A stimulus set of words and pictures matched for visual and semantic similarity.

Floor de Groot[1], Thomas Koelewijn[2], Falk Huettig[3,4] and Christian N. L. Olivers[1]

[1] Department of Cognitive Psychology VU Univeristy, Amsterdam; [2] Section Ear & Hearing, department of Otolaryngology-Head and Neck Surgery and EMGO institute for Health and Care Research, VU University Medical Center, Amsterdam; [3] Max Planck Institute for Psycholinguistics, Nijmegen; [4] Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen

--- in press, *Journal of Cognitive Psychology* --

Affiliation where the research was conducted:

Department of Cognitive Psychology VU University, Amsterdam

Floor de Groot (corresponding author)

Department of Cognitive Psychology VU University, Amsterdam

Van der Boechorststraat 1, 1081 BT, Amsterdam, Netherlands

E-mail: f.de.groot@vu.nl; phone: +31 20 59 83248

Thomas Koelewijn

Otolaryngology-Head and Neck Surgery, section Ear & Hearing

VU University Medical Center

De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands

Email: t.koelewijn@vumc.nl; Phone: +31 20 44 40900

Falk Huettig

Max Planck Institute for Psycholinguistics, Nijmegen

Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen

PO Box 310, 6500 AH Nijmegen, The Netherlands

Email: Falk.Huettig@mpi.nl; Phone: +31-24-3521374

Christian N.L. Olivers

Department of Cognitive Psychology VU University, Amsterdam

Van der Boechorststraat 1, 1081 BT, Amsterdam, Netherlands

E-mail: c.n.l.olivers@vu.nl; phone: +31 20 59 88974

Acknowledgements:

This work was supported by NWO (Netherlands Organization for Scientific Research) under grant 404-10-321 to Christian N. L. Olivers and Falk Huettig. We are grateful to Niko Busch (Institute of Medical Psychology, Charité University Berlin) for sharing the POPORO stimulus set with us, Joni Stolk for her help with selecting pictures, and running part of the naming study, Nicole Boscher for her help with recruiting English native speakers and running studies, Philip Blinko and Faviola Dadis for their help with analysing the data in the English naming study. Finally, we would like to thank Albert Russel for his tremendous help with the internet studies.

Abstract

Researchers in different fields of psychology have been interested in how vision and language interact, and what type of representations are involved in such interactions. We introduce a stimulus set that facilitates such research (available online). The set consists of 100 words each of which is paired with four pictures of objects: One semantically similar object (but visually dissimilar), one visually similar object (but semantically dissimilar), and two unrelated objects. Visual and semantic similarity ratings between corresponding items are provided for every picture for Dutch and for English. In addition, visual and linguistic parameters of each picture are reported. We thus present a stimulus set from which researchers can select, on the basis of various parameters, the items most optimal for their research question.

Keywords: stimulus set, semantic similarity, visual similarity, language-vision interactions

In various fields of psychology researchers have become increasingly interested in how language and visual perception interact (see Ferreira & Tanenhaus, 2007; Hartsuiker, Huettig, & Olivers, 2011; Mishra, Srinivasan, & Huettig, 2015; Henderson & Ferreira, 2004; Myachykov, Scheepers, & Shtyrov, 2013 for reviews). In the visual attention literature, for example, there has been growing interest in the questions if, how, and when linguistic cues - and the semantic information derived from them - can guide visual selection. Specifically, one question that has been extensively investigated is if visual search is as efficient after a verbal instruction as after an instruction that uses a visual depiction of the object (Castelhano, Pollatsek, & Cave, 2008; Maxfield & Zelinsky, 2012; Schmidt & Zelinsky, 2009; Smith, Redford, Gent, & Washburn, 2005; Yang & Zelinsky, 2009; Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004; Wilschut, Theeuwes, & Olivers, 2014). Others have focused on the question whether objects that are semantically related to a search target can capture attention (e.g. a helmet when people are looking for a motor bike, Meyer, Belke, Telling, & Humphreys, 2007; Moores, Laiti, & Chelazzi, 2003; Telling, Kumar, Meyer, & Humphreys, 2010). Moreover, there has been considerable interest in whether people can distinguish visually similar objects purely on the basis of categorical information (Jonides & Gleitman, 1972; however see Duncan, 1983; White, 1977 for initial non-replications; Lupyan, 2008). Finally, in a related field of work, investigators have been interested in the interaction between different types of working memory and vision. For example, a number of researchers have investigated whether vision is biased towards objects that categorically match with an object kept in working memory (Calleja & Rich, 2013; Dombrowe, Olivers, & Donk, 2010), while others have looked at the influence of the retention of verbal material on visual biases (Olivers, Meijer, & Theeuwes, 2006; Soto & Humphreys, 2007; Sun, Shen, Shaw, Cant, & Ferber, 2015). Thus language-vision interactions have become a hot topic in the visual attention literature.

Also in the field of psycholinguistics researchers have been interested in the interaction between language and vision. Unlike in visual attention literature, here the focus is on the processing of linguistic expressions. For example, with the visual world paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; for a recent review, see Huettig, Rommers, & Meyer, 2011) it has been found that participants are more likely to fixate pictures that are semantically related to words they hear, compared to unrelated pictures or words (Dunabeitia, Aviles, Afonso, Scheepers, & Carreiras, 2009; Huettig & Altmann, 2005; Yee & Sedivy, 2006; Yee, Overton, & Thompson-Schill, 2009). Huettig, Quinlan, McDonald, and Altmann (2006) further observed that several corpus-based measures of word semantics (latent semantic analysis, Landauer & Dumais, 1997; contextual similarity, McDonald, 2000) each substantially predicted fixation behaviour. Besides these semantic mapping effects, researchers have also found visual mapping effects. For example, participants are more likely to shift their overt attention to a picture of a cable during the acoustic unfolding of the word "snake" (an effect of shape similarity, Dahan & Tanenhaus, 2005; Huettig & Altmann, 2007; Rommers, Meyer, Praamstra, & Huettig, 2013; Rommers, Meyer, & Huettig, 2015), with similar effects having been observed for colour (Huettig & Altmann, 2011) or conceptually related shape (i.e., a slice of pizza activating the round shape of a whole pizza, Yee, Huffstetler, & Thompson-Schill, 2011). In sum, these studies show that eye gaze during language-vision interactions is influenced by matches between visual input and knowledge retrieved from different levels of representation.

Finally, developmental psychologists have investigated the emergence of perceptual and conceptual knowledge during childhood and in doing so have made use of word-picture stimulus sets. For example, in the so-called preferential looking paradigm (Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987) children are presented with two side-by-side displays containing pictures while listening to linguistic input. Looking preferences are then taken as an indicator of the extent to which the child has processed the input at both semantic and visual levels (Johnson, McQueen, & Huettig, 2011; Johnson & Huettig, 2011; Mani, Johnson, McQueen, & Huettig, 2013; Styles & Plunkett, 2009). So, in many fields within psychology language-vision interactions are a topic of interest.

A new stimulus set

In the past various different stimulus sets have proven to be sufficient to answer several important questions in relation to language-vision interactions. However, we believe that many other important questions have been left unexplored because of current limitations inherent to the existing stimulus sets. For example, studies have either focused on semantic relationships between words and pictures (e.g. Belke, Humphreys, Watson, Meyer, & Telling, 2008; Huettig & Altmann, 2005; Kovalenko, Chaumon, & Busch, 2012; Moores et al., 2003; Telling et al., 2010; Yee & Sedivy, 2006), or on visual relationships (e.g. Dahan & Tanenhaus, 2005; Huettig & Altmann, 2007; Rommers et al., 2013; Rommers et al., 2015) with only few having directly compared visual and semantic relationships (e.g. Huettig and McQueen, 2007). Important commonalities or differences between these relationships can therefore not be explored. Sets that do contain both types of relationship have been small, leading to limited reliability when for instance used in eye movement, EEG and MEG studies that in general use large numbers of trials. In the past, researchers have dealt with this by comparing conditions between rather than within participants (e.g. Huettig & McQueen, 2007). This has some drawbacks, predominantly a lack of power and efficiency. A second solution to the small set sizes has been to allow for the recurrence of specific items (e.g. Moores et al., 2003; Telling et al., 2010). However, stimulus repetition might introduce unwanted effects of learning and familiarity.

Another limitation of the existing stimulus sets is that many use (line) drawings (e.g. Belke et al., 2008; e.g. Huettig & McQueen, 2007; Rommers et al., 2013; Telling et al., 2010). Drawings, however, are reduced depictions of reality as they contain less detailed information than real objects, and are less rich in terms of the number of perceptual dimensions, as they often lack colour and depth cues. The use of drawings therefore possibly reduces not only effect size, but also ecological validity, leading to underestimations of potentially important psychological mechanisms. Also, the use of realistic objects allows for the insertion of these objects in complete scenes, facilitating research into object-based attention in a wider context (e.g. Hwang, Wang, & Pomplun, 2011). There are other, less prominent limitations for individual sets. For example in some studies semantic similarities were only assumed, and not independently normed and rated by a separate set of observers (e.g. Moores et al., 2003).

Given these limitations, we sought to develop a more extensive and independently normed stimulus set. This new set consists of 100 combinations of one word with four different pictures (from now on called *trials*). Each word is combined with one picture that is semantically but not visually related (semantically related pictures), one picture that is visually but not semantically related (visually related pictures) and two pictures that are neither semantically nor visually related (unrelated pictures). Visual and semantic similarity has been rated by naïve participants. To increase ecological validity, and following other existing stimulus sets (e.g. Adlington, Laws, & Gale, 2009; Brady, Konkle, Alvarez, & Oliva, 2008; Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010; Konkle, Brady, Alvarez, & Oliva, 2010; Kovalenko et al., 2012; Moreno-Martinez & Montoro, 2012; Viggiano, Vannucci, & Righi, 2004), we use photographs of real life objects. Furthermore, by including two different types of relationship, one type of relationship can serve as a direct control for the other (in addition to the neutral control pictures) thus showing that not just any relationship is being affected. For example, tasks manipulating semantic discrimination should interact little with visual similarity; while conversely, in tasks stressing visual discrimination semantic similarity could serve as a control. Finally, several visual and linguistic parameters have been estimated so that researchers can select the stimuli that are optimal for their purposes.

An overview of the trials is given in Appendix A (Dutch) and Appendix B (English). Most words have very similar meaning in Dutch and English, except when indicated otherwise. The stimulus set and the corresponding data file containing individual norms and ratings are available via Open Science framework (https://osf.io/6vdys/?view_only=541cd6d599a74f4a99c7411e8cd60b4a).

Method

Picture Selection and Editing

As a starting point we used a subset of the POPORO set developed by Kovalenko et al. (2012), which contains semantic relationships between pictures. We extended this set with both semantic and visual relationships, as inspired by earlier work (Huettig & McQueen, 2007; Rommers et al., 2013). Pictures were mostly taken from the Hemera Photo-Object database (Volumes I, II, III), which has been used for other published stimulus sets (e.g. Brady et al., 2008; Konkle et al., 2010; Kovalenko et al., 2012), extended with a few pictures from the public domain. Each picture in the set is unique and conceptual repetition between pictures was avoided. However, some words return as a picture (and vice versa). These trials are less suitable for studies where all concepts need to be unique across modalities, and therefore these trials have been marked in the data file. Several pilot naming studies were performed to see if people knew what the objects on the pictures were. Objects were complete (so not just parts) and were placed in orientations that one typically encounters from normal viewpoints (so not upside down for example) in the middle of a 400 x 400 pixels picture with a transparent background, and saved in the lossless png format (using Adobe Photoshop CS5, version 5.5; available from: http://www.adobe.com, and PngOptimizer, version 2.3; available from: http://psydk.org/PngOptimizer.php, to further optimize the file structure and remove any unnecessary metadata). All pictures are in colour.

Rating and Naming studies

In separate studies we asked both native Dutch and native English speaking participants to indicate the visual and semantic similarity between a picture and the object the word was referring to, and to name the pictures.

Participants

In total 181 participants took part, of which 60 in the semantic rating study, 61 in the visual rating study, and 60 in the naming study. Of these, 30 native Dutch speakers took part in the Dutch version of the semantic rating study (6 males, average age 22.1, range 18-68), while 31 native Dutch speakers participated in the visual rating study (8 males, mean age 20.9, range 18-28). Thirty native English speakers took part in the English version of the semantic rating study (12 males, mean age 22.4, range 17-35) and another thirty in the visual rating study (15 males, mean age 24.5, range 17-47). This is after six participants had been replaced: In the Dutch semantic rating study one participant was left out before data inspection because of a self-expressed lack of motivation and misunderstanding of the instructions. In the Dutch and English visual rating studies respectively four participants and one participant were omitted after data inspection because of random responding (resulting in an average rating of approximately 5 for all picture groups). Another 30 Dutch native speakers participated in the Dutch version of the naming study (11 males, mean age 22.8, range 18-43) and 30 English native speakers participated in the English version (9 males, mean age 22.6, range 18-34). Two of the Dutch participants were replaced, one because of technical failure and the other because of colour blindness (as was only found out after the experiment). In the English-speaking group, participants had various nationalities (e.g., Great-Britain, USA, Canada, Australia, New Zealand). All participants reported not to be colour blind and they had no history of dyslexia or any other language disorder. All indicated to be native speakers in the language they were tested in. Participants received course credits or were paid for their participation, and none took part in more than one study.

Procedure

In each study half of the participants were tested via the internet through a dedicated testing platform (Max Planck Institute, Nijmegen), whereas the other half was tested in the lab. The procedure was exactly the same in both cases. Each task started with demographic questions (age, bilingualism, educational level, native language, handedness, number of other languages, profession, and sex). Stimuli were presented on a grey background (RGB 230,230,230). During the task participants could not skip a trial or go back to the previous one. Reaction times were not measured so participants were told they could take as much time as they needed.

Rating study. On each trial participants saw one written word and one picture. In the Dutch rating studies there were 520 pictures paired with 130 words and in the English similarity studies 480 pictures paired with 120 words[1]. Each word was repeated four times with different pictures, and pictures were presented in a random order. Participants rated the similarity between the object in the picture and the object the word was referring to. For the semantic rating participants answered the question of how much the objects "had to do with each other", whereas for the visual rating they answered how much the objects "looked alike". Note that these are thus also our definitions of the reported similarities. They made this judgment by clicking on an eleven-point scale that ranged from 0 (no similarity at all) to 10 (very similar). Participants were instructed to only focus on semantics or visual appearance and to ignore the other dimension. They could also indicate if they did not know the object or word, but were instructed to do so only after deliberate thinking.

Naming study. Participants were instructed to name 480 pictures in the Dutch naming study, and 564 pictures in the English study (see Footnote 1). Pictures were presented sequentially in a random order in the middle of the screen. Participants were instructed to identify the displayed object as briefly and unambiguously as possible, using the first name that came to mind, by typing the name on the keyboard. However, they could indicate that they did not know the object, did not know the name or were in a tip of a tongue state (i.e. they know the object and the name but are unable to recall the name at the moment) by typing in abbreviations,

[1] We initially started off with a larger set than the 100 trials reported here, from which we then selected the best trials on the basis of factors such as the semantic and visual similarity ratings. Moreover, some studies included pictures that were intended for other experiments. For these reasons the studies contained more trials than the 100 included in the described set.

respectively “OO”, “OW”, “TOT” in the Dutch version, and “DKO”, “DKN”, “TOT” in the English version. However, participants were told that in such situations to take as much time as they needed to recall the name.

Measures provided in the data file

Ratings of semantic and visual similarity between word and pictures

The main objective of this article was to provide a stimulus set that contained pictures that were either semantically (but not visually) or visually (but not semantically) related to a word. In addition, as a control condition, we added pictures that were neither semantically nor visually related to the word. The strength of the semantic and visual relationship of each individual picture-word pair is given in the data file. Although the majority of the participants indicated that they knew all words and pictures, several indicated that some words or pictures were unknown to them, resulting in less than 0.6% missing data in all studies. To check if our main goal was achieved, we conducted for each rating study a repeated measures ANOVA with the rating data as a dependent variable. This showed that the picture groups (i.e. semantically related, visually related and the average of unrelated pictures) differed significantly in terms of their semantic and visual relationship strength to the word, for the Dutch semantic rating $F_1(1.493,43.290) = 431.068$, $p < 0.001$, $\eta^2_G = 0.891$, and $F_2(1.298,128.537) = 1939.772$, $p < 0.001$, $\eta^2_G =0.927$, for the Dutch visual rating $F_1(1.382, 41.460) = 320.892$, $p < 0.001$, $\eta^2_G = 0.727$, and $F_2(1.554, 153.871) = 654.275$, $p < 0.001$, $\eta^2_G = 0.809$, for the English semantic rating $F_1(1.392, 40.366) = 332.106$, $p < 0.001$, $\eta^2_G = 0.823$, and $F_2(1.155,114.327) = 751.348$, $p < 0.001$, $\eta^2_G =0.840$, and for the English visual rating $F_1(1.134, 32.887) = 208.637$, $p < 0.001$ $\eta^2_G = 0.728$, and $F_2(1.609, 159.249) = 533.456$, $p < 0.001$ $\eta^2_G = 0.773$, (Greenhouse-Geisser corrected values are reported). As Table 1 shows, and as would be expected, visual similarity was rated the highest for the visually related pictures, while semantic similarity was rated the highest for the semantically related pictures, for both Dutch and English native speakers. The relevant inferential statistics are also displayed in Table 1. For the English study, the semantic and visual ratings of the pictures did not correlate, $r = -0.066$, $p = 0.189$, while, if anything the correlation for the Dutch study was negative, $r = -0.102$, $p = 0.041$. Thus, there appeared no positive relationship between semantic and visual similarity. Moreover, the ratings correlated very strongly for both languages, with $r = 0.961$, $p < 0.001$ for the semantic similarity studies, and $r = 0.966$, $p < 0.001$, for the visual similarity studies, indicating that the ratings can be safely pooled across language groups. Furthermore, inter-rater reliability was very high, with Cronbach's alpha exceeding 0.97 for all rating studies, and an average correlation between raters of $r = 0.771$ (with a range from 0.494 to 0.932) and $r = 0.710$ (range: 0.349 - 0.895), for Dutch and English semantic ratings, $r = 0.625$ (range: 0.204 – 0.790) and $r = 0.637$ (range: 0.210 – 0.842) for Dutch and English visual ratings.

"(Table 1 about here)"

Linguistic parameters

Before the proper naming study, each picture was given an intended name on the basis of pilot studies. However, this intended name did not always converge with the dominant name in the final naming study. In that case we changed the intended name to the dominant name of the proper naming study unless (1) raters collectively named the object officially not correct (e.g.

used a colloquialism), (2) the intended name contained an addition that defined the object better, (3) if the intended name was more specific to the object (e.g., euro vs. coin), and (4) the dominant name contained a random detail. In addition, (5) the Dutch word was preferred over the English one in the Dutch study (as here speakers occasionally provided an English name), whereas in the English study US English was preferred over British English. If the intended name and the dominant name were synonyms, we chose the name with the highest word frequency in SUBTLEX-NL (Keuleers, Brysbaert, & New, 2010) or SUBTLEX-US (Brysbaert & New, 2009) for respectively the Dutch and the English study. The same rules were applied when multiple names were given equally often. Due to this procedure and some idiosyncrasies inherent to the different languages, there is not always a direct mapping between the intended name in the Dutch stimulus set and in the English stimulus set. Items were removed when the intended names of the pictures, and the words of the corresponding trials overlapped in the first syllable (i.e. phonological competitors), but on a few trials they do share the first one or two letters. These trials are indicated in the data file. The intended names and dominant names are also listed in the data file, and it is pointed out when they differ (plus the reason why they differ).

Several linguistic properties were computed, for one the *naming agreement*. One way to define naming agreement is as the percentage of participants that gave the dominant name (Brodeur et al., 2010; Moreno-Martinez & Montoro, 2012; Severens, Van Lommel, Ratinckx, & Hartsuiker, 2005; Snodgrass & Vanderwart, 1980). Another way to define naming agreement is as the percentage of people that gave the intended name (Adlington et al., 2009; Bates et al., 2003). Both measures provide different, but important information and were therefore computed for each picture. To compute the naming agreement typos and articles were first removed. In the data file we report for each picture separately the percentages of participants that indicated that they did not know the word or the object and the percentages of participants that were in a tip of the tongue state (over the whole set this was less than 3.65% and 5.18 % of the responses for respectively the Dutch and English group). Only the first word was considered in cases where, despite the instructions, participant gave two names. Clarifications were also removed, (e.g., if participants responded “wheel of a bike” the part “of a bike” was removed). When multiple names were given equally often, this is indicated in the data file.

However, it could be argued that both the intended and the dominant naming agreement are an underestimation of the real naming agreement, as morphological variants, adjectives, elaborations, abbreviations and synonyms were considered as conceptually different. Therefore, we also looked at naming agreement when these variants were considered as conceptually the same, hereafter called the *lenient naming agreement* (Bates et al., 2003; Severens et al., 2005). A native speaker checked if the name a participant gave differed from the intended name in one of the following categories: plural, diminutive (only in Dutch), common abbreviation (e.g., TV instead of television), unnecessary adjectives or elaborations, and/or synonyms. An adjective or elaboration was considered unnecessary if it described some characteristic of the object better (i.e. yellow bag or ugly trousers), but did not change the object (as in figure skate instead of skate). The definition of a synonym was if two objects were interchangeable (i.e. a cup can be considered as a synonym of a mug and vice versa, but a Dalmatian is a dog, but a dog is not a Dalmatian). This was also done for colloquialisms (e.g. in the Dutch naming study a “stoplicht” (which means stop light) was considered a colloquialism for “verkeerslicht” (traffic light), and “plopper” a colloquialism for “ontstopper” (plunger)).

Besides naming agreement we also looked at word frequency measures and age-of-acquisition of the intended name. Word frequency is reported in two ways. Firstly, by using *frequency per million words* (fpmw) from the SUBTLEX-NL (Keuleers et al., 2010) and SUBTLEX-US (Brysbaert & New, 2009) for respectively the Dutch and English intended names. This is a measure that indicates how frequent the name occurs in the database per million words. Secondly, we used the recently proposed *Zipf scale* (van Heuven, Mandera, Keuleers, & Brysbaert, 2014). This is log transformation of the fpmw measure (i.e. log10(fpmw)+3) ranges from approximately 1 to 7, with 1 indicating low frequency. *Age of acquisition* (in years) of each intended picture name is also provided in the data file, with the help of several norms (for Dutch Brysbaert, Stevens, De Deyne, Voorspoels, & Storms, 2014; and English Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012 extended to a total of 51,715 words by Brysbaerts group, http://crr.ugent.be/archives/806). Not all intended names were present in the databases and are therefore marked as missing in the data file (for respectively the Dutch and English group 12.5% and 31.3% of the intended names were missing for word frequency and 20.4 % and 35.5% for age-of-acquisition). Table 2 presents the descriptive statistics of these measures.

“(Table 2 about here)”

Visual parameters

For each picture, we computed luminance, within-object contrast, visual complexity and object size (see Table 3 for descriptive statistics). *Total luminance* was defined as how luminant the picture is as a whole, and calculated by summing the RGB grey scale values for all pixels in the object (i.e. background was not taken into the sum). To arrive at a graspable number, we divided this value by a constant, namely the total number of pixels in the 400 x 400 picture area (i.e.160,000, which is the same for every picture). A low value thus means that the object is darker (0 is black, 255 is white). RGB was chosen over actual luminance (as would be measured by a photometer) since actual luminance is likely to vary considerably for different monitors and settings. Note thought that actual luminance correlates one-to-one with the RGB pixel value. In addition, we calculated the *relative luminance*, i.e. the average luminance per object pixel, again expressed as RGB grey scale value. This was calculated by dividing the sum of the RGB grey scale values by the number of pixels in the object. The average luminance thus corrects for object size.

Within-object contrast was computed as the standard deviation of the RGB grey scale value. Again we computed both total and average contrast. *Total within-object contrast* was calculated by dividing the standard deviation of the RGB grey scale value by the constant total number of pixels in the 400 x 400 picture area (i.e. 160,000), while *average within-object contrast* was divided by the number of pixels in the object (thus correcting for object size).

To provide an indication of *visual complexity* of each object, we took the file size of each grey scale version of the picture. Donderi and McFadden (2005) have shown that the size of ZIP-compressed files is a reasonable predictor for subjective visual complexity judgments. Our files used PNG compression, which is virtually identical to ZIP compression. Both are lossless compression methods that use the same core algorithm (Deutsch, 1996). Finally, average object size was calculated in two ways. Firstly, as *overall surface size* that was simply defined as the total number of pixels in the object (hence corresponding to the object's total surface area). Finally, we report the *radius of the smallest fitting circle* that can be drawn around the object. This measure can be taken as an indicator of the spatial spread of the object.

“(Table 3 about here)”

Discussion

The introduced stimulus set has several advantages compared to previous sets, for one, and most importantly, it does not only contain semantic relationships between words and pictures, but also visual relationships, in larger numbers than in previous sets (e.g. Huettig & McQueen, 2007). The inclusion of both semantic *and* visual relationships makes it possible to directly compare different types of similarities, and explore their independent contributions to behavioural outcomes, such as response times and eye movements, but also how such perceptual and semantic similarities shape the activity and plasticity of different brain regions (e.g. Erez & Yovel, 2014). Neuropsychological research into deficits such as agnosia could also benefit from this stimulus set. For example, Humphreys and Riddoch (2003) reported several subpopulations within a patient group suffering from category-specific agnosia: While some people were more impaired in associative/functional knowledge, others suffered mainly from visual knowledge problems. The introduced set could be used to further explore such dichotomies. This set may also be useful for cross-lingual studies, and for studies investigating bilingualism (e.g. Jones et al., 2012; Verhoef, Roelofs, & Chwilla, 2010), given that the set has been named and rated by both Dutch and English native speakers.

A final advantage of the introduced set is that it consists of photos of real-life objects, rather than line-drawings. This will increase the ecological generalizability. Recently, there has been a tendency to develop ecologically valid stimulus sets. However, there has been a lack of a stimulus set that contained both semantic *and* visual relationships between words and pictures. With the introduced set we will fill this gap in the literature.

We point out that the visually related pictures were also rated as *semantically* more related to the word, compared to the unrelated pictures (see Table 1). Conversely, the semantically related pictures were also regarded as *visually* more related to the word, compared to the unrelated objects. There may be several reasons for this. Instructions emphasized that participants had to pay attention to only one dimension (either visual or semantic), but it may have been difficult to ignore the other relationship. In addition, it is inherently difficult if not impossible to completely separate semantic and visual representations. In general, sensory representations can be regarded as part of the conglomerate of representations that comprise knowledge about a certain object (see for example the grounded /embodied cognition literature, e.g. Barsalou, 2008; Barsalou, 2010; Kiefer, 2001). Knowing what a cat is includes knowing what it looks like. The other way around, visual relationships may sometimes be taken unconsciously as a basis for semantic categorization even though we explicitly instructed participants to ignore possible visual similarities in the semantic rating task. Moreover, both a banana and a canary could be seen as part of a "category of yellow objects" and may thus be regarded as semantically related by an observer. Having said this, the visually related pictures received much higher ratings on the visual ratings scale than the semantically related pictures, whereas on the semantic scale the semantically related pictures received also much higher ratings than the visually related pictures. Thus, semantic and visual relationships are clearly distinguishable within the set. We therefore have no doubt that the current set serves the purpose of measuring different forms of language-vision interactions.

References

Adlington, R. L., Laws, K. R., & Gale, T. M. (2009). The Hatfield Image Test (HIT): a new picture test and norms for experimental and clinical use. *J Clin Exp Neuropsychol, 31*(6), 731-753. doi: 10.1080/13803390802488103

Barsalou, L. W. (2008). Grounded cognition. *Annu Rev Psychol, 59*, 617-645. doi: 10.1146/annurev.psych.59.103006.093639

Barsalou, L. W. (2010). Grounded cognition: past, present, and future. *Topics in Cognitive Science, 2*(4), 716-724. doi: 10.1111/j.1756-8765.2010.01115.x

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., . . . Pléh, C. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review, 10*(2), 344-380.

Belke, E., Humphreys, G. W., Watson, D. G., Meyer, A. S., & Telling, A. L. (2008). Top-down effects of semantic knowledge in visual search are modulated by cognitive but not perceptual load. *Percept Psychophys, 70*(8), 1444-1458. doi: 10.3758/PP.70.8.1444

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proc Natl Acad Sci U S A, 105*(38), 14325-14329. doi: 10.1073/pnas.0803390105

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

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav Res Methods, 41*(4), 977-990. doi: 10.3758/BRM.41.4.977

Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychol (Amst), 150*, 80-84.

Calleja, M. O., & Rich, A. N. (2013). Guidance of attention by information held in working memory. *Atten Percept Psychophys, 75*(4), 687-699. doi: 10.3758/s13414-013-0428-y

Castelhano, M. S., Pollatsek, A., & Cave, K. R. (2008). Typicality aids search for an unspecified target, but only in identification and not in attentional guidance. *Psychonomic Bulletin & Review, 15*(4), 795-801. doi: 10.3758/pbr.15.4.795

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cogn Psychol*.

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review, 12*(3), 453-459. doi: 10.3758/BF03193787

Deutsch, P. L. (1996). DEFLATE Compressed Data Format Specification version 1.3. Retrieved June 25, 2014, from http://tools.ietf.org/html/rfc1951

Dombrowe, I., Olivers, C. N. L., & Donk, M. (2010). The time course of working memory effects on visual attention. *Visual Cognition, 18*(8), 1089-1112. doi: 10.1080/13506281003651146

Donderi, D. C., & McFadden, S. (2005). Compressed file length predicts search time and errors on visual displays. *Displays, 26*(2), 71-78. doi: 10.1016/j.displa.2005.02.002

Dunabeitia, J. A., Aviles, A., Afonso, O., Scheepers, C., & Carreiras, M. (2009). Qualitative differences in the representation of abstract versus concrete words: evidence from the visual-world paradigm. *Cognition, 110*(2), 284-292. doi: 10.1016/j.cognition.2008.11.012

Duncan, J. (1983). Category effects in visual search: A failure to replicate the "oh-zero" phenomenon. *Percept Psychophys, 34*(3), 221-232.

Erez, Y., & Yovel, G. (2014). Clutter modulates the representation of target objects in the human occipitotemporal cortex. *J Cogn Neurosci, 26*(3), 490-500. doi: 10.1162/jocn_a_00505

Ferreira, F., & Tanenhaus, M. K. (2007). Introduction to the special issue on language–vision interactions. *Journal of Memory and Language, 57*(4), 455-459. doi: 10.1016/j.jml.2007.08.002

Hartsuiker, R. J., Huettig, F., & Olivers, C. N. (2011). Visual search and visual world: interactions among visual attention, language, and working memory (introduction to the special issue). *Acta Psychol (Amst), 137*(2), 135-137. doi: 10.1016/j.actpsy.2011.01.005

Henderson, J. M., & Ferreira, F. (2004). *The Interface of Language, Vision, and Action*. New York: Psychology Press.

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: semantic competitor effects and the visual world paradigm. *Cognition, 96*(1), B23-32. doi: 10.1016/j.cognition.2004.10.003

Huettig, F., & Altmann, G. T. M. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition, 15*(8), 985-1018. doi: 10.1080/13506280601130875

Huettig, F., & Altmann, G. T. M. (2011). Looking at anything that is green when hearing "frog": how object surface colour and stored object colour knowledge influence language-mediated overt attention. *Q J Exp Psychol (Hove), 64*(1), 122-145. doi: 10.1080/17470218.2010.481474

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language, 57*(4), 460-482. doi: 10.1016/j.jml.2007.02.001

Huettig, F., Quinlan, P. T., McDonald, S. A., & Altmann, G. T. (2006). Models of high-dimensional semantic space predict language-mediated eye movements in the visual world. *Acta Psychol (Amst), 121*(1), 65-80. doi: 10.1016/j.actpsy.2005.06.002

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: a review and critical evaluation. *Acta Psychol (Amst), 137*(2), 151-171. doi: 10.1016/j.actpsy.2010.11.003

Humphreys, G. W., & Riddoch, M. J. (2003). A case series analysis of “category-specific” deficits of living things: The HIT account. *Cogn Neuropsychol, 20*(3-6), 263-306.

Hwang, A. D., Wang, H. C., & Pomplun, M. (2011). Semantic guidance of eye movements in real-world scenes. *Vision Res, 51*(10), 1192-1205. doi: 10.1016/j.visres.2011.03.010

Johnson, E. K., & Huettig, F. (2011). Eye movements during language-mediated visual search reveal a strong link between overt visual attention and lexical processing in 36-month-olds. *Psychol Res, 75*(1), 35-42. doi: 10.1007/s00426-010-0285-4

Johnson, E. K., McQueen, J. M., & Huettig, F. (2011). Toddlers' language-mediated visual search: they need not have the words for it. *Q J Exp Psychol (Hove), 64*(9), 1672-1682. doi: 10.1080/17470218.2011.594165

Jones, O. P., Green, D. W., Grogan, A., Pliatsikas, C., Filippopolitis, K., Ali, N., . . . Price, C. J. (2012). Where, when and why brain activation differs for bilinguals and monolinguals during picture naming and reading aloud. *Cereb Cortex, 22*(4), 892-902. doi: 10.1093/cercor/bhr161

Jonides, J., & Gleitman, H. (1972). A conceptual category effect in visual search: O as letter or as digit. *Percept Psychophys, 12*(6), 457-460.

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: a new measure for Dutch word frequency based on film subtitles. *Behav Res Methods, 42*(3), 643-650. doi: 10.3758/BRM.42.3.643

Kiefer, M. (2001). Perceptual and semantic sources of category-specific effects: event-related potentials during picture and word categorization. *Mem Cognit, 29*(1), 100-116.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *J Exp Psychol Gen, 139*(3), 558-578. doi: 10.1037/a0019165

Kovalenko, L. Y., Chaumon, M., & Busch, N. A. (2012). A Pool of Pairs of Related Objects (POPORO) for Investigating Visual Semantic Integration: Behavioral and Electrophysiological Validation. *Brain Topography, 25*(3), 272-284. doi: 10.1007/s10548-011-0216-8

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behav Res Methods, 44*(4), 978-990.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol Rev, 104*(2), 211.

Lupyan, G. (2008). The conceptual grouping effect: categories matter (and named categories matter more). *Cognition, 108*(2), 566-577. doi: 10.1016/j.cognition.2008.03.009

Mani, N., Johnson, E., McQueen, J. M., & Huettig, F. (2013). How yellow is your banana? Toddlers' language-mediated visual search in referent-present tasks. *Dev Psychol, 49*(6), 1036-1044. doi: 10.1037/a0029382

Maxfield, J. T., & Zelinsky, G. J. (2012). Searching through the hierarchy: How level of target categorization affects visual search. *Visual Cognition, 20*(10), 1153-1163. doi: 10.1080/13506285.2012.735718

McDonald, S. (2000). *Environmental determinants of lexical processing effort*. Unpublished doctoral dissertation. University of Edinburgh, Scotland. Retrieved from http://www.inf.ed.ac.uk/publications/thesis/online/IP000007.pdf

Meyer, A. S., Belke, E., Telling, A. L., & Humphreys, G. W. (2007). Early activation of object names in visual search. *Psychonomic Bulletin & Review, 14*(4), 710-716.

Mishra, R. K., Srinivasan, N., & Huettig, F. (Eds.). (2015). *Attention and Vision in Language Processing.* Berlin: Springer.

Moores, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nat Neurosci, 6*(2), 182-189. doi: 10.1038/nn996

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

Myachykov, A., Scheepers, C., & Shtyrov, Y. Y. (2013). Interfaces between language and cognition. *Front Psychol, 4*. doi: 10.3389/fpsyg.2013.00258

Olivers, C. N., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: visual working memory content affects visual attention. *J Exp Psychol Hum Percept Perform, 32*(5), 1243-1265. doi: 10.1037/0096-1523.32.5.1243

Rommers, J., Meyer, A. S., & Huettig, F. (2015). Verbal and nonverbal predictors of language-mediated anticipatory eye movements. *Attention, Perception, & Psychophysics, 77*(3), 720-730.

Rommers, J., Meyer, A. S., Praamstra, P., & Huettig, F. (2013). The contents of predictions in sentence comprehension: activation of the shape of objects before they are referred to. *Neuropsychologia, 51*(3), 437-447. doi: 10.1016/j.neuropsychologia.2012.12.002

Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *Q J Exp Psychol (Hove), 62*(10), 1904-1914. doi: 10.1080/17470210902853530

Severens, E., Van Lommel, S., Ratinckx, E., & Hartsuiker, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychol (Amst), 119*(2), 159-187. doi: 10.1016/j.actpsy.2005.01.002

Smith, J. D., Redford, J. S., Gent, L. C., & Washburn, D. A. (2005). Visual search and the collapse of categorization. *J Exp Psychol Gen, 134*(4), 443-460. doi: 10.1037/0096-3445.134.4.443

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of experimental psychology: Human learning and memory, 6*(2), 174.

Soto, D., & Humphreys, G. W. (2007). Automatic guidance of visual attention from verbal working memory. *J Exp Psychol Hum Percept Perform, 33*(3), 730-737. doi: 10.1037/0096-1523.33.3.730

Styles, S. J., & Plunkett, K. (2009). How do infants build a semantic system? *Language and Cognition, 1*(1), 1-24.

Sun, S. Z., Shen, J., Shaw, M., Cant, J. S., & Ferber, S. (2015). Automatic capture of attention by conceptually generated working memory templates. *Attention, Perception, & Psychophysics*, 1-7.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632-1634.

Telling, A. L., Kumar, S., Meyer, A. S., & Humphreys, G. W. (2010). Electrophysiological evidence of semantic interference in visual search. *J Cogn Neurosci, 22*(10), 2212-2225.

van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology, 67*(6), 1176-1190.

Verhoef, K. M., Roelofs, A., & Chwilla, D. J. (2010). Electrophysiological evidence for endogenous control of attention in switching between languages in overt picture naming. *J Cogn Neurosci, 22*(8), 1832-1843.

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*(3), 491-509. doi: 10.1016/s0010-9452(08)70142-4

White, M. J. (1977). Identification and categorization in visual search. *Mem Cognit, 5*(6), 648-657.

Wilschut, A., Theeuwes, J., & Olivers, C. N. (2014). Priming and the guidance by visual and categorical templates in visual search. *Front Psychol, 5*, 148. doi: 10.3389/fpsyg.2014.00148

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Res, 44*(12), 1411-1426. doi: 10.1016/j.visres.2003.11.024

Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. *Vision Res, 49*(16), 2095-2103. doi: 10.1016/j.visres.2009.05.017

Yee, E., Huffstetler, S., & Thompson-Schill, S. L. (2011). Function follows form: activation of shape and function features during object identification. *J Exp Psychol Gen, 140*(3), 348-363. doi: 10.1037/a0022840

Yee, E., Overton, E., & Thompson-Schill, S. L. (2009). Looking for meaning: eye movements are sensitive to overlapping semantic features, not association. *Psychon Bull Rev, 16*(5), 869-874. doi: 10.3758/PBR.16.5.869

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *J Exp Psychol Learn Mem Cogn, 32*(1), 1-14. doi: 10.1037/0278-7393.32.1.1

Appendix A: Trials in Dutch. The last four columns are the intended names of the pictures. The words in the Dutch stimulus set are a direct translation from the English set, except the one marked with an asterisk. The intended names in the Dutch stimulus set are however not always a direct translation of the intended names in the English stimulus set.

Trial number	Word	Semantically Related Picture	Visually Related Picture	Unrelated Picture1	Unrelated Picture2
1	aardappel	maïskolf	bowlingbal	batterij	potlood
2	asbak	pijp	jojo	dennenappel	rozen
3	bad	kraan	slee	honkbalhandschoen	kwast
4	badpak	slippers	kruik	nietjes	koffiezetapparaat
5	bakblik	taart	cassettebandje	schaats	stropdas
6	bal	voetbalschoenen	tomaat	waterpijp	schep
7	ballon	cadeau	kers	kaasschaaf	koffiebonen
8	banaan	aap	kano	tamboerijn	hoed
9	basketbal	badmintonracket	kokosnoot	steekwagen	stanleymes
10	beker	vork	garen	pen	duikbril
11	blokken	hobbelpaard	toffee	saxofoon	beer
12	bolhoed	wandelstok	sinaasappelpers	vlees	olifant
13	boom	bijl	wc-borstel	magnetron	magneet
14	boor	rolmaat	pistool	ballon	bureaustoel
15	boot	anker	klomp	chocolade	honkbal bal
16	bot	puppy	halter	bezem	narcis
17	bril	telescoop	bh	scheermes	sleutel
18	buggy	flesje	tractor	sneeuwschuiver	zonnebloem
19	cd	diskette	reddingsboei	holster	duimstok
20	drol	luier	ijsje	kan	pompoen
21	druiven	wijnglas	biljartballen	kettingzaag	bel
22	drumstel	elektrische gitaar	weegschaal	katapult	speelkaarten

23	ei	haan	wol	tandenborstel	xylofoon
24	fles	kurk	kegel	broek	kerstbal
25	fluit	harp	deegroller	badeend	ton
26	garde	kom	borstel	speldenkussen	pillen
27	gloeilamp	lichtschakelaar	avocado	arend	mandje
28	handboeien	politiepet	trappers	scheerkwast	hijskraan
29	handboog	kanon	ijzerzaag	ananas	nagellak
30	hark	heggenschaar	spatel	dynamiet	zwemband
31	helm	motor	mango	blik	ijshoorn
32	hijskraan	cementwagen	giraffe	kopje	bramen
33	hoefijzer	zadel	koptelefoon	teddybeer	brie
34	ipod	radio	kompas	watermeloen	flesopener
35	jerrycan	benzinepomp	paprika	ventilator	telefoon
36	kleerhanger	kapstok	triangel	luidspreker	driewieler
37	klokhuis	aardbei	vaas	portemonnee	hamer
38	koekje	chips	pleister	boog	thermometer
39	koelkast	ijskristal	mobiel toilet	skeeler	naald
40	koffer	trein	lantaarn	stoel	olijf
41	krijtjes	palet	spelden	kikker	trommel
42	krokodil	uil	augurk	bokshandschoenen	tandartsstoel
43	kussen	schommelstoel	ravioli	leeuw	asbak
44	lampion	zaklamp	bandoneon	peul	hagedis
45	lasso	cowboyhoed	waterslang	stemvork	tas
46	liniaal	perforator	kam	pannenkoeken	waterzak
47	lippenstift	parfum	aansteker	cruiseschip	zak
48	loep	microscoop	tafeltennisbatje	prullenbak	reddingsvest
49	medaille	trofee	bord	garnaal	schroevendraaier
50	meloen	bananen	rugbybal	golfclub	raket
51	mes	theepot	peddel	poederdoos	babybedje
52	microfoon	boxjes	pizzasnijder	ketel	vuilniszakken
53	milkshake	friet	walkietalkie	wetsuit	snelheidsmeter

54	monitor	muis	dienblad	notenkraker	rietjes
55	naald	vingerhoedje	dwarsfluit	fiets	boek
56	oog	haar	wereldbol	broccoli	politieauto
57	oor	voet	croissant	schildersezel	vrachtwagen
58	oven	koekenpan	kastje	honkbalknuppel	tijger
59	pannenkoek	brood	klok	ketting	vijl
60	paraplu	regenlaarzen	kruk	veiligheidsspelden	kruiwagen
61	piano	trompet	streepjescode	riem	bureaulamp
62	pinguïn	ijsbeer	champagne	tissuedoos	bureau
63	pinpas	euro	envelop	blad	zwaan
64	plakband	paperclip	wc-papier	pijl	zonnebril
65	plant	gieter	feesttoeter	nagelknipper	controller
66	potlood	puntenslijper	schroef	skelet	kat
67	radiator	kachel	dranghek	boon	nietmachine
68	raket	tank	vuurtoren	etui	dalmatiër
69	rat	muizenval	stekkerdoos	horloge	brug
70	riem	sokken	slang	dartbord	cappuccino
71	ring	oorbellen	donut	telraam	prei
72	rog	zeepaardje	vliegtuig	bierflesje	discobal
73	schildpad	viskom	noot	vaatwasser	winkelwagen
74	schoen	pet	strijkijzer	propeller	pakket
75	shuttle	tennisbal	gloeilamp	pasta	dunschiller
76	sinaasappel	courgette	golfbal	kalf	snijplank
77	ski's	muts	pincet	ezel	pepervaatje
78	sleutel	kluis	kurkentrekker	basketbal	spinnewiel
79	slof	badjas	cavia	filmrol	strijkplank
80	snijplank	hakmes	laptop	kerstkrans	jas
81	snoep	hamburger	knikkers	wasmachine	fototoestel
82	spaghetti	vergiet	touw	verkeerslicht	klarinet
83	speen	babypakje	pion	picknicktafel	dolfijn
84	spook	grafsteen	shuttle	hondenriem	koffiemolen

85	spuit	stethoscoop	dartpijl	dominostenen	fornuis
86	stijgbeugel	paard	stamper	hotdog	palmboom
87	strijkplank	wasmand	keyboard	bloem	hand
88	surfplank	badpak	veer	bizon	graafmachine
89	sushi	eetstokjes	duct tape	kruisboog	step
90	tamboerijn	viool	pizza	wattenstaafje	kruk
91	televisie	afstandsbediening	schoolbord	trombone	cowboylaarzen
92	theepot	lepel	kandelaar	sportschoenen	bretels
93	toffee	gebit	vlinderdas	agenda	hout
94	trappers	wiel	verfroller	haai	glijbaan
95	visnet	kreeft	zeef	lantaarnpaal	scheerapparaat
96	vlieger	springtouw	voorrangsbord	geweer	printer
97	vliegtuig	label	kruis	worst	muffin bakvorm
98	vlinder	rups	gereedschapskist	rijst	slot
99	zaklamp	kaars	ontstopper	ijsblokjeshouder	flippers
100	zweep	cap	hengel	verrekijker	framboos

Appendix B: Trials in English. The last four columns are the intended names of the pictures. The words in the English stimulus set are a direct translation from the Dutch set, except the one marked with an asterisk. The intended names in the English stimulus set are however not always a direct translation of the intended names in the Dutch stimulus set.

Trial number	Word	Semantically Related Picture	Visually Related Picture	Unrelated Picture1	Unrelated Picture2
1	potato	corn	bowling ball	battery	pencil
2	ashtray	pipe	yoyo	pinecone	roses
3	tub	tap	sleigh	baseball glove	paintbrush
4	swimsuit	flipflops	hot water bottle	staples	coffee machine
5	baking tray	cake	cassette	ice skate	tie
6	ball	soccer shoes	tomato	hookah	spade
7	balloon	present	cherry	cheese slicer	coffee beans
8	banana	monkey	canoe	tambourine	hat
9	basketball	badminton racket	coconut	trolley	stanley knife
10	mug	fork	thread	pen	goggles
11	blocks	rocking horse	toffee	saxophone	bear
12	bowler hat	cane	juicer	meat	elephant
13	tree	axe	toilet brush	microwave	magnet
14	drill	measuring tape	gun	balloon	office chair
15	ship	anchor	clog	chocolate	baseball
16	bone	puppy	dumbbell	broom	daffodil
17	glasses	telescope	bra	razor	key
18	stroller	baby bottle	tractor	snow shovel	sunflower
19	cd	floppy disk	life saver	holster	ruler
20	turd	diaper	ice cream	jug	pumpkin
21	grapes	wine glass	pool balls	chainsaw	bell
22	drum set	electric guitar	scale	slingshot	playing cards

23	egg	rooster	wool	toothbrush	xylophone
24	bottle	cork	bowling pin	pants	bauble
25	flute	harp	rolling pin	rubber duck	barrel
26	whisk	bowl	hairbrush	pin cushion	pills
27	bulb	light switch	avocado	eagle	basket
28	handcuffs	police hat	bike pedals	shaving brush	crane
29	bow	cannon	hack saw	pineapple	nailpolish
30	rake	shears	spatula	dynamite	inner tube
31	helmet	motorcycle	mango	dustpan	ice cream cone
32	crane	cement truck	giraffe	cup	blackberries
33	horseshoe	saddle	headphones	teddy bear	brie cheese
34	ipod	radio	compass	watermelon	bottle opener
35	jerry can	gas pump	bell pepper	fan	telephone
36	hanger	coat rack	triangle	loudspeaker	tricycle
37	apple core	strawberry	vase	wallet	hammer
38	cookie	potato chips	band aid	bow	thermometer
39	fridge	snowflake	portable toilet	rollerblade	needle
40	suitcase	train	lantern	chair	olive
41	crayons	palette	pins	frog	drum
42	crocodile	owl	pickle	boxing gloves	dentist chair
43	cushion	rocking chair	ravioli	lion	ashtray
44	lantern	flashlight	bandoneon	snowpea	lizard
45	lasso	cowboy hat	hose	tuning fork	bag
46	ruler	hole puncher	comb	pancakes	water pouch
47	lipstick	perfume	lighter	cruise ship	paper bag
48	magnifier	microscope	ping pong paddle	trash can	life jacket
49	medal	trophy	plate	shrimp	screwdriver
50	melon	bananas	football	golf club	rocket
51	knife	teapot	paddle	compact blusher	crib
52	microphone	speakers	pizza cutter	tea kettle	trash bags
53	milkshake	french fries	walkie talkie	wetsuit	speedometer

54	monitor	mouse	tray	nutcracker	straws
55	needle	thimble	flute	bicycle	book
56	eye	hair	globe	broccoli	police car
57	ear	foot	croissant	easel	truck
58	oven	frying pan	cupboard	baseball bat	tiger
59	pancake	bread	clock	chain	nail file
60	umbrella	rain boots	stool	safety pins	wheelbarrow
61	piano	trumpet	barcode	belt	desk lamp
62	penguin	polar bear	champagne	tissue box	desk
63	credit card	euro	envelope	leaf	swan
64	tape	paperclip	toilet paper	arrow	sunglasses
65	plant	watering can	party whistle	nail clippers	nintendo controller
66	fountain pen	pencil sharpener	screw	skeleton	cat
67	radiator	furnace	fence	bean	stapler
68	rocket	tank	lighthouse	pencil case	dalmatian
69	rat	mousetrap	power board	watch	bridge
70	belt	socks	snake	dartboard	cappuccino
71	ring	earrings	donut	abacus	leek
72	ray	seahorse	airplane	beer bottle	disco ball
73	turtle	fishbowl	nut	dishwasher	shopping cart
74	shoe	cap	iron	propeller	package
75	shuttlecock	tennis ball	light bulb	pasta	potato peeler
76	orange	zucchini	golf ball	calf	cutting board
77	skis	beanie	tweezers	donkey	pepper shaker
78	key	safe	corkscrew	basketball	spinning wheel
79	slipper	robe	guinea pig	film roll	ironing board
80	cutting board	cleaver	laptop	wreath	jacket
81	candy	hamburger	marbles	washing machine	camera
82	spaghetti	strainer	rope	traffic light	clarinet
83	pacifier	baby onesie	pawn	picnic table	dolphin
84	ghost	tombstone	shuttlecock	leash	coffee grinder

85	syringe	stethoscope	dart	dominoes	stove
86	stirrup	horse	potato masher	hotdog	palm tree
87	ironing board	laundry basket	keyboard	flower	hand
88	surfboard	bathing suit	feather	bison	excavator
89	sushi	chopsticks	duct tape	crossbow	scooter
90	tambourine	violin	pizza	q tip	crutch
91	television	remote control	blackboard	trombone	cowboy boots
92	teapot	spoon	candle holder	sneakers	suspenders
93	toffee	teeth	bow tie	agenda	logs
94	pedals	wheel	paint roller	shark	slide
95	fishnet	lobster	sieve	lamp post	electric shaver
96	kite	skipping rope	traffic sign	rifle	printer
97	airplane	luggage tag	cross	sausage	muffin tray
98	butterfly	caterpillar	toolbox	rice	lock
99	flashlight	candle	plunger	ice cube tray	flippers
100	whip	riding helmet	fishing rod	binoculars	raspberry

Table 1.

Results of the similarity-rating studies with Dutch and English native speakers. The averaged rating, and standard deviation (between brackets) are displayed for each picture group separately with t-tests and Cohen's d_{av} displaying the comparison between different picture groups.

Semantic similarity ratings

		Semantically related	**Visually related**	**Average unrelated**
Mean rating (standard deviation) over participants				
Dutch				
over participants		*6.73 (1.24)*	*0.96 (1.15)*	*0.41 (0.48)*
over trials		*6.73 (1.25)*	*0.96 (0.54)*	*0.41 (0.29)*
English				
over participans		*5.93 (1.68)*	*0.92 (1.05)*	*0.37 (0.51)*
over trials		*5.93 (1.80)*	*0.92 (0.55)*	*0.36 (0.31)*
Comparison with visually related picture				
Dutch	$t_1(29)$ =	19.473, $p < 0.001$.Cohen's d_{av}= 4.824	-	-
	$t_2(99)$ =	42.408, $p < 0.001$. Cohen's d_{av} =6.441	-	-
English	$t_1(29)$ =	17.608, $p < 0.001$ Cohen's d_{av} =3.667	-	-
	$t_2(99)$ =	26.186, $p < 0.001$. Cohen's d_{av} =4.269	-	-
Additional comparison with average of unrelated pictures				
Dutch	$t_1(29)$ =	27.777, $p < 0.001$, Cohen's d_{av} =7.350	3.131, $p < 0.01$. Cohen's d_{av} = 0.679	-
	$t_2(99)$ =	50.050, $p < 0.001$. Cohen's d_{av} =8.196	9.451, $p < 0.001$. Cohen's d_{av} = 1.341	-
English	$t_1(29)$ =	21.187, $p < 0.001$, Cohen's d_{av} =5.077	3.957, $p < 0.001$. Cohen's d_{av} =0.716	-
	$t_2(99)$ =	29.815, $p < 0.001$, Cohen's d_{av} =5.287	9.427, $p < 0.001$. Cohen's d_{av} =1.321	-

Visual similarity ratings

		Semantically related	Visually related	Average unrelated
Mean rating (standard deviation) over participants				
Dutch				
over participants		*1.21 (1.03)*	*5.03 (1.65)*	*0.69 (0.76)*
over trials		*1.21 (0.84)*	*5.03 (1.36)*	*0.69 (0.36)*
English				
over participans		*1.02 (0.92)*	*4.86 (1.80)*	*0.48 (0.58)*
over trials		*1.03 (0.98)*	*4.86 (1.49)*	*0.48 (0.44)*
Comparison with visually related picture				
Dutch	$t_1(30)$ =	18.401, p < 0.001. Cohen's d_{av} =2.850	-	-
	$t_2(99)$ =	24.641, p < 0.001. Cohen's d_{av} =3.472	-	-
English	$t_1(29)$ =	14.096, p < 0.001. Cohen's d_{av} =2.828	-	-
	$t_2(99)$ =	22.754, p < 0.001. Cohen's d_{av} =3.109	-	-
Additional comparison with average of unrelated pictures				
Dutch	$t_1(30)$ =	4.815, p < 0.001. Cohen's d_{av} =0.588	19.410, p < 0.001. Cohen's d_{av} =3.610	-
	$t_2(99)$ =	5.744, p < 0.001. Cohen's d_{av} =0.884	31.417, p < 0.001. Cohen's d_{av} =5.057	-
English	$t_1(29)$ =	6.361, p < 0.001. Cohen's d_{av} =0.714	15.236, p < 0.001. Cohen's d_{av} =3.680	-
	$t_2(99)$ =	5.149, p < 0.001. Cohen's d_{av} =0.764	27.923, p < 0.001. Cohen's d_{av} =4.523	-

Table 2.

Descriptive statistics (minimum, maximum, mean, standard deviation and median) for the psycholinguistc parameters for each competitor group over all 100 trials for the Dutch and English group seperately. See main text for explanation on untis of measurement.

Dutch Group

Variable	Minimum	Maximum	Mean	SD	Median
Intended Naming Agreement					
semantic competitor	3.33	100.00	73.10	25.54	83.33
visual competitor	0.00	100.00	68.93	26.53	73.33
unrelated distractors	0.00	100.00	71.02	24.93	76.67
Dominant Naming agreemeent					
semantic competitor	20.00	100.00	76.40	20.70	85.00
visual competitor	20.00	100.00	72.43	21.35	75.00
unrelated distractors	16.67	100.00	72.73	22.23	76.67
Leninent Naming Agreement					
semantic competitor	40.00	100.00	87.90	14.79	93.33
visual competitor	30.00	100.00	85.57	17.45	93.33
unrelated distractors	30.00	100.00	83.78	17.84	90.00
Word frequency: fmpw					
semantic competitor	0.02	83.63	8.31	15.42	2.92
visual competitor	0.02	102.63	6.06	15.71	1.37
unrelated distractors	0.02	199.91	10.91	26.44	1.67
Word frequency: Zipf scale					
semantic competitor	1.30	4.92	3.35	0.78	3.46
visual competitor	1.30	5.01	3.21	0.66	3.14
unrelated distractors	1.30	5.30	3.28	0.84	3.22
Age of acquisition (in years)					
semantic competitor	4.01	12.70	6.89	1.74	6.86
visual competitor	4.38	12.00	7.84	1.76	7.89
unrelated distractors	3.73	13.29	7.14	1.77	7.03

English Group

Variable	Minimum	Maximum	Mean	SD	Median
Intended Naming Agreement					
semantic competitor	6.67	100.00	65.03	26.23	68.33
visual competitor	0.00	100.00	64.60	27.33	66.67
unrelated distractors	0.00	100.00	62.92	26.76	66.67
Dominant Naming agreemeent					
semantic competitor	16.67	100.00	66.53	24.00	68.33
visual competitor	10.00	100.00	66.27	24.83	66.67
unrelated distractors	13.33	100.00	64.75	24.47	66.67
Leninent Naming Agreement					
semantic competitor	43.33	100.00	87.17	14.60	93.33
visual competitor	26.67	100.00	85.53	16.13	90.00
unrelated distractors	23.33	100.00	84.35	17.99	93.33
Word frequency: fmpw					
semantic competitor	0.04	153.55	18.40	31.91	6.74
visual competitor	0.27	213.20	11.59	27.72	3.745.12
unrelated distractors	0.04	330.02	18.71	41.97	
Word frequency: Zipf scale					
semantic competitor	1.60	5.19	3.74	0.75	3.83
visual competitor	2.43	5.33	3.61	0.60	3.57
unrelated distractors	1.60	5.52	3.67	0.79	3.71
Age of acquisition (in years)					
semantic competitor	2.50	13.89	6.28	2.21	5.84
visual competitor	3.84	12.44	6.91	1.94	6.50
unrelated distractors	2.74	15.13	6.17	2.13	5.86

Table 3.

Descriptive statistics (minimum, maximum, mean, standard deviation and median) for the visual parameters for each competitor group over all 100 trials. See main text for explanation on untis of measurement.

Variable	Minimum	Maximum	Mean	SD	Median
Total luminance					
semantic competitor	2	131	37	29.99	28
visual competitor	3	135	38	29.83	31
unrelated distractors	1	135	34	25.73	27
Relative luminance					
semantic competitor	13	235	107	44.96	105
visual competitor	21	221	110	40.01	109
unrelated distractors	24	217	106	36.94	104
Total within-object contrast					
semantic competitor	2	66	16	10.68	13
visual competitor	2	67	16	12.35	13
unrelated distractors	0.4	71	16	11.55	13
Average within-object contrast					
semantic competitor	15	92	48	17.79	47
visual competitor	18	100	49	18.79	51
unrelated distractors	13	98	52	17.75	51
Visual Complexity: Filsize grayscale picture					
semantic competitor	18,368	112,735	58,777	23,913	57,703
visual competitor	14,608	128,958	56,031	23,752	50.931
unrelated distractors	4,297	122,489	58,349	23,516	56,159
Overall surface size					
semantic competitor	4,836	121,125	53,498	28,916	48,499
visual competitor	7,333	123,373	51,447	30,236	48.636
unrelated distractors	1,877	123,030	50,515	29,102	47,211
Radius smallest fitting circle					
semantic competitor	134	255	210	15.61	206
visual competitor	167	280	210	18.49	204
unrelated distractors	129	255	209	14.63	204