



A standardised set of images for judgements of proportion

Sara Varatojo¹ · Leonor Lavradio¹ · Alexandre Fernandes¹ · Teresa Garcia-Marques¹

Accepted: 28 August 2022
© The Psychonomic Society, Inc. 2022

Abstract

In the present work, we present normative data for a set of 39 original clipart-style images that can be used as material in studies involving judgements of proportion. The original images are drawings that depict different day-to-day scenarios (e.g., lighted windows in a building; books on a shelf) and each has seven variants of different proportions (from 20% to 80%) belonging to different categories (*discrete vs continuous; social vs non-social; natural vs artificial; stimuli physical dimensions; number of referents*). Normative data for these images are presented in an interactive database (available at https://judgment-images-and-norms.shinyapps.io/estimates_interactive/), corresponding to the means of proportion estimates (in percentage form), the perceived ease of making such estimates, the perceived level of familiarity and liking for each image, and the relationships between these variables. In the paper, we analyse the data at an individual level, addressing how the latter judgements are related to the proportion estimates, how those estimates are related to objective proportions, and how these relationships are moderated by image category. The analyses presented in this paper aim to aid readers in selecting images that enable them to better address specific influences on proportional estimates or to control for those influences in their studies.

Keywords Proportion estimates · Images' normative data · Image categories · Ratings of familiarity and liking

Introduction

Various types of judgements and decisions involve quantification processes based on the comparison of items varying in quantity in one or more dimensions (e.g., Mussweiler & Epstude, 2009). Visual stimuli (consisting of quantifiable elements) have been widely used in research on magnitude processing, particularly proportion estimates, a fundamental process in the field of quantitative cognition (Beran & Parrish, 2016).

Proportional reasoning is studied using tasks that are based on the process of understanding a relationship (between quantities) as a simple quantity and being able to perform mental operations with it (Lamon, 2005). Researchers typically manipulate one of two dimensions: the number of elements of multiple types (i.e., *discrete quantity*), or the size of the parts/segments making up a

single element, usually geometric shapes (i.e., *continuous quantity*). Examples of such tasks are those that employ visual delimited displays containing two types of elements varying in relative frequency, with participants being asked to estimate the proportion of one element type in comparison with the other type or to the whole set of elements. A variety of simple geometric elements have been used in these *visual array tasks*, such as coloured dots (Meert et al., 2012; Philip, 1947; Stevens & Galanter, 1957; Varey et al., 1990), squares (Shuford, 1961), rectangles (Spinillo & Bryant, 1999), spheres (Ginsburg & Rapoport, 1967; Jeong et al., 2007), or vertical and horizontal lines (Shuford, 1961). Other tasks use continuous stimuli, whose quantities are processed analogously. Examples include *scaling tasks*, that require matching of a parcel of a target stimulus, which could be a subsegment of a length-line (i.e., *number-line estimation task*) or an area-shape (e.g., divided bars and pie charts), to its symbolic or non-symbolic proportional equivalent (e.g., Cohen & Sarnecka, 2014; Huttenlocher et al., 1999; Möhring et al., 2018), and the *relative fullness tasks*, that require the participant to estimate the percentage of liquid in a container, like water or coloured liquids in straight glass cylinders (e.g., Hollands & Dyre, 2000;

✉ Teresa Garcia-Marques
gmarques@ispa.pt

¹ ISPA - Instituto Universitário; William James Center for Research, Rua Jardim do Tabaco, 34, 1149-041 Lisboa, Portugal

Pearson, 1964; Piaget, 1967; Raghurir & Krishna, 1999; Yang & Raghurir, 2005).

These stimuli and tasks have been used in different types of research exploring the processes and features of proportional reasoning (Hollands & Dyre, 2000; Meert et al., 2012), including studies concerning its relation to acquiring mathematical skills (e.g., Meert et al., 2013; Möhring et al., 2018), development of quantitative cognition in children (e.g., Boyer et al., 2008; Jeong et al., 2007), consumer behaviour (e.g., Attwood et al., 2012; Chandon & Ordayeva, 2009), and clinical deficits in neural processes (e.g., Hayashi et al., 2013; Skagerlund et al., 2016).

Studies in the domain of proportional reasoning are typically carried out using materials built for the study in question. The specific physical features of those stimuli likely motivate differences in the psychophysical functions that establish the relationship between objective and subjective proportions (see claim by Hollands & Dyre, 2000; Spence, 1990; Zhang & Maloney, 2012). For instance, Boyer et al. (2008) show that children's performance is worse when stimuli map discrete rather than continuous quantities. This stimulus variability is welcome, since it allows us to address relevant commonalities in the definition of a psychophysical function associated with proportion estimates (e.g., Hollands & Dyre, 2000). Nevertheless, the differences between studies have not been systematically addressed, and biases found in the description of processes may be amplified by the lack of variability of stimuli within studies themselves. In most cases, experimental stimuli consist only of simple geometric elements (e.g., arrays or parts of circles, squares, spheres, cylinders). More diverse stimuli, comprising different types of features, may allow researchers not only to understand the general mechanism underlying proportion estimates, but also to control features that can interfere with them and to identify different biases that can be promoted by such features.

Although the preferential use of colour and geometric shapes as material allows for greater experimental control (see Smets et al., 2015), these types of stimuli may also lack generalisation to our social world. Stimuli with *social meaning*, on the other hand, have the advantage of generalisation. Studies that examine the development of children's ability to reason about proportions frequently employ experimental paradigms relying on stimuli with social meaning, such as bricks (e.g., Spinillo & Bryant, 1999), marbles (Jeong et al., 2007), drawings of pizza (e.g., Singer-Freeman & Goswami, 2001), pie slices (e.g., Spinillo & Bryant, 1991), chocolate pieces (e.g., Singer-Freeman & Goswami, 2001), or volumes of water and juice (e.g., Fujimura, 2001; Schwartz & Moore, 1998). In adults, the use of stimuli with social meaning has been restricted to consumer behaviour, such as beverages (wine, beer, and lemonade) in glasses or bottles (e.g., Attwood et al., 2012; Troy et al., 2018), food on plates (e.g., Kosıte et al., 2019; Wansink & van Ittersum, 2013), or

supermarket products in packages (e.g., Chandon & Ordayeva, 2009; Li et al., 2018). The use of stimuli with social meaning in these domains relates to the social nature of our cognition and the fact that most quantity estimations occur with this type of stimuli. Quantitative cognition, and proportion estimation research in particular, may be strengthened considerably by taking into account that proportion estimates are usually undertaken in a social environment. Specifically, having social environments depicted in stimuli, in which substantial control of the relevant features of the stimulus is still possible, would be an advantage to this field (see suggestion by Barth et al., 2003, that research on human numerical competence must consider the influence of stimulus properties).

In line with this reading of the field, we found a need for a set of visual stimuli that could be made available to the research community, along with normative data associated with different dimensions, which could be easily used in different proportional reasoning studies, allowing for testing of new hypotheses as well as control of undesirable influences in a context that favours comparisons between studies. These normative data are most relevant when stimuli have a social meaning. Specifically, stimuli with *social meaning* are the most likely to differ in features able to modulate the cues that can be used to infer quantity (Fornaciai et al., 2019), thus needing stronger experimental control.

In this paper, we present normative data for a set of clipart-style social images that can be used as materials in studies involving judgements of proportion, offering the possibility to control for a set of dimensions. We specifically consider dimensions that were previously shown to influence subjective probability estimates, such as stimuli familiarity (e.g., Fox & Levav, 2000), liking (e.g., Bohner et al., 1988; Zajonc, 1968), and response ease (e.g., Suantak et al., 1996). We believe that this set of images will be useful for researchers, allowing them to study different dimensions of proportional reasoning with social stimuli, without having to compromise experimental control. The normative data presented in this paper can also be used in research focusing on the role of these features in the estimation process clarifying, for instance, how perceptual fluency and liking influence proportion estimates (e.g., Doi & Shinohara, 2016; Schwarz, 2004).

Development of norms for quantifiable images

We developed a set of 39 clipart-style images defining different social settings. These images correspond to proportions represented in different day-to-day scenarios (e.g., lighted windows in a building; books on a shelf), which vary systematically in different dimensions and proportion levels.

All the images define different social settings that may be used in a *part-whole* type of task, where researchers ask participants to estimate the proportion of one type of element

(e.g., lighted windows, green books on a shelf) relative to the total set of elements of various types (e.g., green and red books), or the part of an entity (e.g., volume of water in a bottle, or clouds covering the sky) relative to its total (e.g., the total volume of the bottle, or the whole sky).

Each of the 39 images has seven different proportion level variants, where the proportions represented vary from 20% to 80%. The proportional judgements elicited are in percentage format, ranging from 0 to 100% (e.g., percentage of lighted windows in a building, green books on a shelf, water in a bottle, or clouds in the sky). Percentages are more consistently and quickly perceived by participants (see Goldstone, 1993), and provide uniform data for all images.

Descriptive details regarding the data collected for each image are made accessible to readers via an online interactive database (available at https://judgment-images-and-norms.shinyapps.io/estimates_interactive/). Proportional normative data are presented in the interactive database for each of the seven variants of each image in the form of mean percentage estimates, as well as mean estimate bias associated with each image (i.e., the absolute difference between the estimates and the real proportions depicted in the images). In addition, we provide readers with data regarding the mean of the *ease* that participants felt while making the estimates, for each of the seven proportion-level variants of each image. The 39 images are further characterised by their perceived levels of *familiarity* and *liking*, corresponding to the means of all participants' evaluations for the 50% proportion level. Relationships between these judgements at the level of each image are also provided in the interactive database.

It should be further noted that the images were classified according to their different characteristics in order to consider the influence of different stimulus properties on proportion estimates (see Barth et al., 2003). Specifically, this set of images varies in five categories, allowing for different manipulations within the same set of stimuli, besides varying in the proportion to be estimated. The images vary in the *quantity type* represented—either *continuous* or *discrete*—and in terms of other stimulus characteristics: the *presence of social elements*—*social elements* are either *present* or *absent* in the image; the *nature of the stimulus*—either *natural* or *artificial*; the *physical dimensions* to be quantified—*colour*, *shape*, *area*, or *number of elements*; and the *number of subsets of elements*—*dichotomic* (two types of elements) or *multiple* (more than two types).

In the results section of this paper, we provide readers with information at the individual level of analysis, clarifying how the individual's subjective proportion estimates related to objective estimates and the other judgements associated with the images, and also how these are likely to be qualified by the different image features described above. This should allow researchers to select materials knowing

in advance that specific features of the stimuli are likely or unlikely to influence individuals' proportion estimates. The images may be used either for further testing those influences or for controlling for those influences when they are considered undesirable.

Method

Participants

One hundred and eighty English-speaking participants were recruited via the online recruiting platform Prolific (www.prolific.co) [7 January 2020] to take part in this study. Decisions regarding sample size aimed at ensuring sufficient power to perform the analyses at the participant level. This decision was supported by G*Power (Faul et al., 2009), aiming at a .80 power level in detecting a relationship between two variables, and its moderation by a category, with medium effect size ($f=0.25$) and an alpha level of .05. In order to be eligible, Prolific users needed to be native English speakers, aged 18 to 40 years, with an approval rate of over 90%. One set of participant data was not used, as the participant responded to the study twice, thus leaving a total of 179 participants ($M=26.84$, $SD=6.46$; 53.63% female). This study was approved by ISPA's internal research ethics committee. All participants provided informed consent to participate and were informed that they could abandon the study at any time.

Stimuli

Original visual stimuli for judgements of proportions were created in *clipart style* by the authors using Adobe Illustrator CS6 (Adobe Systems Incorporated: San Jose, CA).

The content of the images was developed using a focus group. The group's goal was to put together different ideas for scenarios in which an image may differ in part-whole proportions represented in area or volume, or in specific discrete entities. Images to be selected should provide high variability in scenarios and in different features (see below). An important aspect to be taken into account was that the stimuli should not appear bizarre in one or more of the part-whole proportions. Thirteen *scenarios* resulted from this process (see Fig. 1 for examples). These scenarios were chosen following different criteria. The first criterion was that scenarios would have to allow for stimuli to vary in terms of either *discrete* quantity characteristics, based on multiple arrays of entities (i.e., *proportion of elements*; for instance, percentage of green marbles in a glass container) or *continuous* quantity characteristics, based on area or volume (i.e., *fill proportion*; for instance, percentage of dark fur on a pet).














<i>Scenarios depicted in the images</i>							
<i>Proportion level</i>	20%	30%	40%	50%	60%	70%	80%
<i>Scenario</i>							
<i>Judgement stimulus</i>	Forested area	Baldness	Fruit in bowl	Water in bottle	Cloud-covered sky	Dark fur	Pant-covered leg
<i>Scenario</i>							
<i>Judgement stimulus</i>	Yellow marbles	Lighted windows	Occupied seats	Green books	Star-shaped beads	Occupied parking spaces	

Fig. 1 Scenarios depicted in the images. Examples of one of the versions of each of the 13 scenarios depicted in the images along with the stimulus within the image used for the estimation judgement

Specific criteria were established for the images representing the discrete *part-whole* proportion of elements. First, all images had to include a large number of total elements in order to avoid that participants would resort to counting rather than estimating percentages (participants tend to almost immediately quantify up to six items; Kaufman et al., 1949). Furthermore, all the images designed had to depict non-symbolic quantities, without containing symbolical numerical elements, so that this dimension would not interfere with the judgements made. Discrete images have either *two types of elements* (e.g., percentage of lighted windows against dark windows) or *four types of elements* (e.g., percentage of yellow marbles against blue, green, and red marbles; see how this may affect quantity judgements in Pelham et al., 1994).

For the images representing continuous *part-whole* proportions we ensure that the number of levels of the dimension was either *unitary*—the target and its complementary (e.g., percentage of water in the bottle), or *multiple*—the complementary area is a compound of multiple elements (e.g., percentage of clouds in the sky).

Three different versions of each of the chosen scenarios were designed. More specifically, each version corresponded to a different exemplar of the setting, differing in general colours and stimuli disposition or type (e.g., for the percentage of fruit in a bowl scenario, three different pieces of fruit were used for each version).

For each of the 39 images, seven levels of part-whole proportion of fill or elements (depending on the stimuli

characteristics) were created. The percentages depicted in the images varied from 20% to 80% in intervals of 10%. For the discrete stimuli images, the total number of elements was kept constant, with elements altered to result in the intended percentage of the stimuli. For the continuous stimuli images, we ran a script in Adobe Illustrator which allowed us to calculate the area of a selected vector object. The total area of fill was obtained through this method and the stimuli were altered to correspond to the intended percentage. In sum, a total of 273 original images of equal dimensions (189 × 189 cm) were created.

All images were drawn in the same *clipart style*—elements were derived from simple vector objects and shapes and manipulated using the Adobe Illustrator pen and selection tool, to produce simple images with little detail. Images are coloured, and mainly use tones of red, green, blue, and/or yellow, provided by the Adobe Illustrator colour guide. For some images, these are combined with neutral colours, such as elements of grey (e.g., marble container), brown (e.g., land area; skin), black (e.g., night sky), and white. Images generally contain a white background, except for those images which require an alternative background for emphasising the stimuli (e.g., cloud-covered sky, night sky for lighted windows). The choice of colours and minimalism design was so that the images resembled those we encounter in today's illustration and design industries, which we are exposed to every day.

Design and procedure

Data were collected through *Qualtrics* (Qualtrics: Provo, UT) and participants accessed through *Prolific* (www.prolific.co) [2020, 7 January]. Participants were randomly and equally assigned to one of three groups of images to evaluate. Each group contained all 13 scenarios, but only one version of each with the respective seven proportion levels. Thus, each participant saw 91 images. The order of image presentation was randomised and differed for each participant in each group.

Firstly, participants either consented to participate voluntarily and entered the study or opted to abandon the study and were immediately redirected to the end of the survey.

For the first task, participants were asked to estimate the percentages (0 to 100%) of several visual items as spontaneously as possible. The images were displayed, one at a time, in the centre of the screen below the question “*What is the percentage of (stimuli)?*” (e.g., “*What is the percentage of yellow marbles?*”). Underneath the image, participants were presented with a text box with the instruction “*Estimate the percentage*”, where they were asked to provide a number from 0 to 100. On the same screen, they were also asked to rate how easy it was to give the estimation, using a seven-point scale from 1 = *very hard* to 7 = *very easy*. This was repeated for all 91 images of the group in random order.

Following this task, participants rated the *familiarity* and *liking* of each of the 13 images that they had seen (only the 50% proportion levels of each image were shown), using two seven-point scales, 1 = *unfamiliar* to 7 = *very familiar* and 1 = *dislike* to 7 = *like a lot*, respectively. Participants were explicitly told that the images that they were rating for familiarity had all been presented in the previous task, in order to control for the influence of previous exposure on familiarity ratings¹. Once again, the images were displayed one at a time, in the centre of the screen, in random order. On the same screen, both rating scales appeared below, underneath the question “*how familiar is this item?*” for the familiarity rating, and “*how much do you like this item?*” for the likeability rating.

Final demographic questions were then asked, and participants were thanked. All participants received monetary compensation for their participation. The study lasted approximately 15 min.

¹ Although this procedure allows for controlling differences in familiarity across stimuli, it also likely inflates familiarity ratings for all stimuli, which should be kept in mind when interpreting the data, particularly absolute familiarity values, despite this inflation likely being the same for all stimuli.

Scores in the online interactive database

The 39 images (i.e., three versions of the 13 scenarios) belong to different, but not all orthogonal, categories. Therefore, each image was classified in the interactive database with regard to (a) *the type of quantity*—either *discrete* or *continuous*; (b) *presence of social elements*—either *present*, if the image depicted people (e.g., images portraying a receding hairline in a person) or *absent*, if the image depicted only objects (e.g., images portraying a container with marbles); (c) *nature of the stimulus*—either *natural* (e.g., dark fur) or *artificial* (e.g., green book); (d) the *stimuli dimensions* that varied in proportion—for discrete stimuli, these varied in *amount* (e.g., lighted windows), *colour* (e.g., yellow marbles), or *shape* (e.g., star-shaped beads), and for continuous stimuli, these were coded as varying in a *unitary area* (e.g., baldness—as the images portrayed a receding hairline, and so the bald spot increased as a whole) or *multiple areas* (e.g., forested area—as bare land area would decrease as different patches of trees appeared spread out over the image); (e) the *number of referents* (i.e., total types of elements) present for discrete stimuli images—either *two* (e.g., percentage of lighted windows against dark windows) or *four* (e.g., percentage of yellow marbles against blue, green, and red flowers).

Each of the 273 stimuli (39 images × 7 proportions) was scored for estimated proportion (as a percentage), and the associated estimation ease. Normative scores were calculated for each image by directly averaging individuals' proportion estimates (0–100) and ratings of estimation ease (1–7).

Additionally, the 273 stimuli were scored in terms of the bias associated with the individuals' proportion estimates (i.e., estimate bias). This score corresponded to the mean of the modulo of the value resulting from the subtraction of the estimated proportion provided by each participant from the objective proportion presented in the image.

Each of the 39 images presented in this normative study was also scored in terms of two image features (perceived *familiarity* and *liking*). These scores were derived by averaging the ratings of *familiarity* (1–7) or *liking* (1–7), respectively, provided by individuals for the 50% level of proportion of each image version.

All image information and scores were compiled into an interactive database run as a web application (app). This was accomplished using the R Shiny package. The web app is currently launched and hosted by the Shiny apps platform and can be accessed at https://judgment-images-and-norms.shinyapps.io/estimates_interactive/. Each row of the data table concerns one image, where the image name (i.e., “Image”), image scenarios (i.e., “Group”), amount of proportion fill (i.e., “Proportion”), short description (i.e., “Description”), and stimulus within the image used for the estimation judgement in our experiment (i.e., “Judgement

Stimulus”) are referred to for image identification. The following columns refer to the image classifications, as described above (i.e., “DiscretelContinuous”, “Nature of Stimulus”, “Dimension of Stimulus”, “Total Referents”, and “Social Elements”). Finally, all normative data for mean scores (i.e., “M”), standard deviations (i.e., “SD”), and low and high confidence intervals (i.e., “low CI”, “high CI”) of familiarity (i.e., “Familiarity”), liking (i.e., “Liking”), estimate (i.e., “Estimate”), estimate bias (i.e., “Bias”), and ease (i.e., “Ease”) are presented. The last columns represent the relationship found between the different judgements and estimates (in the form of Pearson correlation coefficients and respective p -values) at each individual image level (e.g., Pearson correlation coefficient of familiarity and liking is represented as “FamiliarityRLiking” and the respective p -value as “FamiliarityRLiking_p”).

The interactive database allows for the following: exact word search in all fields or inside specific columns; numerical or alphabetical ordering of values in each column; restriction of particular score fields between specific numeric intervals using a slider feature; unlimited combination of specific searches and numeric restrictions to identify the best image to be used. All images are available for download in jpeg format in a folder at https://osf.io/4cs6k/?view_only=04e5cedf790f4f88ab04c5e40f70b3ab.

Additional analyses at the participant level

Besides providing normative scores for each image, we also analysed the data in order to understand how specific features and categories of an image interfere with individuals’ estimates, namely, the relationship between objective and subjective proportions (a proxy of a linear psychophysical function). We defined this function with the slope indexing sensitivity (i.e., the ability to discriminate different levels of actual proportions) and the intercept indexing the general response bias (i.e., the tendency to provide high or low estimates).

We first provide descriptive statistics for different measures (i.e., estimate, estimate bias, ease, familiarity, and liking). Secondly, we characterise the relationship between these different measures using a mixed-model approach, with specific images as random factors. Third, we address the relationship between objective and subjective proportions, assessing levels of sensitivity and response bias using a linear mixed-effects analysis, with specific images as random factors. Finally, we focus on how image categories moderate all the previously studied relationships. Our general approach was to estimate different mixed models where categories are used as moderators of the intercepts and of the slopes of the main relationship.

The datasets analysed during the current study are available in the Open Science Framework (OSF) repository, at

Table 1 Descriptive statistics for the variables estimate, estimate bias, ease, familiarity, and liking

	Estimate*	Estimate bias	Ease*	Familiarity	Liking
Mean	51.54	11.54	4.24	5.95	4.04
SD	21.19	5.22	0.51	0.27	0.60
Male	51.2	11.7	4.42	6.03	4.00
Female	51.8	11.4	4.09	5.88	4.07

* signals that gender differences are significant ($p < .05$)

https://osf.io/4cs6k/?view_only=04e5cedf790f4f88ab04c5e40f70b3ab.

Results

Descriptive statistics

Normative data for three different aggregated measures (*estimate*, *estimate bias*, *ease*) of the 272 stimuli and for the aggregated ratings of *familiarity* and *liking* of the 39 images are summarised in Table 1. We further summarise these measures by gender, providing means in Table 1 and signalling when they achieved significance². This allows researchers to take gender differences into account when this could be relevant to their studies.

These summaries reveal that all images were evaluated as generally neutral, $M = 4.04$, with liking ratings ranging from 2.18 to 5.25, and perceived as reflecting very familiar events, $M = 5.95$, with familiarity ratings ranging from 5.03 to 6.40. In general, participants reported the estimation task as not being difficult, $M = 4.24$. However, there are clear differences between the perceived difficulty in estimation of the different images, considering that the values range from 2.43 to 5.39.

Estimates ranged from 13.88 to 92.34, and the mean estimates were correctly around 50%, $M = 51.54$, reflecting the objective range of percentages presented. Congruently, the estimate bias was relatively low, at 11.54, ranging from 2.36 to 28.78.

How do the judgements and estimates relate to one another?

In order to inform researchers as to how individual estimates and judgements relate to one other and also how the different

² Because we found those differences to occur similarly for all images (all interactions with image as a random factor, $p > .05$), we do not differentiate gender ratings for each image in our interactive table.

specific images define those relationships, we ran a set of mixed-model analyses (GAMLj package with support of the *jamovi* platform, version 1.2, The jamovi project, 2020), with the specific images as a random factor. We test for the random intercepts to control for differences between how images were evaluated. We test for random first-order interaction factors (a two-way interaction) to determine whether images vary in terms of the relationships approached in each analysis. We use the likelihood ratio test (*LRT*) to inform readers about possible differences in the focused relationship between different specific images³.

Familiarity and liking As expected, the judgements of familiarity and liking relate to each other; $slope \pm SE = 0.12 \pm 0.02$; $t(33.3) = 5.54$, $p < .001$.

The two-way interaction component shows that images vary in terms of the relationship with which participants evaluate their familiarity and liking; $SD = 0.09$; *LRT: chi-square* (2) = 5.45, $p = .060$. This suggests that it is possible for researchers to choose materials for which these two factors are more strongly or weakly related (image individual data for each relationship are reported on the *interactive database*).

Liking and ease There is a relationship between how much an image is liked and the ease with which the estimation is made; $slope \pm SE = 0.17 \pm 0.02$; $t(37.3) = 8.57$, $p < .001$. This relationship is not moderated by the random factor, suggesting that the relationship tends to be the same across all images.

Familiarity and ease Perceived familiarity and ease are positively related; $slope \pm SE = 0.04 \pm 0.02$; $t(37.3) = 1.89$, $p = .033$. This relationship is also not moderated by the random factor.

Estimate and liking Estimates (i.e., subjective proportions) are related to liking; $slope \pm SE = 0.53 \pm 0.24$; $t(40.4) = 2.21$, $p = .035$, with higher estimates being associated with better-liked images (and vice versa). The effect is moderated by the image random factor; it is less strong for some images than for others (see correlations in the *interactive database*), since the random factor moderates the slopes of those relationships, $SD = 1.07$; *LRT: chi-square* (2) = 13, $p = .002$.

Estimate and ease The estimates relate positively to the ease felt; $slope \pm SE = 3.80 \pm 0.28$; $t(39.2) = 13.6$, $p < .001$, with higher estimates being felt as easier to make. The interaction

with the random image factor shows that this relationship is stronger for some images than others (see correlations in the *interactive database*), $SD = 1.52$; *LRT: chi-square* (2) = 85.3, $p < .001$.

Estimate bias and judgements None of the judgements (i.e., perceived familiarity, liking, and ease) is related to the estimate bias (all $ps > .12$).

Relationship between objective and subjective proportions

Each image was presented in seven different levels of proportions to be estimated. Here, we examine the relationship between these objective proportions and subjective proportions (i.e., estimates) within a mixed-model analysis, with the specific images as a random factor. We test for random effect intercept to determine whether some images lead to higher estimates than others. We test for two-way interaction with the random factor to understand whether some images are more prone to general response bias than others.

Results show that estimates made by participants covary with the *objective proportion* presented in the image; $slope \pm SE = 0.97 \pm 0.03$; $t(39) = 35.20$, $p < .001$, with the *intercept* of this relationship being near the 50% value⁴ (51.49; $SE = 1.21$; $t(40.4) = 2.21$, $p = .035$). The intercept of the random factor is significant; $SD = 7.63$; *ICC* = 0.226; *LRT: chi-square* (1) = 3878, $p < .001$, suggesting that some images promote higher estimates than others. We also found two-way interaction with the random factor, suggesting that the factor interferes with the slope of the relationship ($SD = 0.168$; *LRT: chi-square* (2) = 822, $p < .001$). This suggests that some images promote better estimates across all levels than others (see mean differences in estimate bias). Thus, future researchers should choose their images carefully as those that are more or less susceptible to general response bias and thus to objective proportion sensitivity. For instance, image PD1_50 is the image on the *interactive database* with the lowest mean value of estimate bias, with the mean values of estimate bias for each of the seven proportion variants of this scenario and image version ranging from 2.36 (PD1_50) to 8.49 (PD1_20). On the other hand, image PC3_50 has a mean estimate bias of 8.5, with the mean values of estimate bias for each of the seven proportion variants of this scenario and image version ranging from 6.43 (PC3_80) to 12.09 (PC3_60). Moreover, readers can look at the correlations between estimates and estimate bias (i.e., “EstimateRBias”), along with the respective p -values

³ However, it should be taken into account that the *LRT* for variance parameters can be poorly estimated at times, which indicates that these p -values should be interpreted with caution (Baey et al., 2019).

⁴ The estimated mixed-model recoded objective proportions with 0 representing the 50% proportion level. Thus, this intercept corresponds to the standard 0 of a psychophysical function.

(EstimateRBias_p), which inform the reader about the slopes that operationalise how the degree of bias is dependent on the magnitude of the estimate. More specifically, these are either non-significant (i.e., a constant), positive (i.e., higher estimates are associated with higher bias), or negative (i.e., lower estimates are associated with lower bias). As an example, it can be seen in the interactive database that by selecting the image scenario “PD” and version “1” (PD1) for a future study, we would be selecting an image for which those who provide higher estimates also present less bias, which suggests that the bias promoted by the image is an underestimation. On the contrary, by selecting the image scenario “BA” and version “3” (BA3), we would be selecting an image for which the bias increases for higher estimates, which suggests that the bias promoted by the image is an overestimation.

Differences promoted by image categories

In this section, we analyse differences between image categories in terms of judgements and proportion estimates, and also how they modulate the previously studied relationships, adding them independently, as moderators, to the previously defined mixed models⁵. For this last analysis we also test for random second-order interaction effects (three-way interaction), to determine whether the category moderation effect was the same or different for all the images.

The categories analysed correspond to those described above: discrete versus continuous (*type of quantity*), depicting social elements or not—presence versus absence (*social elements*), artificial versus natural (*nature of stimuli*), as being dichotomic or having more levels (*total types of referents*), and as either varying in amount, colour, or shape (for discrete stimuli), or varying in unitary area or multiple areas (for continuous stimuli) (*stimuli dimensions*).

Below, we summarise our analyses, firstly regarding (a) how each category influences judgements and estimates and (b) how each category moderates the relationships between judgements, and secondly, regarding (c) how each category influences the relationship between objective and subjective proportions (see Fig. 2).

Type of quantity *Discrete* and *continuous* images do not differ in terms of familiarity but differ in terms of liking (*continuous* images are liked best: $effect \pm SE = -0.48 \pm 0.18$; $t(36.9) = -2.71$, $p = .010$). Differences in estimation ease are also significant; $effect \pm SE = -0.230 \pm 0.13$; $t(270) = -3.90$,

$p < .001$, suggesting that estimations are easier for *continuous* images. Discrete and continuous images differ in terms of the mean of proportion of estimates (estimates are lower for *continuous* images; $effect \pm SE = 6.50 \pm 2.55$; $t(270) = 2.55$, $p = .011$) that they promote.

For all of the analyses presented above, the null effect of the image random factor suggests that this effect is similar for any pair of images of both categories.

The image *discrete-continuous* category only interferes with one of the relationships between judgements previously analysed: the relationship between familiarity and liking, $slope \pm SE = -0.09 \pm 0.04$; $t(30.8) = -2.17$, $p = .038$. Simple effects analysis shows that the relationship is stronger for the *continuous* material ($slope \pm SE = 0.15 \pm 0.03$; $t(36.9) = 5.58$, $p < .001$) than the *discrete* material ($slope \pm SE = -0.07 \pm 0.03$; $t(35.5) = 2.22$, $p = .033$). The null second-order interaction component shows that this relationship is equal across all images.

The discrete-continuous factor also interferes with the relationship between objective and subjective proportions; $slope \pm SE = -0.09 \pm 0.01$; $t(16124) = 8.12$, $p < .001$. Simple effects analysis shows that both slopes are close to 1, but that the slope is lower for the *continuous* material ($slope \pm SE = 0.93 \pm 0.007$) than for the *discrete* material ($slope \pm SE = 1.02 \pm 0.008$). This suggests that an increase in the proportion levels reduces the sensitivity to objective proportions to a greater extent for the *continuous* images than for the *discrete* ones. This interaction is further moderated by the image random factor, suggesting that this pattern of differences may be dependent on the specific images selected; $SD_{discrete} = 0.119$, $SD_{continuous} = 0.200$; *LRT: chi-square* (5) = 826, $p < .001$ (see mean correlations in the interactive database).

Presence of social elements Images with *no social elements* and images with *social elements* do not differ in terms of familiarity and ease. However, they differ in liking ($effect \pm SE = -0.478 \pm 0.01$; $t(39) = -2.22$, $p = .032$), with individuals liking images with *no social elements* more than they liked images with *social elements*. A two-way interaction with the random factor, $SD = 1.22$; *LRT: chi-square* (2) = 7.97, $p = .019$, suggests that the effect is not the same for all images, possibly because there is greater variability in the means of liking for *social images* than *non-social images* (see correlations in the interactive database). Again, this variability allows researchers to select images that are equivalent with regard to liking in this category. There are no differences between images with *no social elements* and images with *social elements* regarding the proportion estimates made or estimate bias.

Regarding the previously analysed relationships, this category moderates the relationship found between ease and estimate bias; $slope \pm SE = -0.01 \pm 0.01$; $t(16002) = -2.49$,

⁵ It should be noted that category moderation of the relationship between subjective and objective estimates is akin to differences regarding means in “estimate bias” (difference between the two estimates). Here, we report the effect with this analysis to test image random effects, over both slopes and intercepts.

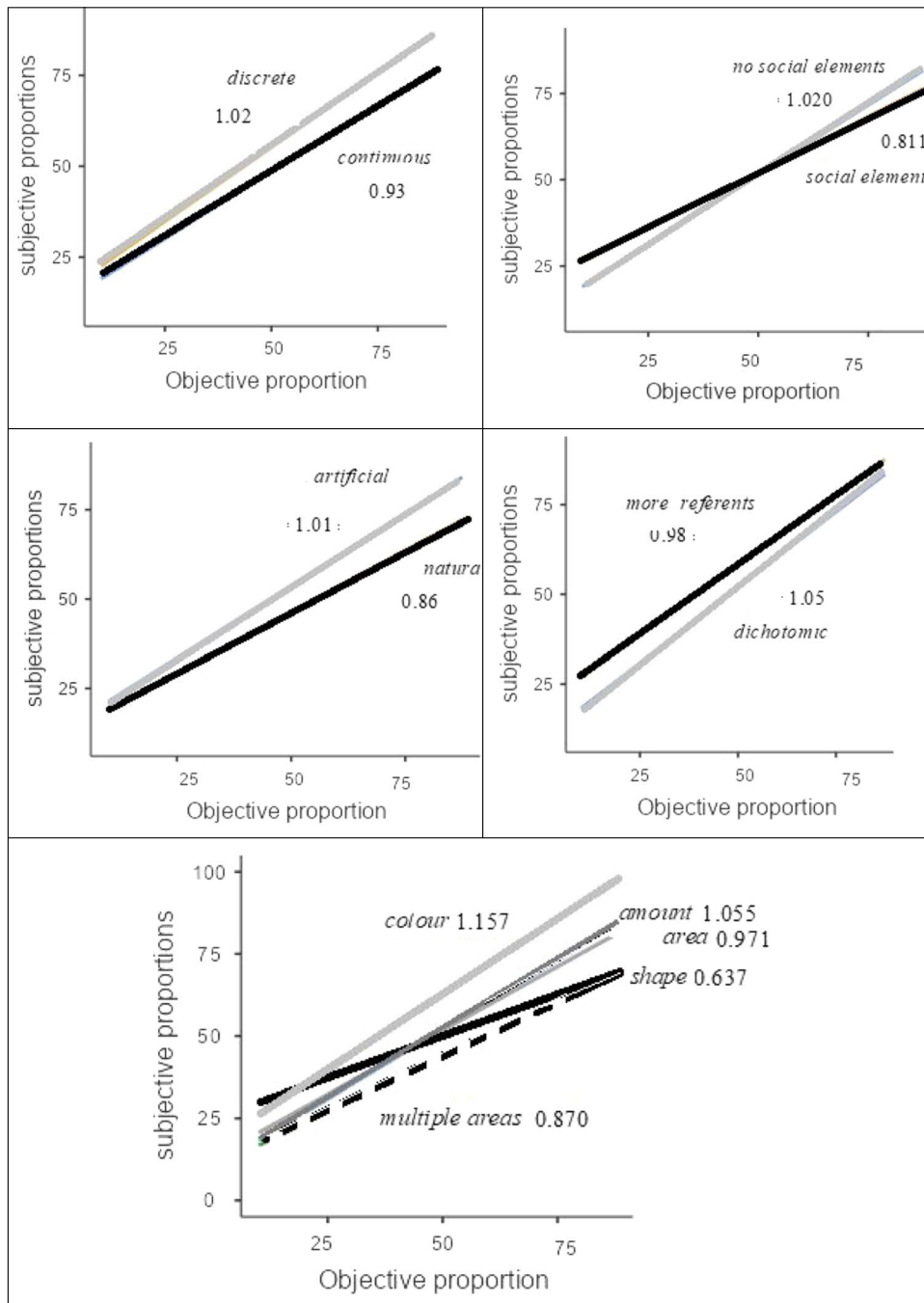


Fig. 2 Relationship between objective and subjective proportions for each category. The slope of the relationship between subjective and objective proportions reflects the amount of bias promoted by each category under analysis. For most of the categories, biases

are not major, but may go in different directions. Higher variability is obtained with different stimuli dimensions, with different shapes reducing accuracy

$p < .013$. Simple effects analysis shows that the slope is positive ($slope \pm SE = 0.01 \pm 0.01$) for images with *no social elements* and the relationship is null ($slope = 0, n.s.$) for images with *social elements*. The interaction is further moderated by the image random factor ($SD_{no\ social\ elements} = 0.02$, $SD_{social\ elements} = 0.01$; $LRT: chi-square (5) = 55.9 p < .001$),

suggesting that this pattern of differences may be dependent on the specific images selected (see correlations in the interactive database).

This category further interferes with the relationship between objective and subjective proportions, $slope \pm SE = -.208 \pm 0.01$; $t(16123.1) = -15.97$,

$p < .001$. Simple effects analysis shows that the slope is smaller ($slope \pm SE = 0.811 \pm 0.01$) for images with *social elements* than for images with *no social elements* ($slope \pm SE = 1.020 \pm 0.006$). More specifically, sensitivity to real proportions is lower for images with *social elements* than for images with *no social elements*. The interaction is moderated by the image random factor ($SD_{no\ presence\ of\ social\ elements} = 0.12$, $SD_{presence\ of\ social\ elements} = 0.20$; $LRT: chi-square(5) = 30.60$, $p < .001$) calling attention to the fact that some images promote more bias than others (see means of bias estimate in the interactive database).

Nature of the stimuli This category did not promote differences in either liking, familiarity, or ease. We found differences in estimates associated with this category; $effect \pm SE = -7.08 \pm 2.31$; $t(22) = -3.06$, $p = .006$. More specifically, participants overestimate *natural* stimuli and underestimate *artificial* stimuli. There was no interaction with the random effect, which indicates that this occurs equally for all images presented.

Regarding the previously analysed relationships, this category moderates the relationship between ease and estimate bias; $slope \pm SE = -0.01 \pm 0.01$; $t(16058) = -4.45$, $p < .001$. Simple effects analysis shows that the slope is positive ($slope \pm SE = 0.01 \pm 0.01$) for *artificial* stimuli and negative ($slope \pm SE = -0.01 \pm 0.01$) for *natural* stimuli. However, the three-way interaction with the random factor shows that this depends on the images considered, with the effect being stronger for some images than others ($SD_{artificial} = 0.02$, $SD_{natural} = 0.01$; $LRT: chi-square(5) = 49.2$, $p < .001$; see correlations in the interactive database).

Natural and artificial stimuli also moderate the relationship between subjective and objective proportions; $slope \pm SE = -0.139 \pm 0.01$; $t(16124.9) = -11.54$, $p < .001$. Simple effects analysis shows that the slope is stronger positive ($slope \pm SE = 1.01 \pm 0.029$) for *artificial* stimuli than ($slope \pm SE = 0.86 \pm 0.044$) for *natural* stimuli. This suggests that *artificial* stimuli promote higher sensitivity to objective proportions than *natural* stimuli. The three-way interaction with the random factor is not significant, suggesting that this effect is generalised for all images.

Number of referents The number of referents did not impact ratings of familiarity nor ratings of liking. However, this category impacted ratings of ease ($effect \pm SE = -0.291 \pm 0.10$; $t(124) = -2.84$, $p = .005$), suggesting that *dichotomic* images made the task easier in comparison with images with *more referents*. The random factor two-way interaction is reliable; $SD = 0.896$; $LRT: chi-square(2) = 18.6$, $p < .001$, and revealed that images with *more referents* can be perceived as either easy to estimate or difficult to estimate (see the interactive database), whereas *dichotomic* images are consistently reported to be easy to estimate ($M = 4.26$; $SD = 0.28$).

This category did not promote differences in subjective estimates.

Regarding the previously analysed relationships, this category moderates the relationship between ease and estimates; $slope \pm SE = 1.40 \pm 0.248$, $t(7379) = 5.64$, $p < .001$. Simple effects analysis shows that this relationship is weaker ($slope \pm SE = 0.86 \pm 0.181$) for *dichotomic* stimuli than for stimuli with *more referents* ($slope \pm SE = 2.26 \pm 0.170$). The three-way interaction shows that the differences promoted by the category are stronger for some sets of images than others; $SD_{dichotomic} = 1.09$, $SD_{more\ referents} = 1.37$; $LRT: chi-square(5) = 35.3$, $p < .001$.

This category further moderates the relationship between ease and estimate; $slope \pm SE = 0.54 \pm 0.180$, $t(7485) = 3.73$, $p = .002$. Simple effects analysis shows that this relationship is stronger ($slope \pm SE = 0.14 \pm 0.132$) for *dichotomic* stimuli than for stimuli with *more referents* ($slope \pm SE = 0.69 \pm 0.123$). This difference is stronger for some images than others ($SD_{dichotomic} = 0.59$, $SD_{more\ referents} = 1.1$; $LRT: chi-square(5) = 40.5$, $p < .002$), suggesting that, depending on the images, this difference may be more or less clear (see correlations in the interactive database).

Finally, this category moderates the relationship between objective and subjective proportions; $slope \pm SE = 6.31 \pm 2.87$; $t(11.5) = 2.20$, $p = .049$. Simple effects analysis shows that the slope is higher ($slope \pm SE = 1.05 \pm 0.051$) for *dichotomic* stimuli than for stimuli with *more referents* ($slope \pm SE = 0.98 \pm 0.051$). This indicates that individuals are better at estimating proportions in *dichotomic* stimuli. This effect is stronger for some images than others; $SD_{dichotomic} = 0.04$, $SD_{more\ referents} = 0.43$; $LRT: chi-square(5) = 22.6$, $p < .001$, suggesting that differences promoted by the number of referents are more likely found with some images than others (see correlations in the interactive database).

Stimuli dimensions Stimuli dimensions promoted differences in familiarity and liking judgements. Specifically, these differences were found to be associated with the shape dimension (see Table 2 and interactive database). *Shape* was perceived to be less familiar ($F(4, 38.9) = 3.25$, $p = .021$) and liked less ($F(4, 38.9) = 12.5$, $p < .001$) than other stimuli dimensions. Image as a random factor did not interfere with these effects Tables 3, 4, 5 and 6.

This category moderates the relationship between ease and estimate, $F(4, 7.23) = 18.8$, $p < .01$. Specifically, this relationship is stronger for the *colour* dimension and weaker for the *amount* dimension (although still significant, $p < .001$). Furthermore, this category moderates the relationship between ease and estimate bias, ($F(4, 1462) = 20.3$, $p < .010$). Specifically, this relationship is stronger for the *colour* dimension and weaker for the *multiple areas* dimension

Table 2 Descriptive statistics for the variables estimate, estimate bias, ease, familiarity, and liking for the category type of quantity

	Type of quantity	Mean	SD
Estimate (0–100%)	Continuous	48.43	23.30
	Discrete	55.01	26.43
Ease (1–7)	Continuous	4.35	1.35
	Discrete	4.12	1.47
Familiarity (1–7)	Continuous	5.98	1.28
	Discrete	5.90	1.30
Liking (1–7)	Continuous	4.26	1.63
	Discrete	3.78	1.65
Estimate bias (0–80)	Continuous	10.30	10.43
	Discrete	13.02	11.78

Table 3 Descriptive statistics for the variables estimate, estimate bias, ease, familiarity, and liking for the category presence of social elements

	Presence of social elements	Mean	SD
Estimate (0–100%)	No social elements	51.34	25.31
	Social elements	51.95	24.00
Ease (1–7)	No social elements	4.27	1.35
	Social elements	4.13	1.58
Familiarity (1–7)	No social elements	5.96	1.25
	Social elements	5.87	1.43
Liking (1–7)	No social elements	4.15	1.58
	Social elements	3.68	1.85
Estimate bias (0–80)	No social elements	11.27	10.45
	Social elements	12.51	13.19

Table 4 Descriptive statistics for the variables estimate, estimate bias, ease, familiarity, and liking for the category nature of the stimuli

	Nature of the stimuli	Mean	SD
Estimate (0–100%)	Artificial	53.60	25.36
	Natural	46.67	23.51
Ease (1–7)	Artificial	4.29	1.44
	Natural	4.12	1.35
Familiarity (1–7)	Artificial	5.98	1.26
	Natural	5.86	1.36
Liking (1–7)	Artificial	4.10	1.66
	Natural	3.91	1.63
Estimate bias (0–80)	Artificial	11.15	11.02
	Natural	12.48	11.42

(although still significant, $p < .001$). Image as a random factor did not interfere significantly with these effects.

Finally, this category moderates the relationship between objective and subjective proportions, $F(4,$

Table 5 Descriptive statistics for the variables estimate, estimate bias, ease, familiarity, and liking, for the category number of referents

	Number of referents	Mean	SD
Estimate (0–100%)	Dichotomic	51.89	25.09
	More referents	58.13	27.36
Ease (1–7)	Dichotomic	4.26	1.33
	More referents	3.97	1.58
Familiarity (1–7)	Dichotomic	5.99	1.18
	More referents	5.81	1.41
Liking (1–7)	Dichotomic	3.91	1.51
	More referents	3.65	1.77
Estimate bias (0–80)	Dichotomic	9.73	9.75
	More referents	16.31	12.68

Table 6 Descriptive statistics for the variables estimate, estimate bias, ease, familiarity, and liking for the category stimuli dimensions

	Stimuli dimensions	Mean	SD
Estimate (0–100%)	Amount	51.89	25.09
	Area	51.51	23.08
	Colour	62.39	26.66
	Multiple areas	44.28	22.96
	Shape	49.61	26.77
Ease (1–7)	Amount	4.26	1.33
	Area	4.61	1.31
	Colour	4.40	1.39
	Multiple areas	4.00	1.34
	Shape	3.12	1.59
Familiarity (1–7)	Amount	5.99	1.18
	Area	6.07	1.22
	Colour	5.94	1.24
	Multiple areas	5.86	1.35
	Shape	5.54	1.67
Liking (1–7)	Amount	3.91	1.51
	Area	4.46	1.61
	Colour	4.14	1.55
	Multiple areas	4.00	1.62
	Shape	2.68	1.80
Estimate bias (0–80)	Amount	9.73	9.75
	Area	8.30	9.44
	Colour	14.98	10.59
	Multiple areas	13.00	11.08
	Shape	18.97	15.75

16124.7) = 150.00, $p < .001$. Simple effects analysis clarifies that only the *amount* ($slope \pm SE = 1.055 \pm 0.01$) and *area* ($slope \pm SE = 0.971 \pm 0.01$) slopes are closer to 1. Slopes are lower than 1 for *multiple areas* ($slope \pm SE = 0.870 \pm 0.01$), *shape* ($slope \pm SE = 0.637 \pm 0.02$), and higher than 1 for *colour* ($slope \pm SE = 1.157 \pm 0.01$). This suggests that researchers

may choose the dimension to be estimated knowing which dimensions increase or decrease estimation correctness. The three-way interaction with the random factor was not significant, suggesting that image as a random factor did not interfere significantly with these effects.

Discussion

In this paper, we have presented normative data associated with a set of visual stimuli that depict social images and allow for judgements of proportion within a part-whole task. These stimuli were developed as an addition to traditional stimuli (mostly compositions of simple geometric elements) used in a broad range of research topics in proportional reasoning studies. The new material adds social relevance to those stimuli (or real-world correspondence), while at the same time providing the capacity to control for other dimensions that are here shown to be relevant for increasing or reducing the bias of estimates. It should be noted, however, that no comparisons were made in the present work between typical stimuli sets used in prior work, as our goal was first and foremost to provide normative data for these newly developed visual stimuli depicting social images. However, future research should take this aspect into consideration, as comparisons between new and traditionally employed stimuli could prove highly valuable in the advancement of the field.

All images under analysis in this paper are readily downloadable, and all information related to them is searchable within an online interactive database. We provide researchers with these materials to facilitate the advancement of research on this topic, since developing these kinds of materials can be time-consuming, even requiring specific skills.

Data available in the online interactive file enable a proper selection of materials for future studies. In selecting materials, researchers should take into account the analysis performed at the individual level in this paper. The category of the image may itself impact either the proportion estimates or their relationship with the objective proportions. Importantly, when the images, added to the analysis as a random factor, are shown to significantly interact with the impact of a category, it means that there is sufficient variability in the image set and that researchers may be able to find images that either strengthen or control the effect at stake. For instance, when data show that discrete and continuous stimuli differ in terms of the sensitivity to real proportions, but that image as a random factor qualifies this difference, researchers can find an image setting for which sensitivity to real proportions is equivalent for continuous and discrete images. For that, authors can rely on the analysis conducted at an individual level in this paper, and on

the correlation analysis at the image level provided in our interactive database.

The analyses conducted at an individual level are further informative in how individuals perceive an image (judgements of liking and familiarity) and the ease felt when providing a response, with the relationships between these variables differing depending on the category of the image. A specific category of images may also moderate the relationship between objective and subjective proportions, suggesting that individuals' sensitivity to the real proportions depends on the image category.

Researchers should consider the fact that the images differ in how they are perceived, and that this perception is related to their estimates and/or estimate bias. Images differ in how much they are liked, and the *level of liking* is associated with the estimate bias. More specifically, images that are less liked are underestimated in their proportions, and images that are more liked are overestimated. Furthermore, images differ in terms of the ease felt by participants when providing estimates. This is relevant because there is a positive relationship between the values estimated and the *perceived ease*. This suggests that higher values are more easily estimated or that the estimation of higher values makes participants feel that the process is easier, or even that if individuals experience processing ease, they estimate higher values. Therefore, it seems that the way individuals perceive a stimulus, associated with how much they like the stimulus and how familiar they feel the stimulus to be, as well as the ease with which they made their estimates, is directly related to the magnitude of estimates. Social cognitive research suggests affect as a relevant modulator of subjective probability estimates (Bohner et al., 1988; Fox & Levav, 2000; Suantak et al., 1996). Since processing of familiar and perceptual fluent stimuli is charged with positive affect (e.g., Garcia-Marques et al., 2010, 2016), it is not surprising that our data directly demonstrate that these features interfere with the proportion estimation process. Considering that the goal of this paper is to provide stimuli for which relevant properties are previously known, when studying proportion reasoning, we do not draw conclusions about the presence of the effects, and neither explore nor offer possible explanations for them. However, we believe that this should be undertaken in future research.

Importantly, this work showed that the relationship established between the subjective and objective proportion estimates was qualified by different types of materials, suggesting that individuals are *less sensitive* to real proportions when (a) these proportions change *continuously* rather than in a *discrete* way; (b) images do not have *social elements* (versus having *social elements*); (c) images represent *natural* (versus *artificial*) scenery; (d) the stimuli represented are *dichotomic* (versus *multiple referents*); and (e) images depict *shapes, colour, and multiple areas*

(versus other elements). This diversity of biases occurring with perceptual stimuli is likely to inform theories that aim to explain proportion estimation processes. The bias detected with our stimuli may have two different sources: either perception differs starkly from explicit expressions, or the numerosity perceptual context modulates perception itself. In the first case, bias occurs in the process of transforming perceptions into responses (see Shepard, 1984), and in the second, bias occurs in the perception of the quantity itself (e.g., Ebbinghaus illusions). An insight provided by the fact that there are differences in the levels of distortions promoted by the different categories is that perceptual mechanisms play a relevant role in promoting these biases. The data suggest that the characteristics of the images themselves create biases, and that these are not promoted exclusively by the process which transforms what is visually perceived into a cognitive response. This perceptual factor should thus be added to the list of factors that are already known to have an impact on the magnitude of systematic estimation errors, such as individual differences (e.g., overconfidence: see Moore & Healy, 2008) and contextual features (e.g., anchoring; see Furnham & Boo, 2011). While it is likely that the same type of bias occurs through different pathways, parsimony would challenge researchers to test the relationship between these different factors—for example, testing whether individual differences or contextual factors are not themselves interfering with how we perceive reality.

Although the material provided makes it easy for researchers to test their hypotheses, additional control procedures are still needed in specific studies. For example, although our study manipulates and measures many features of the images, there are other features, such as perceived visual complexity, that are also relevant in understanding proportion estimation processes (e.g., Madan et al., 2018). Thus, authors may need to assess the degree of perceived complexity of the images that they select for their studies. In addition, it should be kept in mind that, in this normative study, the different images comprising different versions of the same scenario were assessed by different participants (i.e., a between-subjects comparison). Therefore, our data do not allow us to determine whether the scenarios are truly perceived similarly or, in other words, to report on the consistency levels of the evaluations provided to the different images of the same scenario (i.e., their reliability scores).

Extending the use of this set of images to other magnitude estimation tasks

This set of images was addressed here in a *part-whole task* that requires the estimation of the proportion (in percentage) of a type of element relative to the total of the

elements, or of a part of an entity relative to its entirety. However, the use of these stimuli can be extended to other proportion judgement tasks, such as *scaling tasks*, in which the participant is asked to match a target stimulus (composed of two elements with a given proportional relation) to its proportional equivalent (e.g., Barth et al., 2003; Mix et al., 1999; Meert et al., 2012; Sophian, 2000). Tasks of matching non-symbolic fractions (e.g., Boyer et al., 2008; Jacob & Nieder, 2009; Jeong et al., 2007; Meert et al., 2012) or spatial scaling tasks (e.g., Huttenlocher et al., 1999; Möhring et al., 2018) are some examples.

We further believe that the use of these images can also be extended to other magnitude tasks of the *part-part ratio comparisons* type, in which the quantity of an element is evaluated as a *fraction* of another (or its multiple), or of the *ordinal comparisons* type, which involve the comparison of the proportion of a stimulus relative to the proportion of others (B is smaller than/equal to/larger than A).

Acknowledgements This research was supported and funded by the Fundação para a Ciência e Tecnologia under Grant PTDC/PSI-GER/28850/2017 and Grant UIDB/04810/2020 awarded to Teresa Garcia-Marques

References

- Attwood, A. S., Scott-Samuel, N. E., Stothart, G., & Munafò, M. R. (2012). Glass shape influences consumption rate for alcoholic beverages. *PLoS One*, 7(8). <https://doi.org/10.1371/journal.pone.0043007>
- Baey, C., Cournède, P. H., & Kuhn, E. (2019). Asymptotic distribution of likelihood ratio test statistics for variance components in nonlinear mixed effects models. *Computational Statistics & Data Analysis*, 135, 107–122. <https://doi.org/10.1016/j.csda.2019.01.014>
- Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition*, 86(3), 201–221. [https://doi.org/10.1016/s0010-0277\(02\)00178-6](https://doi.org/10.1016/s0010-0277(02)00178-6)
- Beran, M. J., & Parrish, A. E. (2016). Going for more: Discrete and continuous quantity judgments by nonhuman animals. In A. Henik (Ed.), *Continuous issues in numerical cognition: How many or how much* (pp. 175–192). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-801637-4.00008-1>
- Bohner, G., Bless, H., Schwarz, N., & Strack, F. (1988). What triggers causal attributions? The impact of valence and subjective probability. *European Journal of Social Psychology*, 18, 335–345. <https://doi.org/10.1002/ejsp.2420180404>
- Boyer, T. W., Levine, S. C., & Huttenlocher, J. (2008). Development of proportional reasoning: Where young children go wrong. *Developmental Psychology*, 44(5), 1478–1490. <https://doi.org/10.1037/a0013110>
- Chandon, P., & Ordabayeva, N. (2009). Supersize in one dimension, downsize in three dimensions: Effects of spatial dimensionality on size perceptions and preferences. *Journal of Marketing Research*, 46(6), 739–753. <https://doi.org/10.1509/jmkr.46.6.739>
- Cohen, D. J., & Sarnecka, B. (2014). Children's number-line estimation shows development of measurement skills (not number representations). *Developmental Psychology*, 50, 1640–1652. <https://doi.org/10.1037/a0035901>

- Doi, H., & Shinohara, K. (2016). Emotional faces influence numerosity estimation without awareness. *Cognitive Processing*, 17(4), 389–397. <https://doi.org/10.1007/s10339-016-0774-5>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Fornaciai, A., Favalli, M., & Nannipieri, L. (2019). Numerical simulation of the tsunamis generated by the Sciara del Fuoco landslides (Stromboli Island, Italy). *Scientific Reports*, 9, 18542. <https://doi.org/10.1038/s41598-019-54949-7>
- Fox, C. R., & Levav, J. (2000). Familiarity bias and belief reversal in relative likelihood judgment. *Organizational Behavior and Human Decision Processes*, 82(2), 268–292. <https://doi.org/10.1006/obhd.2000.2898>
- Fujimura, N. (2001). Facilitating children's proportional reasoning: A model of reasoning processes and effects of intervention on strategy change. *Journal of Educational Psychology*, 93(3), 589–603. <https://doi.org/10.1037/0022-0663.93.3.589>
- Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, 40(1), 35–42. <https://doi.org/10.1016/j.socec.2010.10.008>
- Garcia-Marques, T., Mackie, D. M., Claypool, H. M., & Garcia-Marques, L. (2010). Is it familiar or positive? Mutual facilitation of response latencies. *Social Cognition*, 28, 205–218. <https://doi.org/10.1521/soco.2010.28.2.205>
- Garcia-Marques, T., Prada, M., & Mackie, D. M. (2016). Familiarity increases subjective positive affect even in non-affective and non-evaluative contexts. *Motivation and Emotion*, 40(4), 638–645. <https://doi.org/10.1007/s11031-016-9555-9>
- Ginsburg, H., & Rapoport, A. (1967). Children's estimates of proportions. *Child Development*, 205–212. <https://doi.org/10.2307/1127141>
- Goldstone, R. L. (1993). Feature distribution and biased estimation of visual displays. *Journal of Experimental Psychology: Human Perception and Performance*, 19(3), 564–579. <https://doi.org/10.1037/0096-1523.19.3.564>
- Hayashi, M. J., Kanai, R., Tanabe, H. C., Yoshida, Y., Carlson, S., Walsh, V., & Sadato, N. (2013). Interaction of numerosity and time in prefrontal and parietal cortex. *Journal of Neuroscience*, 33(3), 883–893. <https://doi.org/10.1523/JNEUROSCI.6257-11.2013>
- Hollands, J. G., & Dyre, B. P. (2000). Bias in proportion judgments: The cyclical power model. *Psychological Review*, 107(3), 500–524. <https://doi.org/10.1037/0033-295x.107.3.500>
- Huttenlocher, J., Newcombe, N., & Vasilyeva, M. (1999). Spatial scaling in young children. *Psychological Science*, 10(5), 393–398. <https://doi.org/10.1111/1467-9280.00175>
- Jacob, S. N., & Nieder, A. (2009). Tuning to non-symbolic proportions in the human frontoparietal cortex. *European Journal of Neuroscience*, 30, 1432–1442.
- Jeong, Y., Levine, S. C., & Huttenlocher, J. (2007). The development of proportional reasoning: Effect of continuous versus discrete quantities. *Journal of Cognition and Development*, 8(2), 237–256. <https://doi.org/10.1080/15248370701202471>
- Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkman, J. (1949). The discrimination of visual number. *The American Journal of Psychology*, 62(4), 498–525. <https://doi.org/10.2307/1418556>
- Kosıte, D., K nig, L. M., De-Loyde, K., Lee, I., Pechey, E., Clarke, N., Maynard, O., Morris, R. W., ... Munaf , M. R., Marteau, T. M., Fletcher, P.C., & Hollands, G. J. (2019). Plate size and food consumption: A pre-registered experimental study in a general population sample. *International Journal of Behavioral Nutrition and Physical Activity*, 16(1), 1-9. <https://doi.org/10.1186/s12966-019-0826-1>
- Lamon, S. J. (2005). *Teaching fractions and ratios for understanding: Essential content knowledge and instructional strategies for teachers* (2nd ed.). Lawrence Erlbaum.
- Li, S., Huang, R., Yuan, H., & Zhou, Q. (2018). Optimising the effectiveness of container packaging design: How shape influences customer proportion judgments. *Applied Marketing Analytics*, 4(1), 79–92.
- Madan, C. R., Bayer, J., Gamer, M., Lonsdorf, T. B., & Sommer, T. (2018). Visual complexity and affect: Ratings reflect more than meets the eye. *Frontiers in Psychology*, 8, 2368. <https://doi.org/10.3389/fpsyg.2017.02368>
- Meert, G., Gr goire, J., Seron, X., & No l, M. P. (2012). The mental representation of the magnitude of symbolic and nonsymbolic ratios in adults. *Quarterly Journal of Experimental Psychology*, 65(4), 702–724. <https://doi.org/10.1080/17470218.2011.632485>
- Meert, G., Gr goire, J., Seron, X., & No l, M. P. (2013). The processing of symbolic and nonsymbolic ratios in school-age children. *PLoS One*, 8(11). <https://doi.org/10.1371/journal.pone.0082002>
- Mix, K. S., Levine, S. C., & Huttenlocher, J. (1999). Early fraction calculation ability. *Developmental Psychology*, 35, 164–174. <https://doi.org/10.1037/0012-1649.35.1.164>
- M hring, W., Frick, A., & Newcombe, N. S. (2018). Spatial scaling, proportional thinking, and numerical understanding in 5- to 7-year-old children. *Cognitive Development*, 45, 57–67. <https://doi.org/10.1016/j.cogdev.2017.12.001>
- Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115(2), 502–517. <https://doi.org/10.1037/0033-295X.115.2.502>
- Mussweiler, T., & Epstude, K. (2009). Relatively fast! Efficiency advantages of comparative thinking. *Journal of Experimental Psychology: General*, 138(1), 1–21. <https://doi.org/10.1037/a0014374>
- Pearson, R. G. (1964). Judgment of volume from photographs of complex shapes. *Perceptual and Motor Skills*, 18(3), 889–900. <https://doi.org/10.2466/pms.1964.18.3.889>
- Pelham, B. W., Sumarta, T. T., & Myaskovsky, L. (1994). The easy path from many to much: The numerosity heuristic. *Cognitive Psychology*, 26(2), 103–133. <https://doi.org/10.1006/cogp.1994.1004>
- Philip, B. R. (1947). Generalization and central tendency in the discrimination of a series of stimuli. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 1(4), 196. <https://doi.org/10.1037/h0084041>
- Piaget, J. (1967). Cognitions and Conservations: Two Views [Review of the book *Studies in cognitive growth*, by J. S. Bruner, R. R. Olver, P. M. Greenfield, J. R. Hornsby, H. J. Kenny, M. Maccoby, N. Modiana, F. A. Mosher, D. R. Olson, M. C. Potter, L. C. Reich & A. M. Sonstrom]. *Contemporary Psychology*, 12(11), 532–533. <https://doi.org/10.1037/007869>
- Prolific. (2020). *Prolific* (Version January 2020) [Online platform]. Retrieved from <https://www.prolific.co>
- Qualtrics. (2020). *Qualtrics* (Version January 2020) [Online platform]. Retrieved from <https://www.qualtrics.com>
- Raghubir, P., & Krishna, A. (1999). Vital dimensions in volume perception: Can the eye fool the stomach? *Journal of Marketing Research*, 36(3), 313–326. <https://doi.org/10.2307/3152079>
- Schwartz, D. L., & Moore, J. L. (1998). On the role of mathematics in explaining the material world: Mental models for proportional reasoning. *Cognitive Science*, 22(4), 471–516. [https://doi.org/10.1016/S0364-0213\(99\)80047-0](https://doi.org/10.1016/S0364-0213(99)80047-0)
- Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology*, 14(4), 332–348. https://doi.org/10.1207/s15327663jcp1404_2
- Shepard, R. N. (1984). Ecological constraints on internal representation: resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological review*, 91(4), 417. <https://doi.org/10.1037/0033-295X.91.4.417>

- Shuford, E. H. (1961). Percentage estimation of proportion as a function of element type, exposure time, and task. *Journal of Experimental Psychology*, *61*(5), 430–436. <https://doi.org/10.1037/h0043335>
- Singer-Freeman, K. E., & Goswami, U. (2001). Does half a pizza equal half a box of chocolates?: Proportional matching in an analogy task. *Cognitive Development*, *16*(3), 811–829. [https://doi.org/10.1016/S0885-2014\(01\)00066-1](https://doi.org/10.1016/S0885-2014(01)00066-1)
- Skagerlund, K., Karlsson, T., & Träff, U. (2016). Magnitude processing in the brain: an fMRI study of time, space, and numerosity as a shared cortical system. *Frontiers in human neuroscience*, *10*, 500. <https://doi.org/10.3389/fnhum.2016.00500>
- Smets, K., Sasanguie, D., Szűcs, D., & Reynvoet, B. (2015). The effect of different methods to construct non-symbolic stimuli in numerosity estimation and comparison. *Journal of Cognitive Psychology*, *27*(3), 310–325. <https://doi.org/10.1080/20445911.2014.996568>
- Sophian, C. (2000). Perceptions of proportionality in young children: Matching spatial ratios. *Cognition*, *75*(2), 145–170. [https://doi.org/10.1016/s0010-0277\(00\)00062-7](https://doi.org/10.1016/s0010-0277(00)00062-7)
- Spence, I. (1990). Visual psychophysics of simply graphical elements. *Journal of experimental psychology. Human Perception and Performance*, *16*(4), 683–692. <https://doi.org/10.1037/0096-1523.16.4.683>
- Spinillo, A. G., & Bryant, P. E. (1999). Proportional reasoning in young children: Part-part comparisons about continuous and discontinuous quantity. *Mathematical Cognition*, *5*(2), 181–197.
- Stevens, S. S., & Galanter, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, *54*(6), 377–410. <https://doi.org/10.1037/h0043680>
- Suantak, L., Bolger, F., & Ferrell, W. R. (1996). The hard–easy effect in subjective probability calibration. *Organizational Behavior and Human Decision Processes*, *67*(2), 201–221. <https://doi.org/10.1006/obhd.1996.0074>
- The jamovi project. (2020). *jamovi* (Version 1.26) [Computer Software]. Retrieved from <https://www.jamovi.org>
- Troy, D. M., Attwood, A. S., Maynard, O. M., Scott-Samuel, N. E., Hickman, M., Woods, A., & Munafò, M. R. (2018). Effect of glass shape on the pouring accuracy of liquid volume. *PLoS One*, *13*(10). <https://doi.org/10.1371/journal.pone.0204562>
- Varey, C. A., Mellers, B. A., & Birnbaum, M. H. (1990). Judgments of proportions. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 613–625. <https://doi.org/10.1037/0096-1523.16.3.613>
- Wansink, B., & van Ittersum, K. (2013). Portion size me: Plate-size induced consumption norms and win-win solutions for reducing food intake and waste. *Journal of Experimental Psychology: Applied*, *19*(4), 320. <https://doi.org/10.1037/a0035053>
- Yang, S., & Raghbir, P. (2005). Can bottles speak volumes? The effect of package shape on how much to buy. *Journal of Retailing*, *81*(4), 269–281. <https://doi.org/10.1016/j.jretai.2004.11.003>
- Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2, Pt.2), 1–27. <https://doi.org/10.1037/h0025848>
- Zhang, H., & Maloney, L. T. (2012). Ubiquitous log odds: A common representation of probability and frequency distortion in perception, action, and cognition. *Frontiers in Neuroscience*, *6*, 1. <https://doi.org/10.3389/fnins.2012.00001>

Open practices statement The data and materials for this study are available at https://osf.io/4cs6k/?view_only=04e5cedf790f4f88ab04c5e40f70b3ab. This study was not preregistered.

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

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.