Original Article

QJEP

Quarterly Journal of Experimental Psychology
2020, Vol. 73(11) 1862–1878

DOI: 10.1177/1747021820932822
qjep.sagepub.com
SAGE

IMABASE: A new set of 313 colourised line drawings standardised in French for name agreement, image agreement, conceptual familiarity, age-of-acquisition, and imageability

Patrick Bonin[1], Bénédicte Poulin-Charronnat[1], Helle Lukowski Duplessy[1], Patrick Bard[1], Annie Vinter[1], Ludovic Ferrand[2] and Alain Méot[2]

Abstract

We provide normative data for a new set of 313 colourised line drawings. The drawings were standardised on name agreement ($N=60$ participants), image agreement ($N=34$), conceptual familiarity ($N=36$), age of acquisition ($N=35$), and imageability ($N=35$). Objective visual complexity measures are given for the pictures, and objective word frequencies are provided for the modal names of the drawings. Reliability measures for the collected norms are very high. There are high levels of agreement between the names given by the participants and the drawings and comparative analyses indicate that the distribution of name agreement scores is very similar in both our own database and the MultiPic database (Duñabeitia et al., 2018). A novel "picture-choice task" used to assess name-image agreement ($N=30$) reveals that the great majority of the IMABASE pictures that are also present in MultiPic are rated as providing better pictorial representations of the corresponding concepts. Finally, most of the correlations are comparable with those reported in other normative studies on colourised drawings. The whole set of pictures is freely available from https://leadserv.u-bourgogne.fr/~lead/imabase/ and the norms are available as Supplementary Material.

Keywords

Line drawings; psycholinguistic norms

Received: 19 November 2019; revised: 6 May 2020; accepted: 7 May 2020

Several normed picture databases are available in the literature. The most popular one is the Snodgrass and Vanderwart (1980) database that provides normative data in American-English on name agreement, image agreement, conceptual familiarity, and visual complexity (see below for a description of these norms) for a set of 260 black-and-white drawings. The goal of Snodgrass and Vanderwart—hereafter SV—was to provide researchers with pictures of drawings and psycholinguistic norms that could be used for the careful design of experiments on perception, memory, and language. The SV database has been impressively adopted by the (mainly psychological) scientific community, and the corresponding paper has been frequently cited (Duñabeitia et al., 2018). SV has encouraged research into the collection of norms for pictured stimuli, and the set of SV pictures has been normed in many different populations and language communities. To name but a few, the SV pictures have been normed for Spanish (Sanfeliù & Fernandez, 1996), British-English (Barry et al., 1997), French (Alario & Ferrand, 1999), Icelandic (Pind et al., 2000), Japanese (Nishimoto et al., 2005), Chinese (Weekes et al., 2007), and more recently, for Croatian (Rogić et al., 2013) and Tunisian

[1]LEAD - CNRS UMR5022, Univ. Bourgogne Franche-Comté, Dijon, France
[2]LAPSCO - CNRS UMR6024, Université Clermont Auvergne, Clermont-Ferrand, France

Corresponding author:
Patrick Bonin, LEAD-CNRS UMR5022, Univ. Bourgogne Franche-Comté, Pôle AAFE - Esplanade Erasme, BP 26513, 21065 Dijon, France.
Email: patrick.bonin@u-bourgogne.fr

Arabic (Boukadi et al., 2016). This set of pictures has also been normed in English-speaking children (Cycowicz et al., 1997) and in Portuguese-speaking children (Pompéia et al., 2001).

The SV pictures have been colourised (Rossion & Pourtois, 2004), and norms for these colourised pictures have been collected in Belgian French (Rossion & Pourtois, 2004), in Modern Greek (Dimitropoulou et al., 2009), in Russian (Tsaparina et al., 2011), in Turkish (Raman et al., 2014), and in Persian (Bakhtiar et al., 2013). Other sets of drawings have been constructed and normed (e.g., Bates et al., 2003 [520 black-and-white line drawings of common objects including 174 pictures from the original SV set]; Bonin, Peereman, et al., 2003 [299 black-and-white drawings of objects]; Khwaileh et al., 2018 [319 black-and-white line drawings of objects]; Martein, 1995 [216 black-and-white line drawings of objects]; Saryazdi et al., 2018 [225 everyday objects, each depicted both as a photograph and a matched clipart image]; Székely et al., 2004 [520 black-and-white line drawings of common objects]). Finally, photograph databases for objects have also been designed and normed (Brodeur et al., 2010 [480 colour photographic images of real objects]; Brodeur et al., 2014 [930 colour photographic images of real objects]; Clarke & Ludington, 2018 [norms in Thai for the Brodeur et al. photo set]; Moreno-Martínez & Montoro, 2012 [360 high-quality colour images of objects]; Salmon et al., 2010 [320 black-and-white photographs of objects]).

The availability of normed pictures has helped in the design of studies investigating the factors influencing naming speed in healthy adults (e.g., Alario et al., 2004; Bonin et al., 2002; Cuetos et al., 1999; Snodgrass & Yuditsky, 1996), as well as naming accuracy in patients (e.g., Cuetos et al., 2002). Likewise, the different processes and representations that are involved in object naming have been delineated. It is generally agreed that object naming involves three main processing stages: (1) object comprehension, which entails the perceptual analysis of the object, access to stored structural representations, and, finally, access to semantic information; (2) name retrieval, which according to certain views of speech production (Levelt et al., 1999), is conceptualised as a two-stage process consisting of (a) lexical selection, that is, activation of abstract lexical entries that are not phonologically specified: lemmas, and (b) word-form encoding; and (3) articulation. In adults, different factors have been found to influence the various stages of picture naming (Bonin, 2013; Roelofs & Ferreira, 2019, for reviews). As claimed by Alario et al. (2004), the most important predictors of naming speed include name agreement and age of acquisition (AoA) of the object names, followed by the objective frequency of the names (objective lexical frequency norms are available from different computerised databases in different languages; for the French language, Lexique.org provides different word frequency measures; New et al., 2004) and, finally, image agreement. Other predictors (e.g., imageability, conceptual familiarity, visual complexity) have been found to exert a reliable influence on naming speed, but in a less consistent manner (Bonin, 2013; see Perret & Bonin, 2019, for a Bayesian meta-analysis of the predictors of naming speed).

Why collect norms on a new set of colourised pictures?

Given the availability of pictures and psycholinguistic norms for a large number of items in the literature, readers may wonder why we decided to design a totally new set of pictures, corresponding largely to the SV concepts, and collect norms in adults on these new pictures and their names.

One important issue relating to the SV pictures, even in their colourised version (Rossion & Pourtois, 2004), is that they are at present not fully optimised for their use in experiments on perception, memory, or language processes for the following reasons. First, many of the SV pictures need to be updated. There are many instances of old-fashioned representations of objects. For example, the way a doll, a phone, a television, a doorknob, or a train is represented in the SV database no longer corresponds to the typical doll, phone, television, doorknob, or train that children or adults see in modern life. Certain items that were very familiar in the past (e.g., tape recorder, spinning wheel, rolling pin, dresser, salt shaker, thimble) are no longer so. Also, many objects present in our environment today are absent in the set of 260 black-and-white drawings published in 1980. It is important that norms collected for a set of stimuli at a given period of time within a given language and/or community be updated, because objects represented in an unusual manner will be difficult to identify. This is especially damaging if one's aim is to investigate the processes underpinning picture naming. Indeed, in adults, Johnston et al. (2010) found that their new and more up-to-date ratings accounted for a larger amount of variance in naming times than did those taken from an earlier study (e.g., Barry et al., 1997). Of course, having norms for objects that are typical of the current environment is also very useful to all researchers who use pictures as stimuli in experiments with children. Second, the different items in the SV database are taken from different semantic categories (15 categories), but for certain categories, there are too few exemplars (e.g., vehicles: $n = 10$; birds: $n = 8$; or insects: $n = 8$). As far as possible, we therefore added some more items for these semantic categories in order to achieve a better representation of each category in the new database. Third, certain SV pictures are typical of the American culture (see, in particular, Duñabeitia et al., 2018 footnote 2; Yoon et al., 2004), and do not correspond to the way most adults and children represent the corresponding concepts in other cultures. To give a few examples, the drawings of plug, bread, barn, key, fridge, sandwich, and bus are all clearly typical of American culture. Fourth, many pictures in SV are difficult to recognise, even for adults, as assessed by name-agreement or image-agreement scores. The French normative study by Alario and Ferrand (1999) revealed that there were 39 out of 260 pictures of objects that had name-agreement scores lower than 75% even for very familiar concepts, such as pig, train, frog, toe, pen, camel, bread, or leopard. Many pictures are rated as being "not typical" of the objects they represent: 68 pictures out of 260 have a rating lower than 3 (on a five-point scale with 5 = *very high agreement* between the drawing and the mental representation of the underlying concept). This represents a clear limitation when it comes to selecting items for the design of experiments involving both children and adults.

It is worth noting that we are not the only authors to point to the need to provide researchers with new databases of colourised line drawings or new sets of photographs. Russo et al. (2018) have provided a database of 640 colourised images of animals (normed on name agreement, picture name agreement, and AoA) because the number of exemplars per category of animals is generally low in the available databases. As claimed by Duñabeitia et al. (2018), there is in fact a limited choice for researchers who want to use (colourised) drawings in their experiments. Fortunately, they have provided a set of 750 drawings that have been normed in six European languages: the MultiPic database. We became aware of this work while we were designing our own database. However, the items in MultiPic have been normed on only two variables—name agreement and visual complexity—which represents a serious limitation, in particular for the design of naming experiments in which a number of influential factors need to be controlled for methodological or statistical purposes (e.g., image agreement, AoA; see Perret & Bonin, 2019).

To design the present picture database, we built on the existing list of depicted SV concepts because these concepts are generally related to everyday common objects. It is indeed useful to have pictures corresponding to these concepts in order to design naming experiments, especially to permit comparisons of the results with those of previous naming studies based on such concepts. We decided to use *colourised* line drawings instead of more traditional black-and-white drawings. The choice of colourised drawings was motivated by the fact that faster and/or more accurate naming performances have been obtained with colourised than with black-and-white drawings (Bonin et al., 2019; Rossion & Pourtois, 2004). Norms on photographs are undoubtedly useful since they represent real instances of objects, that is, idiosyncratic (token) representations of objects. However, we wanted to have generic (type) representations of objects because our database was also designed to be useful for researchers investigating naming in children; generic representations seem to be more suitable to this end. More importantly, some findings in the literature indicate that naming accuracy is lower in response to photographs than to corresponding drawings of the same objects (Clarke & Ludington, 2018). For instance, the overall name agreement is higher in the SV database (88%), as well as in the Bonin, Peereman, et al. (2003) database that was designed to complement the SV set (77%), than in the BOSS database (64% in the Brodeur et al., 2010 study and 59.45% with the BOSS pictures normed in Thai, Clarke & Ludington, 2018), which uses photographs of objects. As claimed by O'Sullivan et al. (2012), photographs include details and colours that prompt participants to rely on idiosyncratic features and, as a result, they elicit a number of different names. This can be a problem when designing experiments in which naming times are the dependent variable (but see Saryazdi et al., 2018).

Turning to the investigation of memory processes, certain findings suggest that the degree of visual iconicity, that is, the degree to which images resemble the real-world objects they depict (Saryazdi et al., 2018), does not play a significant role in memory performance since similar levels of recall and recognition have been found with the use of coloured photographs and black-and-white line drawings of objects (Snow et al., 2014).

The present normative study: norms collected from pictures and their names

Norms correspond either to characteristics of the pictures or to their names. Among the norms collected from pictures are the visual complexity of the pictures, the "typicality" of the depicted drawings as assessed by image agreement, the degree of agreement of the visual representation of objects with their names (i.e., name agreement), and the familiarity of the depicted concepts. Ratings of AoA are also often collected from object names. To collect name agreement, image agreement, visual complexity, as well as conceptual familiarity ratings from pictures, adults are generally tested collectively in tasks in which pictures (or words) are projected onto a large white screen by means of an overhead projector. However, these norms can also be collected individually by presenting pictures on a computer screen. AoA from written words is also collected individually or collectively. Likert-type scales are used for the ratings. In the present study, we collected norms on name agreement, image agreement, and conceptual familiarity from the pictures; and on AoA and imageability from the written names (in the "Procedure" subsection, we provide detailed information about how these norms were collected).

Name agreement corresponds to the degree of agreement among individuals on a specific name to be used to refer to a pictured object, whereas image agreement corresponds to the degree to which the mental images formed by participants in response to an object name match the object's appearance. Name agreement is one of the strongest factors influencing naming performance, and in particular naming speed (e.g., Barry et al., 1997; Bonin et al., 2002; Ellis & Morrison, 1998; Snodgrass & Yuditsky, 1996; Valente et al., 2014; Vitkovitch & Tyrrell, 1995). In word production, name agreement acts either at the comprehension level or at the lexical level (Barry et al., 1997). Image agreement has its effect at the level of structural representations, which correspond to the canonical perceptual representations of objects (Humphreys et al., 1988).

Because different colourised drawings are used to depict the same concepts common to MultiPic (Duñabeitia et al., 2018), Rossion and Pourtois' (2004) database, and the current database, we also included a picture-choice task to assess which drawings best depicted the common concepts. In this task, participants were first presented with a written name and had to form a mental representation of the concept referred to by the name. Depending on the trial, two or three pictures were displayed on a computer screen and the participants had to choose the one which, according to them, was the best depiction of the object name.

Conceptual familiarity is a measure of the acquaintance with the concept depicted by the picture, or referred to by the object name. When pictures are used for these ratings, care is taken to explain that the ratings have to be attributed to the concept itself, and not to the way it is represented. This variable is thought to index the conceptual level in lexical processing tasks. Mixed findings have been reported concerning the impact of this variable in spoken or written naming studies (see for a review Perret & Bonin, 2019).

When evaluating the visual complexity of pictures, participants are required to take account of the number of lines and details in the drawing. Evidence for the influence of this variable is mixed at best, with certain studies reporting that visually complex drawings reliably slow down the naming process compared with visually simple drawings (Alario et al., 2004; Rogić et al., 2013), while others report the reverse pattern (Székely et al., 2005), and, finally, a majority of studies report null effects (e.g., Bonin et al., 2002; Bonin, Peereman, et al., 2003; and see Perret & Bonin, 2019). Székely and Bates (2000) have proposed other measures of visual complexity that do not rely on subjective ratings such as the size of the digitised picture file, that is, objective visual complexity (see also Székely et al., 2004). They claim that using objective visual complexity scores avoids the problem that subjective ratings of visual complexity may be influenced by psycholinguistic variables (e.g., familiarity with the object) that are not directly related to visual complexity. As in Tsaparina et al.'s (2011) study, we decided to report only objective visual complexity scores (the number of bytes in JPEG format) for our set of colourised pictures.[1]

In order to estimate the AoA of the words, participants use Likert scales. For instance, they have to estimate the age at which they think they learned the word *cow* in its written or oral form. In general, the values of the scale correspond to x-year age bands, for instance 2-year age bands. AoA is an important determinant of lexical processing and one of the most important determinants of object naming speed (see Johnston & Barry, 2006; Juhasz, 2005 for reviews).

Imageability corresponds to ratings of the ease with which mental images can be generated in response to words. Imageability is used as an index of semantic code activation in lexical processing (e.g., Cortese et al., 1997; Strain et al., 1995, 2002). Finally, we also reported objective word frequency (subtitle frequency and book frequency measures taken from Lexique.org, New et al., 2004) for the modal names provided for each object since word frequency is certainly the most popular variable in psycholinguistics (see Brysbaert et al., 2019, for a recent review).

In the following, we report several analyses that were performed on the collected data and on the data taken from the Multipic and RP databases. First of all, we describe the reliabilities that were computed for the different collected norms from the new set of items. Second, we report several comparative analyses on name agreement scores to investigate precisely how the naming outcomes from IMABASE differ from those obtained with the Multipic and RP databases. In addition, as indicated above, we used a "novel" task—a "picture-choice task"—to assess further name-image agreement. Here also, we compared the different scores obtained for our pictures with those obtained for the pictures taken from Multipic and RP. Third, descriptive statistics are reported together with the distributions of the norms. Fourth, we present the bivariate correlations that were performed on the collected variables. These analyses were conducted to analyse and summarise the correlational structure of the norms, and to compare it with other normative studies on colourised drawings. Finally, we report multiple regression analyses that were performed to analyse more deeply the relations between each subjective norm and the other dimensions.

Method

Participants

The participants ($N = 230$; 39 males, 190 females and one participant who did not specify sex; mean age: 20.35 years old; range: 18 to 33 years old) were all native speakers of French and had normal or corrected-to-normal vision. Six subgroups were established corresponding to the six collected norms, with the mean ages and standard deviations of the participants being as follows: for name agreement ($N = 60$ because two sublists were used, each involving 30 participants), M = 20.48, SD = 1.76; for the picture-choice task (N = 30), M = 20.17, SD = 1.12; for image agreement (N = 34), M = 20.88, SD = 3.09; for familiarity (N = 36), M = 20.44, SD = 1.92; for AoA (N = 35), M = 20.2, SD = 1.97; and for imageability (N = 35), M = 19.91, SD = 1.76.

The different rating tasks were performed collectively with each participant being randomly assigned to one of the rating tasks. All participants provided fully informed consent to participate and received course credits for their participation. The study was approved by the Statutory Ethics Committee from the University Clermont Auvergne.

Stimuli

The entire set of colourised pictures is available for free download at https://leadserv.u-bourgogne.fr/~lead/imabase/. SV categorised their concepts into 15 semantic categories. We used these SV semantic categories and added more items for certain semantic categories to have a better representation of each category in the new database. We also decided to add exemplars that did not belong to the semantic categories provided by SV. We therefore decided a priori on the different semantic categories that we thought useful to add (e.g., weapons, sports) as well as on the number of examplars to include in each category (which varies from 4 to 21). Likewise, we had 313 items that were classified into 24 (a priori) semantic categories (see Table 1 for the different semantic categories and the corresponding numbers of items). This information about how we classified our items into these semantic categories (see Supplementary Material A) should help researchers to find specific items more easily when designing experiments with drawings. However, in order to establish whether our a priori classification of the items was reliable, we asked 10 independent adults to classify the picture names among the 24 semantic categories (see Table 1). Given the relatively large number of categories, an acceptable interrater reliability was observed as indicated by the value of .77 obtained for Krippendorff's alpha (this index was computed using the SPSS macro developed by Hayes & Krippendorff, 2007). It should be noted that two categories were chosen at an equal level for only five words and that the modal category was chosen by less than 6 participants for 14 items (see details in the Supplementary Material A).[2]

The selected SV concepts (216 concepts in common with SV, overlap = 83%), together with the added concepts, can be seen in the Supplementary Material A. The drawings were designed by a draftsman, graphic artist and illustrator (Bonzai Studio: http://www.bonzaistudio.com/bonzai-studio-dijon-animatin-3d/), who was instructed to use the same graphic style throughout. The drawings were designed with a digital tablet and pen set. The same space was used for the whole set of drawings (dimensions 2362 × 2362 pixels). Two authors of the paper (P.B. and B.P.C.) told the artist exactly what they wanted to be represented for each concept, assessed the quality of the drawings, and improvements were made when needed.

Table 1. Semantic categories (French names in parentheses) and number of items in each category.

Category	Count
Items of furniture (Meubles)	10
School equipment (Fournitures scolaires)	12
Food (Aliments)	16
Aquatic animals (Animaux aquatiques)	12
Farm animals (Animaux de la ferme)	12
Forest and garden animals (Animaux de la forêt et du jardin)	16
Wild animals (Animaux sauvages)	15
Weapons (Armes)	7
Buildings/parts of buildings (Bâtiments/Parties de bâtiments)	10
Household items (Accessoires de la maison)	19
Birds (Oiseaux)	9
Types of vehicle (Transports/Partie de transports)	13
Make-up (Beauté/Toilette)	9
DIY tools (Bricolage/Jardinage)	15
Parts of the (human) body (Corps humain)	16
Childhood (Enfance)	4
Kitchen utensils (Ustensiles de cuisine)	13
Fruits (Fruits)	12
Articles of clothing (Vêtements)	16
Musical instruments (Instruments de musique)	13
Vegetables (Légumes)	11
Natural elements (Nature/Éléments naturels)	13
Sports/hobbies (Sports/Loisirs)	21
Other (Autre)	14

Note. Two categories were chosen at an equal level for five words (arc, bougie, cerf, cerf-volant, chapiteau). These words were not included in the count reported in the table (see Supplemental Materials for details).

Procedure

To collect norms for name agreement, image agreement, conceptual familiarity, and AoA, we closely followed the procedures used in the Alario and Ferrand (1999), Bonin, Peereman, et al. (2003), and Tsaparina et al. (2011) studies. The instructions for each norm are provided in full in the Supplementary Material B.

For name agreement and image agreement, the pictures were projected on a large white screen by means of an overhead projector. For each rating task, the set of pictures was presented in a different random order to the different groups of participants. The participants were tested collectively in small groups (no more than eight individuals). In the name-agreement task, the participants were instructed to identify each picture by writing down the first name that came to mind. Whenever they were not able to provide a name for a given picture, they had to indicate the reason, that is to say, (1) they did not know the object (DKO), (2) they did not know the name of the object (DKN), or (3) they were in a tip-of-the-tongue state (TOT). The name-agreement task was the first task to be conducted in order to compute name agreement scores that then allowed us to select the modal names to be used in the normative measures made on the basis of the object labels (e.g., AoA).

To collect image-agreement ratings, the participants were asked to first generate a mental image from a name that was presented for 1 s. The modal names corresponding to the pictures were displayed in Times New Roman, font size 40, on a large white screen prior to the presentation of the pictures. There was then a 5-s interval during which the adults had to generate a mental image corresponding to the name either, while keeping their eyes closed or while staring at the screen. As soon as the picture was displayed on the screen, the participants had to rate the degree of agreement between the presented picture and their generated mental image on a 5-point Likert-type scale. A rating of 1 indicated *low agreement* and a rating of 5 indicated *high agreement*.

For the conceptual familiarity ratings, the participants had to rate on a 5-point Likert-type scale (1 = *low familiarity*; 5 = *high familiarity*) "how usual or unusual the depicted object was in their realm of experience." More precisely, they were told that familiarity corresponds to "the degree to which you come into contact with or think about the concept." The participants were instructed to rate the concept itself, and not the way it was drawn.

The ratings for AoA and imageability were collected using LimeSurvey from the names assigned to the pictures. In the AoA task, the participants had to estimate the age at which they thought they had learned each of the names in its written or oral form. A 5-point Likert-type scale was used with 1 = *learned between 0 and 3 years* and 5 = *learned at age 12 or after*. Concerning imageability, based on the instructions provided by Paivio et al. (1968) and their French translation (see Bonin et al., 2011), adults had to rate on a 5-point Likert-type scale the ease of generating a mental image from the picture names (from 1 = *difficult* to 5 = *easy*).

Objective measures. Digitised images for all 313 pictures were used to calculate the objective visual complexity scores, using image file size metrics. Following Székely and Bates (2000), the size of the image files in JPEG format was used as an objective measure of visual complexity. To avoid report errors when assessing the number of bytes of each picture, the procedure was automated using a batch script running on Windows.

Objective word frequency was taken from Lexique.org (New et al., 2004) and, more precisely, we used subtitle frequency and book frequency measures of the modal names. Length was measured as the number of letters.

Results

In the Supplementary Material C, the two measures of name agreement corresponding to the percentage of participants giving the most common name (the *%NA*) and the *H* statistic (as described in SV, 1980) are provided. In the Supplementary Material D, the different nonmodal names are listed for each item together with their corresponding frequencies of occurrence. We also indicate DKO, DKN, and TOT responses.

In the Supplementary Material C, means and standard deviations for image agreement, conceptual familiarity, imageability, and rated AoA are provided for the items listed alphabetically according to the French names of the pictures (English translations are provided). The values corresponding to the visual complexity of the pictures are reported. Objective word frequency values and the number of letters of the modal names are also provided.

Reliability

One participant provided unusual details when naming the pictures and her or his data was therefore discarded from the name-agreement task. Another participant, who rated 88% of the pictures as unfamiliar, that is, a score of 1, was excluded from the familiarity-rating task.

Table 2 shows the intraclass correlation coefficients (random effects of both participants and items—ICC[2, k] in Shrout & Fleiss, 1979's terminology) for the different norms obtained with the use of Likert-type scales. Table 2 also reports the correlations between the percentages of name agreement scores—the details concerning the response coding are given below—obtained from even and odd participants for each list of words.[3] The reliability scores turned out to be high for name agreement. With ICC values above .80, reliability was high for *all* the collected norms.

In order to further gauge the reliability of our ratings for AoA and imageability, we correlated these ratings with the ratings taken from previous French studies (Alario & Ferrand, 1999; Bonin et al., 2011; Bonin, Méot, et al., 2003; Bonin, Peereman, et al., 2003; Ferrand et al., 2008). The AoA ratings in the present study were highly correlated with the ratings taken from other French studies (see Table 3). However, the correlations were lower for imageability. These lower correlations are possibly due to the fact that a lot of words are highly imageable.

Name agreement comparisons with the MultiPic, Rossion, and Pourtois databases

Two authors of the paper (H.L.D. and A.M.) checked the answers for spelling errors, corrected them, and recoded

Table 2. Reliability measures for the norms of age-of-acquisition (AoA), conceptual familiarity, imageability, and name agreement (%NA: percentage of name agreement).

Intraclass correlation coefficient				*r* (even, odd)	
				%NA	
AoA	Conceptual familiarity	Image agreement	Imageability	List 1	List 2
.967	.950	.859	.833	.720	.816

Table 3. Correlations between AoA and imageability scores obtained in the present study and in other studies.

	AoA	Imageability
Alario and Ferrand (1999)	.868 (236)	
Bonin, Peereman, et al. (2003)	.929 (32)	
Bonin, Méot, et al. (2003)		.728 (277)
Ferrand et al. (2008)	.923 (99)	
Bonin et al. (2011)		.647 (99)

AoA: age of acquisition.
Note. The numbers of words used in the computations are given in parentheses.

initial responses by merging basic variants of the same name (e.g., hyphenated or pluralised forms). For any given expected name, abbreviations (e.g., *télé* for *télévision*), word additions (e.g., *chapiteau de cirque* for *chapiteau* or *nœud rose* for *nœud*), deletions (e.g., *raquette* for *raquette de tennis*), and synonyms (e.g., *autocar* or *car* for *bus*) were counted as different responses.

Two measures of name agreement were computed: percentage name agreement and the *H* statistic. For each item, the percentage name agreement (%NA) was calculated as the proportion of the modal name across all naming outcomes, including naming failures (DKO, DKN, and TOT). The *H* statistic[4] captures more information than %NA about the distribution of names across participants. An *H* value of 0 indicates that the drawing elicited the same naming outcome from all participants. Increasing *H* values indicate decreasing levels of name agreement (*H* can vary from 0 in the case of perfect agreement to $\log_2[N]$ when all participants produce a different name, where *N* corresponds to the total number of participants giving name responses to the picture).

Figure 1 shows the distributions for the 200 words common to the current database (IMABASE) and the Rossion and Pourtois (2004) picture database (see Table 1A for the descriptive statistics linked to Figure 1 in the Supplementary Material E). It is important to note that, given the serious concerns about the name agreement scores provided by Rossion and Pourtois, as set out in Tsaparina et al.'s (2011) study, the comparisons were made with the name agreement scores that we (Bonin, Guillemard-Tsaparina, & Méot, 2013) collected in French for the Rossion and Pourtois set of colourised SV pictures. The distribution of the name agreement scores for the IMABASE items is somewhat more shifted to the right than the Rossion and Pourtois items (difference in confidence interval of the 95% bootstrapped means: [1.29; 4.82]; Wilcoxon signed-rank test: $Z = 4.05$, $p < .001$).

Figure 2 (see Table 1B for the descriptive statistics linked to Figure 2 in the Supplementary Material E) shows the same information for the 263 words common to MultiPic (Duñabeitia et al., 2018). It is worth noting that for 19 modal names in the MultiPic database, which corresponded to two or even three pictures from this database, we decided to select the picture having the highest %NA. The differences in the distribution of name-agreement scores between IMABASE and MultiPic are very tenuous (difference in confidence interval of the 95% bootstrapped means: [−1.67; 1.35]; Wilcoxon signed-rank test: $Z = -.39$, $p = .696$).

It must be stressed, however, that the criteria used by Duñabeitia et al. (2018) to classify the naming responses were not exactly the same as those used in the present work. Duñabeitia et al. noted that the initial responses were recoded "by [. . .] discarding parts of speech that were not nouns or verbs (e.g., determiners and adjectives)" and that "trials where participants did not know the name of the concept (2.4% of the data across languages) and idiosyncratic responses (i.e., responses provided by a single participant, corresponding to 2.7% of the data) were excluded."[5]

As far as the items from our database that are common to the other two databases are concerned, the percentages of items with a 100% name-agreement score were 40.2% (IMABASE) versus 41.3% (MultiPic), and 42% (IMABASE) versus 30% (Rossion & Pourtois).

Naming failures (including DKO, DKN, TOT, and nonresponses) in our database were somewhat lower than those found in MultiPic when the entire databases were taken into account (1.1% [IMABASE] vs. 2.7% [MultiPic], but they were virtually identical for the shared modal names (1% [IMABASE] vs. 1.1% [MultiPic]) when only taking account of MultiPic pictures yielding the modal names with the highest name-agreement scores, see above). Idiosyncratic responses were comparable in both cases (2.3% [IMABASE] vs. 2.8% [MultiPic] for the entire databases, and 2.2% [IMABASE] vs. 1.6% [MultiPic] for the shared modal names). Naming failures and idiosyncratic

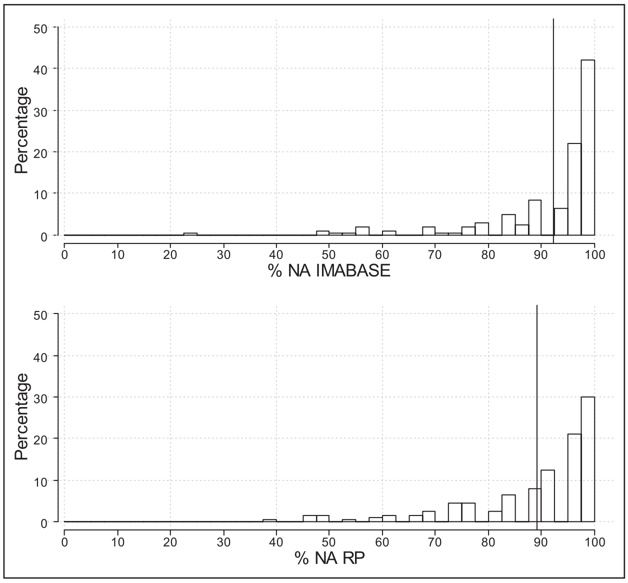

Figure 1. Distributions for words common to IMABASE and Rossion and Pourtois ([2004]; RP) databases (see Table 1A for descriptive statistics linked to this figure in the Supplementary Material).

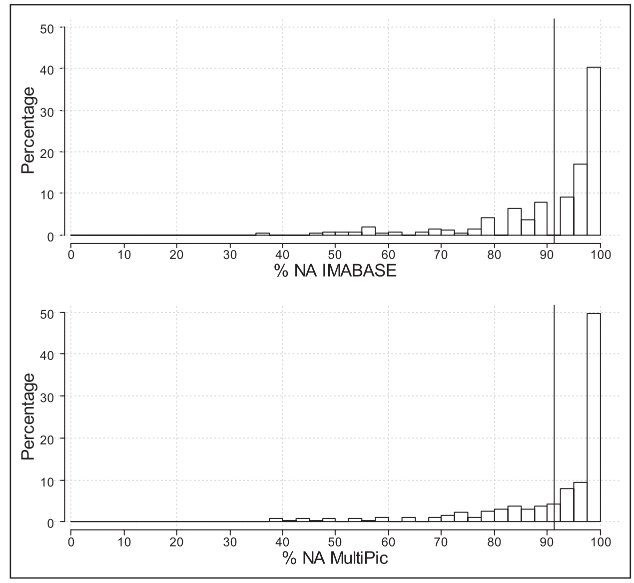

Figure 2. Distributions for words common to IMABASE and MultiPic databases (see Table 1B for descriptive statistics linked to this figure in the Supplementary Material).

responses were also very similar between the current database and the data provided by Bonin, Guillemard-Tsaparina, and Méot (2013) for the Rossion and Pourtois drawings: naming failures (1.6%) and idiosyncratic responses (3.6%) for the entire databases; 1.2% and 2%, respectively, for the words in common.

Picture-choice task

As explained above, to explore more deeply certain differences between the three picture databases, we included a picture-choice task. For each (modal) name in our database, the participants were presented with the different pictures that were available in the three databases, and they had to choose the one they thought best matched the object referred to by the name.

As can be seen from Table 4, a large majority of the pictures taken from our database were chosen as the best match (about 75% of choices), with the MultiPic pictures being the second preference (around 20% of all choices). The Rossion and Pourtois pictures were rarely chosen.

Descriptive statistics

Table 5 provides descriptive statistics for the collected ratings as well as additional characteristics of the items. The distributions corresponding to the collected norms are shown in Figure 3.

Percentage name agreement and *H* were highly negatively and positively skewed, respectively, indicating high levels of agreement between the names given by the participants to the pictures. Even if it was less pronounced, a negative skew was also found for the image agreement variable, which conversely shows that the participants generally agreed with the pictures proposed for the modal names.

With both mean and median roughly equal to 4.5, most of the words were highly imageable, resulting in a high negative skew for imageability ratings. As said earlier, this property is probably the reason why there are relatively low levels of correlations between the current imageability scores and other scores taken from previous published databases.

Most of the words were also judged to be acquired early in life (more than 50% before 6 years of age), whereas only very few (6.1%, $N = 19$) were rated to be acquired after 10 years of age. Finally, the familiarity ratings were distributed along the entire range of the scale except at the extremes. There were two modes: one for very familiar concepts (ratings around 4) and one for relatively unfamiliar concepts (ratings around 2).

Bivariate correlations and multiple regressions with behavioural variables as dependent variables

As shown by the bivariate correlations (Table 6), the correlations between the two name agreement variables and the other psycholinguistic variables were generally low, with higher agreement being obtained for more imageable and early acquired names. Name agreement also increased with image agreement. Low correlations were found for image agreement, which—in addition to name agreement—was significantly correlated only with

Table 4. Results obtained in the picture-choice task.

Picture found in	Number and percentage of pictures to be the best match for the object names			Additional information
	IMABASE	RP	MultiPic	
The three databases (N = 178)	138 (77.5%)	10 (5.6%)	28 (15.7%)	11 = RP > M 11 = M > RP
IMABASE and MultiPic (N = 85)	66 (77.6%)	–	15 (17.6%)	4 equal (4.7%)
IMABASE and RP (N = 22)	21 (95.5%)	0 (0%)	–	1 equal (4.5%)

I: IMABASE; RP: Rossion and Pourtois; M: MultiPic.

Table 5. Descriptive statistics of the collected norms.

	N	Min	Max	*M*	*SD*	Percentiles			Asymmetry
						25	50	75	
Name agreement (%)	313	23.33	100.00	89.64	14.45	83.33	96.55	100.00	−1.80
Name agreement (*H*)	313	0.00	2.24	0.40	0.49	0.00	.22	.66	1.35
Familiarity[a]	312	1.51	4.66	3.07	0.90	2.18	3.06	3.94	0.04
Image agreement	313	2.21	4.82	3.77	0.49	3.46	3.82	4.15	0.54
AoA[a]	312	1.09	4.17	2.05	0.59	1.58	1.97	2.46	0.61
Imageability[a]	312	2.00	4.97	4.46	0.34	4.31	4.54	4.69	−2.22
Visual complexity[b]	313	76	1357	231	140	133	178	307	2.57
Nb of letters[c]	313	3.00	16.00	6.80	2.53	5.00	6.00	8.00	1.17
Nb of phonemes[d]	291	2.00	11.00	4.56	1.58	3.00	4.00	6.00	0.69
Subtitle frequency[d]	291	0.00	570.30	29.56	59.66	3.46	9.83	29.61	4.98
Book frequency[d]	291	0.07	788.72	41.87	89.73	4.05	13.85	32.77	4.61

SD: standard deviation; AoA: age of acquisition.
Notes. N = number of items; objective norms were taken from Lexique 3 (New et al., 2004).
[a]One word (ski) was accidentally omitted.
[b]Expressed in 10^3 bytes.
[c]Blanks were not counted for compound words.
[d]As compound words are generally not included in the Lexique database, the statistics taken from this database are given for only 291 words out the 313 items used in IMABASE.

objective word frequency, and, to a lesser extent, with imageability, with the result, that image agreement was higher for less frequent and more imageable words. Imageability was related to all the other variables, and more particularly to familiarity and AoA, with highly imageable items being judged to be more familiar and acquired earlier in life. Highly imageable items also had higher name-agreement scores and, to a lesser extent, were more objectively frequent. Early acquired words were found to be more imageable, conceptually familiar, frequent, and shorter than late acquired words. Finally, the more imageable, frequent, and early acquired the items were, the more conceptually familiar they were judged to be. Moreover, but to a lesser extent, objectively complex pictures and shorter names were also judged to be more familiar.

Overall, most of the correlations are comparable with those reported in other normative studies on colourised drawings (Bakhtiar et al., 2013; Dimitropoulou et al., 2009; Duñabeitia et al., 2018; Raman et al., 2014; Rossion & Pourtois, 2004 with the name-agreement scores taken from Bonin, Guillemard-Tsaparina, & Méot, 2013; Tsaparina et al., 2011). One notable exception is found in the visual complexity variable, for which the correlation with the other subjective ratings were systematically lower (and generally not significant) in the present database than those found in most normative studies on pictures. This finding is most probably due to the fact that we used objective visual complexity scores in the present work, and not subjective measures of visual complexity. Another noticeable discrepancy concerns the significant negative correlation found between image agreement and objective word frequency. No such correlation was found in the three studies (Bakhtiar et al., 2013; Raman et al., 2014; Rossion & Pourtois, 2004 with the name-agreement scores taken from Bonin, Guillemard-Tsaparina, & Méot, 2013; Tsaparina et al., 2011) that reported this correlation (with, however, non-significant results in two of them). Finally, we did not find a correlation between name agreement and familiarity, whereas such a correlation has generally been found in other studies (Bakhtiar et al., 2013; Raman et al., 2014; Rossion & Pourtois, 2004).

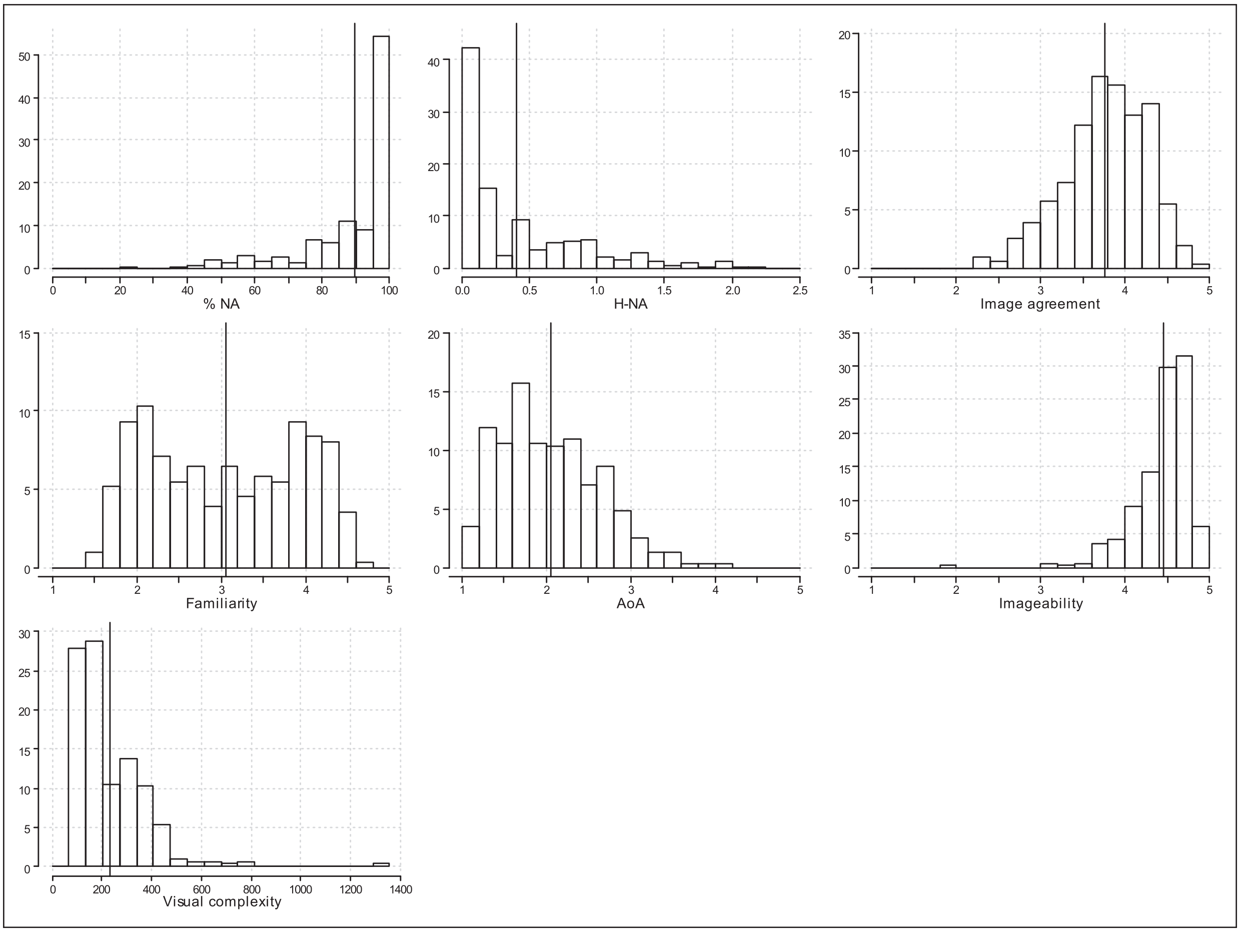

Figure 3. Distributions of the new norms (*y*-axis = word percentages).

To study if the above relationships persisted when other dimensions were controlled for, we included each psycholinguistic variable in turn as the dependent variable and the other variables as independent variables in multiple regression analyses. Given the high correlation found between the two name-agreement measures, only percentage name agreement was retained for the analyses. To make comparisons easier, all the independent variables (IVs) were transformed into *Z*-scores. Nonlinearities were introduced by including restricted cubic splines with three knots for some independent variables using a forward approach: At each step, nonlinear terms were included for the independent variables whose inclusion led to the greatest increase in variance compared with the model that did not include such terms. We opted for a minimum of 1% of added variance, and nonlinearity was not included if the change between the two models was not significant. The results of these analyses are given in Table 7 and Figure 4.

The partial relationships were generally similar to the bivariate relationships. Noticeable differences were as follows. First, three relatively important relations that were found in the bivariate correlations vanished when other IVs were controlled for, such as (a) high name agreement was no longer related to early acquisition; (b) length did not influence AoA ratings (the same was observed for familiarity with, however, a lower bivariate correlation); and (c) word frequency did not have a positive effect on imageability. Second, other relations turned out to be significant: (a) name agreement decreased with higher familiarity (and the reverse); (b) image agreement was influenced by AoA in a nonlinear relationship in such way that it decreased slowly, with higher AoA ratings being observed below the mean, before increasing considerably; (c) the positive relation between image agreement and imageability was more pronounced. Third, nonlinear relations indicated that (a) name agreement increased with imageability ratings below the mean, and then tended to stabilise; (b) familiarity increased only for values situated at the right of the word frequency values, whereas the reverse was true for its relation with visual complexity; (c) imageability was stable for early acquired words and then decreased greatly.

General discussion

The collection of psycholinguistic norms for various types of stimuli (e.g., faces of celebrities: Bonin et al., 2008;

Table 6. Bivariate correlations.

		Name agreement (*H*)	Familiarity	Image agreement	AoA	Imageability	Visual complexity (log)	Nb of letters	Subtitle frequency (log)	Book frequency (log)
Name agreement (%)	*R*	−.942***	.066	.300***	−.288***	.447***	−.017	.088	.029	.057
	95% CI	[−.96, −.92]	[−.06, .18]	[.18, .41]	[−.43, −.11]	[.26, .59]	[−.13, .10]	[−.04, .20]	[−.10, .15]	[−.06, .17]
Name agreement (*H*)	*R*		−.052	−.314***	.239***	−.404***	.026	−.080	.026	−.003
	95% CI		[−.17, .07]	[−.42, −.20]	[.08, .37]	[−.53, −.24]	[−.09, .14]	[−.19, .05]	[−.10, .16]	[−.12, .11]
Familiarity	*r*			−.070	−.588***	.534***	−.176**	−.169**	.457***	.531***
	95% CI			[−.18, .04]	[−.66, −.51]	[.46, .61]	[−.29, −.07]	[−.28, −.06]	[.37, .54]	[.44, .61]
Image agreement	*r*				.079	.144*	.026	.052	−.257***	−.217***
	95% CI				[−.04, .20]	[.03, .24]	[−.10, .14]	[−.06,.15]	[−.37, −.13]	[−.31, −.10]
AoA	*r*					−.710***	.039	.233***	−.533***	−.539***
	95% CI					[−.77, −.64]	[−.07, .14]	[.10, .37]	[−.62, −.44]	[−.61, −.46]
Imageability	*r*						−.108	−.048	.276***	.273***
	95% CI						[−.21, .14]	[−.21, .004]	[.18, .38]	[.18, .37]
Visual complexity (log)	*r*							.05	−.206***	−.246***
	95% CI							[−.06, .16]	[−.32, −.09]	[−.35, −.14]
Nb of letters	*r*								−.399**	−.446**
	95% CI								[−.49, −.31]	[−.53, −.36]
Subtitle frequency (log)	*r*									.901***
	95% CI									[.88, .92]

AoA: age of acquisition; CI: confidence interval.
Notes. First line = Pearson's correlation.
Second line: 1,000 bootstrapped 95% interval for the Pearson's correlation.
Computations were calculated for the 290 words/pictures for which all characteristics were available both in IMABASE and Lexique 3 (New et al., 2004). Frequencies and visual complexity were log-transformed.
*$p < .05$.
**$p < .01$.
***$p < .001$.

Table 7. Multiple regressions including one norm as DV and the others as IVs.

		Dependent variables				
		Name agreement (%NA)	Familiarity	Image agreement	AoA	Imageability
Forward selection of splines	R2lin	.2993	.4395	.1950	.6620	.6308
	1st NL term	Imty (.0195**)	Imty (.0335***)	AoA (.0170*)		AoA (.0338***)
	2nd NL term		VC (.0124**)			
	3rd NL term		FF (.0119*)			
	R2 final	.3188	.4973	.2121	.6620	.6650
Independent variables (IV)	%NA		−.14**	.29***		.18***
	Familiarity	−.18**			−.17***	.26***
	Image agreement	.24***			.09*	.13***
	AoA		−.30***	(+) _6.15_***		(–) _84.74_***
	Imageability	(+) _19.71_***	(+) _20.76_***	.30***	−.53***	
	Visual complexity		(–) _4.28_*		−.09*	
	Nb of letters			−.12*		
	Subtitle frequency		(+) _9.37_***	−.29***	−.27***	

DV: dependent variables; NA: name agreement; NL: nonlinear; VC: visual complexity; FF: subtitle film frequency; AoA: age of acquisition.
Notes. First part: R2lin = R2 with no nonlinearities; R2 final = R2 with nonlinearities included; 1st (2nd, 3rd) NL = first (second, third) IV for which cubic splines were included with the IV name and, in parentheses, the percentage of added variance and its significance. Second part: beta coefficient or in italics F value for IV including nonlinearities. For these latter variables, a (−) or a (+) is added to indicate the direction of change.
Shaded regions signifies dependent variable excluded from the IVs.
$^{*}p < .05$.
$^{**}p < .01$.
$^{***}p < .001$.

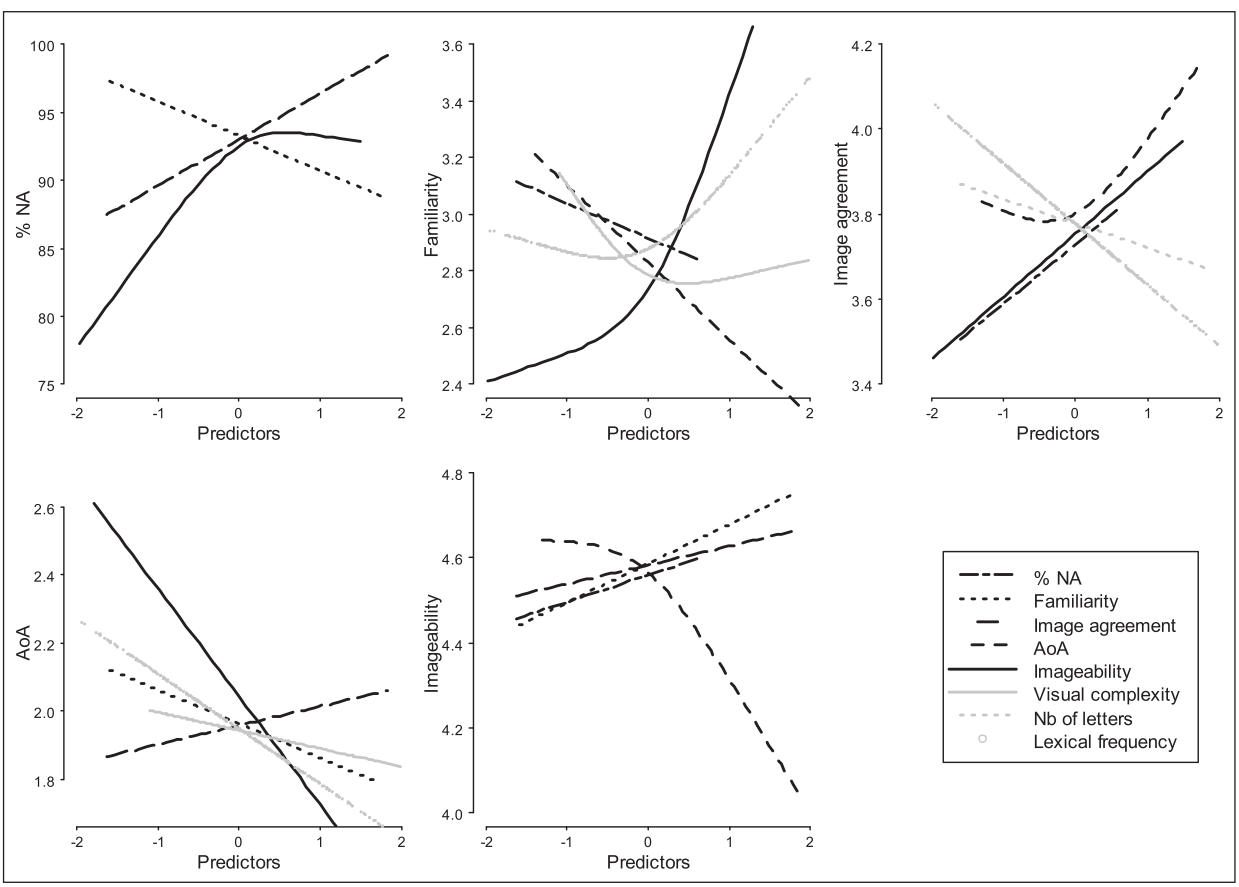

Figure 4. Partial effects of the significant predictors. The effects of subjective variables are indicated by black symbol curves (%NA, familiarity, image agreement, AoA, and imageability), whereas the effects of objective variables (visual complexity, length, and lexical frequency) are shown with grey curves. To obtain clearer figures, independent variables were z-transformed and values below −2 and above 2 were eliminated. These represented less than 5% of the values except in the case of name agreement (7.9%).

Marful et al., 2018; acronyms: Bonin et al., 2015; Izura & Playfoot, 2012; idioms: Bonin et al., 2018; Bonin, Méot, & Bugaiska, 2013) is now a frequent practice in psychology. In the present research work, we provided a new set of 313 colourised drawings, which have been normed on five important psycholinguistic variables: name agreement, image agreement, conceptual familiarity, AoA, and imageability. Importantly, all the collected norms turned out to be highly reliable. Overall, the pictures in our database are consensual as there are high levels of agreement between the names given by the participants and the drawings. Most of the object names are highly imageable and acquired early in life, and the concepts are mostly familiar. Finally, the pattern of correlations among the norms is generally similar to those found in other similar studies (Bakhtiar et al., 2013; Dimitropoulou et al., 2009; Duñabeitia et al., 2018; Raman et al., 2014; Rossion & Pourtois, 2004 with the name agreement scores taken from Bonin, Guillemard-Tsaparina, & Méot, 2013; Tsaparina et al., 2011).

From a general standpoint, having norms for stimuli (1) permits the control—methodological or statistical—of the potential influence of confounding variables when investigating a specific factor in a given lexical processing task, such as object naming; (2) helps to establish the underlying structure of the norms and to better understand what exactly they measure, and (3) last but not least, permits the investigation of their relationships with online (or offline) measures of word processing such as object naming (e.g., Bonin et al., 2002; Bonin, Guillemard-Tsaparina, & Méot, 2013; Bonin, Peereman, et al., 2003). Many studies have attempted to identify the determinants of naming speed in adults, and this has been made possible thanks to the availability of norms for pictures in different languages.

As claimed earlier, there is a limited range of options for researchers who want to use *colourised drawings* in their experiments (Duñabeitia et al., 2018). Although photographs of objects have been thought to be more ecological than line drawings (Moreno-Martínez & Montoro, 2012), they generally elicit less consensus on naming responses than line drawings (Clarke & Ludington, 2018). Consequently, line drawings are to be preferred in the field of speeded object naming because they limit the number of trials that have to be discarded when analysing naming times. As pointed out by Clarke and Ludington, name agreement with line drawings is generally above 80%, with *H* values located between .20 and .80. It is worth remembering that for our picture set, the means of name agreement and *H* were, respectively, 89.64% and .40. Given the finding that the distribution of name agreement scores was very similar between our database and MultiPic, researchers will be able to select pictures from these two databases and combined them. However, as the picture-choice task suggests, whenever researchers have to choose between two pictures to represent one concept, they should select the one from IMABASE rather than the corresponding picture from MultiPic given that the great majority of the IMABASE pictures that were common to MultiPic were rated as providing better pictorial representations of the corresponding concepts. Turning to memory experiments, the format of pictures seems to be a lesser concern (Snow et al., 2014). However, our set of drawings will also be useful for the design of memory experiments because, as evaluated by the picture-choice task, the concepts depicted by the drawings were very easy to recognise. For example, pictures are used in memory experiments in children. To give an example, Fitamen et al. (2019) used pictures from the Rossion and Pourtois database to investigate memory processes in 4- and 5-year-old preschoolers. For this type of population, it is especially important to have drawings that are easy to recognise (and indeed in Fitamen et al.'s study, any given picture had to have more than 80% correct denominations at 4 years of age to be selected). Thus, our drawings could be very helpful for memory researchers working on children.

The colourised version of SV has long been a good option for the design of experiments aimed at exploring perceptual, language, or memory processes in both children and adults (e.g., Bonin et al., 2019; Bonin, Guillemard-Tsaparina, & Méot, 2013; Dimitropoulou et al., 2009; Fitamen et al., 2019; Hachmann et al., 2020; Reis et al., 2006). Thus, this new set of colourised line drawings and the corresponding collected norms should help investigate further the determinants of naming speed and accuracy in adults but also in children, teenagers, and older adults. There have been many studies on picture naming in adults but fewer in children (e.g., Cycowicz et al., 1997; D'Amico et al., 2001) or in older adults (e.g., Au et al., 1995), and we are not aware of any study that has specifically focused on the naming speed and accuracy in teenagers.

As we also pointed out in the "Introduction," with the passing of time, the SV pictures have come to suffer from several limitations: Certain pictures were found not to be "good" representations of familiar concepts; certain pictures were obviously linked to American culture and were not very suitable for other cultures; certain objects needed to be updated. The current database, together with the recent MultiPic database of colourised line drawings, will greatly help researchers to select items when designing experiments that use pictures as stimuli (since its publication, the MultiPic database has already been used to study lexical selection in spoken word production, e.g., Gauvin et al., 2018). Our database is somewhat reduced in terms of the number of pictures it offers for selection compared with MultiPic, but we think that is likely sufficient for many experimental designs (remember that the SV database contains only 260 pictures, and it has been widely used to design many experiments in several psychology domains). However, and more importantly, our database is normed on many more psycholinguistic variables than

MultiPic, which is normed only on two variables: name agreement and visual complexity (the latter of which has been found to be a factor of lesser importance in picture naming than, for instance, conceptual familiarity, see Perret & Bonin, 2019). In the future, we hope to be able to increase the number of pictures proposed in IMABASE. Recently, "new" psycholinguistic variables have been collected for items, such as object manipulability and interestingly, this variable has been found to be a significant predictor of naming latency (Guérard et al., 2015; Lorenzoni et al., 2018). Thus, we also hope in the future to collect additional psycholinguistic variables for our material. Another limitation of the current work is that the norms were collected from French-speaking adults who live in France, and it is possible that the set of norms may not be suitable for use with adults speaking French in other parts of the world such as Belgium, Switzerland, parts of Africa (e.g., Senegal), or Oceania (e.g., New Caledonia). It would be interesting in the future to collect French norms on this set of pictures but with such French speakers and to compare them with the current norms.

Hopefully, as has been the case for the SV picture database, the scientific community will adopt this new database and foreign colleagues will think about collecting norms in their own native languages. To this end, the pictures are open-access and free from copyright restrictions for non-commercial purposes. It is important to point out that IMABASE should not be restricted to the investigation of picture naming. Indeed, all fields of research that use picture naming as an experimental task should be impacted by the publication of such standardised pictures. Likewise, as already said, the database should help researchers who investigate learning and memory processes in children and in both young adults and the elderly. In these areas of research, pictures are often used as stimuli (Bonin et al., 2014). Standardised pictures are, for instance, used to investigate the development of short-term memory coding strategies in children (e.g., Henry et al., 2012), or to address the question of the semantic impairment in early dementia of Alzheimer type (Garrad et al., 2005). In addition, the present database could be useful for the investigation of perceptual processes involved in visual object recognition which was also the case for the SV database (e.g., Snodgrass & Corwin, 1988).

Acknowledgements

The authors thank Patrick Borkowski for his availability and for having provided several versions of many drawings in order for the authors to choose the most suitable for the different concepts present in IMABASE. They thank Manuel Perea and the two anonymous reviewers for helpful comments on a previous version of the manuscript.

Data availability

The different files corresponding to the norms are formatted as .xls. None of the studies were preregistered.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Patrick Bonin https://orcid.org/0000-0003-1576-863X
Patrick Bard https://orcid.org/0000-0001-5464-4766

Notes

1. Certain studies have shown that subjective visual complexity ratings (which are often used as a dimension of visual complexity) are influenced by the familiarity with the objects (e.g., Forsythe, 2009; Forsythe et al., 2017). Following Székely and Bates (2000), we think that using objective visual complexity is a way to attempt to have a "less contaminated" measure of visual complexity. However, we acknowledge that, perhaps, it could be useful in the future to collect subjective visual complexity norms on our drawings because these could be (1) directly compared with the objective ones and (2) used in a "complementary manner" to the objective ones (as far as objective word frequency measures are concerned, a similar debate has arisen concerning the importance of subjective word frequency measures in comparison with objective word frequency measures, e.g., Balota et al., 2001; Brysbaert & Cortese, 2011).
2. We compared, a posteriori, our semantic categories with the semantic categories available for French in a previous study (Bueno & Megherbi, 2009). Using a stringent criterium, that is, use of exactly the same category name, for the 265 items common to Bueno and Megherni and our own study, the overlap is in the magnitude of 67%. Using a more lenient criterium, that is to say, a subcategory is taken as being similar to the main category (e.g., the subcategory "birds" ~ the category "animals"), the overlap amounts to 81%. In addition, it is worth noting that most items belong to several semantic categories. For instance, a hen can be categorised as an animal, or as a bird; it can even be categorised more precisely as a farmyard animal. The different categories provided in the Supplementary Material for our database will make it possible for researchers to select more easily the items they need to build their experiments.
3. Two and six words, from even and odd participants, respectively, were excluded from the computations because of discrepancies between the modal names obtained from even and odd participants.
4. $H = \sum_{i=1}^{k} p_i \log_2(1/p_i)$ where k refers to the number of different names given for each picture and p_i is the proportion of participants giving each name.
5. It is worth noting that Snodgrass and Vanderwart (1980) did not exclude "DK items" and "idiosyncratic responses." The latter procedure leads to an overestimation of name agreement percentages.

Supplementary material

The supplementary material is available at qjep.sagepub.com

References

Alario, F.-X., & Ferrand, L. (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments, & Computers*, *31*, 531–552.

Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers*, *36*, 140–155.

Au, R., Joung, P., Nicholas, M., Obler, L. K., Kass, R., & Albert, M. L. (1995). Naming ability across the adult life span. *Aging, Neuropsychology, and Cognition*, *2*, 300–311.

Bakhtiar, M., Nilipour, R., & Weekes, B. S. (2013). Predictors of timed picture naming in Persian. *Behavior Research Methods*, *45*, 834–841.

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition*, *29*, 639–647.

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *Quarterly Journal of Experimental Psychology*, *50A*, 560–585.

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., Herron, D., Lu, C. C., Pechmann, T., Pleh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., Orozco-Figueroa, A., Tzeng, A., & Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, *10*, 344–380.

Bonin, P. (2013). *Psychologie du langage: La fabrique des mots [Language production: The building of words]*. De Boeck.

Bonin, P., Chalard, M., Méot, A., & Fayol, M. (2002). The determinants of spoken and written picture naming latencies. *British Journal of Psychology*, *93*, 89–114.

Bonin, P., Gelin, M., & Bugaiska, A. (2014). Animates are better remembered than inanimates: Further evidence from word and picture stimuli. *Memory & Cognition*, *42*, 370–382.

Bonin, P., Guillemard-Tsaparina, D., & Méot, A. (2013). Determinants of naming latencies, object comprehension times, and new norms for the Russian standardized set of the colorized version of the Snodgrass and Vanderwart pictures. *Behavior Research Methods*, *45*, 731–745.

Bonin, P., Méot, A., Aubert, L., Malardier, N., Niedenthal, P., & Capelle-Toczek, M.-C. (2003). Normes de concrétude, de valeur d'imagerie, de fréquence subjective et de valence émotionnelle pour 866 mots [Concreteness, imageability, subjective frequency and emotional valence norms for 866 words]. *L'Année Psychologique*, *104*, 655–694.

Bonin, P., Méot, A., Boucheix, J. M., & Bugaiska, A. (2018). Psycholinguistic norms for 320 fixed expressions (idioms and proverbs) in French. *Quarterly Journal of Experimental Psychology*, *71*, 1057–1069.

Bonin, P., Méot, A., & Bugaiska, A. (2013). Norms and comprehension times for 305 French idiomatic expressions. *Behavior Research Methods*, *45*, 1259–1271.

Bonin, P., Méot, A., Ferrand, L., & Roux, S. (2011). L'imageabilité: Normes et relations avec d'autres variables psycholinguistiques [Imageability: Norms and relationships with other psycholinguistic variables]. *L'Année Psychologique*, *111*, 327–357.

Bonin, P., Méot, A., Laroche, B., Bugaiska, A., & Perret, C. (2019). The impact of image characteristics on written naming in adults. *Reading and Writing*, *32*, 13–31.

Bonin, P., Méot, A., Millotte, S., & Bugaiska, A. (2015). Norms and reading times for acronyms in French. *Behavior Research Methods*, *47*, 251–267.

Bonin, P., Peereman, R., Malardier, N., Méot, A., & Chalard, M. (2003). A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. *Behavior Research Methods, Instruments, & Computers*, *35*, 158–167.

Bonin, P., Perret, C., Méot, A., Ferrand, L., & Mermillod, M. (2008). Psycholinguistic norms and face naming times for photographs of celebrities in French. *Behavior Research Methods*, *40*, 137–146.

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*, 585–599.

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*, e10773.

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of Standardized Stimuli (BOSS) Phase II: 930 new normative photos. *PLOS ONE*, *9*, e106953.

Brysbaert, M., & Cortese, M. J. (2011). Do the effects of subjective frequency and age of acquisition survive better word frequency norms? *Quarterly Journal of Experimental Psychology*, *64*, 545–559.

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: A review update. *Current Directions in Psychological Science*, *27*, 45–50.

Bueno, S., & Megherbi, H. (2009). French categorization norms for 70 semantic categories and comparison with Van Overschelde et al.'s (2004) English norms. *Behavior Research Methods*, *41*, 1018–1028.

Clarke, A. J. B., & Ludington, J. D. (2018). 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*, 607–626.

Cortese, M. J., Simpson, G. B., & Woolsey, S. (1997). Effects of association and imageability on phonological mapping. *Psychonomic Bulletin & Review*, *4*, 226–231.

Cuetos, F., Aguado, G., Izura, C., & Ellis, A. W. (2002). Aphasic naming in Spanish: Predictors and errors. *Brain and Language*, *82*, 344–365.

Cuetos, F., Ellis, A. W., & Alvarez, B. (1999). Naming times for the Snodgrass and Vanderwart pictures in Spanish. *Behavior Research Methods, Instruments, & Computers*, *31*, 650–658.

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*, 171–237.

D'Amico, S., Devescovi, A., & Bates, E. (2001). Picture naming and lexical access in Italian children and adults. *Journal of Cognition and Development*, *2*, 71–105.

Dimitropoulou, M., Duñabeitia, J. A., Blitsas, P., & Carreiras, M. (2009). A standardized set of 260 pictures for Modern Greek:

Norms for name agreement, age of acquisition, and visual complexity. *Behavior Research Methods*, *41*, 584–589.
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*, 808–816.
Ellis, A. W., & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 515–523.
Ferrand, L., Bonin, P., Méot, A., Augustinova, M., New, B., Pallier, C., & Brysbaert, M. (2008). Age-of-acquisition and subjective frequency estimates for all generally known monosyllabic French words and their relation with other psycholinguistic variables. *Behavior Research Methods*, *40*, 1049–1054.
Fitamen, C., Blaye, A., & Camos, V. (2019). Five-year-old children's working memory can be improved when children act on a transparent goal cue. *Scientific Reports*, *9*, 15342.
Forsythe, A. (2009). Visual complexity: Is that all there is? In D. Harris (Ed.), *Proceedings of the 8th international conference on engineering psychology and cognitive ergonomics* (pp. 158–166). Springer.
Forsythe, A., Street, N., & Helmy, M. (2017). Revisiting Rossion and Pourtois with new ratings for automated complexity, familiarity, beauty, and encounter. *Behavior Research Methods*, *49*, 1484–1493.
Garrad, P., Lambon Ralph, M. A., Patterson, K., Pratt, K. H., & Hodges, J. R. (2005). Semantic feature knowledge and picture naming in dementia of Alzheimer's type: A new approach. *Brain and Language*, *93*, 79–94.
Gauvin, H. S., Jonen, M. K., Choi, J., McMahon, K., & de Zubicaray, G. I. (2018). No lexical competition without priming: Evidence from the picture–word interference paradigm. *Quarterly Journal of Experimental Psychology*, *71*, 2562–2570.
Guérard, K., Lagacé, S., & Brodeur, M. B. (2015). Four types of manipulability ratings and naming latencies for a set of 560 photographs of objects. *Behavior Research Methods*, *47*, 443–470.
Hachmann, W. M., Cashdollar, N., Postiglione, F., & Job, R. (2020). The relationship of domain-general serial order memory and reading ability in school children with and without dyslexia. *Journal of Experimental Child Psychology*, *193*, 104789.
Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, *1*, 77–89.
Henry, L. A., Messer, D., Luger-Klein, S., & Crane, L. (2012). Phonological, visual, and semantic coding strategies and children's short-term picture memory span. *Quarterly Journal of Experimental Psychology*, *65*, 2033–2053.
Humphreys, G. W., Riddoch, M. J., & Quinlan, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology*, *5*, 67–103.
Izura, C., & Playfoot, D. (2012). A normative study of acronyms and acronym naming. *Behavior Research Methods*, *44*, 862–889.
Johnston, R. A., & Barry, C. (2006). Age of acquisition and lexical processing. *Visual Cognition*, *13*, 789–845.
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*, 461–469.
Juhasz, B. J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin*, *131*, 684–712.
Khwaileh, T., Mustafawi, E., Herbert, R., & Howard, D. (2018). Gulf Arabic nouns and verbs: A standardized set of 319 object pictures and 141 action pictures, with predictors of naming latencies. *Behavior Research Methods*, *50*, 2408–2425.
Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–38; discussion 38–75.
Lorenzoni, A., Peressotti, F., & Navarrete, E. (2018). The manipulability effect in object naming. *Journal of Cognition*, *1*, 30.
Marful, A., Díez-Álamo, A. M., Plaza-Navas, S., & Fernandez, A. (2018). A normative study for photographs of celebrities in Spain. *PLOS ONE*, *13*, e0197554.
Martein, R. (1995). Norms for name and concept agreement, familiarity, visual complexity and image agreement on a set of 216 pictures. *Psychologica Belgica*, *35*, 205–225.
Moreno-Martínez, 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*, e37527.
New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, *36*, 516–524.
Nishimoto, T., Miyawaki, K., Ueda, T., Une, Y., & Takahashi, M. (2005). Japanese normative set of 359 pictures. *Behavior Research Methods*, *37*, 398–416.
O'Sullivan, M., Lepage, M., Bouras, M., Montreuil, T., & Brodeur, M. B. (2012). North-American norms for name disagreement: Pictorial stimuli naming discrepancies. *PLOS ONE*, *7*, e47802.
Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monographs*, *76*, 1–25.
Perret, C., & Bonin, P. (2019). Which variables should be controlled for to investigate picture naming in adults? A Bayesian meta-analysis. *Behavior Research Methods*, *51*, 2533–2545.
Pind, J., Jónsdóttir, H., Tryggvadóttir, H. B., & Jonsson, F. (2000). Icelandic norms for the Snodgrass and Vanderwart (1980) pictures: Name and image agreement, familiarity, and age of acquisition. *Scandinavian Journal of Psychology*, *41*, 41–48.
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.
Raman, I., Raman, E., & Mertan, B. (2014). A standardized set of 260 pictures for Turkish: Norms of name and image agreement, age of acquisition, visual complexity, and conceptual familiarity. *Behavior Research Methods*, *46*, 588–595.
Reis, A., Faísca, L., Ingvar, M., & Petersson, K. M. (2006). Color makes a difference: Two-dimensional object naming in literate and illiterate subjects. *Brain and Cognition*, *60*, 49–54.

Roelofs, A., & Ferreira, V. S. (2019). The architecture of speaking. In P. Hagoort (Ed.), *Human language: From genes and brains to behavior* (pp. 35–50). MIT Press.

Rogić, M., Jerončić, A., Bošnjak, M., Sedlar, A., Hren, D., & Deletis, V. (2013). A visual object naming task standardized for the Croatian language: A tool for research and clinical practice. *Behavior Research Methods*, *45*, 1144–1158.

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*, 217–236.

Russo, N., Hagmann, C. E., Andrews, R., Black, C., Silberman, M., & Shea, N. (2018). Validation of the C.A.R.E. stimulus set of 640 animal pictures: Name agreement and quality ratings. *PLOS ONE*, *13*, e0192906.

Salmon, J. P., McMullen, P. A., & Filliter, J. H. (2010). Norms for two types of manipulability (graspability and functional usage), familiarity, and age of acquisition for 320 photographs of objects. *Behavior Research Methods*, *42*, 82–95.

Sanfeliù, 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*, 537–555.

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*, 2498–2510.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*, 34–50.

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: Learning, Memory, and Cognition*, *6*, 174–215.

Snodgrass, J. G., & Yuditsky, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments, & Computers*, *28*, 516–536.

Snow, J. C., Skiba, R. M., Coleman, T. L., & Berryhill, M. E. (2014). Real-world objects are more memorable than photographs of objects. *Frontiers in Human Neuroscience*, *8*, e837.

Strain, E., Patterson, K. E., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1140–1154.

Strain, E., Patterson, K. E., & Seidenberg, M. S. (2002). Theories of word naming interact with spelling-sound consistency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 207–214.

Székely, A., & Bates, E. (2000). Objective visual complexity as a variable in studies of picture naming. *CRL Newsletter*, *12*, 3–33.

Székely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., Jacobsen, T., Arévalo, A. L., Vargha, A., & Bates, E. (2005). Timed action and object naming. *Cortex*, *41*, 7–25.

Székely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., Lu, C. C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., Orozco-Figueroa, A., Tzeng, A., Tzeng, O., Arévalo, A., Vargha, A., Butler, A. C., Buffington, R., & Bates, E. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language*, *51*, 247–250.

Tsaparina, D., Bonin, P., & Méot, A. (2011). Russian norms for name agreement, image agreement for the colorized version of the Snodgrass and Vanderwart pictures and age of acquisition, conceptual familiarity, and imageability scores for modal object names. *Behavior Research Methods*, *43*, 1085–1099.

Valente, A., Bürki, A., & Laganaro, M. (2014). ERP correlates of word production predictors in picture naming: A trial by trial multiple regression analysis from stimulus onset to response. *Frontiers in Neuroscience*, *8*, Article 390.

Vitkovitch, M., & Tyrrell, L. (1995). Sources of disagreement in object naming. *Quarterly Journal of Experimental Psychology*, *48A*, 822–848.

Weekes, B. S., Shu, H., Hao, M., Liu, Y., & Tan, L. H. (2007). Predictors of timed picture naming in Chinese. *Behavior Research Methods*, *39*, 335–342.

Yoon, C., Feinberg, F., Luo, T., Hedden, T., Gutchess, A. H., Chen, H. Y. M., Mikels, J. A., Jiao, S., & 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*, 639–649.