Open access • Journal Article • DOI:10.1080/17470218.2017.1310261

MultiPic: a standardized set of 750 drawings with norms for six European languages

Jon Andoni Duñabeitia, Davide Crepaldi, Davide Crepaldi, Antje S. Meyer ...+6 more authors

Institutions: University of Milano-Bicocca, International School for Advanced Studies, Radboud University Nijmegen, Max Planck Society ...+5 more institutions

Published on: 01 Jan 2018 - Quarterly Journal of Experimental Psychology (Q J Exp Psychol (Hove))

Topics: Psycholinguistics, British English and German

QJEP

Rapid Communication

Quarterly Journal of Experimental Psychology
2018, Vol. 71(4) 808–816

DOI: 10.1080/17470218.2017.1310261
qjep.sagepub.com
SAGE

MultiPic: A standardized set of 750 drawings with norms for six European languages

Jon Andoni Duñabeitia[1], Davide Crepaldi[2,3], Antje S. Meyer[4], Boris New[5,6], Christos Pliatsikas[7], Eva Smolka[8] and Marc Brysbaert[9]

Abstract

Numerous studies in psychology, cognitive neuroscience and psycholinguistics have used pictures of objects as stimulus materials. Currently, authors engaged in cross-linguistic work or wishing to run parallel studies at multiple sites where different languages are spoken must rely on rather small sets of black-and-white or colored line drawings. These sets are increasingly experienced as being too limited. Therefore, we constructed a new set of 750 colored pictures of concrete concepts. This set, MultiPic, constitutes a new valuable tool for cognitive scientists investigating language, visual perception, memory and/or attention in monolingual or multilingual populations. Importantly, the MultiPic databank has been normed in six different European languages (British English, Spanish, French, Dutch, Italian and German). All stimuli and norms are freely available at *http://www.bcbl.eu/databases/multipic*.

Keywords

Picture database; drawings; psycholinguistic resources; multilingual database; translation equivalents; visual complexity

Received: 15 November 2016; revised: 3 February 2017; accepted: 3 February 2017

When Joan G. Snodgrass and Mary Vanderwart published their standardized set of 260 pictures in 1980, they probably did not anticipate the massive use scientists and practitioners would make of the set: 35 years later, their paper has been cited over 4750 times (Google Scholar), and their materials have been widely used in studies investigating object recognition, word production, aphasia and bilingualism, among others. In addition, the materials have been normed for several other languages, making them the primary item set for cross-linguistic research involving picture naming (e.g. Alario & Ferrand, 1999; Barry, Morrison, & Ellis, 1997; Martein, 1995; Nishimoto, Miyawaki, Ueda, Une, & Takahashi, 2005; Rogić et al., 2013; Sanfeliu & Fernandez, 1996; Wang, Chen, & Zhu, 2014).

Despite its broad use in the field, Snodgrass and Vanderwart's dataset of line drawings is not without limitations. First, as demonstrated by Rossion and Pourtois (2004), the original black-and-white version of the line drawings yields impoverished recognition as compared to colorized and texturized versions of the same items. The latter help participants to retrieve the names of objects faster and more reliably. For this reason, in recent years, researchers have progressively moved to the colored and copyright-free version designed by Rossion and Purtois (e.g. Bonin, Guillemard-Tsaparina, & Méot, 2013; Dimitropoulou, Duñabeitia, Blitsas, & Carreiras, 2009; Raman, Raman, & Mertan, 2014).

Second, and contrary to common belief, the original items from the Snodgrass and Vanderwart set were not freely available. Researchers had to apply for copyright clearance to the authors and at some point had to pay for using the dataset. This was a particular problem for

[1]BCBL – Basque Center on Cognition, Brain and Language, Donostia, Spain
[2]Milan Center for Neuroscience, University of Milano-Bicocca, Milano, Italy
[3]International School for Advanced Studies (SISSA), Trieste, Italy
[4]Max Planck Institute for Psycholinguistics, Radboud University, Nijmegen, The Netherlands
[5]LPNC, Université de Savoie, Chambéry, France
[6]LPNC, CNRS, Grenoble, France
[7]School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK
[8]University of Konstanz, Konstanz, Germany
[9]Ghent University, Ghent, Belgium

Corresponding author:
Jon Andoni Duñabeitia, BCBL – Basque Center on Cognition, Brain and Language, Paseo Mikeletegi 69, 2nd Floor, 20009 Donostia, Spain.
Email: j.dunabeitia@bcbl.eu

authors wanting to develop normative data in other languages, as they could not make the stimuli available together with the norms.

A third limitation was that Snodgrass and Vanderwart's picture databank was limited to 260 drawings, which rapidly became a serious constraint in the design of experiments. As a result, new drawings were added and normed in various languages (see the International Picture-Naming Project, IPNP, by Bates et al., 2003, and Székely et al., 2004, for 520 black-and-white drawings of common objects and 275 black-and-white drawings of actions from multiple sources; see also Bonin, Peereman, Malardier, Méot, & Chalard, 2003; Cycowicz, Friedman, Rothstein, & Snodgrass, 1997; Moreno-Martínez & Montoro, 2012).

As usual, the situation is better for English than for other languages, in particular since the publication of the BOSS dataset, which includes photos of 1468 pictures with various norms (Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010; Brodeur, Guérard, & Bouras, 2014). Unfortunately, these data are limited to American English and, for a subset of the stimuli, to Canadian French (Brodeur et al., 2012).

Given the above limitations, researchers who want to use normed drawings and accumulate data across different European languages have limited choice: They can either choose the Rossion and Pourtois (2004) stimulus set of 260 colored pictures, or they can use the 520 black-and-white drawings from IPNP. The IPNP dataset comes from 10 different sources which, despite being described as "comparable in style" (Bates et al., 2003, p. 350), include items with different pictorial styles and graphic properties related to the artists' individual styles.

The present paper describes the MultiPic databank, which is a new database of standardized pictures that brings together the strengths of existing datasets while avoiding some of their limitations. With 750 pictures, MultiPic is a *large* dataset of *colored* line drawings coming from the *same source* and *normed for multiple languages*. It is the result of a collaborative European project intended to provide the scientific community with a set of 750 publicly available color drawings representing common concrete concepts created by the same artist, standardized for name agreement and visual complexity in several European languages. The languages tested so far are British English, German, Italian, Spanish, French and Dutch (separately normed for the Netherlands and Belgium). In the following sections, we describe how the materials were generated and normed in order to facilitate replication and extension to other languages in future research.

Method

Participants

In all, 620 undergraduate and graduate students (443 females) with a mean age of 22.03 years (SD = 4.26) were tested across sites (see Table 1 for the specific characteristics of the sample tested in each language). Only participants who were native speakers of the target language were selected. Ethical approval for conducting the study had been obtained from ethics boards of the participating institutions as required.

Materials

The materials were commissioned by the first author and drawn by a local artist. An initial set of nearly 600 Spanish words with highly distinctive pictorial characteristics were selected from the Spanish lexeme database ESPAL on the basis of their imageability and concreteness values (ranging between 4 and 7 on the 1-to-7 Likert scales; see Duchon, Perea, Sebastián-Gallés, Martí, & Carreiras, 2013). These words were selected to cover a wide range of frequencies and semantic fields. After consultation with the artist, some items were dropped from the list because they were difficult to express in line drawings, and an extra set of nearly 200 items was added. The artist then started the creation of the pictorial set. She created the drawings with a freehand computer application (using a digital tablet and pen set) that was optimal for making digital, two-dimensional graphics. The artist used the same space for the whole set, so that in its original form each picture occupied a space of 15 x 15 centimeters (DPI = 300 pixels/centimeter). In addition, the artist was asked to use the same graphic style for all the drawings, using strokes of similar width. Finally, we asked for coherence in the coloring of the drawings. Two external informants (research assistants from the BCBL) assessed the quality of the drawings, and improvements were made when needed, resulting in the final set of 750 drawings used in the normative studies.

Procedure

Seven norming studies were carried out, one per language or dialectal variation. Data acquisition was lab-based for all studies, with the exception of the Dutch study (Netherlands), which was carried out over the Internet. The English, Italian, German and Spanish studies were carried out using Experiment Builder (SR-Research, Ontario, Canada). The French study was implemented in OpenSesame 2.8.3 (Mathôt, Schreij, & Theeuwes, 2012). Finally, the Dutch studies (Netherlands and Belgium) were run with Internet presentation using software developed at the Max Planck Institute for Psycholinguistics in Nijmegen.

Participants were tested individually. The 750 drawings were presented one by one in the center of a computer screen. Participants were asked (a) to type in the name that they felt best described the picture and (b) to rate the visual complexity of the drawing on a scale from 1 (very simple) to 5 (very complex). They were specifically asked to provide a single name for each picture, avoiding the use of

Table 1. Characteristics of the samples tested in each language, and means and standard deviations (in parentheses) of the different indices obtained in the norming study in each language.

	LANGUAGES						
	German	Italian	Spanish	English	French	Dutch (Netherlands)	Dutch (Belgium)
Size (n)	100	100	100	100	100	60	60
Females (n)	60	68	61	71	88	42	43
Age	22.2 (3.1)	24.5 (5.7)	23.9 (4.3)	19.5 (3.5)	20.2 (2.8)	21.1 (2.4)	22.8 (4.0)
Percentage of Unknown Items	1.2% (3.2)	2.8% (4.9)	1.3% (3.2)	3.9% (7.2)	2.7% (5.7)	2.0% (5.0)	2.9% (5.0)
Percentage of Idiosyncratic Responses	3.2% (3.3)	1.8% (2.1)	1.4% (1.8)	1.9% (2.0)	2.8% (3.1)	3.8% (3.9)	4.0% (4.3)
H Index	0.86 (0.75)	0.69 (0.67)	0.50 (0.58)	0.78 (0.70)	0.71 (0.68)	0.78 (0.73)	0.83 (0.73)
Percentage of Modal Responses	78.5% (20.8)	82.9% (18.6)	87.6% (16.5)	79.9% (20.2)	82.5% (18.9)	79.7% (20.7)	78.2% (20.9)
Visual Complexity	2.72 (0.59)	2.44 (0.48)	2.43 (0.51)	2.69 (0.50)	2.41 (0.56)	2.77 (0.60)	2.94 (0.67)

longer noun phrases. After both answers had been provided, the next picture was shown. If the participants did not know the object or its name, they were asked to use a specific language string (e.g. "DK" in English, corresponding to "don't know"). For the visual complexity rating, participants were explicitly told not to rate the visual complexity of the object in real life, but the complexity of the drawing, as has been done in previous studies.

The drawings were presented in a different random order to each participant. The experimental session was self-paced and lasted about 90 min.

Data preprocessing

First, a native speaker of each language checked the answers for obvious spelling errors and corrected them. Together with this, an initial response recoding was done for each language by merging basic variants of the same names (e.g. hyphenated or pluralized forms) and discarding parts of speech that were not nouns or verbs (e.g. determiners and adjectives).

Second, 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 (see Table 1 for details). In line with the liberal criterion tested by Snodgrass and Yuditsky (1996), we excluded the responses given by single participants from further analyses, in order not to jeopardize the final outcome by outliers. We felt safe to do so because our stimuli were normed by 60 or 100 participants (rather than the much smaller sample sizes used in other normative studies). Researchers who are particularly interested in idiosyncratic responses are referred to a file with the raw data.[1] The trials with non-idiosyncratic responses constitute the core of MultiPic, and those data were submitted to further analyses. These data ranged from 93.1% of the trials for Belgium Dutch to 97.3% of the trials for Spanish.

Third, name agreement was computed for each of the 750 drawings in each dataset calculating the *H* index and the percentage of modal names. The *H* index is an information statistic that reflects the level of agreement across participants (Shannon & Weaver, 1949). When all participants give the same name to a given drawing the *H* index is 0, and the *H* value increases as a function of the divergence of responses. We also calculated the percentage of participants who gave the most frequent name (i.e. the modal name). As stated above, these indices were based on the agreement for each modal noun regardless of the modifiers (see O'Sullivan, Lepage, Bouras, Montreuil, & Brodeur, 2012). Nonetheless, it is worth mentioning that only 0.81% of the unfiltered responses included more than one word. Visual complexity scores were obtained by averaging the numerical responses to each drawing across participants in each sample and by computing the corresponding means (see Table 1).

Results

Name agreement

As shown in Figure 1, many drawings were named consistently by most participants (in all languages, more than 40% of the items yielded *H* indices below 0.5). The mean *H* index in the different languages ranged from 0.50 (Spanish) to 0.86 (German), and the mean percentage of modal names ranged from 78.5% (German) to 87.6% (Spanish). Across languages, many items yielded a single response (i.e. *H* index of 0). These numbers ranged from 139 in German to 280 items in Spanish. Similarly, the percentages of the most frequent responses for each item were reassuringly high, with a substantial number of drawings yielding values higher than 70% (ranging from 512 in Belgian Dutch to 628 in Spanish).[2]

Native speakers individually inspected the 15,695 different names constituting MultiPic to classify them by

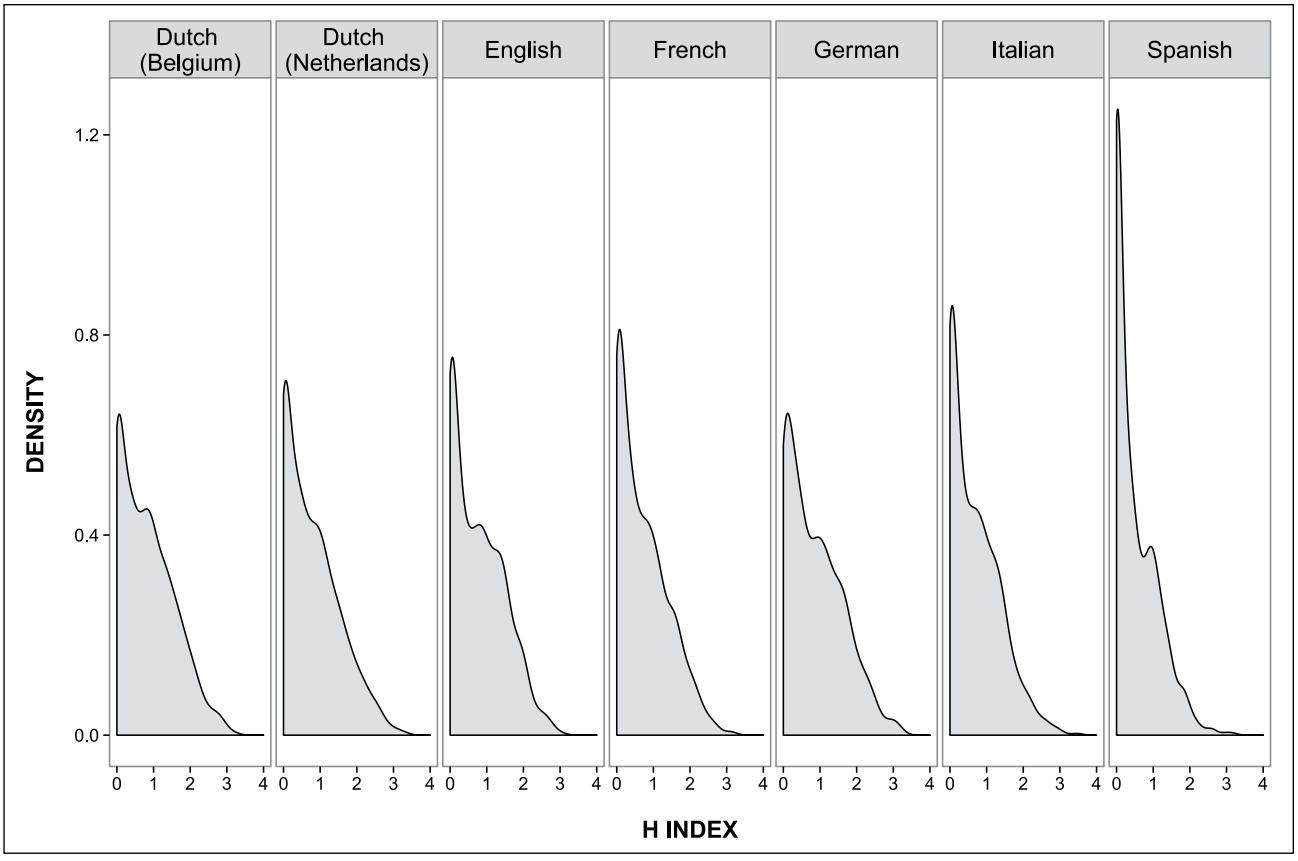

Figure 1. Density plots of the *H* index in each language.

their correctness (i.e., whether or not they could be considered as "reasonable responses").[3] As in preceding studies, this classification was based on the identity of the depicted concept (see O'Sullivan et al., 2012). Responses were classified as correct when the word appropriately reflected the concept presented, and this category included synonyms, near-synonyms, hypernyms and responses, including addition or deletion of some of the words, specifications of the function or part, brands, breeds, types and other forms of hyponymy. In contrast, responses were classified as incorrect when the words corresponded to unrelated concepts, semantically related but distinct concepts, visually similar but semantically unrelated identities, and to incorrect specifications (as in O'Sullivan et al., 2012).

According to this manual coding, only 7.1% of the responses across languages were classified as incorrect (SD = 3.42; range: 0.74-9.93 across languages). At the same time, it became clear that the decision whether a response is acceptable or not is far from trivial. We looked at the inter-rater reliability (IRR) by asking three trained coders to rate the responses to a random subset of 150 drawings from the Spanish dataset. *Kappa* was computed for each coder pair and then averaged to provide a unitary index of IRR (see Light, 1971). The resulting *kappa* indicated partial disagreement between raters as to which responses could be considered correct ($\kappa = .579$; see Landis & Koch, 1977). Readers who want to use the stimuli in a particular language, can find all the unfiltered data and the correctness classification as supplemental materials by referring to *https://figshare.com/s/bf4a360ab3c90537291f*.

Cross-language comparisons

The mean name agreement indices for the current dataset (an *H* index of 0.74 and 80% modal name responses across languages) are in line with those reported in preceding studies (e.g. Alario & Ferrand, 1999: 0.36 and 85%; Bonin et al., 2003: 0.67 and 77%; Brodeur et al., 2014: 1.71 and 60%; Dimitropoulou et al., 2009: 0.55 and 87%; Liu, Hao, Li, & Shu, 2011: 1.32 and 66%; Manoiloff, Artstein, Canavoso, Fernández, & Segui, 2010: 0.68 and 81%; Nishimoto et al., 2005, liberal criterion: 0.74 and 91%; Rossion & Pourtois, 2004, colorized: 0.32 and 90%; Sanfeliu & Fernandez, 1996: 0.27 and 82%; Snodgrass & Vanderwart, 1980: 0.56 and 87%; Tsaparina, Bonin, & Méot, 2011: 0.82 and 81%). In general, name agreement is higher for smaller (often more carefully selected) stimulus sets than for larger stimulus sets (Brodeur et al., 2014).

Because the MultiPic database was tested in the same way across languages, it provides opportunities to look at

language commonalities and differences, and it deals with the problem mentioned by Dell'Acqua, Lotto and Job (2000) that any language difference in previous studies could be due *"to a pure cultural difference in the subjects' population or to a structural difference in the pictures submitted to the subjects' judgment"* (p. 592). Because the MultiPic database includes name agreement data and visual complexity ratings for the same input in different languages, a within-study cross-cultural comparison can be performed.

To do so, a series of correlation analyses were run in order to verify the cross-linguistic commonalities of the measures obtained. These correlations are presented in Table 2, where Pearson's *r* is reported ($\alpha = .05$ after Holm's correction for multiple comparisons). All cross-language correlations between related indices were high, demonstrating the inter-lingual convergence of the results. Not surprisingly, the strongest (close to perfect) correlations were obtained for the visual complexity scores, with *r* values between .90 and .97.

In addition, a hierarchical cluster analysis was performed using Hoeffding's test of independence (30**D* statistic; Hoeffding, 1948). According to this measure of mutual independence ranging from -0.5 to 1, larger values of *D* represent higher dependency between two variables. As shown in Figure 2, the visual complexity measures in all languages clustered together with high *D* scores, being independent of the rest of the variables. The *H* indices and the percentages of modal responses showed high interdependency scores in each language, given that they represent two alternative forms of measuring name agreement. These results showed that the data for each language clustered together, and more interestingly, that the different languages formed coherent clusters on the basis of the language families (i.e. Germanic: Dutch, German and English; Romance: French, Italian and Spanish).

Discussion

In this article, we presented the open-source Multilingual Picture (MultiPic) database, which includes norming data in different European languages for a large set of colored drawings from the same source created for psycholinguistic and (neuro)cognitive research. MultiPic includes 750 drawings of a variety of concepts that have been normed for name agreement and visual complexity in 6 different languages (German, Spanish, British English, French, Italian and Dutch, assessed separately for speakers in Belgium and The Netherlands).

The results revealed a high degree of convergence of the responses across speakers of the same language for each drawing, as shown by the low *H* indices and the high percentages of participants providing modal names. Furthermore, cross-linguistic correlations showed that the name agreement scores for the individual items were relatively similar across languages, while at the same time highlighting interesting language- and culture-dependent differences (see also Brodeur et al., 2012; Dell'Acqua et al., 2000). As shown in Table 2, the correlations across languages in the *H* indices ranged between $r = .51$ and $r = .68$, except for the two samples of Dutch-speaking participants, who showed a correlation of $r = .83$. A similar result was found when comparing the percentages of participants providing the modal names as responses: All bivariate correlations resulted in *r* values between .45 and .63, with the exception of Dutch-speaking groups that showed a much higher correlation of $r=.79$. The moderate strength of the correlations in name agreement across languages is in line with earlier evidence (see in particular Dell'Acqua et al., 2000).

The correlations found between the two Dutch-speaking groups show that a shared language boosts the concordance of the naming, but that it is not enough to get exactly the same names for all pictures. Because the two groups come from two different countries (The Netherlands and Belgium) some names diverge as a result of cultural differences between the two communities. These data can be compared to the results of Brodeur et al. (2012), who showed that norms obtained from different groups of Canadian participants speaking either French or English were highly similar, suggesting that their underlying common cultural setting lessened the impact of their linguistic differences.

MultiPic represents an alternative to the pictorial sets currently used in cross-language research (e.g. Rossion & Pourtois, 2004; Snodgrass & Vanderwart, 1980; Székely et al., 2004). It overcomes the limitations of the existing sets by including considerably more items than any of the preceding databanks and by being created by the same person in order to minimize heterogeneity in format, stroke, color codes and styles of the drawings. Most importantly, the drawings included in MultiPic have been standardized in different languages using a parallel protocol (for a similar approach, see also Bates et al., 2003; Székely et al., 2004; Viggiano, Vannucci, & Righi, 2004; Yoon et al., 2004)). This is an important feature, because it enables researchers from different European countries to use the same materials not only for research on monolingual samples, but also for cross-linguistic research on multilingual participants.

At the same time, we acknowledge that the databank, like any other tool, has its limitations. MultiPic only provides name agreement and visual complexity ratings at the moment. The inclusion of additional subjective ratings on other characteristics of the drawings, such as image agreement, color diagnosticity or emotional valence will be of interest. Another limitation is that the MultiPic databank only includes drawings of individual concrete objects. An extension to drawings of events and actions would be desirable. Researchers should also use the statistics as intended. Given that the *H* indices and the percentages of modal

Table 2. Correlation matrix for the main indices across languages with the corresponding Pearson's *r* coefficients. All bivariate correlations resulted significant after correcting for multiple comparisons, except for those marked with an asterisk. H corresponds to the *H* index. MN corresponds to the percentage of participants giving the modal name as a response. VC corresponds to the visual complexity rating.

		Spanish			Italian			German			Dutch (Netherlands)			Dutch (Belgium)			French			English		
		H	MN	VC	H	MN	VC	H	MN	VC	H	MN	VC	H	MN	VC	H	MN	VC	H	MN	VC
Spanish	H		−0.95	0.22	0.65	−0.61	0.22	0.53	−0.50	*0.10	0.54	−0.50	0.14	0.51	−0.47	0.13	0.60	−0.57	0.17	0.57	−0.54	*0.12
	MN			−0.19	−0.59	0.57	−0.19	−0.47	0.45	*–0.09	−0.48	0.46	−0.12	−0.47	0.45	*–0.11	−0.54	0.53	−0.14	−0.50	0.49	*–0.10
	VC				0.24	−0.22	0.95	0.18	−0.14	0.96	0.15	−0.13	0.96	*0.10	*–0.07	0.97	0.18	−0.14	0.97	0.22	−0.18	0.95
Italian	H					−0.96	0.31	0.55	−0.51	0.16	0.55	−0.51	0.19	0.55	−0.50	0.18	0.58	−0.53	0.21	0.58	−0.54	0.16
	MN						−0.28	−0.51	0.49	−0.14	−0.52	0.49	−0.17	−0.52	0.49	−0.16	−0.54	0.51	−0.18	−0.52	0.49	−0.15
	VC							0.25	−0.19	0.92	0.20	−0.16	0.92	0.17	−0.13	0.93	0.24	−0.18	0.95	0.26	−0.21	0.90
German	H								−0.96	0.13	0.68	−0.64	0.15	0.64	−0.58	0.14	0.60	−0.52	0.18	0.61	−0.56	*0.11
	MN									*–0.10	−0.64	0.63	*–0.11	−0.61	0.58	*–0.10	−0.56	0.51	−0.13	−0.55	0.53	*–0.07
	VC										*0.09	*–0.07	0.96	*0.05	*–0.03	0.97	*0.10	*–0.07	0.95	0.15	*–0.11	0.95
Dutch	H											−0.96	0.12	0.83	−0.78	*0.11	0.60	−0.54	0.13	0.62	−0.60	*0.08
(Netherlands)	MN												*–0.10	−0.80	0.79	*–0.09	−0.58	0.53	*–0.11	−0.57	0.56	*–0.07
	VC													*.08	*–0.06	0.97	0.13	*–0.09	0.95	0.18	−0.15	0.96
Dutch	H														−0.96	*0.08	0.61	−0.56	*0.09	0.58	−0.55	*0.04
(Belgium)	MN															*–0.06	−0.58	0.55	*–0.07	−0.52	0.51	*–0.02
	VC																0.12	*–0.08	0.96	0.17	−0.14	0.96
French	H																	−0.95	0.19	0.59	−0.55	*0.11
	MN																		−0.15	−0.53	0.52	*–0.08
	VC																			0.20	−0.16	0.95
English	H																				−0.95	0.17
	MN																					−0.13
	VC																					

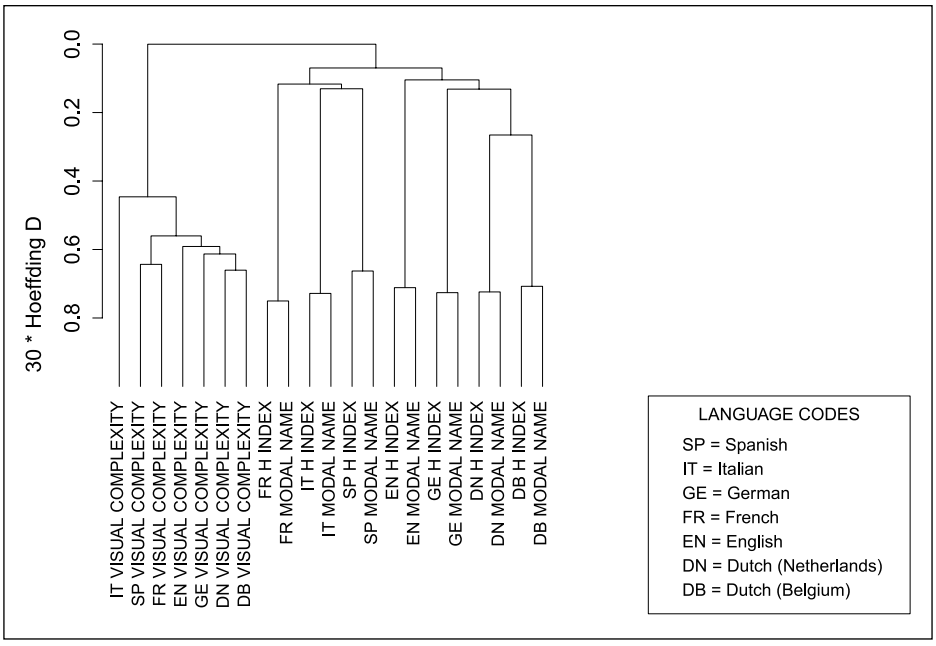

Figure 2. Results of the hierarchical cluster analysis using Hoeffding's test of independence (30**D* statistic).

responses were calculated from the filtered data (i.e. excluding idiosyncratic responses and unknown items), it is advised to take these categories into account when selecting stimulus materials. Drawings with too many "don't know" and idiosyncratic responses can easily be excluded on the basis of the information provided in the various columns of the summary files on the MultiPic website.

Because of the limitations, researchers may find MultiPic particularly relevant in combination with other existing databases. For instance, a comparison of the names of pictures with those of BOSS (Brodeur et al.,2010, 2014) can inform researchers about language differences between speakers of British and American English. Authors can also easily check whether a stimulus used in a bilingual population has the same name agreement in both languages. Also, information about word frequency, word age-of-acquisition, word prevalence and other concept-related features can easily be added to MultiPic from other sources.

In line with current standards in the field, the materials are open-access and free from copyright restrictions for non-commercial purposes, facilitating the use of the data and extensions, for instance to further samples and other languages. Interested researchers can find the materials at *http://www.bcbl.eu/databases/multipic*, where colored and grey-scaled versions of the drawings are presented in different formats together with the corresponding norms for each of the languages we tested.

Acknowledgements

The authors are thankful to Jared Checketts, Silvia Cozzi, Alessandro Ferraris, Janina Fickel, Evelyne Lagrou, Stefano Spinelli, Annelies van Wijngaarden and Lise Van der Haegen for their assistance in the data collection.

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) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research has been partially funded by grants PSI2015-65689-P and SEV-2015-0490 from the Spanish Government, and by the AThEME project funded by the European Union (grant number 613465). This project has been also supported by a 2016 BBVA Foundation Grant for Researchers and Cultural Creators awarded to the first author.

Supplementary Material

The Supplemental Material is available at: https://figshare.com/s/bf4a360ab3c90537291f.

Notes

1. Almost all researchers exclude the "don't know" responses; others exclude tip-of-the-tongue responses, or responses given by single participants. To assess the impact of our exclusions, we calculated the *H* indices for each item in each language (1) including the trials where respondents did not know the name and the trials with idiosyncratic responses and (2) including the trials with idiosyncratic responses but excluding the unknown items. The correlations between the *H* indices calculated in these ways and the *H* indices calculated from the filtered data ranged from $r = .94$ to $r = .98$ across languages, demonstrating that the filtered data agree with the unfiltered data to a great extent. For those

researches interested in the idiosyncratic responses, we have stored the unfiltered data on a publicly available website *https://figshare.com/s/bf4a360ab3c90537291f*, so that colleagues can calculate the exact metric they are interested in.

2. Interestingly, the resulting name agreement values were better (i.e. lower *H* indices and higher percentages of modal names) in Spanish than in the other languages. This is likely due to the facts that the initial list of concepts was created from a Spanish lexical database and that the artist was Spanish. Likewise, it seems reasonable to assume that the Snodgrass and Vanderwart stimulus set has some North-American bias.
3. The authors thank an anonymous reviewer for pointing them to the need of this analysis.

ORCID iD

Christos Pliatsikas http://orcid.org/0000-0001-7093-1773

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.

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*, *50*, 560–585.

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., ... Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, *10*, 344–380.

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., 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.

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*, Article e10773.

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of Standardized Stimuli (BOSS) Phase II: 930 new normative photos. *PLoS ONE*, *99*(9), Article e106953.

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*, 961–970.

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.

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*, 588–615.

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.

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavior Research Methods*, *45*, 1246–1258.

Hoeffding, W. (1948). A non-parametric test of independence. *Annals of Mathematical Statistics*, *19*, 293–325.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.

Light, R. J. (1971). Measures of response agreement for qualitative data: Some generalizations and alternatives. *Psychological Bulletin*, *76*, 365–377.

Liu, Y., Hao, M., Li, P., & Shu, H. (2011). Timed picture naming norms for Mandarin Chinese. *PLoS ONE*, *6*(1), Article e16505.

Manoiloff, L., Artstein, M., Canavoso, M., Fernández, L., & Segui, J. (2010). Expanded norms for 400 experimental pictures in an Argentinean Spanish-speaking population. *Behavior Research Methods*, *42*, 452–460.

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.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *442*, 314–324.

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*(5), Article e37527.

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*(10), Article e47802.

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.

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.

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*, 537–555.

Shannon, C. E., & Weaver, W. (1949). The mathematical theory of communication. In C. E. Shannon & W. Weaver (Eds.), *The mathematical theory of communication*. Urbana: University of Illinois Press.

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 & Memory*, *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.

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*, 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.

Viggiano, M. P., Vannucci, M., & Righi, S. (2004). A new standardized set of ecological pictures for experimental and clinical research on visual object processing. *Cortex*, *40*, 491–509.

Wang, L., Chen, C.-W., & Zhu, L. (2014). Picture norms for Chinese preschool children: Name agreement, familiarity, and visual complexity. *PLoS ONE*, *9*(3), Article e90450.

Yoon, C., Feinberg, F., Luo, T., Hedden, T., Gutchess, A. H., Chen, H. Y., . . . 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.