ISCTE IUL REPOSITÓRIO
INSTITUTO UNIVERSITÁRIO DE LISBOA

Repositório ISCTE-IUL

Deposited in *Repositório ISCTE-IUL*:
2021-10-22

Deposited version:
Accepted Version

Peer-review status of attached file:
Peer-reviewed

Citation for published item:
Souza, C., Garrido, M. V., Saraiva, M. & Carmo, J. C. (2021). RealPic: Picture norms of real-world common items. Behavior Research Methods. 53 (4), 1746-1761

Further information on publisher's website:
10.3758/s13428-020-01523-z

Publisher's copyright statement:
This is the peer reviewed version of the following article: Souza, C., Garrido, M. V., Saraiva, M. & Carmo, J. C. (2021). RealPic: Picture norms of real-world common items. Behavior Research Methods. 53 (4), 1746-1761, which has been published in final form at https://dx.doi.org/10.3758/s13428-020-01523-z. This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

Use policy

Creative Commons CC BY 4.0
The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in the Repository
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Serviços de Informação e Documentação, Instituto Universitário de Lisboa (ISCTE-IUL)
Av. das Forças Armadas, Edifício II, 1649-026 Lisboa Portugal
Phone: +(351) 217 903 024 | e-mail: administrador.repositorio@iscte-iul.pt
https://repositorio.iscte-iul.pt

NOTE: *This is the author´s version of a work that was accepted for publication in the journal Behavior Research Methods. Changes resulting from the publishing process (ie., corrections, structural formatting, etc) may not be reflected in this document. The Version of Record of this manuscript has been published and is available in* *http://doi.org/10.3758/s13428-020-01523*

RealPic: Picture norms of real-world common items

Cristiane Souza

Iscte-Instituto Universitário de Lisboa, CIS-Iscte

Margarida V. Garrido

Iscte-Instituto Universitário de Lisboa, CIS-Iscte

Magda Saraiva

Iscte-Instituto Universitário de Lisboa, CIS-Iscte

Joana C. Carmo

Faculdade de Psicologia, Universidade de Lisboa

Corresponding author:

Cristiane Souza, Iscte-Instituto Universitário de Lisboa, CIS-Iscte

Av. das Forças Armadas, 1649-026, Lisboa, Portugal

Email: Cristiane_Anunciacao_Souza@iscte-iul.pt

Abstract

Pictures are often used as stimuli in several fields, such as Psychology and Neuroscience. However, co-occurring image-related properties might impact their processing, emphasizing the importance of validating such materials to guarantee the quality of research and professional practices. This is particularly pertinent for pictures of common items both because of the high associated knowledge they prompt and their wide applicability potential. Normative studies have already been conducted to create and validate such pictures, yet most of them focused on stimulus without naturalistic elements (e.g., line-drawings). Norms for real-world pictures of common items and rare and their normative examination does not always simultaneously assess affective, semantic and perceptive dimensions, namely in the Portuguese context. Real-world pictures comprehend pictorial representations of the world with realistic details (e.g., natural color or position), thus improving their ecological validity and their suitability for empirical studies or intervention purposes. Consequently, the establishment of norms for real-world pictures is mandatory for exploring their ecological richness and to uncover their impact across several relevant dimensions. In this study, we established norms for 596 real-world pictures of common items (e.g., tomato, drum) selected from existent databases and distributed into 12 categories. The pictures were evaluated on nine dimensions by a Portuguese sample. The results present the norms by item, by dimensions and their correlations as well as cross-cultural analyses. RealPic is a culturally-based dataset that offers systematic and flexible standards and is suitable for selecting stimuli while controlling for confounding effects in empirical tasks and intervention applications.

Keywords: norms; real-world pictures; affective; perceptive; semantic; cross-cultural analysis.

RealPic: Picture norms of real-world common items

Pictures are often used as visual stimuli to access or even improve psychological processes (e.g., Brady et al., 2008; Caramazza & Konkle, 2013). However, pictures are complex stimuli and their characteristics may influence several cognitive and affective processes (Boukadi et al., 2016; Reppa & McDougall, 2015). Therefore, their careful production and validation are essential to guarantee the quality of experimental and interventional designs and to provide comparable results across studies (see Snodgrass & Vanderwart, 1980). Specifically, the assessment of pictures and their characteristics permits the control of their impact on psychological processes, enabling the systematic manipulation of their relevant properties while reducing bias introduced by similar/correlated dimensions (Brodeur et al., 2010; Snodgrass & Vanderwart, 1980).

Critically, validation endeavors require time and precise procedures. In order to overcome this time-consuming task, several databases have been produced and made available to the scientific community. The seminal work by Snodgrass and Vanderwart (1980) constitutes one of the most cited datasets of line-drawing pictures of common items (e.g., animals, fruits, tools), with more than 4.000 and 6.000 citations in SCOPUS and Google Scholar respectively (Souza et al., 2020). Subsequently, several studies replicated and extended this work to different cultures and languages (e.g., Rossion & Pourtois, 2004; Sanfeliu & Fernandez, 1996), to an increased number and types of pictures (e.g., Cycowicz et al., 1997; Rossion & Pourtois, 2004) and to different age groups (e.g., Pompéia et al., 2001; Yoon et al., 2004). Recently, the MultiPic dataset presented an extensive open-access sample of normalized colored line-drawings of common items from the same source, evaluated in name agreement and visual complexity, in six different languages (Duñabeitia et al., 2018).

Notwithstanding the relevance of the existing databases, the importance of using pictures somewhat closer to the real-world in experimental studies has also been acknowledged (e.g., Felsen & Dan, 2005). This concern has motivated the production of more realistic databases (e.g., Foroni et al., 2013; Garrido et al., 2016), which include real-world pictures with vivid and realistic details (e.g., photos) that are suitable for research and intervention.

Common items refer to items of common-name concepts that are easily found in our daily-life. Therefore, pictures of common items are particularly useful for research, such as in semantic memory studies with focus on semantic properties/structure, dissociation of categories as well as in the evaluation of amnesic conditions (e.g., Caramazza & Sheldon, 1998; Farah et al., 1989; Rogers et al., 2015). Considering their high application potential, this type of stimuli may be improved by such ecological concern. However, normative studies that produced and validated real-world pictures of common items are still scarce (e.g., Brodeur et al., 2014; Moreno-Martinez & Montoro, 2012; Shao & Stiegert, 2016). One of the most known databases of real-world pictures of common items is the “Bank of Standardized Stimuli – BOSS” developed by Brodeur and colleagues (2010; 2012; 2014). This database includes a wide range of pictures (930 validated images) of different categories, rated on several attributes (e.g., familiarity, manipulability, visual complexity) and freely available online. Another validated ecological database was offered by Moreno-Martinez and colleagues (2011; 2012), and includes real-world pictures of common items, evaluated, among others, for typicality and manipulability.

Despite the relevance of such databases, the systematic and simultaneous examination of measures from affective, semantic/linguistic and perceptive dimensions of the same set of pictures is not yet available. For example, the BOSS database (Brodeur et al., 2010; 2012; 2014) extensively explored semantic and perceptive dimensions but the affective ones were not investigated. Moreno and colleague’s databases (2011; 2012) present picture norms by-categories but do not address category agreement or any affective dimensions.

In addition, databases with improved ecological validity require careful consideration of important image properties related to their ecological richness (e.g., size, view, color parameters). An example of this concern is provided in the FRIDa - Foodcast Research Image Database (Foroni, et al., 2013), which controlled surface parameters (e.g., brightness and color) while producing norms for real-world pictures of foods and common objects in several important and little explored dimensions, such as aesthetic appeal, valence, arousal, typicality and ambiguity. Rossion and Pourtois (2004) have already shown the advantage in accuracy and reaction times for naming colored line-drawings (vs. black-and-white and grey-scale ones) on a timed vocal naming task. Overall, ignoring such properties implies overlooking additional variables that might be affecting picture processing.

Another important feature to consider in the validation of real-world pictures is the linguistic and/or cultural contexts in which the data is produced. Cross-cultural comparisons have shown that some picture attributes, particularly those related to semantic dimensions (such as familiarity, category agreement, conceptual agreement and name agreement), are culturally-based (Duñabeitia et al., 2018; Kremin et al., 2003; Székely et al., 2004; Yoon et al., 2004). For example, Duñabeitia et al. (2018) provided subjective ratings of name agreement and visual complexity for colored line-drawings in six different European languages across seven European countries. Their findings demonstrated that linguistic similarities are not enough to guarantee the absence of variations in naming (Duñabeitia et al., 2018), since differences were observed for the same language in different cultural contexts (e.g., Dutch-speakers from different countries did not provide the same name for all pictures). Thus, inspecting cultural-based differences is crucial for a better understanding of the way some features of picture processing depend on the cultural background.

To the best of our knowledge, the BOSS is the only real-world pictures database of common items that has been extensively examined in different cultures and languages

(Brodeur et al., 2012; 2014; Clarke & Ludington, 2017). These studies provided interesting inputs regarding culturally based (i.e., English, French, Chinese and Thai) and also linguistic-based differences (i.e., French vs English speakers living in Canada). In the Portuguese context, there are some recently validated picture databases, although they mainly report affective dimensions and none of them focused on real-world pictures of common items (e.g., Garrido et al., 2016; Prada et al., 2016; 2017; Rodrigues et al., 2018). Importantly, the referred studies did not explore cross-cultural differences, nor relevant dimensions, such as typicality, name agreement or category agreement as well as their interaction.

The current work presents a comprehensive, culturally-based, normative study of real-world pictures of common items and includes a systematic validation of several dimensions of picture processing conducted with a Portuguese sample. Specifically, RealPic establishes subjective norms for real-world pictures of 596 common items, selected from existent normalized databases, in nine measures from affective, semantic and perceptive dimensions. These dimensions were selected based on the need to extend existing norms to traditionally less studied dimensions (i.e., arousal, valence, picture-name agreement, and aesthetic appeal) in addition to the most commonly explored ones (e.g., name agreement, familiarity, visual complexity; for a review see Souza et al., 2020).

Dimensions of interest

Category agreement is a relevant indicator that provides general knowledge information about how category membership is processed (see Clarke & Ludington, 2017). The category influence has been observed across several variables, such as familiarity, lexical frequency and typicality (Brodeur et al., 2012; Foroni et al., 2013; Moreno-Martinez, et al., 2011; Rossion & Pourtois, 2004). Categorization may also depend on domain specificities, with living-things processed differently from non-living ones (Caramazza & Sheldon, 1998; Warrington & McCarthy, 1987). Domain effects reflect evolutionary aspects (Caramazza &

Sheldon, 1998) that are expected to influence several variables, such as typicality (Moreno-Martinez et al., 2011) and arousal (Foroni et al., 2013) or even present cultural variance (see Na et al., 2017). Therefore, it seems critical to normalize the stimulus regarding category agreement and to explore the relation that such semantic content presents with other dimensions in a culturally-based manner.

Name agreement refers to the consensus of an individual semantic representation in capturing the most appropriate name as a label for each picture (Pompéia et al., 2001; Snodgrass & Vanderwart, 1980). Name agreement appears to be a consistent measure that is relatively independent of pure language variations as suggested in studies conducted in different languages within the same cultural environment (Brodeur et al., 2012). However, other measures of naming abilities were shown to be affected by linguistic (Kremin et al., 2003; Yoon et al., 2004) and cultural variations (Boukadi et al., 2016; Cycowicz et al., 1997; Duñabeitia et al., 2018). Given its importance to several aspects of pictures and related concept processing (e.g., naming time - Dell´Acqua et al., 2000; reading aloud - Boukadi et al., 2016), the identification of the most common name of the pictures and its variability in a given language assumes particular relevance in picture normalization studies.

Familiarity reflects the degree to which someone interacts or thinks about a specific concept or item-concept in everyday live (concept frequency; Snodgrass & Vanderwart, 1980) and seems to be influenced by characteristics of the respondents such as age, native language and social context (Pompéia et al., 2001). Previous studies suggest that familiarity influences several psycholinguistic measures of picture processing, being positively related with lexical frequency, percentage of name agreement, and typicality, although inversely correlated with visual complexity (see Brodeur et al., 2014; Moreno-Martinez et al., 2011; Snodgrass & Vanderwart, 1980). Familiarity is also a good predictor of affective ratings, showing positive correlations with valence and arousal (Garrido & Prada, 2017; Prada et al., 2016). This dimension has been largely addressed across line-drawing normative studies and may be particularly relevant for real-world pictures of common items.

Typicality refers to how well a given exemplar represents a category (Medin et al., 2007; Murphy et al., 2012). It is dependent of the number of features shared between the item and its own category (e.g., “having feathers”, “having beaks”, into the category “Birds”). Previous studies have shown that less typical items (i.e., items that share less features with their categories) are perceived as less familiar (Moreno-Martinez & Montoro, 2012; Moreno-Martinez et al., 2011, but see Dell´Acqua et al., 2000 for other results), more ambiguous (Foroni et al., 2013), more complex (Moreno-Martinez & Montoro, 2012) and named slower (Dell´Acqua et al., 2000). Although not well explored, typicality is a valuable dimension and examining its interaction with other dimensions may be beneficial to avoid confounding effects.

Arousal represents the emotional activation elicited by an item usually reported in a scale varying from calm to excitatory levels (Foroni et al., 2013; Russell, 1980). In previous studies evaluating symbols, arousal ratings presented a positive correlation with familiarity, aesthetic appeal, visual complexity, concreteness and valence (Prada et al., 2016). Furthermore, previous studies using pictures of food, objects and natural items showed that, overall, arousal presented a positive correlation with valence and also with typicality for natural items but a negative one with familiarity for objects (Foroni et al., 2013). However, normative studies with real-world pictures of common items from different categories have often neglected this dimension.

Aesthetical appeal refers to a preference judgment of beauty based on the capability of an item in attracting interest based on visual liking experience (Prada et al., 2016; Reber et al., 2004). It is a multi-dimensional variable that plays an important role in visual tasks since it entails several features of the aesthetic experience (Reppa & McDougall, 2015), such as surface details of the picture, meaningfulness of the concept or even self-preferences. However, aesthetical appeal is one of the least explored dimensions in picture norms studies.

Valence indicates to which extent an image elicits different degrees of pleasant-unpleasant emotionality (Prada et al., 2014; Russell, 1980). Valence is positively correlated with familiarity, typicality and arousal (Foroni et al., 2013; Prada et al., 2010; Prada et al., 2018) – independently of the item category – and also with aesthetic appeal and visual complexity (Prada et al., 2016), emphasizing the relevance of its inspection in real-world pictures.

Visual complexity is an image-based measure focused on surface features of image quality parameters (i.e., color, shape, brightness, luminosity, contrast, size, complex/simple lines). Snodgrass and Vanderwart (1980) have shown that visual complexity varies as a function of category-specificity. It is also recurrently negatively correlated with familiarity (Brodeur et al., 2012; Brodeur et al., 2014; Pompéia et al., 2001; Prada et al., 2016; Snodgrass & Vanderwart, 1980). Highly complex items modulate category agreement and naming abilities (Brodeur et al., 2014), and are perceived as more appealing, positive and arousing (Prada et al., 2016). It is, therefore, a mandatory dimension in the validation of pictures, particularly real-world pictures due to their realistic surface parameters.

Picture-name agreement refers to the agreement between a concept and its related pictures, often indicated as a viable alternative to measure picture effectiveness in representing the intent concepts (Snodgrass & Vanderwart, 1980). Picture-name agreement is particularly relevant because it allows a direct (based on the concept) way of capturing the agreement between an image and its mental representation (Johnston et al., 2010; Sanfeliu & Fernandez, 1996; Snodgrass & Vanderwart, 1980). Picture-name agreement is positively correlated with categorization (see Sanfeliu & Fernandez, 1996), name agreement (Morrison et al., 1997), and with image agreement (Snodgrass & Vanderwart, 1980), although negatively correlated with familiarity (Sanfeliu & Fernandez, 1996). Its standardization is crucial in real-world pictures as these pictures may not be equally good in visually representing the concepts (e.g., due to different angles and details).

The inspection of such dimensions across languages and cultures may provide important cues about the consistency and generalizability of the norms produced (see Moreno-Martinez & Montoro, 2012; Prada et al., 2017). Therefore, the adaptation of the stimulus sets to different countries enables a more appropriate selection of stimuli regarding linguistic and culturally-dependent aspects, assuring an effective manipulation of stimuli for further empirical or interventional purposes.

The main goals of this research were therefore to (1) establish culturally-based norms of pictures of common-items for the Portuguese context; (2) expand and increase the diversity of parameters standardized in previous studies, namely simultaneously examining affective, semantic and perceptive dimensions using systematic procedures; (3) inspect the consistency of such norms through cross-cultural comparisons.

Methods

Participants

Participants were recruited online through social networks (e.g., Facebook). Participants had to meet all the following criteria: 1) be a native speaker of European Portuguese, 2) be older than 18 years old, 3) have a minimum of four years of formal education; and 4) have their vision preserved or corrected. A sample of 759 participants volunteered to participate in the study. Fifty-nine participants who did not complete at least 50% of the survey and another 16 for not meeting the inclusion criteria were excluded. Overall, the final sample included 684 participants (472 female), with 72.1% completing the entire survey. Participants' age ranged from 18 to 65 years-old, the majority (72.95%) being young adults (age range:18-34), 20.18% mid-aged adults (age range: 35-54) and 6.9% older adults (above 55 years old). The sample reported high education levels (25.4 % post-graduation; 42.1% undergraduates; 32.5% other).

Stimuli

The stimulus set consisted of 718 pictures: 357 were selected from the BOSS database (Version 1, Brodeur et al., 2010; and Version 2, Brodeur et al., 2014), 183 from Moreno-Martinez et al. (2011; 2012) databases, 127 from Konklab database (Brady et al., 2008) and 51 from other free databases licensed for non-commercial usage (e.g., Flirk, Pixabay, Wikipedia). The stimuli were divided into 12 previously defined categories from living (mammal, fruit, vegetable, birds, insects) and non-living (clothing, vehicles, kitchen utensils, musical instruments, furniture, desk materials, tools) domains based on their occurrence in everyday-life, their diversity and their application potential (see Moreno-Martinez & Montoro, 2012, for a similar procedure).

Pictures were resized to 500 X 500 pixels and depicted against a white background. The pictures were previously inspected for their quality during two independent phases using subjective and objective procedures. First, in a pre-selection phase, the most culturally suitable Portuguese name for the item original name was established. Subsequently, four independent raters, native speakers of European Portuguese and completely naïve to the goals of the study, were asked to provide the most appropriate name for the pictures (i.e., two raters named half of the items and the other two the remaining half). Inter-rater agreement was high for both pairs of raters[1] (84% and 79%, respectively). Disagreements between raters were resolved by the first two authors. Overall, these evaluations established the appropriateness of the previously defined name for each item. These two judges also confirmed the suitability of the items for the target categories (see the final distribution of pictures per categories in Table

[1] The agreements were obtained by calculating the percentage of inter-rater agreement for each pair of judges in the cases when they agreed about the target name (i.e., % with which each pair of raters agreed on the name assigned to the picture).

1). Additionally, the first sample of naïve judges was also asked to rate all items regarding their visual quality on a 10-point scale ranging from 1-*very poor quality* to 10- *very good quality*. These procedures lead to the exclusion of 98 pictures (13.64%) that were overall unrecognized/unnamed either due to cultural inadequacy (e.g., the fruit "pecan" or the animal "nyala" are rare or unknown in the Portuguese context), the goodness of the picture in representing the concept (e.g., an image of a "crib" that was not named by any judge) or redundancy (e.g., image of a daddy long leg spider and image of a widow spider being always named as spider). Additionally, twenty-four pictures (3.35%) from the overall sample evaluated as having low quality (i.e., rated below 6 on the quality scale) were excluded. Based on these evaluations 596 (83.01%) out of 718 photographs (119 from BOSS v.1; 175 from BOSS v.2; 158 from Moreno-Martinez & Montoro, 2012; and 144 from other sources) were selected. Each category included about 50 pictures. In a second phase, the color parameters (i.e., RGB and luminance) were also examined to ensure that the visuo-perceptual characteristics were consistent across pictures and to minimize their effect on the ratings of other dimensions. Therefore, a random sample (about 60% of the items) of 356 photographs (from 596) was examined regarding the uniform distribution of RGB and perceived luminance parameters[2] in order to confirm the quality of the selected pictures across domains.

INSERT TABLE 1

Procedure

[2] The surface characteristic of the photographs presented a similar pattern of color (RGB) and luminance distribution (LP) across pictures from different domains. Indeed, planed comparisons revealed that there were no significant differences between the images included in the living and non-living domains [R: $t(403) = 2.31$, $p = .210$; G: $t(403) = 1.53$, $p = .127$; B: $t(403) = .53$, $p = .593$; LP: $t(403) = 1.61$, $p = .109$]. Statistical information regarding these parameters is useful to assure the consistency of the representational quality across the images once it represents an objective measure of visual complexity (see Shao & Stiegert, 2016). For more details see Supplemental Materials.

The study was conducted using the Qualtrics software. After reading the informed consent (including general information, inclusion criteria and ethical information) and agreeing to participate, participants provided sociodemographic information (i.e., age, education, gender and native language). The task instructions were presented, followed by a brief description of each of the dimensions in which pictures should be evaluated. Participants were asked to rate, in seven dimensions, a subset of 40 pictures from different categories randomly selected from a pool of 596 (see Alario & Ferrand, 1999; Brodeur et al., 2014; Cycowicz et al., 1997; Tsaparina et al., 2011 for similar procedures). Additionally, participants were asked to provide a name (name agreement task) and a category (category agreement task) to each picture.

A minimum of 30 evaluations per picture was established, in line with several normative studies using visual stimulus (Brodeur et al., 2010: N = [33, 39]; Brodeur et al, 2014: N= [32, 42]; Johnston et al., 2010: N = [25, 31]; Garrido et al., 2016: N = 30). After treating the data, the number of ratings per picture in each of the seven dimensions ranged from 27 to 34 (M = 30.61, SD = 1.783 to M = 31.20, SD = 1.890). For name agreement and picture name agreement responses per picture ranged from 29 to 57 (M = 32.35, SD = 1.890).

The task was divided into three blocks. Block A included the object-based measures: Familiarity, Arousal and Valence ratings; Block B contained the image-based measures: Visual complexity and Aesthetical Appeal ratings; and Block C consisted of conceptually-based measures such as Name Agreement, Category Agreement, Picture-name Agreement and Typicality. Blocks A and B were randomly presented between participants as well as the order of the dimensions in each Block. Block C was always presented at the end, with a fixed order of dimensions[3]. The dimensions were rated on a 7-point scale (see Table 2), except the

[3] The task order in Block C (conceptually-based) was maintained considering the need to obtain a written modal name and modal category for each item before presenting the target name and target category for each image on Picture-name agreement and Typicality rating tasks, respectively.

naming and the categorization tasks that required a written response (Snodgrass & Vanderwart, 1980). The definition, the scales and the main references for each dimension are presented in Table 2.

INSERT TABLE 2

Results

In this section we present: 1) Preliminary data analysis; 2) Item norms; 3) Descriptive results by evaluative dimension and correlations between dimensions; 4) Linguistic attributes analysis; and 5) Cross-cultural/linguistic data.

Preliminary data analysis

Preliminary analysis of all rated dimensions included the examination of biased inputs and transformations from absolute frequencies to proportional scores. Outliers' analysis followed a criterion of 2.5 standard deviations above or below the mean rating per picture in each dimension (Garrido et al., 2016). Since the occurrence of outliers in all dimensions was very low (range: 1% to 3%), and there was no overall indication of systematic or extremely biased responses, no data were excluded. Missing values were below 5% of the entire database across all rated dimensions. After data treatment, the analysis was run by-item (instead of by participants). The mean ratings (i.e., sum of ratings/N of evaluations per image) and standard deviations were obtained for each image in each dimension. Additionally, a normality test based on curves' peaks and extremities of the distributions indicated that all rated dimensions followed a normal distribution with acceptable values of Kurtosis and Skewness (between ±2; Gravetter & Wallnau, 2014).

Data pre-processing was also conducted for the two linguistic dimensions (i.e., name agreement and category agreement). These dimensions were obtained with free response which provided several linguistic attributes (i.e., modal name-agreement, modal category agreement, alternative valid names/categories, percentage of correct responses and modal responses, and h-value of agreements). Each response was analyzed regarding qualitative (written response) and quantitative (number of references to a given response) parameters. The number of different acceptable responses was quantified for each picture. This procedure included a first inspection for typing variants of the name (e.g., plural, gender, hyphen, composite names with different order, presence of determinants/adjectives/verbs) and spelling mistakes/errors (see Brodeur et al., 2014 for similar procedure). Basic level concepts (e.g., "bird" in reference to "cardinal") and regional variants ("robe", in English *robe*, or "roupão", in English *gown*) were considered as correct. Complete descriptions (e.g., "red orange") were considered different descriptions from summarized ones (e.g., "orange"). Incorrect, don't know and tip-of-the-tongue responses were not considered for further analysis[4].

Item norms

The entire RealPic dataset of norms is provided (Supplemental Materials, Table 1). Detailed information for each item is presented, including: item original database, item original name (i.e., from the original database), item Portuguese target name and item target category. For the seven rating scales, the means and standard deviations, frequencies (number of ratings for each item) and confidence intervals (CI) at 95% are also presented. Additionally, the CI's were used to classify the stimuli as low, moderate or high in each measure (Prada et al., 2016; Rodrigues et al., 2018). Whenever the CI included the scale midpoint (i.e., 4) the items were considered "moderate"; when the upper bound was lower than 4, the items were considered "low"; and when the lower bound of the CI was higher than

[4] The pre-processed responses together with the absolute frequencies for each type of response are available at OSF (https://osf.io/qn35s/?view_only=9c209e9236b94b2cb74f77f47e7ff390) and unfiltered data may be provided upon request to researchers interested in analysing such variations. Considering that images were evaluated in other dimensions, no picture was excluded from the dataset based on low name/category agreement scores.

4, the items were considered “high” (see Supplemental Materials, Table 1). Overall, the obtained normative data is composed of items with a considerable variability in Arousal (175 high, 271 moderated, 150 low), Aesthetic appeal (219 high, 271 moderated, 106 low) and Visual complexity (108 high, 277 moderated, 211 low). The variability of the ratings for Typicality (493 highly typical items), Familiarity (406 highly familiar items) and Picture-name agreement (526 high agreement) was lower. Valence ratings (77 low) were moderate to high.

Descriptive results and correlations by evaluative dimension

Descriptive statistics (mean ratings, standard deviations, confidence intervals, skewness and kurtosis) for each of the seven rated dimensions are provided in Table 3. Overall, the means varied in all the dimensions (see Table 3). Additionally, all the dimensions presented significant differences from the scale midpoint ($p < .05$; see Prada et al., 2018 for further methodological details), with the dimensions of Picture-name agreement presenting the highest mean ratings, and Visual Complexity presenting the lowest mean ratings.

INSERT TABLE 3

Overall, the mean ratings of the seven dimensions presented significant correlations ($p < .05$). Comments on moderate to very strong correlations (Evans, 1996) are provided (see Table 4 for all Pearson´s *r* results). Significant correlations involving less explored dimensions (i.e., Typicality, Arousal, Valence and Aesthetic appeal) in previous normative studies are also reported even if weak.

INSERT TABLE 4

The results showed a positive strong correlation (*r* > .60) between Familiarity and Picture-name agreement. In line with previous findings for photos and line-drawings (Saryazdi et al., 2018), items rated as more familiar also presented increased picture-name agreement. Moreover, moderate correlations (*r* > .40) between Familiarity and Visual Complexity as well as Familiarity and Valence were also observed. Specifically, items rated as less visually complex were considered more familiar (Brodeur et al., 2014; Moreno-Martinez & Montoro, 2012; Sanfeliu & Fernandez, 1996; Shao & Stiegert, 2016; Snodgrass & Vanderwart, 1980; but see Brodeur et al., 2010 for different results) and more positive (see Foroni et al., 2013, for a similar result). Although weak (*r* < .40), some significant correlations presented relevant indicators about the typicality dimension. For instance, typicality was positively correlated with familiarity, confirming previous findings (Moreno-Martinez et al., 2011; Moreno-Martinez & Montoro, 2012), as well as with all the other dimensions (*p* < .05), except visual complexity (*r* < .20).

Visual complexity showed a moderate and positive significant correlation with arousal (*r* = .519). Items rated as complex were also significantly rated as more exciting/arousing. Significant (but weak) correlations between picture-name agreement and valence, typicality, aesthetic appeal (all positive) and visual complexity (negative), were also observed.

The very strong correlation (*r* > .80) observed between valence and aesthetical appeal indicates that the items rated as more positive were also considered more visually appealing. Even though presenting weak correlations (*r* < .40), the significant negative correlations between arousal and aesthetic appeal, valence, and familiarity contrast with the results from previous studies using other type of stimuli in which these correlations were also weak but positive (see Garrido et al., 2016; Prada et al., 2016; Rodrigues et al., 2018). However, the negative correlation between arousal and familiarity is consistent with previous findings using real-world pictures of natural items (see Foroni et al., 2013). The observed correlation between aesthetic appeal and familiarity has also been reported in previous studies using different type of stimuli (e.g., McDougall & Reppa, 2008; Prada et al., 2016; Rodrigues et al., 2018).

Partial correlations were also obtained to control the influence of categories in the correlations between dimensions (see Table 5). Overall, the significant strong correlations reported remained when controlling for categorical effects. Importantly, the positive correlation between typicality and familiarity increased from small to medium. The weak positive correlation between arousal and typicality previously reported without category control was the only one that was not observed with this new analysis.

INSERT TABLE 5

Interestingly, the most powerful correlations were observed among dimensions that were less reported in previous norms of real-word pictures (i.e., aesthetic appeal, valence, arousal and picture-name agreement). Nevertheless, such correlations were reported in normative studies using other type of stimuli (e.g., Prada et al., 2016; Rodrigues et al., 2018), which, together with our findings, emphasize the relevance of exploring these dimensions in real-world pictures.

Linguistic attributes analysis

Name and category agreement included three quantitative measures each: 1) the percentage of correct responses; 2) the percentage of the most common (modal) name/category for the item (e.g., cat / mammal); and 3) the statistic h-value[5]. Overall results are presented in Table 6.

[5] The h-value measure was used to standardize the name or category agreement scores considering the variability of correct names presented for each item. The h-value is inversely related to response-averages of the modal name (Snodgrass & Vanderwart, 1980). In the case, pictures with many attributed names tend to be more complex and each name seems to evoke different mental images. This statistic is sensitive to the diversity of concepts provided, considering the number and the frequency of other possible names (see Brodeur et al., 2014; Pompéia et al., 2001; Snodgrass & Vanderwart, 1980, for more details). To calculate the h-value of name or

INSERT TABLE 6

Regarding name agreement, the percentage of correct responses (92%) was above chance. Participants presented high modal name-agreement (modal NA: *M* = 77.94%, *SE* = 0.92), although considerable variability was observed in valid appropriate names (h-value of NA: *M* = 0.78, *SE* = 0.04). The correspondence between the target name and the modal name was observed in 71% of the 596 pictures. From the responses referring a modal name that was different from the established target name generally, 75.88% reflected culturally accepted general names (e.g., naming different types of spoons with the general concept "spoon", in European Portuguese "colher") or similar names (i.e., naming "tweezers", in European Portuguese "pinça", as an alternative for "tongs" that is "tenaz" in European Portuguese).

The category agreement results indicated an above chance percentage of correct categorization (94%). The modal category agreement was moderate (modal CA: *M* = 65%, *SE* = 0.008), and presented high variability in the valid appropriate categories attributed by the participants (h-value of CA: *M* = 1.40, *SE* = .03), which was expected for this task procedure (i.e., free response). Additionally, the correspondence between the established target category and modal category agreement was observed for 79% of the pictures, with about 7% presenting different but culturally-accepted categories. For example, categorizing "child scooter" as a "toy" instead of "vehicles" or using appropriate non-target categories

---

category, the commonly accepted formula (Brodeur et al., 2014; Snodgrass & Vanderwart, 1980) was used: $H = \sum_{i=1}^{k} P_i log_2\left(\frac{1}{P_i}\right)$, with k referring to the distinct acceptable denominations as correct naming/categorization for each image (excluding the forgettable answers - don't know, don't recognize or don't remember); Pi refers to the proportion of participants that provided an acceptable name/category to the image, excluding errors and forgettable answers. The h-value increases as the number of alternatives of correct names/categories increases. Pictures with a few variations in naming response will present an h-value closer to 0 (Snodgrass & Vanderwart, 1980).

(e.g., naming “legume” for “vegetables”), more specific categories (e.g., “dry fruits” for “fruits”) or more general categories (e.g., “animals” for “mammals” items).

Detailed information about name and category agreement for the entire database and for each image can be found in Table 2 of the Supplemental Materials.

Cross-cultural/linguistic analysis

The current RealPic norms were divided into sub-sets according to their source (original dataset). The mean ratings[6] per item in each sub-set were contrasted with the norms reported in the original datasets: the BOSS dataset (v.1 - Brodeur et al., 2010; v.2 - Brodeur et al., 2014) and the ecological database of Moreno-Martinez and Montoro (2012) obtained with English-Canadian and Spanish samples, respectively (see Table 3, Table 4 and Table 5 of the Supplemental Materials). This analysis was conducted using univariate ANOVAs with 2 Sample (original subsample vs. RealPic) X 2 Domain (living vs. non-living) as factors for each common dimension in both datasets. The variable semantic domain was included in this analysis to provide a more robust inspection of cultural-based effects. Semantic processing involves general knowledge acquired during our life experiences which is related to the environmental context. The processing of non-living items (e.g., tools, furniture, vehicles, etc.) and living ones (e.g., mammals, fruits, birds, etc.) can therefore be influenced by socio-cultural factors, such as cultural values, social needs and evolutionary pressures (see Barbarotto et al., 2002; Na et al., 2017). Domain specificities have been extensively reported in the literature (see Caramazza & Konkle, 2013; Caramazza & Sheldon, 1998; Warrington & McCarthy, 1987; Warrington & Shalice, 1984). Bonferroni adjustment contrasts were used for inspecting main effects and t-tests to explore post-hoc interaction effects.

INSERT TABLE 7

[6] The transformation of scale scores using upper and lower limits in a 0 to 100 scale (see de Vaus, 2002) was applied to compare the means of each common dimension reported in the present norms (7-point scale) and the in the norms of the original databases (5-point scales).

Regarding the comparison of RealPic (Portuguese) versus BOSS v.1 (Brodeur et al., 2010; English-Canadian; item distribution - living items: 31, non-living items: 88), the inspected dimensions were name agreement measures, familiarity and visual complexity. The ANOVA results showed a significant main effect of Sample across dimensions (all *p's* $< .05$), except for visual complexity. Specifically, the Portuguese sample presented higher name agreement (BOSS v.1: $M = 56.58$, $SE = 3.01$; RealPic: $M = 70.14$, $SE = 3.01$) and more consistency in naming (h-value: BOSSv.1: $M = 32.01$, $SE = 2.29$; RealPic: $M = 21.17$, $SE = 2.29$). The Portuguese sample also rated the items as more familiar (BOSSv.1: $M = 60.55$, $SE = 2.31$; RealPic: $M = 76.40$, $SE = 2.31$). The main effect of Domain and its interaction with Sample was not significant for any of the dimensions, indicating consistency across samples by Domain. See Table 7 for detailed results.

The ANOVA results for RealPic (Portuguese) versus BOSS v.2 (Brodeur et al., 2014; English-Canadian; item distribution – living items: 72, non-living items: 103) revealed a significant main effect of Sample across all naming dimensions (all $ps < .05$, see Table 7 for details). Specifically, the Portuguese sample was more accurate in naming (% name agreement- BOSS v.2: $M = 59.01$, $SE = 1.80$, RealPic: $M = 73.65$, $SE = 1.80$) and more consistent in the valid names provided (h-value - BOSS v.2: $M = 38.41$, $SE = 1.54$, RealPic: $M = 17.11$, $SE = 1.54$). In contrast with the above-mentioned comparison with BOSS v.1, the main effect of Domain was observed in all dimensions (all $ps \leq .03$). Living-things were rated as more visually complex (Living: $M = 59.33$, $SE = 1.70$; Non-living: $M = 47.33$, $SE = 1.42$), less familiar (Living: $M = 63.76$, $SE = 1.67$; Non-living: $M = 68.45$, $SE = 1.39$), presented higher name agreement (% of name agreement - Living: $M = 71.21$, $SE = 1.95$; Non-living: $M = 61.45$, $SE = 1.63$) and less variability in naming (h-value - Living: $M = 23.22$, $SE = 1.67$; Non-living: $M = 32.30$, $SE = 1.39$) than non-living things. The interaction effect between

Sample and Domain was significant for most of the dimensions (all $ps \leq .03$; except for Familiarity, $p = .44$), with the Portuguese sample presenting higher name agreement (% of name agreement – Boss v.2: $M = 66.81$, $SE = 2.77$; RealPic: $M = 75.61$, $SE = 2.77$, $t(142) = -2.44$, $p = .016$) and less naming variability (h-value – Boss v.2: $M = 30.29$, $SE = 2.36$; RealPic: $M = 16.15$, $SE = 2.36$, $t(136.499) = 5.09$, $p < .001$) for living things. Living items were also evaluated as less complex by the Portuguese sample (BOSS v.2: $M = 65.31$, $SE = 2.41$; RealPic: $M = 53.34$, $SE = 2.41$, $t(142) = 4.89$, $p < .001$). Regarding the non-living domain, the Portuguese sample showed more agreement in naming (% of name agreement - BOSS v.2: $M = 51.21$, $SE = 2.31$; RealPic: $M = 71.69$, $SE = 2.32$, $t(204) = -5.94$, $p < .001$) and less naming variability in comparison with the English sample (h-value - BOSS v.2: $M = 46.54$, $SE = 1.98$; RealPic: $M = 18.07$, $SE = 1.98$, $t(197.806)= 9.22$, $p < .001$), with no significant differences by sample for the remaining dimensions (all $p > .20$).

The ANOVA results for the RealPic (Portuguese) versus the Ecological database (Moreno-Martinez & Montoro, 2012; Spanish) inspected the dimensions of familiarity, naming agreement, typicality and visual complexity. The results showed a significant main effect of Sample, for familiarity and typicality (all $p < .005$). Portuguese participants rated the items as more typical (Ecological: $M = 63.98$, $SE = 1.87$; RealPic: $M = 76.48$, $SE = 187$) and familiar (Ecological: $M = 62.82$, $SE = 1.86$; RealPic: $M = 70.45$, $SE = 1.86$). Significant main effects of Domain (living: 73 items; non-living: 84 items) for visual complexity and familiarity ($ps < .02$) were also observed, with living things rated as significantly less familiar (Living: $M = 63.40$, $SE = 1.92$; Non-living: $M = 69.87$, $SE = 1.79$) and as visually more complex (Living: $M = 45.03$, $SE = 1.79$; Non-living: $M = 38.04$, $SE = 1.67$) than non-living things. Moreover, significant interaction effects between Sample and Domain were found for name agreement measures (h-value and percentage of NA with $ps \leq .02$). The Portuguese sample presented less variability in naming living-things (h-value - Ecological: $M = 28.45$ SE

= 2.82, RealPic: *M* = 17.02, *SE* = 2.82, $t(132.536) = 2.93$, $p = .004$) but no significant differences between samples were observed for non-living things (all $p > .1$). No differences across cultures were found in the remaining dimensions for living-things and non-living things (all $p > .1$). Statistical details are provided in Table 7.

Discussion

The present study systematically compiled stimuli and extended norms for real-world pictures in nine dimensions comprising the affective, semantic and perceptive domains. RealPic dataset includes a considerable range of pictures distributed across several categories (see Santi et al., 2015). To the best of our knowledge, few normative datasets normed such type of stimuli in the Portuguese context (e.g., Prada et al., 2010; Prada et al., 2014) and none of them includes standards for such a variety of dimensions.

Overall, the results indicated that the RealPic dataset comprises items that are highly familiar, typical, positive, somewhat arousing and visually appealing, medium to low in complexity and presenting high agreement between picture and name. These results are in line with previous studies using real-world pictures of common items, in which those stimuli were rated as relatively complex and presented optimal object agreement (Brodeur et al., 2010; Brodeur et al., 2014). The results also indicate that this type of pictures are less subject to negative feelings (see also Prada et al., 2010), likely because they depict well-known and easily recognizable items. Previous research has shown that the most recognizable and meaningful symbols (high valid responses) were also rated as highly arousing, positive and visually appealing (Prada et al., 2016). Furthermore, the overall high ratings obtained for typicality and familiarity do not constitute a critical issue since real-world pictures of common items are actually expected to be typical and familiar (e.g., Adlington et al., 2009; Brodeur et al., 2014; Moreno-Martinez & Montoro, 2012; Shao & Stiegert, 2016). Congruently, it seems that increasing the quality of the pictures and their proximity to the reality is likely to improve their familiarity, and consequently their typicality ratings, comparatively to line-drawings (see Saryazdi et al., 2018).

The above-chance scores for linguistic attributes (name agreement and category agreement) together with a moderate to high variation of attributed (target and non-target) names and categories, are in line with previous norms using pictures of common items (Brodeur et al., 2010; Brodeur et al., 2014; Snodgrass & Vanderwart, 1980) and also favor the applicability of those stimuli. Moreover, the high variability in category agreement contrasted to the low variation observed in typicality ratings suggests that both dimensions, although part of the categorization processing, may not be identical as considered by Clarke and Ludington (2017). For instance, a picture may be typical even if it is not consistently considered as a member of the target category (e.g., “panini grill”, considered a highly typical item, although presenting high variability in categories attributed and with a CA percentage lower than 40%). In examining such findings, the RealPic dataset is likely to be an useful tool in exploring naming abilities, semantic organization and memory skills[7].

Regarding the correlation results, important insights can be used for a better understanding of the less explored dimensions in previous validation studies, namely arousal, aesthetic appeal, picture-name agreement and valence. The contrast between our correlational results, namely between arousal and aesthetic appeal, valence, and familiarity, and those reported in other normative studies might be related to the specific type of stimuli used across studies. In fact, a previous normative study has shown that the interaction between arousal and other dimensions might depend on the type of stimuli, particularly when they present novelty (see Foroni et al., 2013). In comparison to the distinctiveness of faces (Garrido et al., 2016), symbols (Prada et al., 2016) and emojis (Rodrigues et al., 2018), common items are

[7] In order to increase RealPic usage potential in future studies, the norms per category and their domain are reported in the Supplemental Material.

well-known stimuli related to general knowledge and very frequent in our daily-life, that are likely to be processed in a more semantic manner. The high scores for familiarity, typicality and picture-name agreement observed in RealPic are in line with such perspective.

Original results from our study regarding aesthetic appeal and picture-name agreement showed that such dimensions are positively correlated with all the rated dimensions, except for visual complexity and arousal respectively. Specifically, while aesthetic appeal presented positive correlations with valence (very strong), it was negatively correlated with arousal thus indicating the qualitative differences between these two affective measures as well as their predictive potential. Indeed, aesthetic appeal is a multidimensional variable that seems to capture affective but also the influence of perceptual features (see Reppa & McDougall, 2015), once it focuses on surface image characteristics. Regarding picture-name agreement, the positive correlation (strong) with familiarity (Brodeur et al., 2014; but see Sanfeliu & Fernandez, 1996) and the negative correlation with visual complexity (but see Saryazdi et al., 2018) reflect its multiple influence in both visual and conceptual-based processing (Johnston et al., 2010; Sanfeliu & Fernandez, 1996; Snodgrass & Vanderwart, 1980). Taken together, these findings indicate the relevance of exploring other visual-related attributes of pictures aside from visual complexity to further understand their impact on affective and cognitive processes. The weak/absent correlations between typicality and visual complexity as well as between arousal and typicality and valence still require further examination.

Cross-cultural comparisons indicated that the RealPic items were rated as considerably more familiar than the very same items rated by a Spanish subsample (Moreno-Martinez & Montoro, 2012). Nevertheless, familiarity seems to be the least influenced dimension by Portuguese vs. Canadian cultural differences. Accordingly, strong correlations have been observed across different cultures and languages for familiarity (Boukadi et al., 2016, Brodeur et al., 2012). Such conflicting findings may result from the influence of other variables known to influence familiarity and that were examined simultaneously in our study, such as valence and category agreement (see Foroni et al., 2013; Prada et al., 2018). Moreover, such differences in familiarity ratings could be explained by the fact that the compared items are a subsample of the original datasets used for RealPic which was selected based on their cultural occurrence in the Portuguese environment.

Cultural differences between the Portuguese and Spanish context were also found for typicality ratings. Typicality and familiarity have been presenting positive significant correlations in common items studies (Brodeur et al., 2014; Moreno-Martinez et al., 2011; Snodgrass & Vanderwart, 1980), covarying also by the frequency in which an item or its concept occur. Another possibility is that those findings might have been motivated by the differences in the original items subsamples relatively to living and non-living domains as well as categories, once familiarity and typicality are known to be influenced by category and domain effects (Brodeur et al., 2012; Foroni et al., 2013; Moreno-Martinez et al., 2011; Moreno-Martinez & Montoro, 2012).

The cross-cultural comparison also indicated that name agreement measures (i.e., percentage and h-value) presented significant differences in the Portuguese vs. Canadian samples. However, these measures showed equivalent results for the comparison between Spanish and Portuguese samples, suggesting that similarities in cultural environments associated to the consistent use of pictures may reduce the influence of linguistic differences in naming (see Brodeur et al., 2012). Likewise, linguistic consistency is expected across near-to-Mediterranean cultures and from languages sharing the same linguistic Latin background (Azevedo, 2005). In fact, a previous study reported high correlations of naming measures across languages and/or countries as well as across clustered languages from the same linguistic family (e.g., Germanic or Romance) confirming a reasonable degree of communalities across languages and cultural context (Duñabeitia et al., 2018).

Finally, the main effect of semantic domain (i.e., living and non-living), observed across samples may be also interpreted within a feature distinctiveness approach in which non-living items share less features and present higher correlations with distinctive features than living items (see Moss & Tyler, 1997; Randall et al., 2004). However, the cross-cultural differences (English-Canadian vs. Portuguese and Spanish vs. Portuguese) observed in name agreement, familiarity and visual complexity suggests that cultural background may influence semantic organization. It has been argued that the animacy of the items implies a complex neural network influencing the various stages (i.e., perceptive and semantic) of processing based on their evolutionary weight (see Caramazza & Sheldon, 1998; Nairne et al., 2013). Moreover, the survival issues are susceptible to regions and habits. For instance, it is plausible that cultural characteristics (i.e., climate, accessibility of food, availability and necessity of specific tools or even traditions) may influence the evolutionary-based value of items across the semantic domain in several dimensions which require further cultural examinations. However, the current cross-cultural findings should be interpreted with caution as the current study does not constitute a replication and any methodological differences (i.e., number of assessments, context of data collection, order of presentation of dimensions, etc.) might have influenced the results.

Despite the relevance of such normed dataset, the current study presents a few limitations, namely regarding the number of evaluations per picture, the sample characteristics and the data collection environment. First, a limited number of respondents in psychological studies has driven the production of conflicting findings across studies (Brysbaert, 2019). However, the number of evaluations per item established for the current study was based on previous normative studies that have produced reliable results (Alario &

Ferrand, 1999; Brodeur et al., 2014; Cycowicz et al., 1997; Tsaparina et al., 2011). Second, the sample of our study was fairly homogenous regarding the high levels of participants' education and not equally distributed across age groups, making certain types of comparisons across these variables unfeasible. It well established in the literature that some of the dimensions (e.g., name agreement) assessed in the current study may be influenced by age and education level (Laiacona et al., 2016; Spezzano et al., 2013). For instance, Laiacona and colleagues (2016) have already shown that age and educational level are relevant predictors of naming abilities. Pompéia and colleagues' (2001), also showed differences in normative ratings between children and adults and across different education levels. On the other hand, the demographic characteristics of our sample allowed comparisons with many other normative studies that used highly-educated young adults. Future studies might adopt a developmental approach, contrasting young to older adults with different educational backgrounds in an attempt to grasp potential differences in the explored dimensions. Finally, the use of online resources for collecting data may constitute a challenge in maintaining participants engaged in the study and in establishing some control of the data collection environment. Nevertheless, online data collection procedures allow to overcome a set of constraints regarding the recruitment of participants and has been shown to be as reliable as data collected in lab settings (Saryazdi et al., 2018).

The current norms constitute a useful tool for researchers searching for well-characterized pictures in several dimensions, allowing the manipulation of specific dimensions while controlling others. This enables a better selection of stimuli while avoiding possible confounding effects and ultimately enhancing the quality of the experimental designs. Additionally, the RealPic application potential becomes particularly high if we consider all Portuguese-speaking communities (scattered or territorially distributed) around the world (Godinho & Garrido, 2016) and the rank of the Portuguese language as one of the most spoken languages around the world (see Reto et al., 2016). Future studies should consider the cultural and linguistic diversity of Portuguese speaking communities in non-European Portuguese contexts (i.e., Africa, Asia or South America) as well as expand these norms for additional dimensions (e.g., age-of-acquisition, Johnston et al., 2010; manipulability, Brodeur et al., 2014; image agreement and/or imageability, Snodgrass & Vanderwart, 1980).

In conclusion, the RealPic dataset comprises images of meaningful stimuli commonly encountered in our daily-life. As a particular general class, common items were examined in a more integrative perspective of validating stimuli across a wide range of dimensions emphasizing their independent and combined contributions for picture processing. Furthermore, this research acknowledges a valuable finding about the way we process different types of meaningful information in our "semantic brain" considering cultural diversity. The ecological concern that guided this work and its systematic procedures are likely to make RealPic a promising resource for memory, language and emotion research as well as for interventional settings (e.g., cognitive, linguistic and marketing) requiring more realistic stimuli.

Acknowledgments

The authors would like to thank Luís Teixeira and Catarina Santos for their helpful assistance in stimuli preparation and data collection.

This research was supported by the Fundação para a Ciência e Tecnologia, Portugal, with grants awarded to the first (PD/BD/128249/2016) and the last author (Norma Transitória DL57/2016/CP1439/CT02). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Open Practices Statement

The detailed picture norms and other results are available as supplemental materials (available at https://osf.io/qn35s/?view_only=9c209e9236b94b2cb74f77f47e7ff390) and none preregistration was done for such study.

References

Adlington, R. L., Laws, K. R., & Gale, T. M. (2009). The Hatfield Image Test (HIT): A new picture test and norms for experimental and clinical use. *Journal of Clinical and Experimental Neuropsychology, 31*(6), 731-753. doi:10.1080/13803390802488103

Azevedo, M. (2005). *Portuguese: a linguistic introduction* (1st ed., p. 6-9). Cambridge University Press.

Barbarotto, R., Laiacona, M., Macchi, V., & Capitani, E. (2002). Picture reality decision, semantic categories and gender. Neuropsychologia, 40(10), 1637–1653. doi:10.1016/s0028-3932(02)00029-5

Boukadi, M., Zouaidi, C., & Wilson, M. A. (2016). Norms for name agreement, familiarity, subjective frequency, and imageability for 348 object names in Tunisian Arabic. *Behavior Research Methods, 48*(2), 585-599. doi:10.3758/s13428-015-0602-3

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences, 105*(38), 14325–14329. doi:10.1073/pnas.0803390105

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The bank of standardized stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS ONE, 5*(5): e10773. doi:10.1371/journal.pone.0010773

Brodeur, M. B., Guérard, K., Bouras, M. (2014). Bank of Standardized Stimuli (BOSS) Phase II: 930 New Normative Photos. *PLoS ONE*, 9(9): e106953. doi:10.1371/journal.pone.0106953

Brodeur, M. B., Kehayia, E., Dion-Lessard, G., Chauret, M., Montreuil, T., Dionne-Dostie, E., & Lepage, M. (2012). The bank of standardized stimuli (BOSS): Comparison between French and English norms. *Behavior Research Methods*, *44*(4), 961-970. doi:10.3758/s13428-011-0184-7

Brysbaert, M. (2019). How Many Participants Do We Have to Include in Properly Powered Experiments? A Tutorial of Power Analysis with Reference Tables. Journal of Cognition, 2(1). doi:10.5334/joc.72

Caramazza, A., & Konkle, T. (2013). Tripartite organization of the ventral stream by animacy and object size. *The Journal of Neuroscience, 33*(25), 10235-10242.

Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain: The animate-inanimate distinction. *Journal of Cognitive Neuroscience*, *10*(1), 1–34. doi:10.1162/089892998563752

Clarke, A. J. B., & Ludington, J. D. (2017). Thai norms for name, image, and category agreement, object familiarity, visual complexity, manipulability, and age of acquisition for 480 Color photographic objects. *Journal of Psycholinguistic Research*, *47*(3), 607–626. doi:10.1007/s10936-017-9544-5

Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology, 65*(2), 171–237. doi:10.1006/jecp.1996.2356

Dell´Acqua, R., Lotto, L., & Job, R. (2000). Naming times and standardized norms for the Italian PD/DPSS set of 266 pictures: Direct comparisons with American, English, French, and Spanish published databases. *Behavior Research Methods, Instruments, & Computers, 32*(4), 588-615. doi:10.3758/BF03200832

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, *71*(4), 808–816. doi:10.1080/17470218.2017.1310261

Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences.* Pacific Grove, CA: Brooks/Cole Publishing.

Farah, M. J., Hammond, K. M., Mehta, Z., & Ratcliff, G. (1989). Category-specificity and modality-specificity in semantic memory. *Neuropsychologia, 27*(2), 193–200. doi:10.1016/0028-3932(89)90171-1

Felsen, G., & Dan, Y. (2005). A natural approach to studying vision. *Nature Neuroscience*, *8*, 1643-1646. doi:10.1038/nn1608

Foroni, F., Pergola, G., Argiris, G., & Rumiati, R. I. (2013). The Foodcast Research Image Database (FRIDa). *Frontiers in Human Neuroscience*. 7:51. doi:10.3389/fnhum.2013.00051

Garrido, M. V., Lopes, D., Prada, M., Rodrigues, D., Jerónimo, R., & Mourão, R. P. (2016). The many faces of a face: Comparing stills and videos of facial expressions in eight dimensions (SAVE database). *Behavior Research Methods, 49*(4), 1343–1360. doi:10.3758/s13428-016-0790-5

Garrido, M. V., & Prada, M. (2017). KDEF-PT: Valence, Emotional Intensity, Familiarity and Attractiveness Ratings of Angry, Neutral, and Happy Faces. *Frontiers in Psychology*, 8. doi:10.3389/fpsyg.2017.02181

George, A., & Mathuranath, P. (2007). Community-based naming agreement, familiarity, image agreement and visual complexity ratings among adult Indians. *Annals of Indian Academy of Neurology*, *10*(2), 92-99. doi:10.4103/0972-2327.33216

Godinho, S., & Garrido M. V. (2016). Oral approach-avoidance: A replication and extension for European–Portuguese phonation. *European Journal of Social Psychology*, *46*(2), 260-264. doi:10.1002/ejsp.2172.

Gravetter, F., & Wallnau, L. (2014). *Essentials of statistics for the behavioral sciences* (8th ed.). Belmont, CA: Wadsworth.

Johnston, R. A., Dent, K., Humphreys, G. W., & Barry, C. (2010). British-English norms and naming times for a set of 539 pictures: The role of age of acquisition. *Behavior Research Methods*, *42*(2), 461-469. doi:10.3758/BRM.42.2.461

Kremin, H., Akhutina, T., Basso, A., Davidoff, J., De Wilde, M., Kitzing, P., …Weniger, D. (2003). A cross-linguistic data bank for oral picture naming in Dutch, English, German, French, Italian, Russian, Spanish, and Swedish (PEDOI). *Brain and Cognition*, *53*(2), 243-246. doi:10.1016/s0278-2626(03)00119-2

Laiacona, M., Barbarotto, R., Baratelli, E., & Capitani, E. (2016). Revised and extended norms for a picture naming test sensitive to category dissociations. *Neurological Science*, *37*(9), 1499-1510. doi:10.1007/s10072-016-2611-0

Martin, A., & Chao, L. (2001). Semantic memory and the brain: structure and processes. *Current Opinion in Neurobiology*, *11*(2), 194-201. doi:10.1016/S0959-4388(00)00196-3.

McDougall, S. J. P., & Reppa, I. (2008). Why do I like It? the Relationships between Icon Characteristics, user Performance and Aesthetic Appeal. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *52*(18), 1257-1261. doi:10.1177/154193120805201822

Medin, D. L., Unsworth, S. J., & Hirschfeld, L. (2007). Culture, categorization, and reasoning. In S. Kitayama & D. Cohen (Eds.), *Handbook of cultural psychology*, (pp. 615-644). New York, NY: Guilford Press.

Moreno-Martinez, F. J., & Montoro, P. R. (2012). An ecological alternative to Snodgrass & Vanderwart: 360 high quality colour images with norms for seven psycholinguistic variables. *PLoS ONE 7*(5): e37527. doi:10.1371/journal.pone.0037527.

Moreno-Martinez, F. J., Montoro, P. R., & Laws, K. R. (2011). A set of high quality colour images with Spanish norms for seven relevant psycholinguistic variables: the Nombela naming test. Aging, *Neuropsychology, and Cognition*, *18*(3), 293-327. doi:10.1080/13825585.2010.540849

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *50*(3), 528-559. doi:10.1080/027249897392017

Moss, H. E., & Tyler, L. K. (1997). A category-specific deficit for non-living things in a case of progressive aphasia. *Brain and Language*, *60*, 55-58. doi:10.1016/S0028-3932(99)00044-5

Murphy, G. L., Hampton, J. A., & Milovanovic, G. S. (2012). Semantic memory redux: An experimental test of hierarchical category representation. *Journal of Memory and Language*, *67*, 521-539. doi:10.1016/j.jml.2012.07.005.

Na, J., Huang, C.-M., & Park, D. C. (2017). When Age and Culture Interact in an Easy and Yet Cognitively Demanding Task: Older Adults, But Not Younger Adults, Showed the Expected Cultural Differences. Frontiers in Psychology, 8. doi:10.3389/fpsyg.2017.00457

Nairne, J. S., VanArsdall, J. E., Pandeirada, J. N. S., Cogdill, M., & LeBreton, J. M. (2013). Adaptive memory: The mnemonic value of animacy. *Psychological Science*, *24*, 2099-2105. doi:10.1177/0956797613480803.

Pompéia, S., Miranda, M. C., & Bueno, O. F. A. (2001). A set of 400 pictures standardised for Portuguese: Norms for name agreement, familiarity and visual complexity for children and adults. *Arquivos de Neuro-Psiquiatria*, *59*(2 B), 330-337. doi:10.1590/s0004-282x2001000300004

Prada, M., Cunha, C., Garcia-Marques, T., & Rodrigues, D. (2010). Continuação do FIM – Normas de valência e familiaridade das imagens do Ficheiro de Imagens Multicategoriais. *Laboratório de Psicologia, 8*, 185-211. doi:10.14417/lp.851

Prada, M., Fonseca, R., Garcia-Marques, T., & Fernandes, A. (2014). Se correr o bicho pega...Normas de avaliação de imagens de animais negativos. *Laboratório de Psicologia*, *12*(1): 41-56. doi:10.14417/lp.851.

Prada, M., Garrido, M. V., Camilo, C., & Rodrigues, D. L. (2018). Subjective ratings and emotional recognition of children’s facial expressions from the CAFE set. *PLOS ONE*, 13(12), e0209644. doi:10.1371/journal.pone.0209644

Prada, M., Rodrigues, D., Garrido, M. V., & Lopes, J. (2017). Food-pics-PT: Portuguese validation of food images in 10 subjective evaluative dimensions. *Food Quality and Preference*, *61*, 15-25. doi:10.1016/j.foodqual.2017.04.015.

Prada, M., Rodrigues, D., Silva, R. R., & Garrido, M. V. (2016). Lisbon Symbol Database (LSD): Subjective norms for 600 symbols. *Behavior Research Methods*, *48*, 1370-1382. doi:10.3758/s13428-015-0643-7.

Randall, B., Moss, H. E., Rodd, J. M., Greer, M., & Tyler, L. K. (2004). Distinctiveness and Correlation in Conceptual Structure: Behavioral and Computational Studies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 393–406. doi:10.1037/0278-7393.30.2.393

Reber, R., Schwarz, N., &, Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, *8*, 364-382. doi:10.1207/s15327957pspr0804_3

Reppa, I., & McDougall, S. (2015). When the going gets tough the beautiful get going: aesthetic appeal facilitates task performance. *Psychonomic Bulletin & Review*, *22*(5), 1243-1254. doi:10.3758/s13423-014-0794-z

Reto, L. A., Machado, F. L., & Esperança, J. P. (2016). *New atlas of the Portuguese language*. Lisbon, Portugal: Editora Imprensa Nacional-Casa da Moeda.

Rodrigues, D, Prada, M., Gaspar, R., Garrido, M. V., & Lopes, D. (2018). Lisbon Emoji and Emoticon Database (LEED): Norms for emoji and emoticons in seven evaluative dimensions. *Behavior Research Methods*, *50*(1), 392-405. doi:10.3758/s13428-017-0878-6.

Rogers, T. T., Patterson, K., Jefferies, E., & Lambon Ralph, M. A. (2015). Disorders of representation and control in semantic cognition: Effects of familiarity, typicality, and specificity. *Neuropsychologia*, *76*, 220–239. doi:10.1016/j.neuropsychologia.2015.04.015

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's Object Pictorial Set: The role of surface detail in basic-level object recognition. *Perception*, *33*(2), 217-236, doi:10.1068/p5117

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161-1178. doi:10.1037/h0077714

Sanfeliu, M. C., & Fernandez, A. (1996). A set of 254 Snodgrass-Vanderwart pictures standardized for Spanish: Norms for name agreement, image agreement, familiarity, and visual complexity. *Behavior Research Methods, Instruments, & Computers*, *28*(4), 537-555. doi:10.3758/BF03200541

Santi, A., Raposo, A., & Marques, J. F. (2015). Superordinate and domain category structure: Evidence from typicality ratings. *Revista Portuguesa de Psicologia*, *44*, 81-108. doi:10.21631/rpp44_81

Saryazdi, R., Bannon, J., Rodrigues, A., Klammer, C., & Chambers, C. G. (2018). Picture perfect: A stimulus set of 225 pairs of matched clipart and photographic images normed by Mechanical Turk and laboratory participants. *Behavior Research Methods, 50*(6), 2498-2510. doi: 10.3758/s13428-018-1028-5

Shao, Z., & Stiegert, J. (2016). Predictors of photo naming: Dutch norms for 327 photos. *Behavior Research Methods, 48*(2), 577-584. doi:10.3758/s13428-015-0613-0

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174-215. doi:10.1037/0278-7393.6.2.174

Souza, C. , Garrido, M. V., & Carmo, J. (2020). Common items standardization in adult samples: A Systematic Review of Normative Studies Using Images of Common items. [Unpublished manuscript].

Spezzano, L. C., Mansur, L. L., & Radanovic, M. (2013). Aplicabilidade da "Bateria de Nomeacão de Objetos e Verbos" no Português Brasileiro. *CoDAS, 25*(5), 437-443. doi:10.1590/S2317-17822013000500007

Székely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., … Bates, E. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language*, *51*(2), 247-250. doi:10.1016/j.jml.2004.03.002.

De Vaus, D. (2013). *Surveys In Social Research*. London: Routledge. doi:10.4324/9780203519196

Warrington, E. K., & McCarthy, R. (1987). Categories of knowledge: Further fractionation and an attempted integration. *Brain*, *110*, 1273-1296. doi:10.1093/brain/110.5.1273

Warrington, E. K., & Shallice, T. (1984). Category-specific semantic impairments. *Brain*, *107*, 829-853. doi:10.1093/brain/107.3.829

Yoon, C., Feinberg, F., Luo, T., Hedden, T., Gutchess, A. H., Chen, H. Y. M., …Park, D. C. (2004). A cross-culturally standardized set of pictures for younger and older adults:

American and Chinese norms for name agreement, concept agreement, and familiarity. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 639-649. doi:10.3758/bf03206545

Tables and Figures

Table 1

Distribution of Items by Categories and Domains

Living things (242)	Non-living things (354)
Birds (50)	Clothing (50)
Fruits (47)	Desk material (50)
Insects (47)	Furniture (48)
Mammals (49)	Kitchen utensils (57)
Vegetables (49)	Musical instruments (50)
	Tools (49)
	Vehicles (50)

Table 2

Instructions and their References for Each Dimension

Dimension	*English version*	Main references
	Block A (random)	
Familiarity	*Object familiarity:* you should consider how often you find the picture represented in the image in your daily life, indicating how familiar this stimulus is. Frequently encountered items are usually considered more familiar. For example, an “Apple” is a very familiar fruit, but not a “Guava”. Scale: 1-*unfamiliar* to 7-*very familiar*	Brodeur et al. (2010); Foroni et al. (2013); Prada et al. (2016); Snodgrass & Vanderwart (1980);
Arousal	*Activation capacity of the object:* you should indicate to which extent you consider that this object represents something active/intense or passive/calm Scale: 1-*very passive/calm* to 7-very active/intense	Foroni et al. (2013); Garrido et al. (2016)
Valence	*Valence of the object:* you should consider to which extent this item refers to something positive/pleasant or negative/unpleasant. Scale: 1-*very negative/unpleasant* to 7-*very positive/ pleasant*	Prada et al. (2014); Prada et al. (2016)
	Block B (random)	
Visual complexity	*Visual complexity of an image:* you should evaluate the degree of picture elaboration regarding its visual details (quantity of details, lines patterns, quantity of colors), You should consider the complexity of the visual characteristics of the picture presented, but not the actual object or concept represented. The greater the amount of details/elaboration the more complex the image is. Scale: 1-*very simple* to 7-*very complex*	Brodeur et al. (2010); Cycocwcz et al. (1997); Pompéia et al. (2001); Prada et al. (2016); Snodgrass & Vanderwart (1980);
Aesthetic appeal	*Pleasantness of the image:* you should consider how visually appealing the image is, considering its visual characteristics and not the associated concept or object. Scale: 1-*visually unpleasant/unappealing* to 7-*visually pleasant/appealing*	Prada et al. (2016)

	Block C (fixed order)	
Name agreement and Category agreement	First, you will be asked to identify the item represented on the picture (write the first name that comes to your mind) and its category. Be succinct and write only one name, without ambiguity. For example, when see an image of a “sunflower”, you should write “sunflower” as name response and “flower” as category response. If you do not know the object/category, you should respond "I do not know the object/category". In situations where you identify the object/category, but do not remember the name, answer "I do not know the name of the object/category". However, if you recognize the object/category and know the name but cannot remember it at the moment, say "I do not remember the object/category name".	Pompéia et al. (2001); Snodgrass & Vanderwart (1980)
Picture-name agreement	*Congruence between image and name:* you should evaluate to which extent the image corresponds to a good representation of the name presented. Scale: 1-*very poor representation of the name* to 7-*excellent representation of the name*	Morrison et al. (1997)
Typicality	*Typicality:* you should evaluate to which extent the object is a good example of the indicated category. Consider the representativeness of the stimulus relative to the category, regardless of the frequency you encounter the object in your daily life or your personal preferences. For example, a "Church" can be found frequently, but it will not be a very representative item of the "Buildings" category. The objects considered the best exemplars are the most typical. Scale: 1-*very bad example of its category* to 7-*excellent example of its category*	Foroni et al. (2013); Moreno-Martinez et al. (2011; 2012)

Table 3

Descriptive Statistics for All Items in Each Dimension

	Familiarity	**Typicality**	**Arousal**	**Valence**	**Aesthetical Appeal**	**Visual Complexity**	**Picture-name Agreement**
Mean	5.394	5.747	4.077	4.604	4.255	3.756	6.036
SE of Mean	.045	.040	.037	.039	.037	.036	.036
SD	1.120	.971	.908	.958	.913	.890	.890
Range	2.00-7.00	2.24-7.00	2.24-6.44	1.56-6.70	1.59-6.70	1.69-6.32	2.00-7.00
Skewness	-.738	-.954	.370	-.676	-.129	-.044	-1.143
SE of Skewness	.100	.100	.100	.100	.100	.100	.100
Kurtosis	-.187	.143	-.725	.436	-.032	-.601	.665
SE of Kurtosis	.200	.200	.200	.200	.200	.200	.200
95% IC Low	5.304	5.668	4.004	4.527	4.182	3.684	5.964
95% IC Upper	5.484	5.825	4.150	4.681	4.329	3.827	6.107

Note. Means, Standard Error (SE), Standard Deviation (SD), Range interval (minimum and maximum), Normality estimation (kurtosis and skewness) and Confidence Intervals at 95% for low and upper cut-offs are provided.

Table 4.

Pearson´s r Correlation Values for all Rated Dimensions

	1	**2**	**3**	**4**	**5**	**6**	**7**
1. Familiarity	--						
2. Typicality	.255**	--					
3. Arousal	-.188**	.107*	--				
4. Valence	.431**	.139**	-.288**	--			
5. Aesthetic Appeal	.342**	.190**	-.092*	**.906****	--		
6. Visual Complexity	-.459**	-.044	.519**	-.053	.097*	--	
7. Picture-name agreement	**.686****	.172**	.039	.333**	.310**	-.205**	--

**p* < .05 (two-tailed); ** *p* < .01 (two-tailed).

Note. Significant and strong correlations are presented in bold.

Table 5.

Partial Correlation for all Rated Dimension Controlled by Category

	1	2	3	4	5	6	7
1. Familiarity	--						
2. Typicality	.312**	--					
3. Arousal	-.168**	.072	--				
4. Valence	.421**	.173**	-.277**	--			
5. Aesthetic Appeal	.340**	.209**	-.087*	**.907****	--		
6. Visual Complexity	-.454**	-.104	.515**	-.045	.101*	--	
7. Picture-name agreement	**.679****	.172**	.059	.324**	.308**	-.197**	--

*p < .05(two-tailed); **p < .01 (two-tailed).

Note. Significant and strong correlations are presented in bold.

Table 6.

Descriptive Statistics for all Items in Each Linguistic Attribute

	Correct Naming (%)	%NA	H-VALUE NA	Correct Categorization (%)	%CA	H-VALUE CA
Mean	92.16	77.94	0.78	94.32	65.17	1.40
SE	0.480	0.924	0.04	0.004	0.008	0.034
SD	10.88	20.97	0.90	0.089	20.42	0.81
Skewness	-144.34	-66.35	1.26	-184.13	-9.38	0.44
Kurtosis	104.21	-78.91	1.00	357.14	-89.77	-0.20
95% CI Low	91.22	76.13	0.70	93.59	63.49	1.33
95% CI Upper	93.11	79.76	0.85	95.05	66.84	1.47

Note. NA- Name agreement; CA – Category Agreement; Means, Standard Error (SE), Standard Deviation (SD), Normality estimation (kurtosis and skewness) and Confidence Intervals at 95% for low and upper cut-offs are provided.

Table 7

Main Effects and Interaction Effects Between Sample and Domain Across Rated Dimensions

	BOSS v.1 (Eng) X RealPic (PT)			BOSS v.2 (Eng) X RealPic (PT)			Ecological database (Spanish) X RealPic (PT)		
	Sample $F(1,237)$ 90% CI	Domain $F(1,237)$ 90% CI	Sample X Domain $F(1,237)$ 90% CI	Sample $F(1, 349)$ 90% CI	Domain $F(1, 349)$ 90% CI	Sample X Domain $F(1, 349)$ 90% CI	Sample $F(1, 313)$ 90% CI	Domain $F(1, 313)$ 90% CI	Sample X Domain $F(1, 313)$ 90% CI
NA (%)	10.168^{**} $\eta_p{}^2$=.042 [.00, .09]	n.s.	n.s.	32.889^{***} $\eta_p{}^2$=.087 [.045, .13]	14.623^{***} $\eta_p{}^2$=.041 [.01, .07]	5.229^{*} $\eta_p{}^2$=.015 [.00, .04]	n.s.	n.s.	5.449^{*} $\eta_p{}^2$=.017 [0.00, 0.05]
H-value of NA	11.237^{***} $\eta_p{}^2$= .046 [.01, .09]	n.s.	n.s.	95.419^{***} $\eta_p{}^2$=.216 [.15, .27]	17.341^{***} $\eta_p{}^2$=.048 [.02,.08]	10.797^{***} $\eta_p{}^2$=.030 [.00, .06]	n.s.	n.s.	6.453^{*} $\eta_p{}^2$=.020 [0.00, 0.05]
Visual Complexity	n.s.	n.s..	n.s.	n.s.	29.291^{***} $\eta_p{}^2$=.078 [.03, .13]	13.130^{***} $\eta_p{}^2$=.037 [.01, .07]	n.s.	8.140^{**} $\eta_p{}^2$=.026 [0.00, 0.06]	n.s.
Familiarity	23.408^{***} $\eta_p{}^2$= .091 [.04, .15]	n.s.	n.s.	n.s.	4.643^{*} $\eta_p{}^2$=.013 [.00, .03]	n.s.	8.447^{**} $\eta_p{}^2$=.027 [0.00, 0.06]	6.080^{*} $\eta_p{}^2$=.019 [0.00, 0.05]	n.s.
Typicality	-----	-----	----	----	----	----	23.655^{***} $\eta_p{}^2$=.071 [.03, .12]	n.s.	n.s.

$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

Note. n.s. – non-significant; Eng – English language; PT – Portuguese language; NA – Name Agreement.

Statistical F(F), eta partial effect size ($\eta_p{}^2$) and Confidence Intervals (90%CI) in brackets [ ] are provided for significant differences.