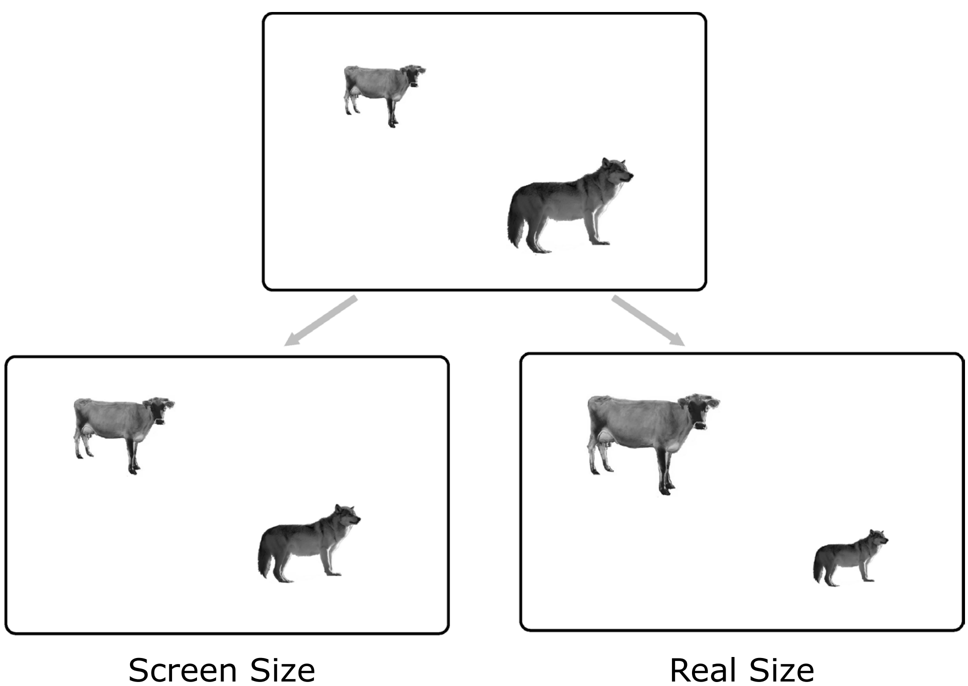


Fig. 2 Experimental tasks. In each trial, observers were presented with two images belonging to the same category. In the first session (Screen Size), they were required to change the relative size of the two images (by pressing the up or down key) until they appeared to have the same size on the screen. In the second session (Real Size), they did the same to change the relative size of the images until the depicted animal or vehicle had the same relative size as when standing next to each other in the real world. Notice that for copyright reasons the two images in this depiction are not the ones used in the actual experiments (these are available from the author upon request)

Size condition only once was due to the assumption that the differences between the stimuli would be more reliable in this condition. In the end, it is conceivable that all observers will consistently adjust the real size of a hippo as being way larger than the one of a cat.

After the experiment, observers completed a questionnaire where they were asked to report, next to the picture of each animal and vehicle, the estimate minimum and maximum height in cm that it could have in the real world.

Data analysis

Data analysis involved first translating the final adjusted size of each picture in each trial into five of the size measures from the lookup table I mentioned earlier (Major Axis Length, Area, Width, Height, Perimeter). For each measure, I then computed the ratio between the two images. Notice that this step is necessary because the raw sizes cannot be compared across trials, given that the size adjusted in one trial is always relative to the size of the other item. The ratios were converted to logarithm values before proceeding with the analysis. This was done both because the absolute value of the raw ratios depends on the image which is picked as the numerator in the division, and because using logarithms also reduced the asymmetry in their distribution (see Fig. 3). In a preliminary analysis, I computed the ratios between the image presented in the upper half of the screen and the one presented in the lower half of the screen for both conditions. In the main analysis, I evaluated how the real-size adjustments predicted the screen size adjustments through a regression analysis. This choice was motivated by the fact that the items were not picked to be directly classifiable into small and large, and to account for possible inter-observer differences in the size associated with a given item. For each possible combination of images, I picked one ordering of the two items to compute the ratio, so as to be able to compare it across trials and conditions. Finally, I identified the trials that pertained to each possible combination of items, that is one trial from the Real Size condition and 1 or 2 trials from the Screen Size condition. To have a uniform ordering of the ratios in the further analyses, I changed the sign of the log ratio of all matched trials if the sign was negative in the Real Size condition, which is equivalent to computing the inverse of the ratios.

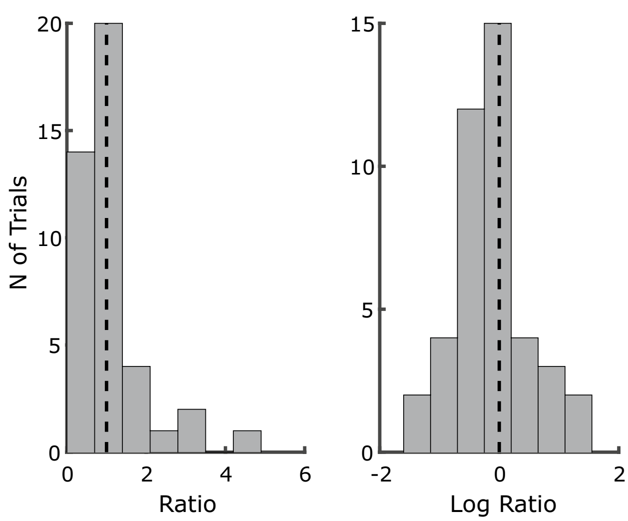


Fig. 3 Example histograms of data (ratios of major axis length) obtained from the first observer in the Real Size condition of Experiment 1. Ratios are computed between stimuli in the upper and lower half of the screen. The vertical dashed line indicates the situation where size was adjusted to be the same. The distribution of raw ratios (left) is highly skewed and becomes symmetrical once the values are converted to log ratios (right)

When I performed statistical analyses on the log ratios, I used parametric tests, given that they are normally distributed (Fig. 3). For analyses performed on regression slopes, I report nonparametric tests instead. This choice is, however, not crucial. All results are identical when parametric tests are performed, probably due to the fact that all slope values were quite low, given that the size ratios were closer to 1 in the Screen Size condition as opposed to the Real Size condition.

Results

The first preliminary analysis was concerned with the question whether observers adjusted the image in the upper half of the screen as being larger than the one in the lower half. The average log ratios of the different size measures, in the Screen Size and Real Size conditions are shown in Fig. 4. Despite the larger variability in the real ratio condition, it is evident that log ratios are overwhelmingly negative, indicating that observers adjusted the stimulus in the upper half of the screen as being smaller than the one in the lower half. This is consistent with a residual tendency by the observers to perceive the stimulus in the upper half of the screen as being further away and thus larger. Observers probably adjusted the upper stimulus as smaller to null this effect and produce an equal apparent screen size.

The main experimental question I asked is whether observers are able to ignore the real-world size of two animals or objects when judging the relative size of their image on a computer screen. To answer this question, I computed a linear regression of the ratio of size for each couple of animals or objects in the Screen Size condition relative to the ratio of size for the same animals or objects obtained from the Real Size condition. This was performed for each observer and stimulus category. An example of this procedure for one observer and two example measures is depicted in Fig. 5.

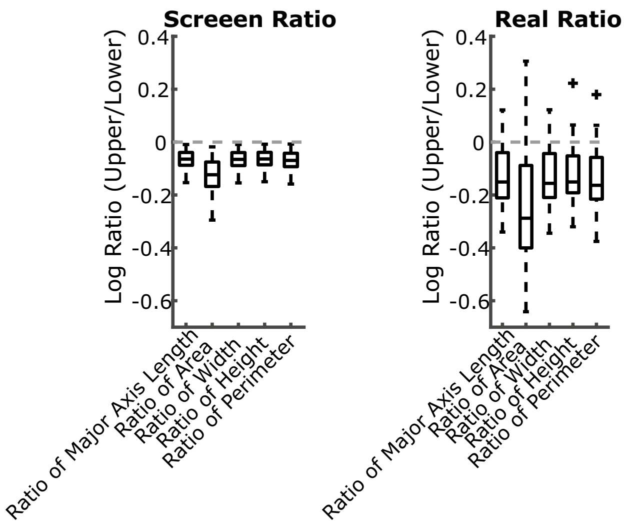


Fig. 4 Upper half of the screen bias. The box plots depict the distribution of average log ratios of the different size measures across observers. The ratios are computed in each trial dividing the size measure for the stimulus in the upper half of the screen by the size measure for the stimulus in the lower half. All measures yield negative log ratios (Bonferroni corrected *t* tests showed that values are significantly smaller than 0 in all cases with $p < 0.01$), indicating that the stimulus presented on top was adjusted smaller

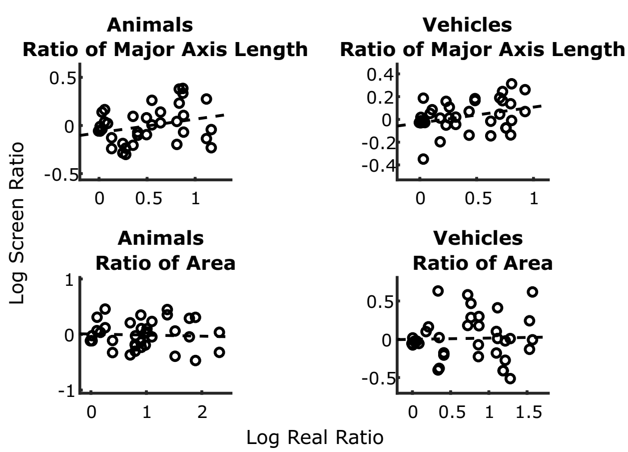


Fig. 5 Example of regression between log ratios in the Real Size condition and in the Screen Size condition. Each circle represents a trial from the Screen Size condition. A positive slope indicates that the screen matches were biased towards the same ratio that was reported to exist in the world. Notice that the log ratios are closer to 0, and thus the original ratios are much closer to 1 for the Screen Size condition, which is expected given that the observers were asked to directly match the depiction of the stimuli

An overall depiction of the resulting slope values as a function of the image category and descriptor can be found in Fig. 6. The slopes are in most cases positive. To evaluate this observation statistically, and in the light of the fact that some of the distributions are skewed and some clear outliers are present, I opted for a nonparametric testing approach. For each descriptor and category, I performed a Wilcoxon signed ranked test, subsequently performing a Bonferroni correction for ten tests on the *p* values. Subsequently, I compared the slopes across categories, this time with a Bonferroni correction for five tests (Table 1). The results show that in the case of every significant test, the slopes were positive. This indicates that in general, observers tended to produce matches in the Screen Size condition that were biased towards the matches that they had performed in the Real Size condition. To get an idea of the strength of the effect, let us consider a typical slope of 0.05 as observed for the animal category in the case of the Ratio of Major Axis length descriptor (similar also to the results of Experiment 2), and a typical log screen ratio of 0.5 (meaning that one stimulus' major axis is adjusted 64.9% larger than the other). The regression would predict a corresponding log ratio in the Screen Size condition of 0.025 (corresponding to a ratio of major axis length difference of 2.5%). The results also suggest a possibly larger effect for vehicles, limitedly to the ratios of Height and Perimeter.

The fact that the results are partially different between the different measures might be surprising at first, given that one would expect all of the measures to increase when the observer increases the size of one picture relative to the other. In fact, if the manipulations actuated by the observers are of limited amplitude, it can be the resulting correlation between the different measures is small or even negative, yielding the overall different results between the different measures in Fig. 6.

Given that the results are not univocal, I tried to understand which of the parameters was most relevant perceptually to the observers while doing the task. I reasoned that if observers were trying to match a specific parameter, its ratios in the Screen Size conditions would be closer to 1, and thus the log ratios closer to 0, whereas the ratios would be larger for the parameters that were being ignored. To estimate how close to one the ratios were, I isolated the observations corresponding to the trials in the Screen Size condition where the two depicted items were different, separately for each observer, category and parameter (30 observations for each combination). I then fit the Standard Deviation of a Gaussian distribution with mean 0 to the observed log ratios. The values are presented in Fig. 7. As can be seen, the parameters that did not show a significant bias by Real Size (area for both categories and height and perimeter for Animals), showed relatively large Standard deviations, suggesting that they were less well equalized and thus less relevant to the observers. Conversely, the parameters that consistently

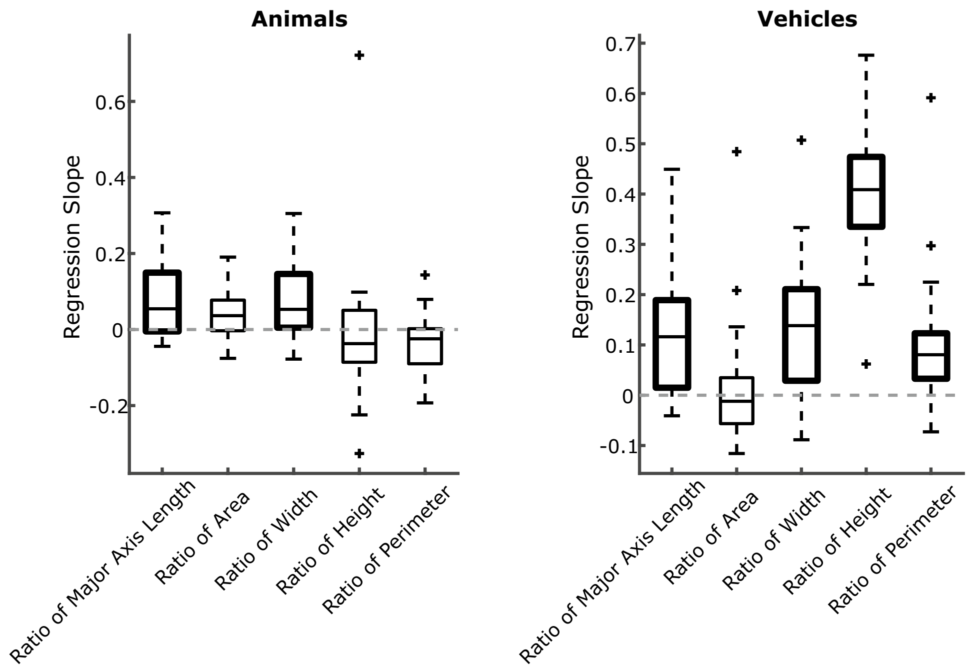


Fig. 6 Average regression slopes for all measure ratios across observers and categories. The thicker boxes depict ratios where median slopes were different from zero ($p < 0.05$ after Bonferroni correction Wilcoxon signed rank test, see Table 1). Wherever significantly different from 0, the slopes are positive, indicating that observers' Screen Size matches were biased towards the matches that they performed in the Real Size condition

Table 1 Summary of statistical tests on the regression slopes in Experiment 1

Category	Ratio	Median	*Z*	Signed rank	*p* (Bonferroni corrected)
Animals	Major axis length	**0.055**	**3.076**	**204**	**0.021**
	Area	0.036	2.277	181	0.228
	Width	**0.053**	**2.972**	**201**	**0.030**
	Height	− 0.037	− 1.130	83	1
	Perimeter	− 0.024	− 1.964	59	0.496
Vehicles	Major axis length	**0.116**	**3.563**	**218**	**0.004**
	Area	− 0.012	− 0.191	110	1
	Width	**0.138**	**3.771**	**224**	**0.002**
	Height	**0.409**	**4.015**	**231**	**< 0.001**
	Perimeter	**0.080**	**3.354**	**212**	**0.008**
Animals vs. vehicles	Major axis length		− 1.025	86	1
	Area		1.720	165	0.426
	Width		− 1.407	75	0.796
	Height		**− 3.319**	**20**	**0.004**
	Perimeter		**− 3.215**	**23**	**0.006**

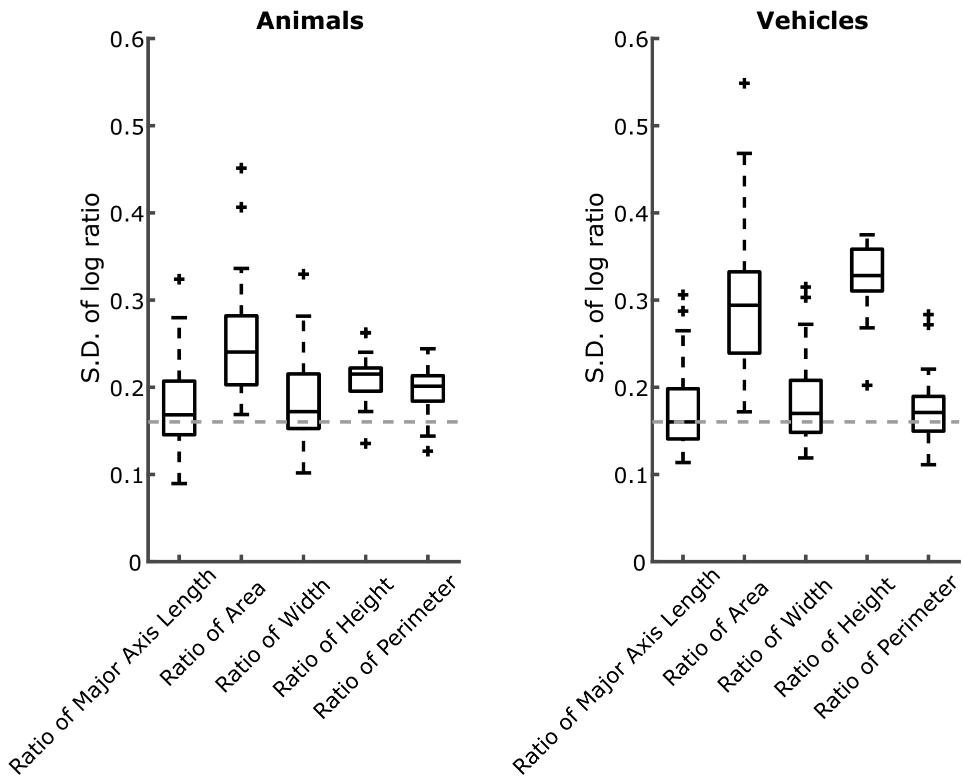
All significant results are marked in bold. All significant tests in the individual object categories pertain to cases where the median slope was positive. The comparison between categories suggests a possibly larger association between real and screen size adjustments for vehicles in the case of the ratios of height and perimeter

showed the positive bias (Ratio of Major Axis Length and Ratio of Width) were matched more precisely. This confirms that the positive slopes in the regression analysis were associated with the dimensions that were most relevant to the observers.

Experiment 2

The results of Experiment 1 suggest that there is a general tendency for observers to be biased in their screen size matches by the real size of the objects whose images they adjust. My design, however, limited the number of items

Fig. 7 Box plots of standard deviation values corresponding to Gaussian fits of the log ratio distributions in the Screen Size condition. The ratios of Major axis length and the ratios of width consistently show the lowest SD (dashed line is the minimum value for comparison), indicating that observers took them into account when matching the size. Trials where the adjustment was between two same images were removed from this analysis



that could be used for testing, which might question the universality of my finding. Moreover, there were some aspects of the image structure that showed some degree of correlation with the real-world size of the items. In experiment 2, I decided to replicate the results that were obtained in Experiment 1 with animals, using a different and larger set of stimuli.

Materials and methods

Participants

16 observers participated in Experiment 2 (14 female, mean age 25.6 years). None of the observers had participated in Experiment 1.

Stimuli

Stimuli were one set of 9 images belonging to the category "animal", different from the ones used in Experiment 1, but selected with the same criteria. Additionally to the preprocessing which I used in Experiment 1, in Experiment 2, I equalized across the images the average luminance (24.83 Cd/m^2) and the pixel-wise standard deviation of luminance (19.07 Cd/m^2) within the image boundary.

The different descriptors of size were extracted using the same lookup table approach procedure as was used in Experiment 1. Average ratio of major to minor axis length and average ratio of equivalent diameter to perimeter, computed with screen texture size of 301 pixels, as well as the rectilinearity index are depicted in Fig. 8. Once again the descriptors of global structure show little correlation with real-world size, consistent with the fact that the items were picked to have the same global structure. Instead, the index of rectilinearity shows the expected (Long et al., 2016) relationship, i.e., the images of larger animals are more rectilinear.

Procedure

The procedure was identical to the one of Experiment 1. The larger number of items and the fact that I only tested animal images, so that each item was compared to each other, again resulted in an overall duration of between 30 and 60 min for Experiment 2.

Each combination of the 9 images was tested twice in the Screen Size session, except for the combinations of two identical pictures, which were only tested once, yielding a total of 81 trials. In the Real Size condition, I only tested each combination of pictures once, yielding a total of 45 trials.

After the experiment, observers completed a questionnaire where they were asked to report for each animal (next to its picture) the estimate minimum, maximum and average height in cm that it could have in the real world.

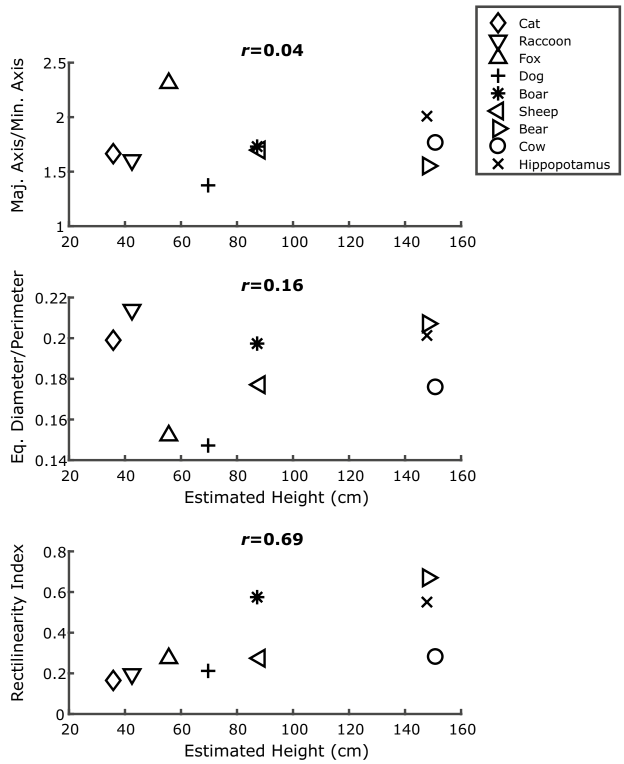


Fig. 8 Ratio of major and minor axis, ratio of equivalent diameter to perimeter and rectilinearity index for each image used in Experiment 2, as a function of estimated real-world animal height. While the global contour structure is uncorrelated with size, this sample presents the expected correlation of size and rectilinearity, whereby the images of larger items are more rectilinear

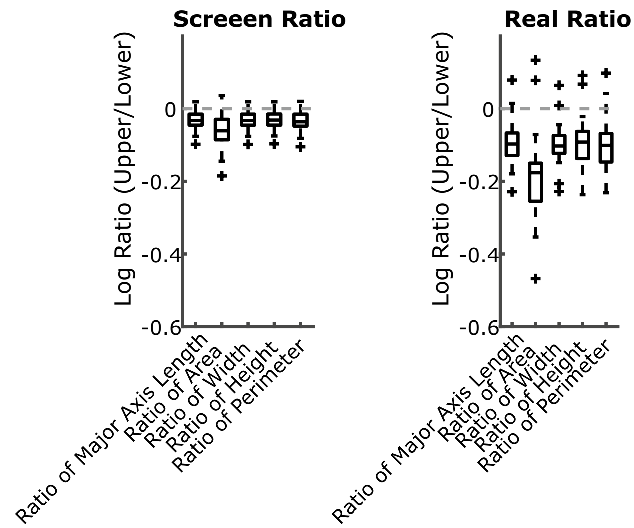


Fig. 9 Upper half of the screen bias. The box plots depict the distribution of average log ratios of the different size measures across observers in Experiment 2. The ratios are computed in each trial dividing the size measure for the stimulus in the upper half of the screen by the size measure for the stimulus in the lower half. All measures yield negative log ratios (Bonferroni corrected *t* tests showed that values are significantly smaller than 0 in all cases with $p<0.01$), indicating that the stimulus presented above was adjusted smaller

Data analysis

Data were analyzed exactly like in Experiment 1, first computing log ratios between the images presented in the upper and lower halves of the screen, and subsequently by computing the correlation of log ratios between the two conditions for matching pairs of items (after changing both signs if the sign was negative for the Real Size condition).

Results

Similarly to the first experiment, a preliminary analysis showed consistent tendency of observers to adjust the size of the stimuli in the upper half of the screen as smaller than the one of the stimuli in the lower half (Fig. 9).

Similar to Experiment 1, the key step in the analysis of the data was to compute a linear regression of the log ratio of size for each couple of animals in the Screen Size condition relative to the ratio of size for the same animals obtained from the Real Size condition. An example of this procedure for one observer and two example measures is depicted in Fig. 10.

The average regression slopes can be seen in Fig. 11. Contrary to the Animal Category results of Experiment 1 (Fig. 6), the median value across observers is positive for all five size descriptors. Wilcoxon signed ranked test, with Bonferroni correction for 5 tests on the *p* values (Table 2) showed that the difference from 0 is significant for Major Axis Length and Width, as it was in Experiment 1, and additionally also for the Ratio of Areas, which failed to reach significance in Experiment 1.

Again similar to Experiment 1, I conducted an additional analysis to try and understand which descriptors were more relevant to the observers while they were performing the screen size matching task. This involved fitting the distribution of log ratios in the Screen Size condition, for the trials where two different items were matched (Fig. 12). The results are consistent with the ones of the Animal category in Experiment 1 in the sense that the ratios were generally closest to 1 for the ratio of major axis length, whereas the ratios of area tended to be the most variable. In this case, the ratio of width had an intermediate variability, more similar to the one of the ratio of perimeter. Generally speaking, the results confirm that the descriptor that shows the lowest discrepancy in the matches (Major Axis Length) also shows a consistent positive bias between the adjustments in the Real Size

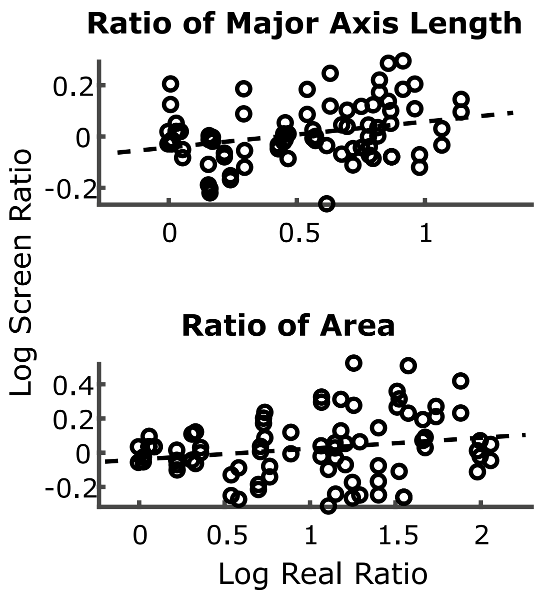


Fig. 10 Example of regression between log ratios in the Real Size condition and in the Screen Size condition. Each circle represents a trial from the Screen Size condition. A positive slope indicates that the screen matches were biased towards the same ratio that was reported to exist in the world. Notice that the log ratios are closer to 0, and thus the original ratios much closer to 1 for the Screen Size condition, which is expected given that the observers were asked to directly match the depiction of the stimuli

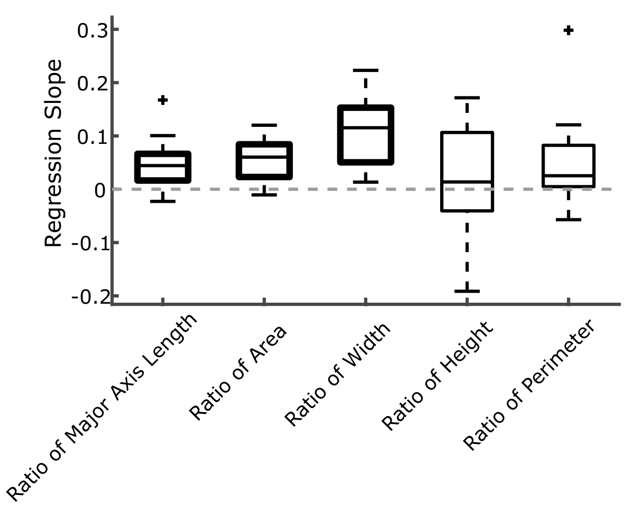


Fig. 11 Average regression slopes for all measure ratios across observers. The thicker boxes depict ratios where median slopes were different from zero ($p < 0.05$ after Bonferroni correction Wilcoxon signed rank test, see Table 2). Wherever significantly different from 0, the slopes are positive, indicating that observers' Screen Size matches were biased towards the matches that they performed in the Real Size condition

Table 2 Summary of statistical tests on the regression slopes in Experiment 1

Ratio	Median	*Z*	Signed rank	*p* (Bonferroni corrected)
Major axis length	**0.044**	**3.258**	**131**	**0.006**
Area	**0.060**	**3.413**	**134**	**0.003**
Width	**0.115**	**3.516**	**136**	**0.002**
Height	0.014	0.465	77	1
Perimeter	0.025	2.172	110	0.149

All significant results are marked in bold and pertain to tests where the median slope was positive

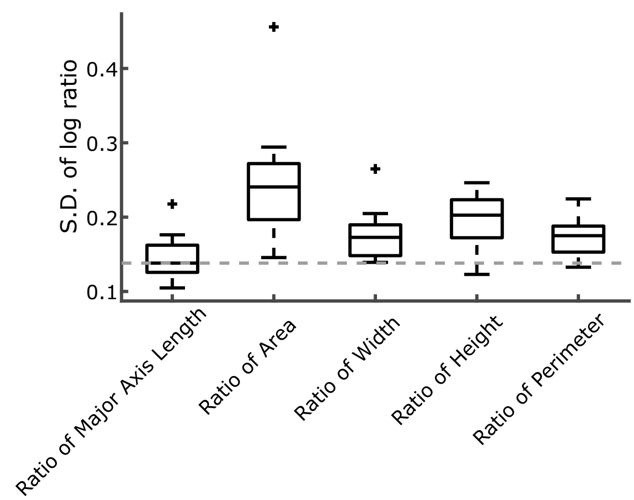


Fig. 12 Boxplots of standard deviation values corresponding to Gaussian fits of the log ratio distributions in the Screen Size condition. Similar to the findings for Animals in Experiment 1, the ratios of Major axis length consistently show low SD (dashed line is the minimum value for comparison), whereas the ratios are most variable for the area. Trials where the adjustment was between two same images were removed from this analysis

condition and in the Screen Size condition, although other descriptors might also show it.

Discussion

In two experiments, I investigated whether something akin to a size memory effect exists. Both experiments seemed to indicate that this is not the case. When asked to match the size of two depictions of animals or vehicles on the screen, observers were consistently biased in the direction of the relative size which they used to match the two items in terms of world size. If a perceptual bias would have been produced by the memory size, the sign of the correlations should have been the opposite. Matches would have been biased in the opposite direction of real-world ratios, so as to compensate for the perceptual effect, as is observed in the color memory effect.

One difference between the paradigm I used in this experiment and the gray setting procedure that was used to demonstrate the memory color effect (e.g., Hansen et al., 2006 and Olkkonen et al., 2008) is that in this study observers directly compared two images, whereas in the previous ones they adjusted the color of one single item to gray. It is an empirical question whether using a single item paradigm for size would have produced the same results, say if observers were asked to equate the size of an item to a memorized standard of 10 cm^2. What seems to suggest that this aspect is not crucial is the observation that the memory color effect can also occur when two images are compared side by side (Witzel et al., 2016).

I also do not believe that finding size adjustments positively correlated to the memorized size of the items was due to the specific limited set of stimuli that I used. The same result was present in basically three independent sets of stimuli and two independent groups of observers. Notice that a direct comparison of the results seems to suggest that the effect might be even larger for vehicles as compared to animals. This result should be taken with caution, because the sets of stimuli were very limited and specifically picked for their image structure within but not across categories. This implies that any difference between categories might be due to the structural differences between the images, rather than to their animacy status. Moreover, the specific size descriptors which showed a significant difference between categories were not among the most precisely adjusted, indicating that observers might have been concentrating on different aspects of the object when matching its size.

One crucial difference might be that observes might be unable to generate a genuine 2D representation of the images and use distance as a way to discount the memory size effect. In a way, by seeing the image of the cow as being further in a 3D scene, I could match its screen size correctly while still perceiving the image as being larger. Observers were told to just consider the image in the screen match and not to consider it a scene, but it might be that the 3D distance representation based on memory size is triggered automatically in the absence of a 3D context (Epstein, 1963; Paivio, 1975; Smith, 1952). Indeed, the possibility that observers were not entirely compliant with the instruction to represent both objects at the same distance is supported by the finding that in both experiments they consistently adjusted the stimuli in the upper half of the screen as smaller. Elevation is well known to be a cue to distance within scenes (e.g., Epstein, 1963 and Gibson, 1950), and objects with a higher elevation tend to appear more distant. Observers might have adjusted the upper object as smaller to nullify the perceptual effect of perceived distance on size. My choice of not using a pictorial context was aimed at avoiding providing observers with cues that could be used strategically when matching the screen size, e.g., by counting the tiles in a representation like the one used by Murray et al. (2006). It is possible that providing very clear cues to distance, for instance in a VR display, could prevent the discounting of the size effect through distance and reveal a memory size effect.

This interpretation, however, shifts the question of why color and size are treated differently when observers try to match them between two depictions of objects, to the question of why we use differently distance and illumination to compensate for the effects of memory size and memory color. While I cannot provide a functional explanation of why it should be the case, at least anecdotically it appears to me that assumptions about distance in impoverished displays are very unstable and to a certain extent subject to voluntary control, think for instance of the relative distance of the two faces of a Necker Cube (e.g., Pelton & Solley, 1968 and van Ee, van Dam, & Brouwer, 2005), whereas interpretations of the color of illumination, such as those that subtend individual differences in the perception of the #TheDress (Witzel, Racey, & O'Regan, 2017) seem to be much more stable (Lafer-Sousa, Hermann, & Conway, 2015).

This study found the opposite of a memory size effect. This bears quite a resemblance to the canonical size (Konkle & Oliva, 2011; Linsen et al., 2011). What is meant by this concept is that human observers prefer to view and to produce depictions of real-world objects in a size which is positively related to their real-world size. In the context of my experiments, I can extrapolate that observers would prefer and would spontaneously produce a configuration where the depiction of the cow is larger than the one of the cat (as they did in the experiment) as opposed to one where the cat is larger than the cow (which would have been predicted by a memory size effect). One possible explanation for the difference between the current results and the results that supported the existence of a memory color effect would be the absence of canonical color preferences. My impression is that known objects are preferred in their canonical color, at least in the case of fruits and vegetables (Schifferstein, Wehrle, & Carbon, 2019; Siple & Springer, 1983), although I believe a thorough investigation of this question with animate and man-made objects is still missing.

One final point that needs to be discussed is the role of image structure. Image structure can affect perceived size of geometrical shapes (e.g., Warren & Pinneau, 1955), so that interpreting the results of the study would be problematic if image structure were to be systematically confounded with real-world size. The sets of stimuli were constructed so as to have approximately the same structure and an analysis of the image structure as a function of estimated real-world size confirmed that they were not confounded. Another important aspect of image structure is, however, rectilinearity. Rectilinearity is normally associated with real-world size in object images (Long et al., 2016) because objects that are larger tend to be heavier and their shape needs to be able to withstand gravity. Rectilinearity was not associated with real-world size in the animal set that I used in Experiment 1, and it was associated with real-world size in the vehicle image set that I used for Experiment 1 (although mostly because of one specific item) and in the animal set that I used for Experiment 2. The reason for the discrepancy is probably due to the fact that using a larger item set in Experiment 2 produced a set of stimuli that was more representative of the expected association. Combining the results from all experiments confirms that equivalent screen size adjustments of familiar items are positively correlated with real-world size both when structure and rectilinearity are controlled for and when they are distributed as expected from a random sample of items.

In conclusion, the results show that when observers are asked to match the size of two objects, their adjustments are dominated by their tendency to produce a configuration matching the canonical size, rather than being influenced by a perceptual bias similar to a memory size effect. While I cannot exclude that a memory size effect could be evidenced under conditions that eliminate the possibility of seeing the test configuration as two images at different distances in a 3D scene, the results once again confirm that the representation of memory size is automatically triggered as observers see a depiction of a known object, and this representation cannot be ignored when attempting to evaluate the size of the depiction itself.

Acknowledgements Matteo Valsecchi is supported by the SFB TRR 135/Project A8 Grant from the Deutsche Forschungsgemeinschaft. I would like to thank Theresa Geier and Benita Nmakwe for help with data collection and stimulus construction, Cesar Echavarria and Shahin Nasr for providing the rectilinearity toolbox and Christoph Witzel for his comments and suggestions during the preparation of the manuscript. The dataset pertaining to this study is available at zenodo.org (https://doi.org/10.5281/zenodo.3402735).

Compliance with ethical standards

Conflict of interest The author declares that he has no conflict of interest.

Ethical standards All the procedures performed were in accordance with the ethical standards of the institutional research committee (the study protocol was approved by the local ethics committee at the University of Giessen, approval number: LEK FB6 2017-08) and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

Baird, J. C. (1963). Retinal and assumed size cues as determinants of size and distance perception. *Journal of Experimental Psychology, 66*(2), 155–162. https://doi.org/10.1037/h0046554.

Baird, J. C. (1965). Stimulus and response factors in size instruction effects. *Perceptual and Motor Skills, 21*(3), 915–924. https://doi.org/10.2466/pms.1965.21.3.915.

Bannert, M. M., & Bartels, A. (2013). Decoding the yellow of a gray banana. *Current Biology, 23*(22), 2268–2272. https://doi.org/10.1016/j.cub.2013.09.016.

Collegio, A. J., Nah, J. C., Scotti, P. S., & Shomstein, S. (2019). Attention scales according to inferred real-world object size. *Nature Human Behaviour, 3*(1), 40–47. https://doi.org/10.1038/s41562-018-0485-2.

Coutanche, M. N., & Koch, G. E. (2018). Creatures great and small: Real-world size of animals predicts visual cortex representations beyond taxonomic category. *NeuroImage, 183,* 627–634. https://doi.org/10.1016/j.neuroimage.2018.08.066.

Coutanche, M. N., & Thompson-Schill, S. L. (2019). Neural activity in human visual cortex is transformed by learning real world size. *NeuroImage, 186,* 570–576. https://doi.org/10.1016/j.neuroimage.2018.11.039.

Epstein, W. (1963). The influence of assumed size on apparent distance. *The American Journal of Psychology, 76*(2), 257–265. https://doi.org/10.2307/1419162.

Fisher, C., & Sperandio, I. (2018). Familiar size effects on reaction time: When congruent is better. *Journal of Experimental Psychology Human Perception and Performance, 44*(9), 1414–1425. https://doi.org/10.1037/xhp0000543.

Gibson, J. J. (1950). *The perception of the visual world*. Oxford: Houghton Mifflin.

Gilinsky, A. (1955). The effect of attitude upon the perception of size. *American Journal of Psychology, 68*(2), 173.

Gliksman, Y., Itamar, S., Leibovich, T., Melman, Y., & Henik, A. (2016). Automaticity of conceptual magnitude. *Scientific Reports, 6,* 21446.

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience, 9*(11), 1367–1368. https://doi.org/10.1038/nn1794.

Higashiyama, A. (1984). The effects of familiar size on judgments of size and distance: An interaction of viewing attitude with spatial cues. *Perception and Psychophysics, 35*(4), 305–312. https://doi.org/10.3758/BF03206333.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., et al. (2007). What's new in Psychtoolbox-3. *Perception, 36*(14), 1.

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

Konkle, T., & Oliva, A. (2011). Canonical visual size for real-world objects. *Journal of Experimental Psychology Human Perception and Performance, 37*(1), 23–37. https://doi.org/10.1037/a0020413.

Konkle, T., & Oliva, A. (2012). A familiar-size Stroop effect: Real-world size is an automatic property of object representation. *Journal of Experimental Psychology Human Perception and Performance, 38*(3), 561.

Lafer-Sousa, R., Hermann, K. L., & Conway, B. R. (2015). Striking individual differences in color perception uncovered by 'the dress' photograph. *Current Biology, 25*(13), R545–R546. https://doi.org/10.1016/j.cub.2015.04.053.

Linsen, S., Leyssen, M. H. R., Sammartino, J., & Palmer, S. E. (2011). Aesthetic preferences in the size of images of real-world objects. *Perception, 40*(3), 291–298. https://doi.org/10.1068/p6835.

Long, B., & Konkle, T. (2017). A familiar-size Stroop effect in the absence of basic-level recognition. *Cognition, 168,* 234–242. https://doi.org/10.1016/j.cognition.2017.06.025.

Long, B., Konkle, T., Cohen, M. A., & Alvarez, G. A. (2016). Mid-level perceptual features distinguish objects of different real-world sizes. *Journal of Experimental Psychology General, 145*(1), 95–109. https://doi.org/10.1037/xge0000130.

Murray, S. O., Boyaci, H., & Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. *Nature Neuroscience, 9*(3), 429–434. https://doi.org/10.1038/nn1641.

Naor-Raz, G., Tarr, M. J., & Kersten, D. (2003). Is color an intrinsic property of object representation? *Perception, 32*(6), 667–680. https://doi.org/10.1068/p5050.

Nasr, S., Echavarria, C. E., & Tootell, R. B. H. (2014). Thinking outside the box: rectilinear shapes selectively activate scene-selective cortex. *The Journal of Neuroscience, 34*(20), 6721. https://doi.org/10.1523/JNEUROSCI.4802-13.2014.

Olkkonen, M., Hansen, T., & Gegenfurtner, K. R. (2008). Color appearance of familiar objects: Effects of object shape, texture, and illumination changes. *Journal of Vision, 8*(5), 13. https://doi.org/10.1167/8.5.13.

Paivio, A. (1975). Perceptual comparisons through the mind's eye. *Memory and Cognition, 3*(6), 635–647. https://doi.org/10.3758/BF03198229.

Pelton, L., & Solley, C. (1968). Acceleration of reversals of a Necker cube. *The American Journal of Psychology, 81*(4), 585–588. https://doi.org/10.2307/1421064.

Reynvoet, B., Vos, H., & Henik, A. (2018). Comparative judgment of familiar objects is modulated by their size. *Experimental Psychology, 65*(6), 353–359. https://doi.org/10.1027/1618-3169/a000418.

Rubinsten, O., & Henik, A. (2002). Is an ant larger than a lion? *Acta Psychologica, 111*(1), 141–154. https://doi.org/10.1016/S0001-6918(02)00047-1.

Schifferstein, H. N. J., Wehrle, T., & Carbon, C.-C. (2019). Consumer expectations for vegetables with typical and atypical colors: The case of carrots. *Food Quality and Preference, 72,* 98–108. https://doi.org/10.1016/j.foodqual.2018.10.002.

Scholte, H. S., Meuwese, J. D. I., Lamme, V. A. F., Vandenbroucke, A. R. E., & Fahrenfort, J. J. (2014). Prior knowledge about objects determines neural color representation in human visual cortex. *Cerebral Cortex, 26*(4), 1401–1408. https://doi.org/10.1093/cercor/bhu224.

Shen, M., Xie, J., Liu, W., Lin, W., Chen, Z., Marmolejo-Ramos, F., & Wang, R. (2016). Interplay between the object and its symbol: The size-congruency effect. *Advances in Cognitive Psychology, 12*(2), 115–129. https://doi.org/10.5709/acp-0191-7.

Siple, P., & Springer, R. M. (1983). Memory and preference for the colors of objects. *Perception and Psychophysics, 34*(4), 363–370. https://doi.org/10.3758/BF03203049.

Smith, W. M. (1952). Past experience and the perception of visual size. *The American Journal of Psychology, 65*(3), 389–403.

Tyer, Z. E., Allen, J. A., & Pasnak, R. (1983). Instruction effects on size and distance judgments. *Perception and Psychophysics, 34*(2), 135–139. https://doi.org/10.3758/BF03211338.

van Ee, R., van Dam, L. C. J., & Brouwer, G. J. (2005). Voluntary control and the dynamics of perceptual bi-stability. *Vision Research, 45*(1), 41–55. https://doi.org/10.1016/j.visres.2004.07.030.

Warren, J. M., & Pinneau, S. R. (1955). Influence of form on judgment of apparent area. *Perceptual and Motor Skills, 5*(1), 7–10. https://doi.org/10.2466/pms.1955.5.g.7.

Witzel, C., & Gegenfurtner, K. R. (2018). Color perception: Objects, constancy, and categories. *Annual Review of Vision Science, 4*(1), 475–499. https://doi.org/10.1146/annurev-vision-091517-034231.

Witzel, C., Olkkonen, M., & Gegenfurtner, K. R. (2016). Memory colours affect colour appearance. *Behavioral and Brain Sciences, 39,* e262. https://doi.org/10.1017/S0140525X15002587.

Witzel, C., Racey, C., & O'Regan, J. K. (2017). The most reasonable explanation of "the dress": Implicit assumptions about illumination. *Journal of Vision, 17*(2), 1. https://doi.org/10.1167/17.2.1.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.