Predicting naming latencies for action pictures: Dutch norms

Zeshu Shao · Ardi Roelofs · Antje S. Meyer

Abstract The present study provides Dutch norms for age of acquisition, familiarity, imageability, image agreement, visual complexity, word frequency, and word length (in syllables) for 124 line drawings of actions. Ratings were obtained from 117 Dutch participants. Word frequency was determined on the basis of the SUBTLEX-NL corpus (Keuleers, Brysbaert, & New, *Behavior Research Methods,* 42, 643–650, 2010). For 104 of the pictures, naming latencies and name agreement were determined in a separate naming experiment with 74 native speakers of Dutch. The Dutch norms closely corresponded to the norms for British English. Multiple regression analysis showed that age of acquisition, imageability, image agreement, visual complexity, and name agreement were significant predictors of naming latencies, whereas word frequency and word length were not. Combined with the results of a principal-component analysis, these findings suggest that variables influencing the processes of conceptual preparation and lexical selection affect latencies more strongly than do variables influencing word-form encoding.

Keywords Action naming · Age of acquisition · Imageability · Image agreement · Name agreement · Visual complexity

Z. Shao · A. Roelofs · A. S. Meyer
Max Planck Institute for Psycholinguistics,
Nijmegen, The Netherlands

A. Roelofs · A. S. Meyer
Donders Institute for Brain, Cognition and Behaviour,
Centre for Cognition, Radboud University,
Nijmegen, The Netherlands

Z. Shao (✉)
P.O. Box 310, 6500 AH, Nijmegen,
The Netherlands
e-mail: Zeshu.Shao@mpi.nl

The picture-naming task is a well-established research tool for studying word production. In this task, participants are required to name a set of pictures as quickly and accurately as possible. A number of different models have been proposed for the cognitive processes that lead from the perception of a picture to the articulation of its name (e.g., Glaser, 1992; Humphreys, Riddoch, & Quinlan, 1988; Johnson, Paivio, & Clark, 1996). The models differ in many ways, but they all agree that picture naming involves four main processing steps, namely (1) perceptual and conceptual identification of the depicted object or event, hereafter *conceptual preparation*; (2) selection of a suitable name for the object or event, hereafter *lexical selection*; (3) encoding of the corresponding word form (i.e., encoding morphological, phonological, and phonetic representations), hereafter *word-form encoding*; and (4) articulation (e.g., Caramazza, 1997; Levelt, Roelofs, & Meyer, 1999; Rapp & Goldrick, 2000; Roelofs, 1992, 1997).

For many applications of the picture-naming task, it is useful to know beforehand which picture names speakers are likely to choose and how difficult the naming task will be. Following Snodgrass and Vanderwart's (1980) seminal study, norms have been developed for various sets of pictures and for different languages. These norms often include not only the preferred picture names and the corresponding naming latencies, but also indicators of various properties of the pictures and their names, including, for instance, their visual complexity, the age of acquisition of the dominant name, and the frequency of the name. Comparisons of these data sets have shown that the norms are, to some extent, specific to the language tested. For instance, Van Schagen, Tamsma, Bruggemann, Jackson, and Michon (1983) found no significant correlation for name agreement or familiarity between their Dutch norms for object pictures and the English norms for the same picture set provided by Snodgrass and Vanderwart.

Most norming studies have concerned pictures of objects (e.g., Snodgrass & Vanderwart, 1980, for English; Sanfeliu & Fernandez, 1996, for Spanish; Alario & Ferrand, 1999, for

French; Nisi, Longoni, & Snodgrass, 2000, for Italian; Wang, 1997, for Chinese; and Severens, Van Lommel, Ratinckx, & Hartsuiker, 2005, for Dutch). Yet, norms for action pictures are needed in many research, educational, and clinical contexts—for instance, when verb- or action-specific knowledge or the underlying cortical representations are to be investigated, or when morphologically complex forms, such as past tense forms of verbs, are to be elicited (e.g., Gentner, 1981; Kemmerer & Tranel, 2000).

Norms for action pictures have been provided for some languages, including English (e.g., Druks & Masterson, 2000; Szekely et al., 2005), French (Schwitter, Boyer, Méot, Bonin, & Laganaro, 2004), and Spanish (Cuetos & Alija, 2003), but, to the best of our knowledge, not for Dutch. The present study fills this gap by providing Dutch norms for a set of line drawings of actions—namely, the 100 drawings used in the English naming battery of Druks and Masterson, as well as 24 additional drawings (Konopka & Meyer, 2012).

The present study was carried out in three steps: The first step was an online study, in which participants rated the action pictures on a number of dimensions described below. We largely followed Druks and Masterson (2000) in choosing the rating dimensions and preparing the materials, and we determined how well the ratings of the drawings given by English and Dutch participants correlated with each other. The second step was a laboratory study, in which participants named the pictures as quickly and accurately as possible. We determined name agreement and the average naming latency for each picture. Subsequently, we determined how well each of the variables assessed in the rating study, as well as name agreement and indicators of the length and frequency of the picture names, predicted the naming latencies. Since Druks and Masterson did not record naming latencies, such analyses could not be carried out for the English data. Steps 1 and 2 above are referred to as the "main study." The third step was a partial replication of the rating study, referred to as the "supplementary study."

In the main rating study, the participants were asked to rate two properties of the drawings: namely, their visual complexity—defined as the amount of detail in the drawings—and image agreement—defined as the degree to which the visual image evoked by the picture name corresponded to the drawing. Visual complexity probably affects the ease of recognizing the actions, and we therefore expected the picture-naming latencies to increase with the complexity of the pictures (see also Attneave, 1957; Ellis & Morrison, 1998; but see Snodgrass & Yuditsky, 1996). Image agreement should also affect object recognition, with poor imagine agreement leading to slow object recognition, and hence to long object-naming latencies (e.g., Barry, Morrison, & Ellis, 1997).

Three further rating scales—of imageability, familiarity, and subjective age of acquisition—concerned the dominant names of the actions. *Imageability* refers to the degree to which a word can evoke mental images. It can be used to index semantic richness and to study the activation of semantic codes during lexical processing (e.g., Cortese, Simpson, & Woolsey, 1997). *Familiarity* indicates the subjective frequency of exposure to a word. Several studies have shown that familiar words are produced and recognized faster than less familiar ones (e.g., Balota, Pilotti, & Cortese, 2001; Connine, Mullennix, Shernoff, & Yelen, 1990). We expected negative correlations of imageability and familiarity with the action naming latencies.

Age of acquisition (AoA hereafter) refers to the age at which words are learned. It can be measured through estimates given by adults, or by more objective measures (e.g., parental or teacher ratings or formal tests) of when children can name pictures or read words. Several studies have shown that subjective ratings of AoA and objective measures correlate well (e.g., Carroll & White, 1973; Gilhooly & Gilhooly, 1980; Morrison, Chappell, & Ellis, 1997). In the present study, ratings were used, as in the study of Druks and Masterson (2000). A large body of research has demonstrated the strong impact of AoA in a variety of linguistic tasks (e.g., Brysbaert & Cortese, 2011; see Barry et al., 1997, for a review). It has been shown that naming latencies increase with age of acquisition in both object (e.g., Carroll & White, 1973; Ellis & Morrison, 1998) and action naming (e.g., Bogka et al., 2003; Bonin, Boyer, Méot, Fayol, & Droit, 2004; Morrison, Hirsh, & Duggan, 2003). The precise origin of the AoA effect is still a matter of debate (see Johnston & Barry, 2006, for a review), although most of the evidence points toward an early origin—that is, toward object recognition or conceptual preparation processes (e.g., Belke, Brysbaert, Meyer, & Ghyselinck, 2005; Catling & Johnston, 2006; but see Brown & Watson, 1987). In any event, we expected, on the basis of the existing empirical evidence, to see a positive correlation of AoA with the picture-naming latencies.

In addition to the ratings above, we included estimates of the frequencies of the action names in the language (using the SUBTLEX Dutch database; Keuleers, Brysbaert, & New, 2010) and their lengths (in terms of number of syllables) as predictors of the naming latencies. A large literature has demonstrated the impact of frequency on performance in both word and picture naming (e.g., Barry et al., 1997; Cuetos, Ellis, & Alvarez, 1999; Ellis & Morrison, 1998). In the literature, some debate has centered on the origin of the frequency effect in picture naming, but overall, the emerging consensus is that frequency affects both lexical selection and word-form encoding (see Jescheniak & Levelt, 1994; Piai, Roelofs, & Van der Meij, 2012; Strijkers, Costa, & Thierry, 2010).

Several studies have shown that long words take longer to initiate than shorter words (e.g., Eriksen, Pollack, & Montague, 1970; but see Ferrand & Segui, 2003; Henderson, 1982, for a review), most likely because speakers generate the phonological and phonetic codes of successive syllables in sequence

during word-form encoding (e.g., Meyer, Belke, Häcker, & Mortensen, 2007; Roelofs, 2002). However, speakers can initiate the articulation of a word on the basis of a partial form representation (e.g., the representation of one syllable), and in this case no word length effect is found. Thus, whether or not a word length effect is obtained may depend on speakers' response strategies (Damian, Bowers, Stadthagen-Gonzalez, & Spalek, 2010; Meyer, Roelofs, & Levelt, 2003).

From the naming study, we obtained indicators of name agreement for the pictures, which is the extent to which participants agree on the name of a picture. Name agreement is a robust predictor of naming difficulty. Many studies have shown that pictures with a single dominant name are named more quickly than pictures with multiple plausible names (Barry et al., 1997; Paivio, Clark, Digdon, & Bons, 1989; Vitkovitch & Tyrrell, 1995). The origin of the name agreement effect is not entirely clear. Some evidence suggests that the effect arises during lexical selection. For instance, Johnson et al. (1996) found that name agreement affected naming latencies but not object identification latencies for the same object pictures. Alario et al. (2004; see also Alario & Ferrand, 1999) proposed that the effect of name agreement reflects how strong the connections are between the pictured object and its possible names, and that competition among the multiple lexical candidates causes a delay in selecting target lexical representations.

In natural languages, many properties of objects and their names are related. For instance, names acquired early will typically be short and frequent and will refer to familiar concepts (see, e.g., Brown & Watson, 1987; Severens et al., 2005). Consequently, we expected to see intercorrelations between the predictors of naming latencies. We used multiple regression to determine how well each variable independently predicted the naming latencies, and we used principal-component analysis to explore which predictors clustered together.

To summarize, the present study provides Dutch norms for action naming. We report naming latencies and indices of name agreement in order to characterize the overall difficulty of naming each item. In addition, we provide a range of often-used measures that characterize the ease of conceptual preparation (including object recognition), lexical selection, and word-form encoding, and report their impacts on naming latencies.

Main study

Method of the rating study

Participants A group of 117 native Dutch speakers (20 males, 97 females; mean age = 28.57 years, range from 16 to 67 years) participated in the rating study. They were selected from the participant pool of the Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands. They completed a consent form before the study and were paid for their participation. All of the participants had normal or correct-to-normal vision.

Materials A total of 124 black-and-white line-drawings of actions were used in the study (see Fig. 1 for an example). These were the 100 drawings used by Druks and Masterson (2000), along with 24 drawings used by Konopka and Meyer (2012).[1] In a pilot study carried out with ten native speakers of Dutch, we established dominant names for all of the pictures. The rating scales for image agreement and visual complexity were presented in combination with action pictures, whereas the scales for imageability, image agreement, subjective age of acquisition and familiarity were presented with the corresponding written verbs.

All pictures were sized to fit into frames of 150 × 150 pixels on the computer screen. Words were presented in lowercase, black Tahoma font with a large font size. The rating scales were presented simultaneously below the pictures. The study was programmed using the Oracle Application Development Framework 11.1.1.4.0.

Design and procedure All of the participants rated all items on each of the five dimensions, completing one dimension before moving on to the next one. The order of the dimensions for all participants was (1) imageability, (2) image agreement, (3) familiarity, (4) subjective age of acquisition, and (5) visual complexity. Items were rated by all participants in the same order on each dimension, but different orders were used for different dimensions. The rating scales and the instructions for the ratings are reproduced in the supplementary materials.

On each trial of the image agreement rating task, a verb was presented in the center of the computer screen for 1 s and followed by the corresponding picture, which remained in view until the participant responded. In the remaining tasks, a single item (a picture or word) was presented per trial until the participant responded. All of the tasks were self-paced, but the program automatically terminated if no response was recorded within 1 h. Participants could not revise their ratings.

Seven-point rating scales were used for all of the dimensions except age of acquisition. In the *imageability* rating task (adapted from Paivio, Yuille, & Madigan, 1968), participants were presented with written verbs and instructed to rate how readily each verb evoked mental images. A value of 1 on the scale indicated that the verb *evoked mental images with greatest difficulty or not at all*, and a value of 7 indicated that the verb *evoked mental images most readily*.

---

[1] The pictures from Konopka and Meyer (2012) can be obtained by writing to a.s.meyer@mpi.nl or Agnieszka.Konopka@mpi.nl.

Fig. 1 One of the drawings taken from Druks and Masterson (2000), showing the action "digging" (Dutch *graven*)

In the *image agreement* rating task (adapted from Snodgrass & Vanderwart, 1980), participants were instructed to rate how closely each picture resembled their mental images of the action. A value of 1 indicated that the picture provided *a very poor match*, and a value of 7 indicated *a very good match* to the participant's mental image of the action.

In the *familiarity* rating task (adapted from Gilhooly & Logie, 1981), participants were required to rate the familiarity of each word, estimating how often they had experienced it. A value of 1 indicated that the participant had *never* seen, heard, or used the word, and a value of 7 indicated that the participant saw, heard, or used the word *very often*.

In the *age of acquisition* rating task (adapted from Carroll & White, 1973), participants were instructed to rate on a 9-point scale the age at which they thought they had acquired the verb. On this scale, 1 indicated *0–2 years*, 2 indicated *3 years*, 3 indicated *4 years*, 4 indicated *5 years*, 5 indicated *6 years*, 6 indicated *7–8 years*, 7 indicated *9–10 years*, 8 indicated *11–12 years*, and 9 indicated *13 years or older*. If the participants did not know the meaning of a verb, they should choose *Ik ken het woord niet* ("I don't know the word"). "Learning" a verb was defined as the age at which the participants thought that they would have understood the verb if somebody had used it in front of them, even if they did not yet say, read, or write the verb.

In the *visual complexity* rating task (adapted from Snodgrass & Vanderwart, 1980), participants were instructed to rate the complexity of each picture. A value of 1 indicated that the picture was *very simple*, and a value of 7 indicated that it was *very complex*. Participants were required to rate the complexity of the drawing itself rather than the complexity of the action that it represented. Here, "complexity" was identified as the amount of detail or intricacy of the lines in a picture.

Method of the picture-naming study

Participants A group of 74 native Dutch speakers (six males, 68 females; mean age = 35.23 years, range from 16 to 63 years) participated in picture-naming task. They had not participated in the rating study. Two different samples of participants were used so that providing the ratings could not affect the naming latencies, or vice versa. The participants in the picture-naming study were also selected from the participant pool of the Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands. They completed a consent form before the study and were paid for their participation. They had normal or correct-to-normal vision.

Materials Twenty of the rated items (six from the Druks and Masterson, 2000, corpus) were excluded from the set because their name agreement was insufficient (i.e., five or fewer participants used the same name), because the pictures elicited multiword descriptions rather than verbs (two cases), or because another picture had the same dominant name (one case). Although these items were not included in the naming study, they are listed in a separate section of the supplementary materials, together with the most common responses given by the participants in the pilot study. The remaining 104 pictures were resized to fit in 4 × 4 cm frames on the computer screen (2.29º of visual angle).

Procedure Each trial began with the presentation of a fixation cross (+), which was presented for 800 ms in the center of the screen. Then a picture was shown for 600 ms, followed by a red flashing exclamation mark, which was presented until the end of the trial (see also Shao, Roelofs, & Meyer, 2012). The exclamation mark was used to remind participants to respond quickly. The trial was terminated as soon as the voice key was triggered or, if the participant did not respond, 2,000 ms after the onset of the picture. The interstimulus interval was 1,500 ms. The order of items was random and different for each participant, and participants were tested individually.

Apparatus and data analyses The naming task was controlled by a HP 8540P laptop with the Presentation software package (Version 14.3, www.neurobs.com). Naming latencies were recorded online using a voice key and were later confirmed or corrected using the speech analysis program Praat (Boersma, 2001). Responses were coded as errors when participants used names that were different from the dominant names or when the response contained a repair or disfluency (i.e., stutters or starting with filler words).

Scoring The participants' responses were transcribed in order to determine name agreement. We computed two indicators of name agreement. The first was the proportion of

participants who used the dominant name determined in the pilot study. This proportion represented the degree of agreement on a name but did not reveal how many different names the participants used. Information about the spread of responses was captured in a second indicator, the *H* statistic, introduced by Snodgrass and Vanderwart (1980). In Formula 1 below, *k* represents the number of different names given to each picture, *i* represents the number of names assigned to one picture, and P_i represents the proportions of each assigned name.

$$H = \sum_{i=1}^{k} P_i \log_2 \left(1 / P_i\right). \tag{1}$$

If only one name is given for a picture, *H* is zero. If two names occur with equal frequency, *H* is 1. *H* increases with the number of responses given, and generally decreases with the proportions of each response.

Results and discussion

Table 1 provides descriptive statistics for each of the variables assessed in the rating study, along with the word frequency and word length (i.e., number of syllables) of the dominant names. Because of technical problems, five participants could not provide ratings of AoA, imageability, and image agreement. The ratings for each item are listed in the supplementary materials. The ratings show that the actions have high imageability, and that the pictures have good image agreement and low visual complexity.

Table 2 provides descriptive statistics about name agreement (represented by both the proportion of dominant names and the *H* statistic), naming latencies, and the percentages of responses excluded from the analyses of naming latencies ("error rates," including nondominant names, and missing and disfluent responses). The corresponding values for each picture are listed in the supplementary materials. The mean naming latencies were based only on correct responses using the dominant names.

As we mentioned above, the order of the items in each rating task was constant across participants. Given that the set of items was not large and that participants worked at

their own pace, it is unlikely that the ratings were strongly affected by fatigue or practice. To assess this, we correlated the ratings on each scale with the serial positions of the items. No significant correlation was found for any of the five dimensions (for familiarity, $r = .05$; for AoA, $r = .00$; for visual complexity, $r = .11$; for image agreement, $r = .07$; and for imageability, $r = .06$).

Table 3 shows the correlations among all variables. All of the average ratings, as well the indicators of name agreement, word frequency, and length (i.e., syllable number) showed the expected correlations with the action-naming latencies. As can be seen, the two variables specifically pertaining to word-form encoding (i.e., word frequency and word length) showed the weakest correlations with the naming latencies, and the variables pertaining to name agreement showed the highest correlation with the latencies. The table also shows that, as predicted, some of the ratings were strongly correlated: Familiarity correlated with imageability, AoA, and name frequency; the latter two variables also correlated strongly with each other; and, unsurprisingly, the two indicators of name agreement were highly correlated.

Stepwise multiple regression was used to assess how much of the variance in the naming latencies was explained uniquely by each of the predictor variables. Since both the *H* statistic and the proportion of dominant names were used to indicate name agreement, only the *H* statistic was included in the analysis. A similar pattern was found when replacing the *H* statistic with the proportion of dominant names. The results, summarized in Table 4, show that AoA, imageability, image agreement, *H* statistic, and visual complexity were all significant predictors of naming latencies. The regression analysis yielded $R^2 = .55$, $F(6, 94) = 24.37$, $p < .001$.

Table 2 Descriptive data for the *H* statistic, proportion of dominant names, naming latency (in milliseconds), and error rate

	H Statistic	Proportion of Dominant Names	Naming Latency	Naming Error Rate (%)
Mean	0.32	.86	886	23
SD	0.53	.14	135	20
Range	0–3.09	.45–1.00	446–2,331	0–82

SD = standard deviation

Table 1 Descriptive statistics for 124 Dutch action pictures for age of acquisition (AoA), familiarity, imageability, image agreement, visual complexity, word frequency, and word length

	AoA	Familiarity	Image-ability	Image Agreement	Visual Complexity	Word Frequency	Word Length
Mean	5.03	4.54	5.80	6.08	3.49	1.41	2.18
SD	1.23	1.01	048	0.68	0.89	0.71	0.44
Median	5.00	4.43	5.91	6.33	3.45	1.42	2.00
Range	1–9	1–7	1–7	1–7	1–7	0–3.11	2–4

SD = standard deviation

Table 3 Matrix of correlations between naming latencies and the predictor variables

	AoA	Familiarity	Image-Ability	Image Agreement	Name Agreement	H Statistic	Word Length	Visual Complexity	Word Frequency
Naming latency	.43**	−.39**	−.57**	−.38**	−.53**	.43**	.27**	.36**	−.24*
AoA		−.64**	−.36**	.14	−.29**	.21*	.35**	.31**	−.55**
Concept familiarity			.49**	−.03	.21*	−.11	−.22*	−.20*	.61**
Imageability				.26**	.27**	−.27*	−.13	−.20*	.10
Image agreement					.27**	−.14	.04	−.11	−.15
Name agreement						−.71**	−.12	−.11	.12
H statistic							.06	.12	−.07
Word length								−.05	−.26**
Visual complexity									−.21*

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

Principal-component analysis (PCA) was used to explore how many underlying factors were reflected by the predictors. The PCA with varimax rotation was computed on all eight original variables for the sample of 104 Dutch verbs. The Kaiser–Meyer–Olkin measure verified the sampling adequacy for the analysis, KMO = .68. Bartlett's test of sphericity, $\chi^2(28) = 212.96$, $p < .001$, indicated that the correlations between variables were sufficiently high for PCA. Three components had eigenvalues greater than 1 over the Kaiser criterion, and in combination these components explained 67.51 % of the variance. Table 5 shows the factor loadings after rotation. Three factors were extracted. In detail, Factor 1 loaded on AoA, familiarity, and word frequency; Factor 2 loaded on imageability, image agreement, and name agreement; and Factor 3 loaded on name length and visual complexity. Similar results were found in the study of action pictures by Bonin et al. (2004), with one factor loading on AoA, familiarity, and visual complexity, and another loading on imageability, image agreement, and name agreement. However, unlike in the study of Bonin et al. (2004), we found that visual complexity grouped with name length (forming a third factor), rather than with AoA and familiarity (our first factor).

As noted, the majority of the pictures used in the present study were taken from the corpus provided by Druks and Masterson (2000).[2] Table 6 shows the descriptive statistics for the items included in both studies, as well as the results of t tests comparing the average ratings from the two languages and the correlations among the ratings. The averages were significantly different for all variables except AoA. Ratings of AoA were transformed to a common scale of estimated age by using the midpoints of the ranges used in the two studies. As compared with the British-English speakers of Druks and Masterson, our Dutch speakers thought that the pictures were

visually less complex, $t(88) = 14.64$, $p < .01$, and that the words were more familiar, $t(88) = 5.67$, $p < .01$, and harder to imagine, $t(88) = 23.35$, $p < .01$. The Dutch target words were longer than the English equivalents, $t(88) = 1.72$, $p < .05$. However, the table also shows that, in spite of these differences in the average ratings, the ratings for the individual items in the two languages were highly correlated. Only a weak correlation emerged for number of syllables, which is not surprising, given that different languages were tested.

Supplementary study

In the rating study described above, all of the participants were presented with the rating dimensions in the same order. A potential concern is that the increasing experience with the items may have affected the ratings. To assess whether this was the case, a supplementary study was carried out in which the order of eliciting the ratings on three of the five dimensions—namely, familiarity, AoA, and visual complexity—was counterbalanced across participants. The first two dimensions of the main study, imageability and image agreement, were not included because the ratings on the very first dimension

Table 4 Results of the multiple regression analysis with naming latency as the criterion variable and age of acquisition (AoA), familiarity, imageability, image agreement, visual complexity, word frequency, word length, and H statistic as predictor variables

Variable	Beta	SE	t Value
AoA	.25	8.67	3.12**
Imageability	−.32	21.54	−4.15**
Image agreement	−.29	14.42	−3.95**
H statistic	.24	18.10	3.31**
Visual complexity	.16	10.98	2.22*

Only variables that made significant contributions are listed in the table. SE = standard error. $^{*} p < .05.$ $^{**} p < .01.$ N = 102

---

[2] For the AoA ratings, the English norming study had used a 7-point scale. Each value on the scale represented a two-year age band, except that 1 represented *below 2 years* and 7 represented *13 years and older*.

Table 5 Summary of results of the principal-component analysis

Variable	Rotated Factor Loading		
	Factor 1	Factor 2	Factor 3
AoA	.33		
Familiarity	−.31		
Word frequency	−.35		
Image agreement		.48	
Imageability		.38	
H statistic		−.44	
Word length			.68
Visual complexity			−.70
Eigenvalue	2.82	1.56	1.02
% Variance	35.25	19.49	12.77

Table 6 Comparison and correlation between Dutch and English action normative data

Variable	Mean/SD	Dutch	English	Correlation
AoA	Mean	5.03 (5–6 years)	2.56 (5–6 years)	.71[**]
	SD	1.23	0.68	
Imageability	Mean	5.85	4.26	.63[**]
	SD	0.48	0.58	
Familiarity	Mean	4.54	3.96	.74[**]
	SD	1.01	1.41	
Word length	Mean	2.18	2.06	.27[*]
	SD	0.44	0.23	
Visual complexity	Mean	3.48	4.26	.80[**]
	SD	0.89	0.77	

[*] $p < .05$. [**] $p < .01$

presented could not be affected by the following ratings, and because assessing image agreement after assessing imageability was entirely appropriate and should not be reversed; participants were first asked to indicate whether they could form images of the actions, and then to decide how well the pictures matched their images.

Method

Participants A group of 63 native Dutch speakers (11 males, 52 females; mean age = 35.66 years, SD = 20.32) participated in the study. They were recruited from the same pool and in the same ways as the participants of the first rating study. They completed a consent form before the study and were paid for their participation. All of the participants had normal or correct-to-normal vision.

Materials and design The materials were the same ones that had been used to assess familiarity, AoA, and visual complexity in the first rating study. A Latin-square design was used to counterbalance the order of administering the rating tasks across participants. Three groups were created of 21 participants each, and the different groups worked on the tasks in the orders familiarity–AoA–visual complexity, AoA–visual complexity–familiarity, and visual complexity–familiarity–AoA.

Results and discussion

The supplementary study aimed to determine how the ratings of familiarity, AoA, and visual similarity were affected by prior experience with the items. Table 7 shows the average ratings for each dimension depending on the serial position of the corresponding rating task. The average visual complexity rating of the items was somewhat lower when visual complexity ratings were given last (third position), as

compared with when they were given earlier (first and second positions). Similarly, the average familiarity rating for the items was higher when those ratings were provided last. Thus, as one might expect, the participants perceived the items as being less complex and more familiar when they had seen them twice before than when they were novel. The AoA ratings did not show this pattern, but were instead lowest when AoA was the second dimension to be evaluated.

To assess the statistical significance of the serial order effects, analyses of variance were carried out for each scale using participants or items as random variables. This analysis yielded a significant main effect only for AoA [by participants, $F_1(2, 61) = 3.47$, $p < .05$; by items, $F_2(2, 122) = 339.74$, $p < .001$].[3] Tukey post-hoc tested showed that only the difference in the ratings between the second and third positions was significant ($p < .01$). How this effect arose is unclear. It is important to keep in mind that different participants rated the items in each serial position. The group of participants who assigned the highest AoA ratings to the items rated the items lowest for familiarity and highest for visual complexity. Thus, the observed pattern might be due to one group of participants perceiving the items as "harder" than the other participants, and consequently assigning them lower familiarity and higher AoA and visual complexity ratings. Though we found some effects of the ordering of the dimensions or/and differences across participant groups that affected the ratings, it is important to note that the ratings given by the three groups of participants for each dimension were highly correlated (all rs > .90, ps < .001).

In further analyses, we compared the ratings obtained in the supplementary study to those of the main study, in which all participants first rated the items for imageability and

---

[3] For visual complexity and familiarity, we found a significant main effect by items [$F(2, 122) = 78.65$, $p < .001$, and $F(2, 122) = 35.67$, $p < .001$, respectively], but not by participants (ps > .39).

Table 7 Mean rating scores for visual complexity, AoA, and familiarity in each serial position of the rating scale (standard deviations are given in parentheses)

Serial Position	Dimensions		
	Visual Complexity	AoA	Familiarity
First	3.89 (0.94)	5.88 (1.19)	4.24 (1.08)
Second	3.84 (0.76)	5.41 (1.22)	4.19 (0.86)
Third	3.61 (0.85)	6.13 (1.10)	4.53 (0.99)
Average	3.78	5.81	4.32

image agreement and then rated them, in fixed order, for familiarity, AoA, and visual complexity. A comparison of Tables 1 and 7 shows that, on average, the items were rated as being "harder" (i.e., less familiar, with later AoA, and visually more complex) in the supplementary than in the main study. These differences were significant for familiarity, $t(123) = 6.45$, $p < .01$; for AoA, $t(123) = 8.73$, $p < .01$; and for visual complexity, $t(123) = 3.52$, $p < .01$. This suggests that the repeated presentation of the materials in the main study indeed did affect the ratings. However, the correlations of the average ratings for each dimension in the two studies were high (for familiarity, $r = .95$, $p < .001$; for AoA, $r = .93$, $p < .001$; and for visual complexity, $r = .76$, $p < .001$). Furthermore, we correlated the ratings given for each dimension when the specific dimension was presented in first serial position in the supplementary study to the corresponding ratings in the main study. Again, high correlations were found: for familiarity, $r = .88$, $p < .001$; for AoA, $r = .94$, $p < .001$; and for visual complexity, $r = .90$, $p < .001$. In sum, the repeated presentation of the items may have affected the absolute ratings for the items, but it did not systematically affect how the items were rated relative to each other.

General discussion

The present study is, to the best of our knowledge, the first norming study of action pictures in Dutch. The ratings of visual complexity, AoA, imageability, image agreement, and familiarity, the average naming latencies for the dominant name of each picture, the total number of alternative names, and two indicators of name agreement (i.e., proportion of dominant names and H statistic) are listed for each picture in the supplementary materials. These norms should be useful for future studies in which, for instance, action drawings are required that are fast or slower to be named, or in which sets of pictures are to be constructed that are matched for, say, visual complexity or AoA. We found that the ratings of AoA, imageability, familiarity, and visual complexity obtained here correlated well with the ratings in the norming study of Druks

and Masterson (2000) for British English. Given the similarity of the two languages and the similarity of the speakers' cultural and educational background, this is perhaps not too surprising, but it does demonstrate the reliability of the ratings.

Additionally, we found high correlations among ratings for the three dimensions within the supplementary study and between the supplementary study and the main study. This demonstrates the reliability of the ratings and indicates that repeated presentation of the items in our studies may have affected the absolute ratings for the items, but not the relative rankings, which will probably be most important for the majority of researchers who will use the norms.

Furthermore, all predictor variables included in the present study were significantly correlated with the picture-naming latencies. Multiple regression analyses showed that only imageability (how easy it was to create a mental image, given the action name), image agreement (how well the drawing corresponded to the raters' mental images of the action), the H statistic (indicative of name agreement), AoA, and visual complexity all independently predicted naming latencies. Taken together, these variables accounted for 55 % of the variance in the naming latencies. By contrast, we found no independent contribution of familiarity, word frequency, and word length. Given the likely origins of these effects that we discussed in the introduction, these results suggest that the time that speakers need to produce suitable names for action drawings depends to a large part on the amounts of time that they require to identify the actions (i.e., conceptual preparation) and to select appropriate lexical units (i.e., lexical selection), and much less on the time that is needed to retrieve the corresponding morphological, phonological, and phonetic forms (i.e., word-form encoding). Of course, it is important to keep in mind that the pattern seen here may not generalize to other sets of pictures. For instance, how much time speakers need to identify the actions shown and how variable this time might be will depend on the kinds of actions and the quality of the pictures. Similarly, whether lexical variables such as word frequency and word length will correlate with naming latencies may depend on the variability in the word frequencies and lengths related to the pictures present in the set. Nevertheless, the present results suggest that in action naming, differences in the duration of early cognitive processes, related to visual and conceptual processes (i.e., conceptual preparation) and the selection of a suitable name (i.e., lexical selection), contribute more to latency differences than do differences in the duration of later word-form encoding processes.

This conclusion is in line with earlier action-norming studies in other languages. For instance, Bonin and colleagues (2004; see also Schwitter et al., 2004) found that name agreement, image agreement, and AoA significantly predicted naming latencies in French. Cuetos and Alija (2003) found that AoA and name agreement were significant predictors of

action-naming latencies in Spanish. Additionally, the absence of a frequency effect in the present study is consistent with the other mentioned normative studies (e.g., Bonin et al., 2004; Cuetos & Alija, 2003; Schwitter et al., 2004). Apparently name agreement, image agreement, and AoA are relatively stable predictors of action-naming latencies across languages.

Studies of object naming have also highlighted the impact of early visual/conceptual variables on naming latencies. Specifically, Alario et al. (2004) showed that more-complex drawings took more time to be recognized than did less-complex drawings (see also Ellis & Morrison, 1998), and pictures with higher image agreement ratings were named faster than pictures with lower ratings. Moreover, object-naming studies have also shown that the frequency effect disappeared when AoA was controlled for (Bonin, Peereman, Malardier, Méot, & Chalard, 2003; Carroll & White, 1973; but see Barry et al., 1997). These findings suggest a commonality between object and action naming. However, the effects of early visual/conceptual variables appear to be less pervasive in object than in action naming. For instance, many studies have failed to show effects of visual complexity on object-naming latencies (Barry et al., 1997; Bonin, Chalard, Méot, & Fayol, 2002; Bonin et al., 2003; Cuetos et al., 1999; Snodgrass & Yuditsky, 1996). A possible cause for this might be that object pictures are generally easier to process at the recognition and conceptualization levels than are action pictures. Consistent with this assumption, we found in an earlier object- and action-naming study (Shao et al., 2012) that inhibitory control was more systematically involved in action naming than in object naming, possibly because more interference among competing conceptual representations occurred during conceptual preparation in the action-naming task. Overall, higher-level cognitive processes may be more influential determinants of the speed of action than of object naming.

Conclusions

To recapitulate, the present study provides Dutch normative data for 124 line drawings of actions. Naming latencies for a subset of 104 drawings were also collected. Multiple regression analyses showed that name agreement, image agreement, imageability, visual complexity, and age of acquisition were all significant predictors of the naming latencies.

Author note   Correspondence should be addressed to Zeshu Shao. P. O. Box 310, 6500 AH Nijmegen, the Netherlands. Tel: 0031 24 3521 204. Email: Zeshu.Shao@mpi.nl.

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. doi:10.3758/BF03200732

Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers, 36,* 140–155. doi:10.3758/BF03195559

Attneave, F. (1957). Physical determinants of the judged complexity of shapes. *Journal of Experimental Psychology, 53,* 221–227. doi:10.1037/h0043921

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition, 29,* 639–647. doi:10.3758/BF03200465

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *Quarterly Journal of Experimental Psychology, 50A,* 560–585. doi:10.1080/783663595

Belke, E., Brysbaert, M., Meyer, A. S., & Ghyselinck, M. (2005). Age of acquisition effects in picture naming: Evidence for a lexical–semantic competition hypothesis. *Cognition, 96,* 45–54.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5,* 341–345.

Bogka, N., Masterson, J., Druks, J., Fragkioudaki, M., Chatziprokopiou, E. S., & Economou, K. (2003). Object and action picture naming in English and Greek. *European Journal of Cognitive Psychology, 15,* 371–403.

Bonin, P., Boyer, B., Méot, A., Fayol, M., & Droit, S. (2004). Psycholinguistic norms for action photographs in French and their relationships with spoken and written latencies. *Behavior Research Methods, Instruments, & Computers, 36,* 127–139. doi:10.3758/BF03195558

Bonin, P., Chalard, M., Méot, A., & Fayol, M. (2002). The determinants of spoken and written picture naming latencies. *British Journal of Psychology, 93,* 89–114. doi:10.1348/000712602162463

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. doi:10.3758/BF03195507

Brown, G. D. A., & Watson, F. L. (1987). First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory & Cognition, 15,* 208–216. doi:10.3758/BF03197718

Brysbaert, M., & Cortese, M. J. (2011). Do the effects of subjective frequency and age of acquisition survive better word frequency norms? *Quarterly Journal of Experimental Psychology, 64,* 545–559. doi:10.1080/17470218.2010.503374

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology, 14,* 177–208.

Carroll, J. B., & White, M. N. (1973). Age-of-acquisition norms for 220 picturable nouns. *Journal of Verbal Learning and Verbal Behavior, 12,* 563–576. doi:10.1016/S0022-5371(73)80036-2

Catling, J. C., & Johnston, R. A. (2006). Effects of age-of-acquisition and short term priming on picture naming. *Quarterly Journal of Experimental Psychology, 9,* 1443–1453.

Connine, C. M., Mullennix, J., Shernoff, E., & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1084–1096. doi:10.1037/0278-7393.16.6.1084

Cortese, M. J., Simpson, G. B., & Woolsey, S. (1997). Effects of association and imageability on phonological mapping. *Psychonomic Bulletin & Review, 4,* 226–231. doi:10.3758/BF03209397

Cuetos, F., & Alija, M. (2003). Normative data and naming times for action pictures. *Behavior Research Methods, Instruments, & Computers, 35,* 168–177. doi:10.3758/BF03195508

Cuetos, F., Ellis, A. W., & Alvarez, B. (1999). Naming times for the Snodgrass and Vanderwart pictures in Spanish. *Behavior Research Methods, Instruments, & Computers, 31,* 650–658. doi:10.3758/BF03200741

Damian, M. F., Bowers, J. S., Stadthagen-Gonzalez, H., & Spalek, K. (2010). Does word length affect speech onset latencies when producing single words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 892–905. doi:10.1037/a0019446

Druks, J., & Masterson, J. (2000). *Object and action naming battery.* Hove, UK: Psychology Press.

Ellis, A. W., & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 515–523. doi:10.1037/0278-7393.24.2.515

Eriksen, C. W., Pollack, M. D., & Montague, W. E. (1970). Implicit speech: Mechanisms in perceptual encoding? *Journal of Experimental Psychology, 84,* 502–507.

Ferrand, L., & Segui, J. (2003). Reading aloud polysyllabic words. In E. M. H. Assink & D. Sandra (Eds.), *Reading complex words: Cross language studies* (pp. 295–314). New York, NY: Kluwer.

Gentner, D. (1981). Some interesting differences between verbs and nouns. *Cognitive & Brain Theory, 4,* 161–178.

Gilhooly, K. J., & Gilhooly, M. L. M. (1980). The validity of age-of-acquisition ratings. *British Journal of Psychology, 71,* 105–110. doi:10.1111/j.2044-8295.1980.tb02736.x

Gilhooly, K. J., & Logie, R. H. (1981). Word age-of-acquisition, reading latencies and auditory recognition. *Current Psychological Research, 1,* 251–262.

Glaser, W. R. (1992). Picture naming. *Cognition, 42,* 61–105.

Henderson, L. (1982). *Orthography and word recognition in reading.* London, UK: Academic Press.

Humphreys, G. W., Riddoch, M. J., & Quinlan, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology, 5,* 67–103.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 824–843.

Johnson, C. J., Paivio, A., & Clark, J. M. (1996). Cognitive components of picture naming. *Psychological Bulletin, 120,* 113–139. doi:10.1037/0033-2909.120.1.113

Johnston, R. A., & Barry, C. (2006). Age of acquisition and lexical processing: A review. *Visual Cognition, 13,* 789–845.

Kemmerer, D., & Tranel, D. (2000). Verb retrieval for action naming in brain-damaged subjects: 1. Analysis of stimulus, lexical, and conceptual factors. *Brain and Language, 73,* 347–392.

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

Konopka, A., & Meyer, A. S. (2012). *Priming sentence planning: The role of words and structures.* Unpublished manuscript.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *The Behavioral and Brain Sciences, 22,* 1–38. disc. 38–75.

Meyer, A. S., Belke, E., Häcker, C., & Mortensen, L. (2007). Use of word length information in utterance planning. *Journal of Memory and Language, 57,* 210–231.

Meyer, A. S., Roelofs, A., & Levelt, W. J. M. (2003). Word length effects in object naming: The role of a response criterion. *Journal of Memory and Language, 48,* 131–147. doi:10.1016/S0749-596X(02)00509-0

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

Morrison, C. M., Hirsh, K. W., & Duggan, G. B. (2003). Age of acquisition, ageing, and verb production: Normative and

experimental data. *Quarterly Journal of Experimental Psychology, 56A,* 705–730. doi:10.1080/02724980244000594

Nisi, M., Longoni, A. M., & Snodgrass, J. G. (2000). Italian measurement on the relation of name, familiarity, and acquisition age for the 260 figures of Snodgrass and Vanderwart. *Giornale Italiano di Psicologia, 27,* 205–218.

Paivio, A., Clark, J. M., Digdon, N., & Bons, T. (1989). Referential processing: Reciprocity and correlates of naming and imaging. *Memory & Cognition, 17,* 163–174. doi:10.3758/BF03197066

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, 76*(1, Pt. 2), 1–25. doi:10.1037/h0025327

Piai, V., Roelofs, A., & Van der Meij, R. (2012). Event-related potentials and oscillatory brain responses associated with semantic and Stroop-like interference effects in overt naming. *Brain Research, 1450,* 87–101.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review, 107,* 460–499.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42,* 107–142. doi:10.1016/0010-0277(92)90041-F

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition, 64,* 249–284.

Roelofs, A. (2002). Syllable structure effects turn out to be word length effects: Comment on Santiago et al. (2000). *Language & Cognitive Processes, 17,* 1–13.

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. doi:10.3758/BF03200541

Schwitter, V., Boyer, B., Méot, A., Bonin, P., & Laganaro, M. (2004). French normative data and naming times for action pictures. *Behavior Research Methods, Instruments, & Computers, 36,* 564–576. doi:10.3758/BF03195603

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

Shao, Z., Roelofs, A., & Meyer, A. S. (2012). Sources of individual differences in the speed of naming objects and actions: The contribution of executive control. *Quarterly Journal of Experimental Psychology, 65,* 1927–1944. doi:10.1080/17470218.2012.670252

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,* 174–215. doi:10.1037/0278-7393.6.2.174

Snodgrass, J. G., & Yuditsky, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments, & Computers, 28,* 516–536. doi:10.3758/BF03200540

Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking lexical access in speech production: Electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex, 20,* 912–928.

Szekely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., . . . Bates, E. (2005). Timed action and object naming. *Cortex, 41,* 7–25.

Van Schagen, I., Tamsma, N., Bruggemann, F., Jackson, J. L., & Michon, J. A. (1983). Namen en normen voor plaatjes [Names and norms for pictures]. *Nederlands Tijdschrift voor de Psychologie, 38,* 236–241.

Vitkovitch, M., & Tyrrell, L. (1995). Sources of disagreement in object naming. *Quarterly Journal of Experimental Psychology, 48A,* 822–848. doi:10.1080/14640749508401419

Wang, M. Y. (1997). The evaluation of perceptual and semantic characteristics for a set of object contour pictures. *Chinese Journal of Psychology, 39,* 157–172.