Psychological Assessment
2018, Vol. 30, No. 2, 259–273

1040-3590/18/$12.00 http://dx.doi.org/10.1037/pas0000472

Computerized Scoring Algorithms for the Autobiographical Memory Test

Keisuke Takano
University of Leuven

Charlotte Gutenbrunner
Victoria University of Wellington

Kris Martens
University of Leuven

Karen Salmon
Victoria University of Wellington

Filip Raes
University of Leuven

Reduced specificity of autobiographical memories is a hallmark of depressive cognition. Autobiographical memory (AM) specificity is typically measured by the Autobiographical Memory Test (AMT), in which respondents are asked to describe personal memories in response to emotional cue words. Due to this free descriptive responding format, the AMT relies on experts' hand scoring for subsequent statistical analyses. This manual coding potentially impedes research activities in big data analytics such as large epidemiological studies. Here, we propose computerized algorithms to automatically score AM specificity for the Dutch (adult participants) and English (youth participants) versions of the AMT by using natural language processing and machine learning techniques. The algorithms showed reliable performances in discriminating specific and nonspecific (e.g., overgeneralized) autobiographical memories in independent testing data sets (area under the receiver operating characteristic curve > .90). Furthermore, outcome values of the algorithms (i.e., decision values of support vector machines) showed a gradient across similar (e.g., specific and extended memories) and different (e.g., specific memory and semantic associates) categories of AMT responses, suggesting that, for both adults and youth, the algorithms well capture the extent to which a memory has features of specific memories.

Public Significance Statement

The present study proposes newly developed computerized algorithms to score specificity of autobiographical memories. As reduced memory specificity is recognized as a hallmark of depressive cognition, these algorithms can be used to detect people with a cognitive risk factor for depression. Furthermore, implementing this basic technology as part of computerized online cognitive training can help improve memory functioning in depression.

Keywords: autobiographical memory, natural language processing, machine learning, overgeneralized autobiographical memory

Past decades of studies have highlighted reduced specificity of autobiographical memories (AMs) as a hallmark of depressive cognition (Williams et al., 2007). These studies typically suggest that patients with depression have difficulty recalling single events experienced in the past, even when they are explicitly asked to do so. Rather, depressed individuals tend to respond with general aspects of events, referring to extended periods of time or categories of similar events that occurred repeatedly in the past (Griffith et al., 2012). Such reduced AM specificity (rAMS; also known as *overgeneral autobiographical memory* [OGM]) is associated with increased levels of depressive symptoms (e.g., van Vreeswijk & de Wilde, 2004; Williams et al., 2007). Findings also suggest that rAMS is a cognitive vulnerability factor for depression, which predicts increases in depressive symptoms and poor prognosis

This article was published Online First April 3, 2017.

Keisuke Takano, Center for the Psychology of Learning and Experimental Psychopathology, University of Leuven; Charlotte Gutenbrunner, School of Psychology, Victoria University of Wellington; Kris Martens, Center for the Psychology of Learning and Experimental Psychopathology, University of Leuven; Karen Salmon, School of Psychology, Victoria University of Wellington; Filip Raes, Center for the Psychology of Learning and Experimental Psychopathology, University of Leuven.

Keisuke Takano was supported by Japan Society for the Promotion of Science Postdoctoral Fellowships for Research Abroad. Karen Salmon was supported by Grant VUW1308 from the Royal Society of New Zealand Marsden Fund awarded to her and Paul Jose. Filip Raes and Kris Martens were supported by the KU Leuven Research Council Grant PF/10/005. We thank Eva Rens and Ruth Glynn for their assistance in data collection and transcribing memories.

Correspondence concerning this article should be addressed to Keisuke Takano. Center for Learning and Experimental Psychopathology, University of Leuven, Tiensestraat 102, Leuven 3000, Belgium. E-mail: keisuke.takano@kuleuven.be

after treatment for depression in both adult and youth populations (Brittlebank, Scott, Williams, & Ferrier, 1993; Hermans et al., 2008; Liu et al., 2016; Peeters, Wessel, Merckelbach, & Boon-Vermeeren, 2002; Rawal & Rice, 2012; for a review, see Sumner, Griffith, & Mineka, 2010).

To assess AM specificity, researchers typically use the Autobiographical Memory Test (AMT; Williams & Broadbent, 1986), in which respondents are presented with emotional cue words (e.g., happy, sad) and asked to describe memories that the cue words remind them of. Each reported memory is manually coded by researchers on the basis of pre-established criteria; a *specific* memory is operationally defined as a memory that refers to an event that occurred at a particular time and place and lasted less than 24 hr (see Williams et al., 2007, for a review). The total number of memories that are identified as specific is used as an index of AM specificity. Although scores coded by experts are highly reliable, manual coding is sometimes a challenge for large epidemiological studies and big data analytics; for example, Heron et al. (2012) analyzed 40,000 memories generated on the AMT, which required two researchers to undertake a full day of training before the actual coding of memories could commence.

One possible solution for this limitation is to develop a computerized algorithm that automatically scores memory specificity, which would contribute to standardization of the coding rules of the AMT. Standardization is crucially important in multisite administration of the AMT, where multiple raters across different laboratories are coding autobiographical memories. A computerized scoring system could also enable immediate "online" feedback in memory specificity training for people with depression (Raes et al., 2009), which involves repeated practice retrieving specific AMs as homework between sessions.

In line with this notion, one recent study has proposed a computerized classifier for Japanese-written autobiographical memories (Takano et al., 2016). In a series of language analyses on more than 10 thousand memories, the researchers identified linguistic features that are particularly important for discrimination between specific and nonspecific memories. For example, specific memories are often written in past tense with adverbial words and phrases pertaining to time and location (e.g., yesterday, in the park). Nonspecific memories, on the other hand, are often written in present tense without temporal and spatial details or with expressions of repetition of an event (e.g., always, sometimes). Based on these linguistic features, the researchers trained a support vector machine (SVM; Cortes & Vapnik, 1995), which is a machine learning model that "learns" a pattern of input information to classify new samples into two distinct categories. Figure 1 shows visual illustrations of an SVM at work (see also Noble, 2006). The goal of training an SVM is to find a "hyperplane" that best separates samples between two classes. Panel A is an example of two-dimensional linear separation, in which a hyperplane is depicted as a "line" discriminating between circle and square samples (e.g., plotting word frequencies of two words to separate between specific and nonspecific memories). When defining such a separating line, the SVM algorithm tries to maximize the margin, which is the distance from any one of the given samples to the hyperplane (Panel A). Many cases, however, are more complex and cannot be resolved by a linear separation. Therefore, the SVM algorithm uses a mathematical trick, by projecting data into a higher dimensional space (Panel B). In this way, the algorithm can identify a hyperplane to separate the samples, even when they are not separable in the original lower dimension. Past research suggests that SVMs are useful in text-classification and often achieve high accuracy comparable to that of human raters (e.g., Pestian, Nasrallah, Matykiewicz, Bennett, & Leenaars, 2010; Sebastiani, 2001).

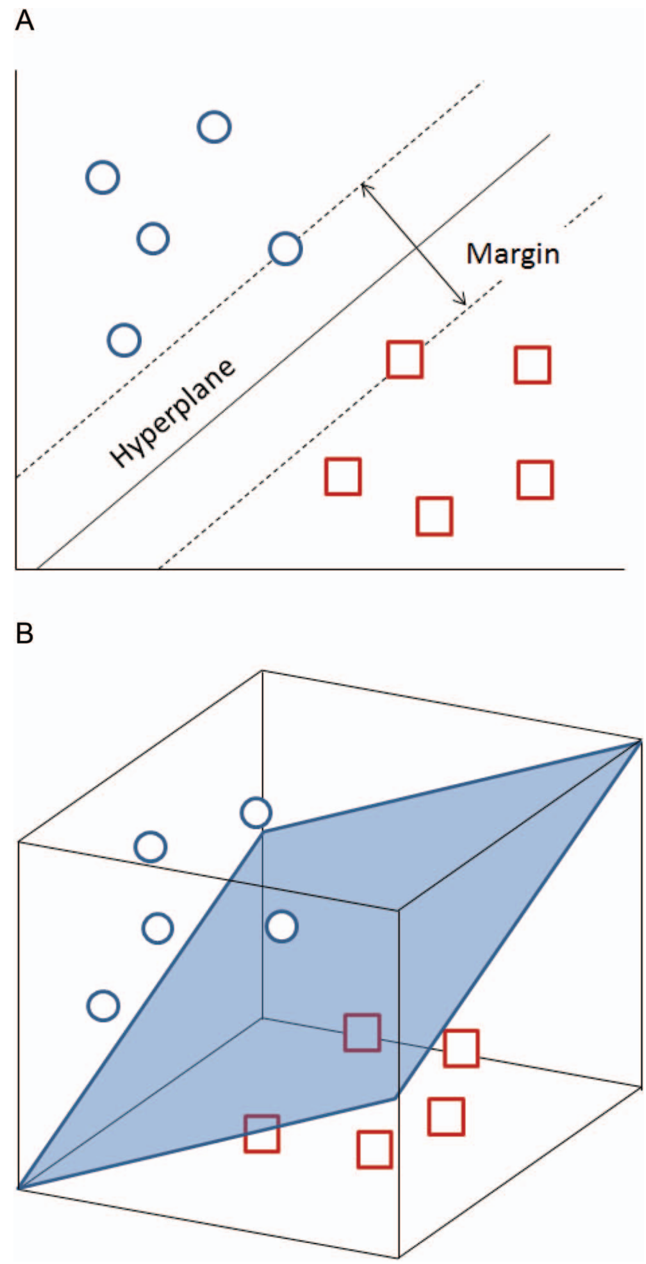

Figure 1. Hyperplane and margin (Panel A) and projection to three dimensional space (Panel B). See the online article for the color version of this figure.

In the AM context, an SVM can be seen as an algorithm to predict AM specificity from particular language features that are observed in individual AMs. In Takano et al. (2016), performance tests showed that the trained SVM reliably classified "unknown" new memory samples at an accuracy level of .92 as the area under the curve in receiver operating characteristic (ROC) analysis.[1] One crucial limitation of this study is that the trained SVM can only be used for Japanese written memories, but not for other languages. Moreover, the SVM was trained with adult memories, which limits the applicability of the SVM scoring algorithm for youth memories.

The current study, therefore, aimed at (a) training SVMs for memories of Dutch-speaking adults and English-speaking adolescent youth and (b) testing their prediction accuracies against manually coded AM specificity. The performance tests of the SVMs entailed two steps: (a) testing the SVM performances in discriminating specific and nonspecific memories that were not used in the SVM training processes and (b) testing the versatility of the SVMs in predicting the specificity of memories that were collected by using different cue words from the memories that were used for SVM training. The rational for the latter test is that cue words presented in the AMT were different across studies (Griffith, Sumner, et al., 2012). Because psychometric studies have shown that the AMT has a unifactorial structure x(i.e., single dimension of AM specificity) across different cue words (Griffith, Kleim, Sumner, & Ehlers, 2012; Griffith et al., 2009) and age groups (cf. Heron et al., 2012), we predicted that the SVM performances should be retained across the two testing data sets regardless of the cue-word differences.

Method

Overview of Data

We prepared three data sets of Dutch written memories, and two data sets of English written memories for SVM training and performance testing. For the Dutch written memories, we used two sets of memory corpus data that had been collected independently using the same procedure (i.e., the written version of the AMT) but different AMT cue words. Therefore, these two sets of corpus data covered different contents of memories due to the cue-word differences. One of the corpus data sets was randomly split into two parts: the training data, which were used for training an SVM, and the testing data, which were used for testing the SVM performance in predicting "unknown" memories that were not used in the SVM training processes. The other corpus data, which consisted of memories that were collected using a different set of cue words, were used for the versatility test of the SVM in predicting AM specificity for different contents of memories. For the English corpus data, we analyzed memories in two data sets, which had been collected using the same procedure (i.e., the written version of the AMT), and the same cue words, but at different time points of a longitudinal study with youth aged 10 to 15 years at the first wave (Gutenbrunner, Salmon, & Jose, 2016). The first corpus (Wave 1) data were used for SVM training, and the second (Wave 2) data were used for performance testing. We did not perform the versatility test for the English SVM.

Dutch Corpus Data

The Dutch written memories were collected using the written, self-report version of the AMT (e.g., Henderson, Hargreaves, Gregory, & Williams, 2002; Raes, Williams, & Hermans, 2009), in which participants wrote down their autobiographical memories in response to 10 emotional cue words. Participants were explicitly instructed to recall a specific memory (i.e., an experience of an event that happened on one particular day) in response to each cue word. No time constraint was imposed when retrieving memories. The same version of the AMT was used for the two Dutch corpus data sets, but cue words differed. For the first corpus data, following 10 emotional (five negative, five positive) cues were used: zelfverzekerd (confident), alleen (alone), bekwaam (capable), wanhopig (desperate), geslaagd (succeeded), jaloers (jealous), verrast (surprised), beschaamd (ashamed), tevreden (satisfied), and gefaald (failed). For the second corpus data, the cue words were as follows: Aangenaam (pleasant), boos (angry), belangstellend (interested), gekwetst (hurt), trots (proud), kwaad (evil), sociaal (social), onhandig (awkward), enthousiast (excited), and ontgoocheld (disappointed).

The first corpus data consisted of 6,390 memories, which were collected as part of a paper-and-pencil survey on community adults in Leuven and its environs. Participants (n = 639; 212 males, 413 females, and 14 unknown) were recruited from social networks. The mean age was 33.6 (SD = 13.7; range: 17–84) years (cf. Raes et al. 2017). These memories were split into two parts: the training (5,390 memories) and testing dataset (1,000 memories). The second corpus data consisted of 4,080 memories, which were collected through an online survey on Dutch-speaking community (n = 408; 56 males and 352 females) with a mean age of 31.8 (SD = 13.5) years. Participants were recruited via social media outlets (i.e., Twitter and Facebook) and email lists of potential participants derived from personal networks of student research assistants. After the completion of the survey, participants drew a lottery for a chance to win a shopping coupon (€20). The second corpus data were used for the versatility test of the SVM.

English Corpus Data

Two sets of English corpus data were imported from a longitudinal study of adolescents living in New Zealand (Gutenbrunner et al., 2016). The memories collected during the first wave of this study were used as a training dataset, which consisted of 2,880 memory responses made by 288 adolescents (153 boys and 135 girls; mean age = 13.87; SD = 1.16 years) on an adapted version of the written minimal instructions version of the AMT (Mi-AMT; Debeer, Hermans, & Raes, 2009) with the following 10 cue words: happy, guilty, proud, scared, excited, angry, lucky, lonely, relaxed, and sad. Data from the second wave, which took place 12 months after the first wave, were used as a testing dataset. The testing dataset consisted of 1,000 memories from 100 adolescents (41 boys, 59 girls) who participated in the first wave. As for the training dataset, the Mi-AMT (with the same set of cue words) was administered at the second wave. All written memories were transcribed for further text analyses. All study protocols were approved by the Social and Societal Ethics Committee of the University of Leuven (for the Dutch corpus data), and by the School of Psychology Human Ethics Committee, under delegated authority to the Victoria University of Wellington's Ethics Committee (for the English corpus data).

[1] The ROC analysis is a diagnostic tool to evaluate the performance of a two-class classifier. A ROC curve is obtained by plotting the hit (sensitivity) rate as a function of the false alarm (1 – specificity) rate for various threshold values (e.g., decision values in an SVM). The area under the ROC curve is used as a performance index of a classifier with a greater value indicating better performance.

Manual Coding of the AMT

All Dutch written memories were manually categorized into five classes following the pre-established criteria (e.g., Griffith, Sumner, et al., 2012; Heron et al., 2012): *specific memory*, which describes a personal event that happened on a particular day; *extended memory*, which describes one particular event that lasted longer than a day; *categoric memory*, which refers to events that happened repeatedly or a category of similar events; *semantic associate*, which is not a memory response but a mere association to the cue word; *omission*, in which participants did not make any response. This manual coding identified 4,105 specific and 2,285 nonspecific memories (652 extended memories, 242 categoric memories, 439 semantic associates, 952 omissions) in the first Dutch corpus data; 3,301 specific and 779 nonspecific memories (66 extended memories, 271 categoric memories, 344 semantic associates and 98 omissions) in the second corpus data. For SVM training and testing, we used dichotomous labels of *specific* versus *nonspecific* memories instead of the full five-class categories. This is because (a) training an SVM for five-class classification requires more samples, particularly for categorical memories (see also Study 3 of Takano et al., 2016); (b) specific memories are the sole "correct" response category in the AMT, whereas all other (four) categories are different types of error responses. Sometimes it is informative to examine distinctive nonspecific (e.g., categorical) memories, but mathematically, the number of a particular type of nonspecific memories is inversely correlated with the number of specific memories (Griffith, Sumner, et al., 2012). To assess interrater agreement, two independent raters manually scored 200 memories. Cohen's kappa for the five-class categorization was .73, and the concordance rate for the binary (specific vs. nonspecific) classification was 92%.

Similarly, each memory in the English corpus data had been manually coded into 11 (5 base + 6 extra) categories by researchers before language analyses. The coding rule followed the same pre-established criteria (Heron et al., 2012), although some extensions were made for the purpose of identifying detailed differences among the five base categories (i.e., specific memory, extended memory, categorical memory, semantic associate, omission). The additional six categories were as follows: *specific* or *extended* memory, referring to responses where the duration of the reported event was ambiguous; *specific* or *category memory,* referring to responses where uniqueness of the reported event was ambiguous; *extended* or *categoric memory*, referring to responses containing a categorical memory within an extended time period (e.g., "On holiday I read a book by the pool"); *future-oriented,* referring to events that would be experienced in the future; *incomplete response,* referring to responses with sufficient detail to suggest participants had thought of a memory but not sufficient detail to code; *repeated memory,* which is a memory that participants have already provided to a different cue.

As in the Dutch data sets, we used dichotomous labels of specific versus nonspecific memories and integrated (a) specific (n = 1,777), (b) specific or extended (n = 139), and (c) specific or categorical memories (n = 82) into a single category of *specific* memories. Note that the AMT coding scheme by Heron et al. (2012) has the "benefit of doubt" rule for ambiguous memories; if raters are uncertain about whether a memory is specific or not, the memory is coded as specific. Other categories (except for repeated memories) were relabeled as *nonspecific* memories in the following analyses. The repeated memories were excluded from the SVM training and performance tests because such memories can include both features of specific and nonspecific memories depending on whether the original memories were specific or not. This manual coding identified 1,998 specific and 873 nonspecific memories in the training data (excluding 9 repeated memories), and 728 specific and 262 nonspecific memories (excluding five repeated memories) in the testing dataset. Cohen's kappa for 200 memories, coded independently by two raters, was .81.

Factor Structure and Language Processing

For the following language analyses, memories were collapsed across cue words. This was considered appropriate on the basis of past research, which suggests that (a) the AMT has a robust unidimensional structure in terms of AM specificity (Griffith et al., 2009, 2012; Heron et al., 2012) and that (b) cue words do not have additive information in predicting AM specificity (Takano et al., 2016). In our first Dutch dataset, we performed exploratory factor analysis on AM specificity across 10 cue words with the principal component analysis method.[2] The first three eigenvalues were 7.2, 0.5, and 0.4, which led us to accept a one-factor solution. For the English dataset, factor analysis also indicated a unifactorial structure, as the first three eigenvalues were 4.4, 1.3, and 0.9.

By using software for text annotating (Frog, for Dutch; Van den Bosch, Busser, Daelemans, & Canisius, 2007; TreeTagger, for English; Schmid, 1994, 1995), we performed tokenization (i.e., breaking a sentence into words), lemmatization (i.e., grouping words in different inflected forms), and part-of-speech (POS) tagging for all Dutch and English memories in the training and testing data sets.[3,4] The tokenization and lemmatization provided a document-token matrix, which reflects how many times individual words were used in each memory response. Similarly, we also created a document-POS matrix, which describes how many times each POS tag appeared in a memory. After this language processing, we identified 5492 different words and 219 POS tags in the first Dutch corpus and 2,243 words and 55 POS tags in the first English corpus.

Feature Selection

To extract words that are most relevant to AM specificity, we compared the frequency of words and POS tags between specific and nonspecific memories. This technique is referred to as *feature selection*, which is of practical importance for text-classification by machine learning to save computational resources needed for model training and to improve classification accuracy by eliminating features that may be noise (e.g., Ikonomakis, Kotsiantis, & Tampakas, 2005; Manning, Raghavan, & Schütze, 2008). We selected words that had (a) sufficient frequency in the training data (i.e., used more than four times) and (b) a large absolute difference in frequency between specific and non-specific memories (i.e., $\chi^2 > 3$). Similarly, we also computed chi-squared statistics for POS tags to find most predictive grammatical structure of AM

[2] We used tetrachoric correlations among the AMT items.

[3] Frog and TreeTagger are free software, distributed via following links: https://languagemachines.github.io/frog/ and http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/.

[4] Through lemmatization, grammatically modified forms of a word (i.e., inflection) were reduced to a common base form (e.g., dogs–dog; plays, played, playing–play).

specificity. Because of the relatively small repertoire of POS tags in TreeTagger (58 tags at maximum), we included all the tags in the SVM training for the English version.

Model Training

We trained a support vector machine (SVM) with Gaussian kernel and fivefold cross-validation, wherein (a) token and type numbers and (b) document-term and document-POS matrices were used as input information in predicting observer-rated AM specificity (i.e., specific or nonspecific). Because both Dutch and English training data were not balanced between specific and nonspecific memories, we put class weights (1:2 for specific and non-specific memories) in the SVM training to reduce potential bias due to the unbalanced sample sizes. The SVM models were tuned through a grid search, which suggested optimal parameters ($\gamma = 0.001$ and cost = 100) both for Dutch and English data sets. This parameter tuning resulted in models with prediction accuracies of 87.7% for Dutch and 86.2% for English memories (averaged across five iterations of cross-validation) in predicting the training samples.

Performance Tests

Performances of the trained SVMs were evaluated on the testing data on the basis of ROC analysis. In ROC analysis, we computed a hit (a ratio of memories that were correctly identified as specific by the SVM relative to memories that were manually scored as specific), correct rejection rate (a ratio of memories that were correctly identified as nonspecific by the SVM relative to memories that were manually scored as nonspecific), and accuracy (all correct predictions by the SVM relative to all samples). Note that the accuracy is biased because of the unbalanced sample sizes between specific and nonspecific memories in the testing data sets; for example, a model predicting all samples as specific should achieve a high accuracy, although it always has a zero correct rejection rate. Therefore, we presented the hit and correct rejection rates in the performance tests. Next, we depicted a ROC curve with plotting hit rates as a function of false alarm rates (= 1 – correct rejection rates) for every possible cutoff point of decision values of the SVM. An SVM returns a numeric decision value as an initial output for each testing sample, which is transformed into a binary output (labels of *specific* and *nonspecific*) via a sign function. The ROC curve provides another performance index, namely area under the curve (AUC), which describes a general performance of a classifier that is independent of cut-off points, with 0.5 indicating a random separation and with 1.0 indicating a perfect separation. SVM training and performance testing were conducted by using the R packages, e1071 (Meyer, Dimitriadou, Hornik, Weingessel, & Leisch, 2014) and ROCR (Sing, Sander, Beerenwinkel, & Lengauer, 2005).[5]

Results

Feature Selection

We identified 227 tokens and 55 POS tags that meet the feature-selection criteria in the Dutch training dataset (see Tables A1 and A2 in the Appendix). For the English training data, 179 tokens and 26 POS tags met the selection criteria (see Tables A3 and A4 in the Appendix). Each of the training data nominated adverbs (e.g., sometimes [soms], always [altijd], yesterday [gisteren]) and conjunctions (e.g., when [toen]) that convey temporal information as important features to distinguish between specific and nonspecific memories. Another important feature was the verb tense; the past tense verbs (e.g., was, did, played) were strong indications of specific memories, whereas the present tense (e.g., is, do, play) was a typical verb form in nonspecific memories. Although there were substantial overlaps in the selected features between Dutch and English data sets, this feature selection process identified some corpus-specific (or region-specific) proper nouns; for example, "Australia" and "Rorotonga" were only found in English (New Zealand) data sets.

Model Performances in Predicting AM Specificity

First, we tested the prediction accuracy of the SVM for Dutch written memories. The model performance was evaluated on the two testing data sets: one with memories collected using the same cue words as the training dataset (n = 1,000), and one with different cue words (n = 4080). The ROC analysis (see Table 1) revealed that the SVM performed well in classifying memories of the same-cue data (AUC = .936), as well as different-cue data (AUC = .901). Similarly, we repeated the ROC analysis for the English SVM, which also demonstrated good performance in discriminating between specific and nonspecific memories (AUC = .904). These results indicate that the trained SVMs demonstrate reliable performances across different languages, different sets of cue words, and different age groups. This suggests that AM specificity can be determined on the basis of basic vocabularies (e.g., sometimes, always, yesterday) and grammatical information (e.g., past tense) across different languages and memory contents.

Second, we tested whether the outputs (i.e., decision values) of the trained SVMs capture the gradient of AM specificity across different AMT categories. Given that the difference between specific and extended memories is mention of an event's duration (within or longer than 24 hr), features of specific memories should be more similar to those of extended memories than other AMT categories. We depicted a boxplot illustrating distributions of the decision values of the Dutch SVM across the five AMT categories in Figure 2. The Dutch SVM gave higher scores for observer-rated specific than categoric memories, and in-between scores for extended memories. Note that "omissions" were typically "nothing," "no memory," "not applicable," and empty responses. Because of this little variation in the vocabulary (i.e., few response patterns), the variance of the decision value in "omissions" tends to be smaller than in the other categories. We also depicted a similar boxplot for the boundary memories, namely *specific* or *extended* and *specific* or *categorical* memories in the English testing dataset. Figure 3 shows that the boundary memories are scored in between specific versus extended or categorical memories. These results imply that the trained SVMs well capture the extent to which each memory has the features of specific memories.

Discussion

In the present study, we demonstrated the feasibility of computerized algorithms to score AMT responses across different languages. Both SVMs that we trained on Dutch and English corpus data showed

[5] The models trained here can be found via Open Science Framework: https://osf.io/3p9g6/?view_only=dc2aa0f9d6024bf0ac8d0e33cb392d40.

Table 1
Performance Tests of support vector machines (SVMs) for Dutch and English Memories

Testing data	*N* (*N* of specific memories)	Hit rate	Correct rejection rate	Accuracy	AUC
Dutch testing data (same cues)	1,000 (649)	.895	.846	.878	.936
Dutch testing data (different cues)	4,080 (3301)	.885	.769	.863	.902
English testing data (same cues)	995 (728)	.901	.742	.858	.904

Note. AUC = area under the curve.

good performances (AUC > .90) against observer-rated scores in discriminating between specific and nonspecific memories. Furthermore, the model performances were retained in the versatility test, in which the SVM predicted the specificity of memories that were generated in response to different sets of cue words from the training dataset. These results suggest that performances of the trained SVMs are reliable and robust across different contents of memories.

Along with the previous findings from Japanese written memories (Takano et al., 2016), the feature selection processes in Dutch and English corpus data identified around 200 tokens and 50 POS tags as potentially relevant features to discriminate between specific and nonspecific memories. Many of the selected words pertain to temporal details of a memory (e.g., always, yesterday, birthday), and furthermore, the past (and present) verb tense is a critical feature to deem a memory more (or less) specific. There were also other factors that contributed to memory specificity; for example, verbs pertaining to actions (i.e., get, come, go), nouns indicating places (e.g., pool) and events (e.g., party, festival), and cardinal and ordinal numbers (e.g., 2009, first day of school). Because a specific memory has to describe an event that happened on one particular day, it should be written in past tense with temporal and spatial information. Such information about when, where, and what would be a critical factor in AMT scoring, although the contribution of "with whom" information is less clear in the current data (e.g., brother [broer] appears in the Dutch list, which was not in the English list; see Tables A1 and A3 in the Appendix). Overall, as the SVMs trained on these selected features showed good performances in categorizing test samples, we can conclude that AM specificity is mostly determined by a fairly small number of words and phrases in the current study's Dutch adult and English youth samples.

This is a particularly important point when it comes to scoring memory responses of young people, who typically have a smaller vocabulary and provide fewer specific details than adults (e.g., Bloom & Markson, 1998; Willoughby, Desrocher, Levine, & Rovet, 2012). Given that our English corpus data consisted of memories provided by young adolescents, our findings regarding the determinants of AM specificity are likely to be robust across different levels of language abilities. Related to this point, a study that investigated response format (written or oral) of the AMT in children showed that the written-oral difference (Glynn, Salmon, & Jose, 2016) does not influence AM specificity, although spoken memories were longer and contained more details than written memories (see also Marinellie, 2009). These findings suggest that AM specificity can be determined independently of the age of participants and the associated features of their memories and also of reporting format. It seems, therefore, that AM specificity may be influenced by basic linguistic expressions that are evident across ages and reporting modes.

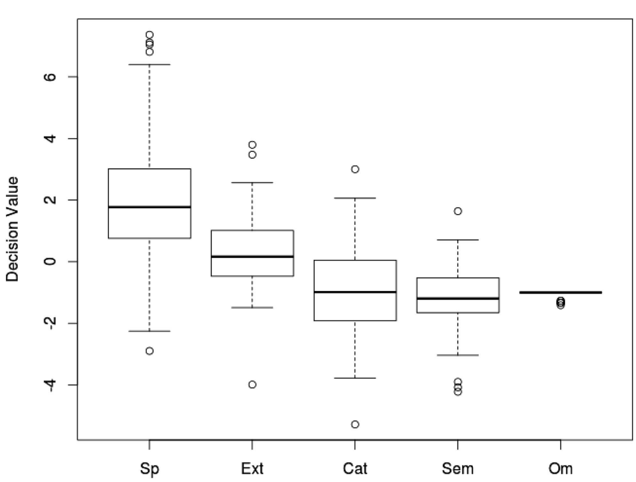

Figure 2. Distribution of Decision Values of the Support Vector Machine (Dutch) for Manually Coded Specific Memories (Sp; n = 649), Extended Memories (Ext; n = 36), Categoric Memories (Cat; n = 99), Semantic Associates (Sem; n = 71), and Omission (Om; n = 145). In binary coding, memories with positive decision values should be classified into the specific category, whereas those with negative values should be into the nonspecific category.

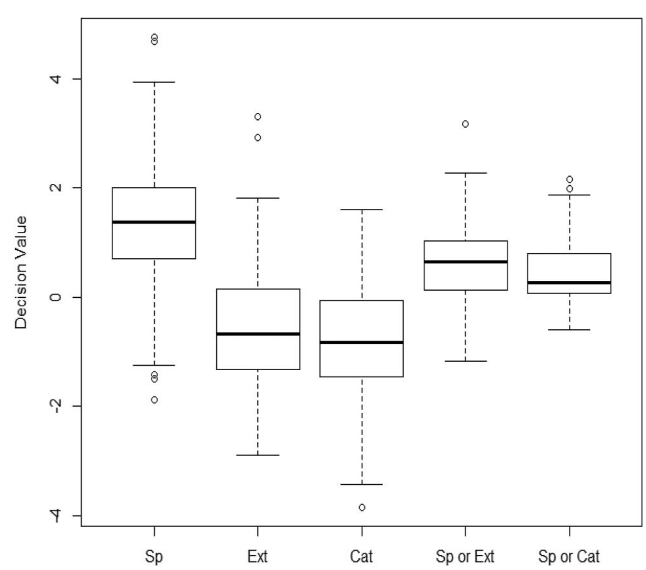

Figure 3. Distribution of Decision Values of the Support Vector Machine (English) for Manually Coded Specific (Sp), Extended (Ext), Categoric (Cat), Specific-or-extended, and Specific-or-categoric Memories.

Although overall results of the feature selection were similar between the Dutch and English corpus data, we also found language-specific (region-specific, more precisely) features in proper nouns; for example, Australia and Rarotonga, which are only used in memories collected in New Zealand, not in Belgium. This finding suggests that fine tuning of the SVM (e.g., by installing usage patterns of proper nouns that are commonly used in the region where a survey or experiment is conducted) might be important to achieve the highest model performance. Thus, it would be of practical importance to clarify the scope of application of the computerized algorithms that we proposed here; (a) if new memories to be scored are collected from Flemish people, the Dutch version of the algorithm can be directly applied without any calibration, which should achieve a comparable level of prediction accuracy to the current validation data; (b) the Dutch algorithm seems to be robust across different cue words, as long as negative and positive emotional adjectives are used as AMT cues (not covering neutral words or nonadjective words); (c) the English algorithm was validated in an adolescent sample in New Zealand using cue words identical to the training data. Calibration will be needed when researchers apply the algorithm to scoring memories of people in other regions (e.g., US, U.K.) and/or with using different sets of cue words. In general, if a new study deviates from the current validation conditions (e.g., AMT instructions, cue words, background of respondents), we recommend testing the consistency between machine and hand scoring for a subsample of memory data. In addition, we believe that manual coding remains important and informative, particularly when researchers want to analyze the "error" patterns in the AMT (e.g., categoric memories); the computerized algorithms can discriminate between specific versus other response categories, but not between nonspecific memories.

Decision values of the SVMs showed a gradient corresponding to the similarities and differences among the AMT categories. The Dutch SVM gave highest values for observer-rated specific memories, and often scored extended memories in-between specific and categorical memories. The English SVM also showed a similar gradient, in which "ambiguous" memories, namely *specific* or *extended* and *specific* or *categorical* memories, were scored in-between specific and the other two memory categories. Indeed, the difference between the specific and "extended" categories is whether a memory refers to an event that occurred within 24 hr or that lasted longer than one day. In this regard, a specific memory should have similar linguistic features (e.g., written in paste tense) to an extended memory, except for the temporal information; it should refer to one particular day (e.g., on May 1st at 7 o'clock) but not an extended period of time (e.g., on vacation, during summer). Therefore, it seems that decision values may reflect the gradient of AM specificity, that is, the extent to which each memory has specific language features.

It should be noted, however, that due to the complexities inherent to SVM calculation processes, decision values do not provide psychologically meaningful information above and beyond indication of dichotomous class membership. Moreover, decision values are not standardized, which makes it difficult to interpret the values and may be a challenge when aggregating the SVM outputs across AMT items to calculate a single test index of the AMT. As a possible solution for this scaling issue, researchers have proposed the logit transformation (Platt, 2000), which is useful for interpretation of decision values as probabilities that each testing sample belongs to a given category. It is also claimed that an SVM is optimized for binary classification, and thus, decision values would not retain numeric or probabilistic information beyond the dichotomous class membership—whether a sample belongs to Class A or B (cf. Franc, Zien, & Schölkopf, 2011). Therefore, although the decision values and transformed probabilities have potential as a continuous measure of AM specificity, further validation is needed.

Although the performance tests indicated high reliability of machine-coded scores, the trained SVMs made classification errors for approximately 10% of memories in the testing data sets. These errors can mainly be attributed to false alarms, as the correct rejection rates were relatively low ($< .80$) for the Dutch testing dataset with different cues and the English testing dataset. Because our training data sets had a smaller number of nonspecific than specific memories, the amount of input information may not have been sufficient for the SVMs to fully learn the unique language patterns for nonspecific memories. Future research needs to improve the prediction accuracy by adding corpus data, particularly for nonspecific memory samples.

To conclude, the current study demonstrated the feasibility of machine scoring for Dutch and English versions of the AMT. The trained algorithms demonstrated good performances in discriminating specific and nonspecific memories. Furthermore, decision values of the SVMs indicated a gradient across various pre-established AMT response categories, potentially representing a measure of degree of AM specificity. We hope that the algorithms will facilitate big-data research in the field of autobiographical memory and psychopathology (or epidemiology) by means of reducing time spent on manually scoring transcribed memory responses. Furthermore, the computerized classification algorithms open up new possibilities for online clinical interventions that specifically target overgeneralized autobiographical memory, such as memory specificity training (Raes et al., 2009).

References

Bloom, P., & Markson, L. (1998). Capacities underlying word learning. *Trends in Cognitive Sciences, 2,* 67–73. http://dx.doi.org/10.1016/S1364-6613(98)01121-8

Brittlebank, A. D., Scott, J., Williams, J. M., & Ferrier, I. N. (1993). Autobiographical memory in depression: State or trait marker? *The British Journal of Psychiatry, 162,* 118–121. http://dx.doi.org/10.1192/bjp.162.1.118

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20,* 273–297. http://dx.doi.org/10.1023/A:1022627411411

Debeer, E., Hermans, D., & Raes, F. (2009). Associations between components of rumination and autobiographical memory specificity as measured by a Minimal Instructions Autobiographical Memory Test. *Memory, 17,* 892–903. http://dx.doi.org/10.1080/09658210903376243

Franc, V., Zien, A., & Schölkopf, B. (2011). Support vector machines as probabilistic models. *Proceedings of the 28th International Conference on Machine Learning*. Retrieved from https://pdfs.semanticscholar.org/e87a/69b643dd35d1f7e7da4cc11c239311d63d78.pdf

Glynn, R., Salmon, K., & Jose, P. (2016). The influence of reporting mode on children's cued personal memories. *Memory, 24,* 1224–1230. http://dx.doi.org/10.1080/09658211.2015.1088034

Griffith, J. W., Kleim, B., Sumner, J. A., & Ehlers, A. (2012). The factor structure of the Autobiographical Memory Test in recent trauma survivors. *Psychological Assessment, 24,* 640–646. http://dx.doi.org/10.1037/a0026510

Griffith, J. W., Sumner, J. A., Debeer, E., Raes, F., Hermans, D., Mineka, S., . . . Craske, M. G. (2009). An item response theory/confirmatory factor analysis of the Autobiographical Memory Test. *Memory, 17,* 609–623. http://dx.doi.org/10.1080/09658210902939348

Griffith, J. W., Sumner, J. A., Raes, F., Barnhofer, T., Debeer, E., & Hermans, D. (2012). Current psychometric and methodological issues in the measurement of overgeneral autobiographical memory. *Journal of Behavior Therapy and Experimental Psychiatry, 43*(Suppl. 1), S21–S31. http://dx.doi.org/10.1016/j.jbtep.2011.05.008

Gutenbrunner, C., Salmon, K., & Jose, P. E. (2016). Do overgeneral autobiographical memories predict increased psychopathological symptoms in community youth? A 3-year longitudinal investigation. *Journal of Abnormal Child Psychology*. Advance online publication. http://dx.doi.org/10.1007/s10802-017-0278-5

Henderson, D., Hargreaves, I., Gregory, S., & Williams, J. M. G. (2002). Autobiographical memory and emotion in a non-clinical sample of women with and without a reported history of childhood sexual abuse. *British Journal of Clinical Psychology, 41,* 129–141. http://dx.doi.org/10.1348/014466502163921

Hermans, D., Vandromme, H., Debeer, E., Raes, F., Demyttenaere, K., Brunfaut, E., & Williams, J. M. (2008). Overgeneral autobiographical memory predicts diagnostic status in depression. *Behaviour Research and Therapy, 46,* 668–677. http://dx.doi.org/10.1016/j.brat.2008.01.018

Heron, J., Crane, C., Gunnell, D., Lewis, G., Evans, J., & Williams, J. M. G. (2012). 40,000 memories in young teenagers: Psychometric properties of the Autobiographical Memory Test in a UK cohort study. *Memory, 20,* 300–320. http://dx.doi.org/10.1080/09658211.2012.656846

Howe, M. L., & Courage, M. L. (1997). The emergence and early development of autobiographical memory. *Psychological Review, 104,* 499–523. http://dx.doi.org/10.1037/0033-295X.104.3.499

Ikonomakis, M., Kotsiantis, S., & Tampakas, V. (2005). Text classification using machine learning techniques. *WSEAS Transactions on Computers, 4,* 966–974.

Liu, Y., Zhang, F., Wang, Z., Cao, L., Wang, J., Na, A., . . . Zhao, X. (2016). Overgeneral autobiographical memory at baseline predicts depressive symptoms at follow-up in patients with first-episode depression. *Psychiatry Research, 243,* 123–127. http://dx.doi.org/10.1016/j.psychres.2016.06.029

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511809071

Marinellie, S. A. (2009). The content of children's definitions: The oral-written distinction. *Child Language Teaching and Therapy, 25,* 89–102. http://dx.doi.org/10.1177/0265659008098662

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2014). *e1071: Misc Functions of the Department of Statistics (e1071)*. R package version 1.6–4. Retrieved from http://CRAN.R-project.org/package=e1071

Noble, W. S. (2006). What is a support vector machine? *Nature Biotechnology, 24,* 1565–1567. http://dx.doi.org/10.1038/nbt1206-1565

O'Carroll, R. E., Dalgleish, T., Drummond, L. E., Dritschel, B., & Astell, A. (2006). Effects of age, dysphoria, and emotion-focusing on autobiographical memory specificity in children. *Cognition and Emotion, 20,* 488–505. http://dx.doi.org/10.1080/02699930500341342

Peeters, F., Wessel, I., Merckelbach, H., & Boon-Vermeeren, M. (2002). Autobiographical memory specificity and the course of major depressive disorder. *Comprehensive Psychiatry, 43,* 344–350. http://dx.doi.org/10.1053/comp.2002.34635

Pestian, J., Nasrallah, H., Matykiewicz, P., Bennett, A., & Leenaars, A. (2010). Suicide note classification using natural language processing: A content analysis. *Biomedical Informatics Insights, 2010,* 19–28. http://dx.doi.org/10.4137/BII.S4706

Platt, J. C. (2000). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In A. J. Smola, P. Bartlett, B. Schölkopf, & D. Schuurmans (Eds.), *Advances in large margin classifiers*. Cambridge, MA: MIT Press.

Raes, F., Griffith, J. W., Vanbrabant, K., Nelis, S., Smets, J., Craeynest, M., . . . Hermans, D. (2017). *Negative overgeneralization to the self and overgeneral autobiographical memory: Independent predictors of a more severe course of depression*. Manuscript in preparation.

Raes, F., Williams, J. M. G., & Hermans, D. (2009). Reducing cognitive vulnerability to depression: A preliminary investigation of MEmory Specificity Training (MEST) in inpatients with depressive symptomatology. *Journal of Behavior Therapy and Experimental Psychiatry, 40,* 24–38. http://dx.doi.org/10.1016/j.jbtep.2008.03.001

Rawal, A., & Rice, F. (2012). Examining overgeneral autobiographical memory as a risk factor for adolescent depression. *Journal of the American Academy of Child and Adolescent Psychiatry, 51,* 518–527. http://dx.doi.org/10.1016/j.jaac.2012.02.025

Schmid, H. (1994). *Probabilistic part-of-speech tagging using decision trees*. Stuttgart, Germany: University of Stuttgart. Retrieved from https://pdfs.semanticscholar.org/bd0b/ab6fc8cd43c0ce170ad2f4cb34181b31277d.pdf

Schmid, H. (1995). Improvements in part-of-speech tagging with an application to German. In *Proceedings of the ACL SIGDAT-Workshop* (pp. 47–55). Dublin, Ireland: European Chapter of the Association for Computational Linguistics.

Sebastiani, F. (2001). Machine learning in automated text categorization. *ACM Computing Surveys, 34,* 1–47. http://dx.doi.org/10.1145/505282.505283

Sing, T., Sander, O., Beerenwinkel, N., & Lengauer, T. (2005). ROCR: Visualizing classifier performance in R. *Bioinformatics, 21,* 3940–3941. http://dx.doi.org/10.1093/bioinformatics/bti623

Sumner, J. A., Griffith, J. W., & Mineka, S. (2010). Overgeneral autobiographical memory as a predictor of the course of depression: A meta-analysis. *Behaviour Research and Therapy, 48,* 614–625. http://dx.doi.org/10.1016/j.brat.2010.03.013

Takano, K., Ueno, M., Moriya, J., Mori, M., Nishiguchi, Y., & Raes, F. (2016). Unraveling the linguistic nature of specific autobiographical memories using a computerized classification algorithm. *Behavior Research Methods*. Advance online publication. http://dx.doi.org/10.3758/s13428-016-0753-x

Van den Bosch, A., Busser, G. J., Daelemans, W., & Canisius, S. (2007). An efficient memory-based morphosyntactic tagger and parser for Dutch. In F. van Eynde, P. Dirix, I. Schuurman, and V. Vandeghinste (Eds.), *Selected papers of the 17th Computational Linguistics in the Netherlands Meeting* (pp. 99–114). Leuven, Belgium: University of Leuven.

Van Vreeswijk, M. F., & De Wilde, E. J. (2004). Autobiographical memory specificity, psychopathology, depressed mood and the use of the Autobiographical Memory Test: A meta-analysis. *Behaviour Research and Therapy, 42,* 731–743. http://dx.doi.org/10.1016/S0005-7967(03)00194-3

Williams, J. M. G., Barnhofer, T., Crane, C., Herman, D., Raes, F., Watkins, E., & Dalgleish, T. (2007). Autobiographical memory specificity and emotional disorder. *Psychological Bulletin, 133,* 122–148. http://dx.doi.org/10.1037/0033-2909.133.1.122

Williams, J. M. G., & Broadbent, K. (1986). Autobiographical memory in suicide attempters. *Journal of Abnormal Psychology, 95,* 144–149. http://dx.doi.org/10.1037/0021-843X.95.2.144

Willoughby, K. A., Desrocher, M., Levine, B., & Rovet, J. F. (2012). Episodic and semantic autobiographical memory and everyday memory during late childhood and early adolescence. *Frontiers in Psychology, 3,* 53. http://dx.doi.org/10.3389/fpsyg.2012.00053

Appendix

Feature Selections in Dutch and English Corpus Data

Words and part-of-speech (POS) tags that should be most relevant for discriminating between specific and non-specific memories were selected on the basis of chi-squared statistics, which indicate the extent to which the occurrence of the term (i.e., words or POS tags) and class (i.e., specific or non-specific memories) are independent (see Tables A1 through A4 in the Appendix). Odds ratios were also computed as relative frequencies of the words (and POS tags) used in specific versus nonspecific memories, with values greater than 1 indicating that the words were used more frequently in specific than in nonspecific memories.

Table A1
Words Used in Support-Vector-Machine Training (Dutch) with Chi-Square Statistics and Odds Ratios

Token	English translation	Chi-square	Odds ratio
Als	If, as	345.90	.17
Toen	When	187.48	4.06
Soms	Sometimes	150.17	.03
Nooit	Never	128.07	.09
Kind	Child	74.60	.28
Altijd	Always	72.05	.08
Niet	Not	62.25	.50
Mens	Man	59.18	.19
Werk	Work	57.91	.37
Kunnen	Can	52.77	.42
Je	You	51.65	.18
Herinneren	Remember	49.66	.06
Leven	Life	48.69	.16
Steeds	Ever	45.87	.07
Dat	That	41.63	1.85
Elk	Each	41.51	.07
(Niet) van toepassing	(Not) applicable	39.28	.00
Een	A	38.09	1.47
De	The	37.77	1.54
Jaloers	Jealous	34.12	.37
Wat	What	31.55	.31
Alleen	Alone	29.91	.49
Krijgen	Get	26.99	2.82
Bekwaam	Competent	26.78	.42
Vorig	Last	26.48	3.18
Dag	Day	23.72	2.22
Doen	Do	21.88	.49
Zelden	Rarely	21.82	.00
Tevreden	Satisfied	20.97	.46
In	In	20.80	.69
Dagelijks	Daily	20.76	.04
Telkens	Constantly	20.76	.04
Gezond	Healthy	20.75	.07
Iets	Something	20.54	.42
Nu	Now	19.39	.24
Over	About	19.39	.51
Ding	Thing	19.31	.17
Pijn	Pain	18.25	.11
Zijn	Be/his	17.31	.79
Gebeurtenis	Event	16.76	.08
Geleden	Ago	16.24	3.44
Gisteren	Yesterday	16.06	
Hun	Their	16.01	.30
Verjaardag	Birthday	15.34	9.75
Partner	Partner	15.05	.34
Gewoon	Just	14.92	.18

(*Appendix continues*)

Table A1 (*continued*)

Token	English translation	Chi-square	Odds ratio
Horen	Hear	14.87	4.91
Weer	Again, weather	14.56	.29
Graag	Gladly	14.26	.26
Wanhopig	Desperate	13.04	.49
Echtgenoot	Husband	12.97	.23
Zoveel	So many	12.97	.23
Sommige	Some	12.86	.09
Bijvoorbeeld	For example	12.86	.09
Beletselteken (. . .)	Ellipsis (. . .)	12.74	.36
Wanneer	When	12.60	.62
Alles	All	12.45	.42
opvoeding	Education	12.41	.06
Meestal	Mostly	12.41	.06
Pesterij	Harassment	12.41	.06
Week	Week	12.22	2.19
Diploma	Diploma	12.08	4.31
Leren	Learn	12.06	.30
Of	Or	12.06	.39
Wonen	Live	11.91	.20
Mezelf	Myself	11.67	.23
Geen	No	11.04	.53
Eens	Once	10.88	.35
Voor	Before	10.69	1.49
Bezig	Busy	10.62	.20
Trekken	Pull	10.62	.20
Ben	(I) am	10.38	.23
Vriend	Friend	10.25	1.65
Studie	Study	10.20	.33
Sport	Sport	10.15	.14
Mijzelf	Myself	10.15	.14
Dit	This	9.92	.54
Tijd	Time	9.60	.43
Lijken	Appear	9.35	.19
Schoolresultaat	Grades	9.12	.11
Druk	Busy, crowded	9.08	.23
Feest	Festival	8.98	4.13
Kijken	Look	8.70	.33
Van	Of, from	8.49	1.25
Laat	Late	8.49	2.35
Ander	Other	8.44	.58
Jaloezie	Jealousy	8.09	.18
Tijdens	During	8.02	.65
Halen	Get	7.80	4.51
Vaak	Often	7.78	.23
Schamen	Ashamed	7.78	.23
We	We	7.70	2.43
Relatie	Relation	7.59	.48
Voelen	Feel	7.54	.77
Stoppen	Stop	7.52	.34
Vroeg	Early	7.51	.38
Thuis	Home	7.22	.53
Keer	Time	7.09	2.20
Komen	Come	7.09	1.99
Hij	He	6.91	2.07
Avond	Evening	6.90	3.17
Geslaagd	Succeeded	6.90	.31
Voldoende	Sufficiently	6.84	.17
Bereiken	Reach	6.84	.17
Algemeen	General	6.84	.17
Zeker	Certainly	6.84	.40
Rijbewijs	License	6.65	8.71
Job	Job	6.64	.42
Zelfzeker	Confident	6.37	.36

(*Appendix continues*)

Table A1 (*continued*)

Token	English translation	Chi-square	Odds ratio
Gevoel	Feeling	6.30	.51
Verrassen	Surprise	6.09	2.03
Praten	Talk	6.03	.39
1ste	1st	5.96	
Jaar	Year	5.94	1.51
Veel	Many	5.83	.68
Falen	Fail	5.79	.62
Wel	Well	5.78	.56
Zaterdag	Saturday	5.73	
Tegen	Against	5.68	3.21
Nog	Yet	5.68	.60
Blijven	Stay	5.65	.44
Stil	Quiet	5.64	.15
Gedacht	Thought	5.64	.15
Hobby	Hobby	5.64	.15
Verblijven	Stay	5.64	.15
Rustig	Quietly	5.64	.15
Voldoen	Pay	5.64	.15
Emotioneel	Emotionally	5.64	.15
Verpleegkundige	Nurse	5.64	.15
Ervaren	Experience	5.64	.15
Redden	Save	5.64	.15
Achter de rug	Behind	5.64	.15
Vragenlijst	Questionnaire	5.64	.15
Professioneel	Professional	5.64	.15
Sportief	Sportive	5.64	.15
Brengen	Bring	5.63	.38
Blijken	Prove	5.55	4.70
Sterven	Die	5.53	7.57
Aflopen	Expire	5.50	
Vakantie	Vacation	5.45	.50
Teveel	Too much	5.41	.31
Al	Already	5.41	.63
Herexamen	Re-examination	5.34	4.59
Vrijdag	Friday	5.27	
Zondag	Sunday	5.27	
Om	To	5.24	.70
Gezin	Family	5.24	.37
In orde	Alright	5.21	.27
Doorbrengen	Spend	5.19	.23
Verlof	Leave	5.19	.23
Resultaat	Result	5.07	3.52
Tot	Until	5.07	.48
Statistiek	Statistics	5.04	
Staan	Stand	4.94	2.14
Middelbaar	Middle	4.89	2.98
Periode	Period	4.87	.36
Zwanger	Pregnant	4.86	6.88
Mijn	My	4.74	1.13
Auto	Auto	4.69	2.92
Beseffen	Realize	4.58	
Buizen	Tube, fail	4.58	
Me	Me	4.54	.82
Gevoel	Feeling	4.51	.34
Situatie	Situation	4.51	.34
Vertellen	Tell	4.44	2.06
Bellen	(Make a phone) call	4.36	
Rijexamen	Driving test	4.28	4.01
Probleem	Problem	4.27	.46
Slagen	Succeed	4.26	1.48
Afstuderen	Graduation	4.20	6.19
Kamp	Camp	4.20	.32
Dan	Than	4.17	.58
Ver	Far	4.15	.39

(*Appendix continues*)

Table A1 (*continued*)

Token	English translation	Chi-square	Odds ratio
Enkel	Just	4.15	2.33
Haar	Her, hair	4.14	1.77
Vernemen	Learn	4.13	
Evaluatie	Evaluation	4.13	
Gaan	Go	3.98	1.39
Onlangs	Recently	3.97	5.96
Fietsen	Bicycle	3.95	.29
Kleinkind	Grandchild	3.95	.29
Hangen	Hang	3.95	.29
Nicht	Niece	3.90	
Onderscheiding	Award	3.90	
Zetel	Seat	3.89	.23
Minder	Less	3.89	.23
Kids	Kids	3.89	.23
Rij	Drive	3.89	.23
Figuur	Figure	3.89	.23
Best	Best	3.89	.23
Dieet	Diet	3.89	.23
Erop	On	3.89	.23
Jongen	Boy	3.89	.23
Richting	Direction	3.89	.23
Moment	Moment	3.78	2.13
Opleiding	Education	3.75	.38
Lukken	Do/go well	3.75	.38
2009	2009	3.67	
Broek	Pants	3.67	
Eindelijk	Finally	3.67	
Broer	Brother	3.66	2.98
Kopen	Buy	3.66	2.98
Zeggen	Say	3.66	1.75
Onderwijs	Education	3.54	5.50
Overlijden	Death	3.47	2.90
Behalen	Achieve	3.46	2.18
Het	The	3.44	1.15
Opmerking	Remark	3.44	
Reis	Journey	3.42	.46
Huwelijk	Wedding	3.37	.52
Rond	Around	3.37	.37
Positief	Positive	3.24	3.44
Laatst	Last	3.21	
Bus	Bus	3.21	
Genoeg	Enough	3.21	
2e	2nd	3.21	
Iedereen	Everybody	3.19	.57
Aan het	(be do)-ing	3.18	2.46
Zelf	Self	3.12	.56
Houden	Hold, keep	3.11	.42
Vrij	Free	3.11	.42
Plannen	Plan	3.11	.42
Duidelijk	Clear	3.01	.34
Gelukken	Success	3.01	.34
Reden	Reason	3.01	.34
Liefde	Love	3.01	.34
Direct	Directly	3.01	.34

(*Appendix continues*)

Table A2
List of Part-of-Speech Tags Used in Support-Vector-Machine Training (Dutch) with Chi-Square Statistics and Odds Ratios

Part-of-Speech tag	Chi-square	Odds ratio	Examples
Verb, present, 1st person, singular	1002.49	.18	kom, speel
Punctuation	599.14	.52	".", ". . .", "?"
Verb, past, 1st person, singular	517.80	3.24	kwam, speelde
Noun, singular	82.05	1.38	die stoel, elke avond, deze muziek
Verb, present, 3rd person singular	81.55	.33	komt, speelt
Adverb	77.22	.70	zoals, gisteren
Article, definite	54.74	1.67	de hond(en), de kinderen
Verb, present, plural	52.91	.36	komen, spelen.
Conjugation	50.88	1.43	ze komt niet, omdat ze zich niet goed voelt
Ordinal number	46.86	3.23	de vierde man
Noun, plural	44.14	.68	stoelen, kinderen, hersenen
Pronoun, 2nd person, reduced	42.15	.16	Je
Article, indefinite	36.20	1.48	een kind
Verb, past participle, without "e"	30.93	1.48	is gekomen. Een boom versierd met slingers.
Proper noun	29.64	2.03	Den Haag
Wh-pronoun, interrogative	28.98	.29	wat ik niet begrijp is
Noun, singular, diminutive	26.54	2.33	dit stoeltje, op 't nippertje
Adjective, free position	25.47	.78	die stok is lang. Lang slapen.
Verb, infinitive	25.45	.76	zal komen
Cardinal number	24.92	3.10	veertig worden. Zoveel sneller, pagina vijf, de jaren zestig, zes juli
Verb, past, plural	23.44	2.31	kwamen, speelden.
Pronoun, indefinite, determined	21.27	.14	elke hond, iedere keer
Pronoun, indefinite, pronominal	20.90	.54	alles, (n)iets, niks, wat, zoiets
Interjections	18.81	.07	oei, amai, uh hoera
Preposition + Noun	17.90	.26	Niet van toepassing, met een lepeltje. Met Jan in het hospitaal. Met zo te roepen.
Reflexive pronoun, 1st person, singular	17.16	.23	mezelf, mijzelf
Possessive pronoun, 1st person, singular	13.48	1.22	mijn paard(en)
Nominal adjective, plural (with -n)	12.30	.30	de rijken
Proper noun, abbreviation	12.06	.36	inc. ph.d.
Possessive pronoun, 3rd person, plural	10.59	.35	hun paarden
Adjective, pre-noun, superlative	10.05	1.99	de mooiste keuken, het grootste paard
Possessive pronoun, 3rd person, singular	9.67	2.09	zijn paard(en), haar kind
Pronoun, 1st person, plural, reduced	9.38	2.65	we
Pronoun, 3rd person, singular, male	8.74	2.25	hij
Noun, singular, 'onzijdig'	8.61	1.16	het kind, ons huis, dat brood
Indefinite pronoun	8.18	.58	geen kind(eren)
Demonstrative pronouns	7.78	3.81	dat boek, dit dier,
Relative pronoun	7.69	.65	de man die daar staat
Verb, past participle, pronominal	7.52	.37	een getemde feeks
Cardinal number, pre-noun	7.37	1.48	vier cijfers
Article, indefinite	7.04	1.25	het kind, in 't geniep
Adverb + adverb	6.68	.29	niet alleen
Pronoun, 3nd person, female, reduced	6.32	2.79	ze
Possessive pronoun (2nd person, reduced)	5.64	.25	je paard(en)
Adjective, diminutive, free position	5.01	.17	het is hier stilletjes, stilletjes wegsluipen
Pronoun, undetermined, orderable	4.72	.57	dat is te weinig, veel groter
Pronoun, undetermined, determiner without -n	4.51	.25	elk huis, ieder kind, enig benul, een enkel woord, sommig bier
Nominal adjective	4.34	.45	het leuke is dat
Pronoun, 1st person, singular	4.19	1.09	ik
Preposition + Article	3.87	2.69	aan het
Adjective, comparative, free position	3.47	.65	deze stok is langer, langer slapen
Preposition + Noun	3.38	.25	in orde
Demonstrative pronoun	3.19	.72	dat, dit, zulks
Foreign word	3.19	2.12	ad, hoc, wishful
Pronoun, 3rd person, plural	3.07	.42	zij

(*Appendix continues*)

Table A3
List of Words Used in Support-Vector-Machine Training (English) With Chi-Square Statistics and Odds Ratios

Token	Chi-square	Odds ratio
NA(".","/","?","-")	215.90	.00
Die	157.22	.11
Or	87.08	.13
Sad	62.66	.24
Always	55.82	.06
The	34.95	1.61
Can	34.43	.12
Family	34.31	.33
Holiday	34.04	.26
To	31.33	.68
Like	29.55	.27
Sometimes	29.19	.05
Lonely	27.90	.37
Find	27.46	9.05
Get	24.83	1.70
Win	23.21	3.56
Relax	22.87	.41
A	22.80	1.48
Have	19.21	.64
Break	19.03	6.88
Every	18.32	.13
If	18.28	.14
Know	17.93	.35
Mean	16.64	.20
Pass	16.38	.19
Will	16.11	.11
Anything	16.03	.16
Team	14.72	7.03
Someone	13.87	.36
Grandma	13.83	.17
Because	13.45	.60
Often	13.30	.00
Into	12.64	4.52
Move	12.33	.31
No	11.63	.44
Different	11.54	.11
Ordinal numbers (1st, 2nd, etc)	11.51	7.91
Life	11.42	.14
nobody	11.42	.14
away	11.28	.46
great	11.19	.26
Use	11.19	.26
Nothing	10.76	.25
Live	10.36	.30
Scared	10.36	.30
People	10.26	.41
Thing	10.23	.33
Yesterday	10.04	5.40
Do	10.02	.70
Stuff	9.72	.33
Run	9.68	6.96
Its	9.48	.08
Your	9.48	.08
Soccer	9.44	11.67
Grandad	9.41	.28
Cross	9.40	
I	9.31	1.13
Something	8.91	.50
As	8.81	.38
Wrong	8.50	.26
Party	8.43	3.26
Proud	7.91	2.03
How	7.67	.23
Sport	7.67	.23
With	7.63	.70
Real	7.31	.21
Right	7.31	.21
An	7.24	3.31
Annoy	7.16	.34
Competition	7.15	
Hear	7.15	
Love	7.02	.19
First	6.99	1.87
Everyday	6.99	.09
Tonga	6.99	.09
Travel	6.99	.09
Girlfriend	6.86	.15
Stop	6.64	.34
Chocolate	6.48	8.65
Much	6.27	.28
Now	6.27	.28
Out	6.22	1.74
Music	6.13	.34
Top	6.12	8.28
Day	5.96	1.86
Nice	5.81	.27
Pet	5.81	.27
Australia	5.66	.43
Hurt	5.66	.43
Some	5.65	2.94
Age	5.64	
Hour	5.64	
Though	5.64	
Lose	5.62	2.70
At	5.48	1.32
About	5.41	.62
Month	5.39	.41
Girl	5.39	7.52
Mate	5.39	7.52
Visit	5.38	.25
Off	5.32	2.63
Good	5.32	.61
Feel	5.31	.82
Ditch	5.27	
Glass	5.27	
Not	5.23	.78
Well	5.18	.38
Steal	5.03	2.78
Chill	4.99	.23
Job	4.99	.23
Still	4.99	.23
Say	4.93	2.11
Time	4.91	.65
It	4.89	1.37
Tournament	4.89	
Would	4.85	.57
Birthday	4.77	1.69
Pool	4.77	3.01
Rarotonga	4.68	.19
Show	4.68	.19
Summer	4.68	.19
Why	4.68	.19
Money	4.66	6.77
Allow	4.51	
Grade	4.51	
Parent	4.47	.55
Think	4.46	.56

Table A3 (*continued*)

Token	Chi-square	Odds ratio
Score	4.46	2.92
In	4.44	1.21
Few	4.36	.41
Award	4.30	6.39
Toy	4.30	6.39
Home	4.17	.68
Climb	4.14	
Dream	4.14	
Under	4.14	
Most	4.11	.31
Dad	4.08	.68
Goal	4.02	3.95
Spa	4.02	3.95
Catch	3.94	6.02
Jump	3.94	6.02
Eat	3.89	3.13
On	3.80	1.26
Scooter	3.76	
Which	3.76	
Window	3.76	
Yell	3.76	
Our	3.66	2.15
You	3.64	.45
Everything	3.61	.30
Heap	3.61	.30
Blame	3.58	5.64
Dollar	3.58	5.64
Exam	3.58	5.64
Guilty	3.43	1.64
Tell	3.43	1.64
Come	3.39	.67
Leader	3.38	
Road	3.38	
Stand	3.38	
Train	3.38	
Ball	3.35	3.57
Buy	3.35	3.57
Cardinal numbers	3.33	1.23
Cat	3.27	.61
Football	3.25	2.54
Arm	3.22	5.27
Favourite	3.22	5.27
Player	3.22	5.27
Street	3.14	.28
Who	3.02	.49
Black	3.01	
Cut	3.01	
Distinction	3.01	
Save	3.01	
Shop	3.01	
Speech	3.01	
Water	3.01	

Note. Odds ratios were calculated as relative frequencies of the words used in "specific" vs. "non-specific" memories, with values greater than 1 indicating that the words were used more frequently in "specific" than "non-specific" memories.

Table A4
List of Part-of-Speech Tags Used in Support-Vector-Machine Training (English) With Chi-Square Statistics and Odds Ratios

Part-of-Speech tags	Abbreviations in TreeTagger	Chi-squared	Odds ratio
Verb, present, non-3rd person.	VVP	488.19	.13
Verb be, pres non-3rd p.	VBP	304.53	.07
Verb, past tense	VVD	211.98	2.14
Verb, present 3d p. sing.	VVZ	151.55	.14
Verb be, pres, 3rd p. sing	VBZ	126.18	.07
Verb be, past	VBD	108.72	2.27
Verb do, pres, non-3rd per.	VHP	93.22	.10
To	TO	92.08	.54
Determiner	DT	49.40	1.45
Adverb	RB	48.75	.69
Noun, singular or mass	NN	45.10	1.24
Noun plural	NNS	23.48	.74
Verb, base form	VV	20.59	.72
Modal	MD	12.66	.59
Verb have, past participle	VHN	10.65	.00
Verb have, pres 3rd per.sing	VHZ	9.18	.13
Currency symbol ($)	$	9.05	4.23
Verb have, base form	VH	8.01	.47
Adjective, superlative	JJS	7.70	.50
Coordinating conjunction	CC	7.57	.84
Verb be, base form	VB	7.52	.47
Adverb, comparative	RBR	6.15	.34
Predeterminer	PDT	4.88	.59
Personal pronoun	PP	4.85	1.08
Wh-pronoun	WP	3.26	.59
Particle	RP	3.15	1.28
Possessive ending	POS	2.96	1.50
Verb be, gerund/participle	VBG	2.84	.59
Closing Bracket	)	2.83	1.57
Cardinal number	CD	2.65	1.17
Comma	,	2.49	.79
Verb have, past	VHD	2.35	1.25
Verb, gerund/participle	VVG	2.16	.90
Quotation	"	2.12	2.16
Verb be, past participle	VBN	1.63	.60
Proper noun, plural	NPS	1.51	3.38
End Quotation	"	1.37	1.88
Verb, past participle	VVN	1.25	1.10
Adjective, comparative	JJR	1.18	1.63
Adjective	JJ	1.08	.95
Adverb, superlative	RBS	1.04	.38
General joiner (. . ., -)	:	.87	.64
Foreign word	FW	.75	—
Interjection	UH	.75	—
Preposition/subord. conj.	IN	.62	1.03
Proper noun, singular	NP	.29	1.04
End punctuation	SENT	.29	.97
Verb have, gerund/participle	VHG	.23	1.31
Existential there	EX	.15	1.18
Opening Bracket	(	.13	1.09
Complementizer	IN/that	.09	1.07
Possessive pronoun	PP$	.08	1.01
Wh-determiner	WDT	.06	.92
Symbol	SYM	.03	.94
Wh-adverb	WRB	.00	1.00

Note. Odds ratios were calculated as relative frequencies of the words used in "specific" vs. "non-specific" memories, with values greater than 1 indicating that the words were used more frequently in "specific" than "non-specific" memories.

Received September 5, 2016
Revision received February 14, 2017
Accepted February 16, 2017 ■