Dream content analysis using Artificial Intelligence

Patrick McNamara[1], Kevin Duffy-Deno[2,3], Tom Marsh[3], and Thomas Marsh Jr[3]

[1]Department of Psychology, Northcentral University, Department of Neurology, Boston University School of Medicine, Boston, MA, USA
[2]School of Engineering and Computer Science, National University, San Diego, CA, USA
[3]Boulder Equity Analytics LLC, Boulder, CO, USA

Summary. We developed a dream content analysis system (DCAS) based on an artificial intelligence (AI) algorithm that was trained using a relatively large corpus of over 35,000 dreams. This sample of dreams were supplied by 424 female and 211 male users over 4 years who had posted them at the dream posting website and app Dreamboard.com. Building upon previous dream content ontologies developed by Hall, Van de Castle, Domhoff and Bulkeley, forty-seven reliably identified dream themes emerged from repeated application of algorithm and agent training procedures. DCAS reproduced most of the key dream content themes from these previous ontologies but also returned some unexpected findings. Mixed-model estimation detected significant male-female content differences for 34 dream themes, with female dreams evidencing higher incidence percentages for most themes, but effect sizes were small. Mixed-model logistic regression identified those themes that best predicted self-reported positive or negative mood associated with dreams. We conclude that the AI-based DCAS algorithm developed here is a promising tool for detailed analyses of dream content patterns.

Keywords: Dreams, dream content, artificial intelligence, content analysis, text analysis, dream theories, threat simulation, social simulation, mood function, memory consolidation

1. Introduction

Dreaming is a universal human phenomenon. The scientific study of dream content began in 1893 when Calkins (Calkins, 1893) tabulated 10 parameters in 170 dream reports from a 32-year old man and 205 reports from a 28-year old woman (Calkins' own dreams). After Freudian psychoanalysis emerged in the early part of the 20th century, most studies of dream content were conducted within the theoretical framework of Freudian theory and without controlled tabulation of various content categories.

After the discovery of REM (rapid eye movement) sleep in 1953 by Kleitman and his students (Aserinsky & Kleitman, 1953; Dement & Kleitman, 1957) there was a rebirth of the empirical approach to dream content studies in the US. Hall and Van De Castle attempted to establish standardized scoring procedures for tabulation of key dream features (Hall and Van de Castle, 1966) (e.g., people, objects, places, social interactions, activities, emotions, etc.). The Hall Van de Castle scoring system has been used to study many different dream samples derived from many different populations (e.g., Domhoff, 1996; Domhoff & Schneider, 1998; Domhoff, 2003; www.dreambank.net). Dream content norms for male vs female dreams (see also Domhoff, 1996) were derived after coding the dreams of 100 men and 100 women at Case Western Reserve University in the 1940s and 50s (5 dreams per person, yielding two samples of 500 dreams each). Subsequent dream content studies using the Hall Van de Castle system were largely congruent with the original norms unless special populations were being studied (reviews in Domhoff, 1996; 2003). Although the Hall Van de Castle system has proven invaluable in standardizing coding approaches for dream content studies, it is also quite labor-intensive requiring many hours of manual coding for even small datasets. It is very difficult to use with large datasets.

Automatic textual analysis of dream reports may be a technique that could overcome the labor-intensive drawback of more conventional dream content coding techniques. Initial attempts to utilize automated systems to tabulate the major Hall Van de Castle categories were carried out by Domhoff and by Bulkeley. Domhoff and colleagues (reviewed in Domhoff, 1996; 2003) established a dreambank (dreambank.net) of some 20,000 dream reports and a spreadsheet program to assist in calculation of the dream content ratios derived from the major Hall Van de Castle categories. While the dreambank.net spreadsheet was immensely helpful in dream content studies, investigators still had to learn the elaborate scoring system associated with the Hall Van de Castle categories and then manually code all dream reports before the normative ratios could be calculated within the spreadsheet. Bulkeley (2014) built upon the key Hall Van de Castle categories to establish a category list of about 40 word strings that could be used for automated queries and word-frequency counts in dream reports. He established a website dream archive, the Sleep and Dream Database (SDDb; http://sleepanddreamdatabase.org/) that houses several thousand dream reports and a word search facility for dream content analyses. He (Bulkeley, 2014) demonstrated that many of the Hall Van de Castle normative categories could be reliably reproduced using the keyword

Corresponding address:
Patrick McNamara, Department of Neurology, Boston University School of Medicine, Boston, MA 02118, USA
Email: pmcnamar3@gmail.com

Submitted for publication: June 2018
Accepted for publication: December 2018

search facility and a set of word templates/categories such as perception (see, hear, touch etc.), emotion (fear anger sadness etc.), characters (family animals, fantastic beings etc.), social interactions (friendliness, aggression, sex), movement (walking/running, falling, etc.) and cognition (thinking, speaking etc.) (see Appendix for words used within templates). Similar textual analytic tools have been used for dream content studies with encouraging results. For example, several investigators (e.g. see McNamara et al., 2015) have profitably used the Linguistic Analysis and Word Count (liwc.net; Pennebaker, et al., 2007) program which allows for identification of some 70 plus word categories in narratives and texts validated against a wide array of texts typically found in the public domain from novels to newspaper articles. While these word search techniques help to solve the problem associated with traditional labor intensive manual coding techniques, they are nevertheless based on simple tabulations of words that fall semantically within superordinate categories. These word tabulation approaches can be supplemented with techniques, such as sentiment and contextual analysis, borrowed from artificial intelligence (AI).

Nadeau et al (2018), for example, developed a text analytic technique for automatic dream sentiment analysis that combined word augmented count capacities as in the Bulkeley and the LIWC approaches along with artificial intelligence or AI-related sentiment analyses. They attempted to label dreams on a four-point negative/positive sentiment scale representing dreams as word vectors and including dynamic features to represent sentiment changes in intensity in the dream story. They also attempted fuzzy assignment of five emotion categories to dream descriptions, based on semi-automatically compiled emotion word dictionaries, thus demonstrating potential fruitful AI approaches to dream content analyses. Hendrick et al (2018) used three different AI and text analytic techniques (automatic text classification, topic modeling and text coherence analysis) to identify distinguishing features of dreams. They found that when compared against personal reports, dream reports were mildly less coherent than personal stories and contained a variety of themes not markedly different than personal stories. Amini et al. (2011) used word associations to improve accuracy of identification of emotion words in dream texts. They showed that with a sample of 458 dreams, the word association-enriched model demonstrated a significant improvement in identification of dream emotion for both negative and positive emotion tone scales. Razavi et al. (2014) analyzed leading themes of the dream and the sequential unfolding of associated sentiments during the dream to produce an algorithm for emotion identification in dreams. Agreements between the machine algorithm and human judges regarding identification of these sentiments reached 64 %.

While the foregoing papers represent an advance in application of automated text analytic techniques to dream reports, they did not build upon the earlier dream content ontologies developed by Hall Van de Castle, Domhoff and Bulkeley, so it is difficult to evaluate their findings against previously reported normative findings for dream content in men and women.

In what follows we built upon previous dream content coding ontologies using a more recent AI approach to develop an automated text analytic tool for dream content analysis for men and women. We call this new dream content analysis system tool "DCAS" (D-CAS). We were interested in describing typical dream content patterns for men vs women in a relatively large dream dataset. Our AI approach specifically expands upon Bulkeley's keyword approach (Bulkeley, 2014). However, the tool developed here not only tabulates words within target categories but also considers semantic "context." This is accomplished by training the AI algorithm not only on keywords but also on text surrounding the keyword and on sample text (sentences and paragraphs) that most closely resembles the concept that we wish the agent to find in the dream dataset. In addition, DCAS can also handle time-related dynamic changes in word meanings or the entry of new words into the dream lexicon. We used the AI analytic tool we developed here to describe what people dream about, comparing our results with previous work based on smaller dream datasets as well as reporting new findings never before discussed in the dream content literature.

2. Method

Dreamboard is an internet forum in which users can record and track their dreams over time via a phone application or online. Dreamboard.com began in 2012, and without any advertisement, grew tremendously thereafter. Users post their dreams in narrative form and are then invited to categorize or comment upon their dreams. Posted entries consist of a combination of standardized fields and free text to capture individual dream narrative and themes. Once entered into the database they are assigned an identification code to protect the anonymity of the dreamer. In 2017, the company made available to the authors a database of approximately 38,150 dreams posted from 2013 to 2016. Dreamboard.com itself is available for use all over the world. While most dreams are posted in English, a substantial proportion are not, thus, significantly increasing the cross-cultural data available for studies of dreams and nightmares. We, however, focused on dreams posted in English in this study.

2.1. Sample selection criteria

The initial dataset of 38,150 dreams were derived from 4 year-long postings from 641 anonymous users of Dreamboard from 2013-2016. Because Dreamboard does not collect any personal identifiers on its users we have no other information about the users/dreamers whose dreams we are analyzing other than gender and birthdate. A few of these users were very active on Dreamboard and posted over 280 dreams over these 4 years. After examining the distribution of the number of dreams submitted per user, so as not to introduce any bias from these heavy users, we eliminated the top 1% of users from the sample (i.e. those that submitted more than 280 dreams). There were 6 such users.

We also eliminated any "multi-part" dreams. Sometimes users of the Dreamboard app would batch enter their dreams from multiple dates. To facilitate the categorization of the dreams, we kept only dream entries that appeared to be associated with a single dream. We also selected first entries as the preferred dream report when multi-entries occurred as the recollection of the dream was likely more accurate immediately on waking. These two sample selection criteria took the sample size from 38,150 dreams to 35,138. This final sample of dreams were supplied by 424 female and 211 male users.

2.2. Training "agents" to classify dreams by theme

A proprietary artificial intelligence (AI) technology, created by the company ai-one (www.ai-one.com), was used to classify the 35,138 dreams based on the presence of 47 "themes." Forty of these themes were taken directly from Bulkeley's digital dream analysis word search system (Bulkeley 2014) tables 1-9. Bulkeley had demonstrated that his 40 themes captured the basic Hall Van de Castle system categories and thus were congruent with the Hall Van de Castle ontology for normative dream content. In addition to Bulkeley's 40 themes we added an additional 7-word categories from the LIWC system to capture cognitive process indicators (e.g., insight, causation, inhibition etc., see McNamara et al 2016). For each of these 47 themes, an "agent" was created which crawled through the dream text and scored the dream for how closely it matched the theme upon which the agent was trained. Agent training consisted of supplying the agent with keywords (taken from Bulkeley's 2.0 word search templates, and LIWC as described above) and samples of the theme from the dreams database, whole sentences as well as paragraphs. Each of the 47 agents then analyzed all 35,138 dreams and returned a similarity score (0 to 1.5) for each dream depending on how closely it reflected the theme.

The technology we used in training agents abstracts or generalizes the concept being scored from the words and their patterns in language. The technology is part of a platform for the analysis of language in documents, databases, research papers and other content accessible on both the web and a user's local drive. Specifically, this technology is focused on the detection of concepts/topics/themes at the paragraph and/or sentence level as a means to accurately define the context(s) of the text before performing more specific analyses such as sentiment, grammar, entity extraction, etc. The technology for concept detection is our derivative from work in neural networks, NLP and computational linguistics.

Given that training provided each agent with a large number of sample texts (approximately 100 per agent), the agent learned additional words and patterns, stored as an array, that indicated the concept/theme. The agent training process also allowed us too boost the score (by 50% in this case) if any of a list of specific unambiguous target words (Bulkeley 2014) are found in the narrative. As an example, the Bulkeley target words for the agent Fantastic Beings included "monster, witch, spirit, fairy and alien". From the training samples provided, the agent learned the words: "santa claus, supernatural, wizard, vampire, magic, knight, king, queen, prince, zombie, dwarf, and werewolves." These additional word patterns picked up narratives with ghosts, Harry Potter characters and references to Walking Dead scenes. Given that this particular classifier could be influenced by current cultural references, movies and video games, prior techniques using word lists limited to fixed templates would appear to under report this class and others like it. Agents can also be tested and retrained periodically to accommodate relevant new terms and language patterns as language evolves

The array for an agent was then compared with the array for each dream and scored for similarity. Since we generated a similarity score instead of a binary classifier (word counts), the resulting system is more flexible across large variances in narrative language. Additionally, the DCAS algorithm listed herein and those used by other researchers can be incorporated into the workflow and attributes added to the data so both approaches can be compared and/or extended by other researchers to provide a superior tool for dream research.

The similarity score returned by each agent is then converted to a final classification of theme presence. This is accomplished using a cutoff value. If the similarity score is greater than or equal to this cutoff value, then the theme is deemed to be present in the dream. If the similarity score is less than this cutoff value, then the theme is not present in the dream. Before settling on the cutoff value for this study, different cutoff values were evaluated to ensure the classifiers reflected the most relevant results for this type of text. We used a cutoff value of .95 making all 47 theme "variables" used in the following analysis 1/0 dummy variables indicating presence or non-presence of the theme. The 47 agents (themes) are presented in the Appendix along with their keyword definitions.

As part of this process in creation/training of the classifiers for each agent, the results of each agent are tested against a corpus and scored manually for false positives and false negatives. The above described agent similarity scores provide a ranking relevance which is reviewed both for the appropriate cutoff value (0/1 cutoff value) and also to view the performance of the agent against different types of dream content i.e. short, long, punctuation, slang, etc.

2.3. Statistical Analysis

Our first statistical analysis explores the differences in theme norms or "incidence" (i.e. the percentage of dreams characterized by a given theme) across the Dreamboard users, especially between female and male users. The purpose is to determine whether what we are finding in this dream sample is consistent with what previous research has shown. For example, Hall and Van de Castle (1966) reported that there was a higher proportion of male dream characters, unknown characters, more physical aggression, weapons, and sexuality in men's vs women's dreams. But these results and all subsequent studies of gender differences in dream content have been conducted on relatively small datasets with number of dreams rarely exceeding a few hundred (see review in Schredl et al, 1998). So further exploration of the reliability of gender differences in dream content is justified. It is known from previous studies that the dream reports of females are often longer than those of males (e.g., Domhoff, 1999, 2000; Mathes & Schredl, 2013).

We accomplish this analysis through a generalized linear mixed-model estimation. Dreamboard users posted multiple dreams, some more than others, over the 2013 to 2016 period. To control for the differential numbers of dreams contributed by different users/posters at Dreamboard and for the fact that dream content for any given user can be correlated across dreams, and since the presence of a given theme is binary, we employ a conditional-logit estimation where each user's intercept is treated as a random variable (i.e. "random effect") while the dream theme is treated as a fixed effect. Since the length of the reported dream narrative can positively impact the theme incidence (i.e. the longer the dream, the greater the opportunity for it to reflect multiple themes) our base model controls for narrative length. Thus, our base mixed, conditional-logit model is given, in simplistic form, by equation (1).

$$log(p/(1-p)) = \alpha + \beta_0 * THEME + \beta_1 * THEME * LENGTH + \mu \qquad (1)$$

where p is the probability that a given theme characterizes a dream, α varies randomly across users, THEME is the presence of a given theme (1/0), LENGTH is the dream narrative length, and μ is the normally distributed error term. Thus, this generalized linear mixed-model (GLMM) is an extension of logistic regression to include both fixed and random effects (hence the term “mixed”). Allison (2012) does a good job of explaining the basics of logistic regression. He explains the basics of the more complex GLMMs in Allison (2005) which we find to be one of the more accessible treatments of mixed modeling. In a typical GLMM estimation, the intercept consists of the mean intercept (across users) coupled with a random effect for each user. To yield an estimate for every 1/0 theme, however, estimation of the mean intercept was suppressed, leaving just the random effect for each user. Also, since dream narrative length is highly skewed, we use the natural log of LENGTH. To investigate whether theme incidence varies across genders, we add another fixed effect to equation (1) which is theme interacted with gender (THEME*GENDER) where GENDER = 1 if female, 0 if male:

$$\log(p/(1-p)) = \alpha + \beta_0{*}THEME + \beta_1{*}THEME{*}LENGTH + \beta_2{*}THEME{*}GENDER + \mu \qquad (2)$$

A second analysis focused on the relationship between the 47 dream themes and the users’ self-report on the Dreamboard app as to whether the dream was “pleasant”, “neutral” or “unpleasant” (i.e. “mood”). We wanted to assess whether specific dream content elements/themes could predict self-reported dream mood. We employed a similar mixed-model estimation approach. User intercepts are modeled as random effects while the 47 themes are modeled as fixed effects, each as their own dummy variable. A conditional logistic regression was again used for this analysis. In simplistic form, our estimation equation is:

$$\log(p/(1-p)) = \alpha + \beta_{0i}{*}\sum THEMEi + \beta_1{*}GENDER + \beta_2{*}NO\_CHARACTERS + \beta_3{*}NO\_LOCATIONS + \beta_4{*}AGE + \beta_{5j}{*}\sum TIMEj + \beta_6{*}LENGTH + \mu \qquad (3)$$

where p is the probability the self-reported dream mood was “pleasant” (vs. “unpleasant”); α varies randomly across users; ∑THEME are the 47 dummy variables for dream themes; GENDER = 1 if female, 0 if male; NO_CHARACTERS is the number of characters in the dream narrative; NO_LOCATIONS is the number of locations in the dream narrative; AGE is the user age in years; ∑TIME are dummy variables for the years 2013, 2014 and 2015; LENGTH is the natural log of narrative length; and μ is the normally distributed error term. Again, to facilitate an estimate for every 1/0 THEME, estimation of the mean intercept was suppressed leaving just the random effect. Also, NO_LOCATIONS and NO_CHARACTERS were determined through textual analysis of the dream narratives.

3. Results

3.1. Thematic incidence in the dream corpus

The first column of Table 1 shows the percentage of the 35,138 dreams characterized by each of the 47 themes, adjusting for narrative length (i.e. the theme means resulting from the estimation of equation (1)). We used SAS PROC GLIMMIX to estimate equation (1). Kierman (2018) discusses the use of GLIMMIX to model categorical outcomes with random effects. For example, 64% of the dreams were characterized by the theme Speaking (the highest theme incidence) while just 1.5% were characterized by the theme Tentativeness. The top 5 occurring themes are Speaking, Architecture, Vision, and Thinking each with better than 40% incidence scores. Rather than present the full statistical results of the conditional logit model estimation, we present the estimated theme incidences derived from these estimations.

3.2. Male vs female content differences for 47 agents/themes

We turn next to whether there is a significant difference between females and males in terms of what they dream about. The second column of Table 1 shows the percent of dreams characterized by each of the 47 themes across genders. For example, 13.2% of female dreams were characterized by the theme Animals while a smaller proportion of male dreams were so characterized (11%). This difference is statistically significant but the effective difference, shown by Cohen’s h, is small (.07). Statistical difference in means is given by the p-value on the THEME*GENDER coefficient, the estimated differential effect of GENDER on THEME presence, accounting for other fixed effects and the intra-user correlation in dream content (i.e. random effects). Cohen’s h is measured as 2*(arcsine(sqrt(female mean) – arcsine(sqrt(male mean))). In fact, this is a consistent finding across all 47 themes: though there are 38 statistically significant differences (at the 95% level of confidence or better) with females consistently demonstrating higher percentages of the themes in their dreams, the effective difference is consistently small. We will discuss these results more in the discussion section below.

The third column of Table 1 shows the estimated theme incidences that result when user age is also included as a fixed effect. Not all users reported a birthdate, although most did report a “valid” birthdate (i.e., one that did not imply they were unreasonably old or young). We excluded all users who reported themselves to be less than 12 years old (the oldest user in the sample is 66). This reduced the sample size to 571 users and 31,387 dreams. However, the estimated themes norms remain essentially unchanged after controlling for age. Across the three estimated models, the applicable fixed effects were always significant at the 1% level (THEME, THEME*LENGTH, THEME*GENDER, AGE). We also examined whether explicitly controlling for time by including dummy variables for 2013, 2014 and 2015 affect theme norms (they do not) as well as whether interacting narrative length with gender (also does not). The female vs male norms, once narrative length is accounted for, appear to be stable in the face of alternative specifications of equation (1) among this sample of dreamers.

3.3. Does dream content predict dream mood?

A second analysis focused on the relationship between the 47 dream themes and the users’ self-report on the Dreamboard app as to whether the dream was “pleasant”, “neutral” or “unpleasant” (i.e. “mood”). We focused on the dichotomous self-report of “pleasant” vs “unpleasant”, also including dreams categorized as “neutral” into the analysis. Also including users only if they reported an age of 12 years or older resulted in a sample of 17,247 dreams, roughly evenly split between pleasant (53%) and unpleasant (47%).

Table 1. 47 Dream Theme Norms: Mixed-model Logistic Estimation

Fixed Effects	Theme, Theme*Length	Theme, Theme*Length, Theme*Gender				Theme, Theme*Length, Theme*Gender, Age			
Theme	Norm	Female Norms	Male Norms	p	Effect Size	Female Norms	Male Norms	p	Effect Size
Air	6.4%	6.0%	7.1%	<.001***	-0.044	6.0%	6.9%	0.003***	-0.035
Anger	10.9%	12.0%	9.0%	<.001***	0.098	12.0%	9.4%	<.001***	0.086
Animals	12.4%	13.2%	10.8%	<.001***	0.075	13.3%	10.6%	<.001***	0.085
Architecture	55.5%	55.7%	55.0%	0.346	0.014	55.7%	56.0%	0.723	-0.006
Art	12.0%	12.5%	11.1%	<.001***	0.043	12.5%	11.2%	0.003**	0.039
Causation	4.2%	4.1%	4.5%	0.051*	-0.020	4.1%	4.4%	0.144	-0.017
Clothing	16.7%	17.4%	15.2%	<.001***	0.062	17.5%	15.4%	<.001***	0.056
Color	29.6%	29.7%	29.2%	0.427	0.011	29.6%	29.3%	0.666	0.007
Death	11.2%	11.5%	10.5%	0.007***	0.033	11.6%	10.7%	0.027**	0.029
Discrepancy	5.7%	5.4%	6.2%	0.002***	-0.033	5.3%	6.1%	0.004***	-0.034
Earth	14.8%	13.7%	17.0%	<.001***	-0.093	13.7%	17.4%	<.001***	-0.102
Exclusion	7.5%	7.5%	7.4%	0.887	0.002	7.5%	7.3%	0.604	0.006
Falling	9.9%	9.7%	10.2%	0.117	-0.018	9.6%	10.3%	0.105	-0.021
Family	37.9%	40.2%	33.5%	<.001***	0.140	40.4%	33.6%	<.001***	0.140
Fantastic_Beings	6.4%	6.5%	6.1%	0.114	0.018	6.5%	6.2%	0.256	0.014
Fear	16.8%	18.1%	14.2%	<.001***	0.107	17.8%	14.1%	<.001***	0.101
Female_References	38.1%	39.7%	35.0%	<.001***	0.096	40.0%	35.4%	<.001***	0.095
Fire	4.9%	4.6%	5.3%	0.002***	-0.034	4.7%	5.4%	0.002***	-0.036
Flying	8.7%	8.1%	9.9%	<.001***	-0.064	8.1%	10.2%	<.001***	-0.073
Food_Drink	21.0%	21.6%	19.8%	<.001***	0.044	21.6%	20.0%	0.006**	0.039
Friendliness	30.7%	32.3%	27.4%	<.001***	0.108	32.6%	27.8%	<.001***	0.105
Happiness	10.6%	11.3%	9.3%	<.001***	0.067	11.1%	9.0%	<.001***	0.070
Hearing	8.6%	8.4%	8.9%	0.096*	-0.019	8.4%	9.2%	0.025	-0.028
Inclusion	5.3%	4.8%	6.0%	<.001***	-0.050	4.9%	6.0%	<.001***	-0.048
Inhibition	6.7%	6.5%	7.0%	0.098*	-0.018	6.5%	7.3%	0.009**	-0.031
Insight	3.0%	2.8%	3.3%	0.003***	-0.029	2.7%	3.5%	<.001***	-0.043
Male_References	31.8%	32.9%	29.5%	<.001***	0.074	33.3%	30.5%	<.001***	0.059
Money_Work	22.4%	21.5%	24.2%	<.001***	-0.065	21.7%	24.7%	<.001***	-0.072
Physical_Aggression	27.1%	26.6%	28.1%	0.017**	-0.033	26.8%	28.1%	0.041**	-0.031
Reading_Writing	9.4%	9.3%	9.3%	0.952	0.001	9.3%	9.3%	0.998	0.000
Religion	5.9%	6.3%	5.0%	<.001***	0.056	6.1%	5.0%	<.001***	0.050
Sadness	7.7%	8.6%	6.1%	<.001***	0.096	8.6%	6.0%	<.001***	0.101
School	20.1%	21.1%	18.0%	<.001***	0.077	21.3%	17.7%	<.001***	0.089
Sexuality	10.6%	10.8%	10.0%	0.03**	0.026	11.0%	10.1%	0.024**	0.030
Smell_Taste	3.2%	3.4%	2.8%	<.001***	0.038	3.4%	2.8%	0.002***	0.036
Speaking	63.6%	63.9%	62.8%	0.115	0.023	64.1%	62.8%	0.095*	0.027
Sports	10.9%	10.2%	12.3%	<.001***	-0.067	10.3%	12.5%	<.001***	-0.071
Technology_Science	14.5%	14.0%	15.6%	<.001***	-0.045	14.2%	16.0%	<.001***	-0.052
Tenativeness	1.5%	1.4%	1.6%	0.038**	-0.019	1.4%	1.7%	0.013**	-0.025
Thinking	42.5%	41.9%	43.6%	0.02**	-0.034	41.8%	42.8%	0.206	-0.020
Touch	36.8%	37.5%	35.4%	0.003***	0.043	37.5%	35.7%	0.02**	0.037
Transportation	28.8%	27.7%	30.9%	<.001***	-0.071	27.9%	31.8%	<.001***	-0.086
Vision	42.9%	42.3%	43.9%	0.037**	-0.031	42.4%	44.4%	0.014**	-0.040
Walking_Running	38.4%	38.3%	38.6%	0.677	-0.006	38.4%	39.9%	0.051*	-0.031
Water	18.8%	19.0%	18.4%	0.28	0.014	18.9%	18.6%	0.571	0.008
Weapons	4.3%	3.9%	5.1%	<.001***	-0.060	3.9%	5.3%	<.001***	-0.064
Wonder_Confusion	13.5%	13.9%	12.7%	0.005***	0.034	13.9%	12.7%	0.009***	0.035
Dreamers	635	424	211			399	172		
Dreams	35,138	23,119	12,019			21,829	9,558		
Dream Length (avg)	952	994	872			1,000	883		

p-value is from the regression coefficient on THEME*GENDER. *** indicates significance at the 1% level; ** at the 5 % level; and * at the 10% level of confidence.

Table 2 shows the estimated mixed-model conditional logit regression. Since the fourth year (2016) is the omitted time variable, the coefficients of the included time variables are interpreted relative to 2016. The interpretation of the model coefficients is straightforward. A positive coefficient indicates that the relationship is more likely to be pleasant than unpleasant, while a negative coefficient indicates the opposite.

Table 2. Dream Theme and Self-Reported Mood: Mixed-model Logistic Estimation

Effect	Regression Estimate Coefficient	t-value	p	
Air	0.076	0.071	0.280	
Anger	-1.013	0.058	<.001	***
Animals	0.071	0.058	0.220	
Architecture	-0.163	0.043	<.001	***
Art	0.546	0.060	<.001	***
Causation	0.002	0.084	0.980	
Clothing	0.211	0.052	<.001	***
Color	0.049	0.046	0.287	
Death	-1.063	0.061	<.001	***
Discrepancy	0.251	0.075	<.001	***
Earth	0.170	0.055	0.002	***
Exclusion	-0.067	0.068	0.324	
Falling	-0.100	0.062	0.106	
Family	-0.107	0.046	0.020	**
Fantastic_Beings	-0.111	0.073	0.131	
Fear	-0.988	0.050	<.001	***
Female_Reference	-0.128	0.045	0.005	***
Fire	0.127	0.082	0.122	
Flying	0.348	0.065	<.001	***
Food_Drink	0.220	0.049	<.001	***
Friendliness	0.151	0.044	<.001	***
Happiness	0.618	0.059	<.001	***
Hearing	-0.270	0.062	<.001	***
Inclusion	0.104	0.078	0.182	
Inhibition	0.048	0.071	0.497	
Insight	0.148	0.091	0.105	
Male_References	-0.054	0.045	0.231	
Money_Work	0.057	0.047	0.225	
Physical_Aggress	-0.461	0.047	<.001	***
Reading_Writing	0.264	0.063	<.001	***
Religion	-0.128	0.078	0.099	
Sadness	-0.757	0.064	<.001	***
School	0.023	0.050	0.647	
Sexuality	0.797	0.060	<.001	***
Smell_Taste	0.394	0.095	<.001	***
Speaking	-0.029	0.045	0.521	
Sports	0.420	0.062	<.001	***
Technology_Scien	-0.098	0.055	0.075	*
Tenativeness	0.063	0.125	0.617	
Thinking	-0.025	0.044	0.565	
Touch	0.046	0.044	0.300	
Transportation	-0.099	0.045	0.026	**
Vision	-0.040	0.043	0.361	
Walking_Running	-0.079	0.045	0.078	*
Water	0.149	0.050	0.003	***
Weapons	-0.765	0.086	<.001	***
No_characters	0.112	0.013	<.001	***
No_locations	0.162	0.033	<.001	***
2013_dum	-0.088	0.084	0.295	
2014_dum	-0.137	0.075	0.069	*
2015_dum	-0.105	0.062	0.090	*
Age	-0.008	0.005	0.112	
Female	-0.437	-3.990	<.001	***
Length	0.139	4.750	<.001	***

*** indicates significance at the 1% level; ** at the 5 % level; and * at the 10% level of confidence.

For example, the coefficient on the theme Clothing is positive meaning that dreams characterized by this theme are more likely to be self-reported as "pleasant" than are dreams not characterized by this theme, all else constant. As another example, the coefficient on the theme Physical Aggression is negative. This means that dreams characterized by this theme are less likely to be self-reported as "pleasant" than are dreams not characterized by this theme, all else constant.

The themes most strongly associated with self-reported mood in a dream, based on the size of the coefficients, are the themes of anger (negative), death (negative), fear (negative), happiness (positive), sadness (negative), sexuality (positive) and weapons (negative). These findings are not too surprising. But other more surprising elements were also predictive of negative mood in dreams: female references, family and architecture.

Relative to the omitted year (2016) the negative coefficients on 2013_dum, 2014_dum and 2015_dum indicate that users were less likely to self-report their dream as "pleasant" in those years, all else constant. However, the effects are not statistically significant at the 95% level of confidence. Also, dreams with a greater number of characters and locations were more likely to be rated as "pleasant." As compared to males, female dreamers were less likely to rate their dreams as "pleasant." And, perhaps not surprisingly, dream length is positively associated with a reportedly pleasant dream. Finally, older users were less likely to report their dreams as pleasant although the relationship is not statistically significant.

4. Discussion

We developed a Dream Content Analysis System (DCAS) based on an AI algorithm that was trained using a relatively large corpus of some 35,138 dreams supplied over a four-year period by 424 female and 211 male dreamers. Our study may be the largest dream content analysis study ever published. Using an online dream reporting portal Nielsen (2012) collected reports on dreams from some 28,883 respondents. Though he did not analyze dream content per se, respondents answered a questionnaire concerning diversity of their dream themes. Nielsen found that diversity of dream themes declined linearly with age for both sexes up to age 50–59 and then dropped even more sharply for the 60–79 age group. To our knowledge Nielsen's study was the largest such study to date that addressed dream theme diversity, but again he did not analyze dream content in any detail.

Building on previous dream content ontologies (e.g. Hall van de Castle's, Domhoff's and Bulkeley's) we have demonstrated that forty-seven reliably identified dream themes derived from these standardized dream content ontologies, can be reliably captured by the DCAS algorithm and agent training procedures. Incidence/incidence analyses of these dream themes agreed substantially with previous dream content scoring systems such as the Hall Van de Castle and the Bulkeley systems. Our DCAS typically returned incidence rates within 5% of rates reported by Bulkeley for his analysis of the original Hall Van de Castle sample of dreams. For example, for Bulkeley's "perception" template, Bulkeley found (for males) 38.7% incidence for vision, 12.8% hearing, 12.2% touch, 2.6% smell/taste and 14.5% for color; for females 48.5% incidence for vision, 13.2% hearing, 13.0% touch, 2.4% smell/taste and 27.3% for color. DCAS returned incidence rates for several of these themes (males and females combined) within 5% of Bulkeley's rates: 43% incidence for vision, 9% hearing, 30% for color and 3% smell/taste. On the other hand, DCAS reported 37% for touch—a much higher rate than that reported by Bulkeley for this theme. Similar findings were obtained for each of the other Bulkeley templates (mood, characters social interactions, movements, cognition etc). DCAS incidence rates were within 5% of Bulkeley's for about 2/3rds of

Bulkeleys' themes. On the other hand, large discrepancies between Bulkeley's and DCAS rates were noted for several significant themes such as male references, female references, family, and friendliness with DCAS reporting lower incidence for these themes. Similarly, DCAS reported higher incidence rates for speaking, physical aggression and fantastic beings.

It appears that the DCAS algorithm reliably reproduces the majority of key dream content indices derived from previous content scoring systems based on Bulkely and Hall Van de Castle systems, but also returns some unexpected new findings as well. Regarding the noted discrepancies in incidence rates for certain content themes, we suggest that the DCAS may be the more reliable estimate given that that they are based on a larger dream corpus and that the DCAS algorithm not only tabulates words within the content category or Bulkeley template itself but searches surrounding context for related words as well and these are then added to the similarity score and ultimately the incidence scores. For example, speaking, physical aggression and fantastic beings are perhaps much more frequent occurrences in dreams than previously reported.

The set of themes that emerged in our analyses appear to be consistent with Social Simulation Theory or SST of dreaming. Many authors have remarked on the probable social functions of dreaming. Anthropologists have long treated the dream as a strategic social act; that is, dreams are used in traditional societies to facilitate negotiations in social alliances and to facilitate change in the social status of the dreamer. Freud (1900/1950) and many authors in the psychoanalytical tradition can be read as supporting a kind of SST given that they often interpret dreams in terms of emotional conflict in families of origin or in current families as well as between sexual partners and romantic targets. McNamara (2004) and Revonsuo, Touminen and Valli (2015) have marshalled some of the data and arguments that support the SST. The SST postulates that dreams virtually simulate socially significant interactions for the dreamer; that is, they simulate human social reality, including the social skills, bonds, interactions, and networks that we engage in during our waking lives. Brereton (2000; see also Franklin and Zyphur, 2005) presented a similar idea in his "Social Mapping Hypothesis" which suggests that dreaming allows for rehearsal of emotional and perceptual abilities needed for relating the dreamer to emotionally significant others and social groups. Our data are also partially consistent with the continuity hypothesis on dreams (Schredl and Hoffman, 2003; Domhoff, 1996) which suggests that dreams simulate the kinds of things we do and encounter in everyday life, including, of course, those social interactions that SST captures. Our data at present cannot decide between these theories.

a. Our findings regarding gender differences in dream content while largely consistent with previous reports in this area (Schredl et al., 1998; Mathes and Schredl, 2013) evidenced some surprises. For example, previous investigators have reported that unfamiliar, outdoor settings were present more often in men's vs women's dreams and that there was a higher proportion of male dream characters, unknown characters, more physical aggression, weapons, and sexuality in men's vs women's dreams. We found that incidence rates for physical aggression levels were similar in men (28%) and women's (27%) dreams and in fact anger and sexuality levels were slightly higher in women's vs men's dreams. Nor did we find that males dreamt of other male characters more frequently than did women (33% male reference in female dreams v 29.5% male references in male dreams). Despite the ubiquity of gender differences in dream content we found that the differences were relatively small for most of the 47 themes we analyzed. Interestingly, men evidenced higher incidence rates for themes on several of the cognitive indicators (discrepancy, inclusion, inhibition, insight and thinking) and a few other items such as sports, technology, transport and weapons. It is unclear exactly what this mean. Men appear to do more cognitive processing of the content they have in their dreams.

Mixed-model logistic regression demonstrated that themes best predicting negative mood in a dream were themes like anger, death, fear, physical aggression, sadness and weapons while the themes best predicting positive moods in dreams were themes like food/drink, happiness, friendliness, art, clothing, sexuality, etc. While all these associations seem to square with common sense concerning the things which make us feel good or bad, there were other findings that were more difficult to interpret. For example, other themes predictive of negative mood in dreams included female references, family, and architecture; and other themes predictive of positive mood excluded the senses such as vision touch and hearing (hearing was associated with unpleasant mood). In addition, previous work by other investigators has reported no or rather low correlations between self- and external ratings of dream emotions (see e.g., Schredl & Doll, 1998; Sikka et al., 2014, 2017). Further work will be required to explore the reliability of these associations. These findings appear to be somewhat consistent with continuity theory insofar as the same things which make us unhappy in waking life (weapons, aggression etc.) also make for a bad dream. On the other hand, other themes associated with bad dreams such as female references and architectural themes do not necessarily make us unhappy during waking life, so these findings may not be consistent with continuity theory or SST. Mood function theory suggests that dream content elements are being used to reduce affective loads to promote long term memory encoding, yet dream content elements themselves help to create dream mood states. The social simulation theory (SST) can accommodate the finding that things that make us unhappy in waking life also do so in dream life but again it is difficult to see how SST could handle the findings that female references or architecture references are associated with negative mood in dreams as they are not typically unpleasant in waking life. Therefore, these dream elements do not faithfully simulate social life. Nevertheless, these findings on what dream elements predict bad dreams could be important clinically. For example, reducing the occurrence of content elements that consistently produce negative mood in dreams could help in reducing recurring bad dreams or nightmares. Cognitive behavioral techniques such as imagery rehearsal therapy can effectively change dream content patterns (Krakow and Zadra, 2010).

5. Limitations

Although our study is innovative in that it uses AI techniques to automate dream content analyses and it is the largest dream content study ever published, it has significant limitations that need to be considered when interpreting results. The major limitation of the study of the study was that the data was obtained via internet postings, so we must trust that the posters themselves were really posting dream narratives. While there is no definitive way to check the reliability of the narratives posted without interviewing each poster, it is now widely accepted that internet samples constitute a valid source of information about self-reported human behavior (Nielsen, 2012). We noted above that Nielsen used dreams posted at his website to generate estimates of age-related changes in dream diversity. Indeed, there are significant advantages to data collection over the internet including ease of data collection and the possibility of substantially increasing statistical power. In addition, anonymous collection of data may also promote more complete and honest descriptions of dream narratives and allow us to reach populations other than the typical upper-class white north American college student from whom most dream samples have been drawn since the beginning of the scientific study of dream content.

6. Conclusions

We developed a dream content analysis system (DCAS) based on an AI algorithm that was trained using a relatively large corpus of over 35,000 dreams. Forty-seven reliably identified dream themes emerged from repeated application of algorithm and agent training procedures. Mixed-model estimation detected significant male-female content differences for most dream themes, with female dreams evidencing higher incidence percentages for most themes, but effect sizes were small. Mixed-model logistic regression identified those themes that best predicted self-reported positive or negative mood associated with dreams. We conclude that the DCAS algorithm developed here is a promising tool for detailed analyses of dream content patterns.

Disclosures and Acknowledgements

Patrick McNamara was science officer of Dreamboard Inc from 2016-2017; Kevin Duffy-Deno, Tom Marsh and Thomas Marsh Jr. are research scientists for ai-one inc and partners at Boulder Equity Analytics.

References

Allison, P. (2005). Fixed Effects Regression Methods for Longitudinal Data Using SAS. SAS Press, Cary, North Carolina.

Allison, P. (2012). Logistic Regression Using the SAS System. SAS Press, Cary, North Carolina.

Amini, R., Sabourin, C., & De Koninck, J. (2011). Word associations contribute to machine learning in automatic scoring of degree of emotional tones in dream reports. Consciousness and Cognition, 20, 1570-1576.

Aserinsky, E., & Kleitman, N. (1953). Regularly occuring periods of eye motility, and concomitant phenomena, during sleep. Science, 118, 273-274.

Brereton, D. (2000). Dreaming, adaptation, and consciousness: The Social Mapping Hypothesis. Ethos, 28 (3), 379-409. 10.1525/eth.2000.28.3.379

Bulkeley K. (2014). Digital dream analysis: a revised method. Conscious Cogn. 2014 Oct;29:159-70. doi: 10.1016/j.concog.2014.08.015. Epub 2014 Oct 3.

Calkins, M. W. (1893). Statistics of dreams. American Journal of Psychology, 5, 311-343.

Domhoff, G. W. (1996). Finding meaning in dreams: A quantitative approach. New York, NY: Plenum Press.

Domhoff, G. W. (2003). The Scientific Study of Dreams: Neural Networks, Cognitive Development, and Content Analysis. American Psychological Association, New York.

Domhoff, G. W., & Schneider, A. (1998). New rationales and methods for quantitative dream research outside the laboratory. Sleep, 21(4), 398-404.

Franklin, M. S. & Zyphur, M. J. (2005). The role of dreams in the evolution of the human mind. Evolutionary Psychology, 3, 59-78.

Freud, S. (1995). The interpretation of dreams; and On dreams (J. Strachey, Trans. Vol. 4-5). London: Hogarth Press.

Hall, C. S., & Van de Castle, R. L. (1966). The content analysis of dreams. New York: Appleton-Century-Crofts.

Hendrick, Iris, Louis Onrust, Florian Kunneman, Ali Hürriyetoğlu, Antal van den Bosch and Wessel Stoop; (2017). Unraveling reported dreams with text analytics. Digital Humanities Quarterly, V. 11, No. 4

Kiernan, K. (2018). Insights into Using the GLIMMIX Procedure to Model Categorical Outcomes with Random Effects. SAS Institute, SAS2179-2018.

Krakow, B., and Zadra, A. (2010). Imagery Rehearsal Therapy: Principles and Practice. Sleep Medicine Clinics, 4(2), 289-298.

Mathes, J., & Schredl, M. (2013). Gender differences in dream content: Are they related to personality? International Journal of Dream Research, 6, 104-109.

McNamara, P., Pae, V., Teed, B., Tripodis, Y., & Sebastian, A. (2016) Longitudinal studies of gender differences in cognitional process in dream content. Journal of Dream Research. 9(1). doi: hhtp://dx.doi.org/10.11.588/ijord.2016

McNamara, P., Minsky, A., Pae, V., Harris, E., Pace-Schott, E., & Aurbach, S. (2015). Aggression in nightmares and unpleasant dreams and in people reporting recurrent nightmares. Dreaming. 25(3), 190-205. http://dx.doi.org/10.1037/a0039273.

McNamara, P., Johnson, P., McLaren, D., Harris, E., Beauharnais, C., & Auerbach, S. (2010). REM and NREM sleep mentation. International Review of Neurobiology, 92, 69-86.

McNamara, P. (2004). An evolutionary psychology of sleep and dreams. Westport, CT: Praeger/Greenwood Press.

McNamara, P., McLaren, D., Smith, D., Brown, A., & Stickgold, R. (2005). A "Jekyll and Hyde" within: Aggressive versus friendly social interactions in REM and NREM dreams. Psychological Science, 16(2), 130-136. PMID: 15686579

Nadeau, D., Sabourin, C., De Koninck, J., Matwin, S., & Turney, P.D. (2006). Automatic dream sentiment analysis. In Proceedings of the workshop on computational aesthetics at the twenty-first national conference on artificial intelligence (AAAI-06), Boston, USA.

Nielsen, T (2012). Variations in dream recall frequency and dream theme diversity by age and sex. Frontiers in Neurology, July 2012 | Volume 3 | Article 106, 1-9.

Nielsen, T. A., & Levin, R. (2007). Nightmares: A new neurocognitive model. Sleep Medicine Reviews, 11, 295-310.

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). The development and psychometric properties of LIWC2007 [LIWC manual]. Austin, TX: LIWC.net.

Razavi, A., Matwin, S., De Koninck, J., and Amini, R. (2014). "Dream sentiment analysis using second order soft co-occurrences (SOSCO) and time course representations." Journal of Intelligent Information Systems, 42(3), pp. 393–413.

Revonsuo, A. (2000). The reinterpretation of dreams: An evolutionary hypothesis of the function of dreaming. Behavioral and Brain Sciences, 23(6), 877-901.

Revonsuo, A., Tuominen, J. & Valli, K. (2015). The Avatars in the Machine - Dreaming as a Simulation of Social Reality. In T. Metzinger & J. M. Windt (Eds). Open MIND: 32(T). Frankfurt am Main: MIND Group. Pps 1-28; doi: 10.15502/9783958570375

Schredl, M. & Hofmann, F. (2003). Continuity between waking activities and dream activities. Consciousness and Cognition, 12 (2), 298-308. 10.1016/S1053-8100-(02)00072-7

Schredl, M., Sahin, V., & Schäfer, G. (1998). Gender differences in dreams: Do they reflect gender differences in waking-life? Personality and Individual Differences, 25, 433-442.

Schredl, M. & Doll, E. (1998). Emotions in diary dreams. Consciousness and Cognition, 7, 634-–646.

Schwartz, S. (2004). What dreaming can reveal about cognitive and brain functions during sleep? A lexico-statistical analysis of dream reports. Psychologica Belgica, 44-1/2, 5-42.

Sikka, P., Valli, K., Virta, T. & Revonsuo, A. (2014). I know how you felt last night, or do I? Self- and external ratings of emotions in REM sleep dreams. Consciousness and Cognition, 25, 51-66.

Sikka, P., Feilhauer, D., Valli, K. & Revonsuo, A. (2017). How you measure is what you get: Differences in self- and external ratings of emotional experiences in home dreams. American Journal of Psychology, 130, 367-384.

Strauch, I., & Meier, B. (1996). In search of dreams: Results of experimental dream research. Albany, NY: State University of New York Press.

Appendix: Theme Keyword Definitions

Theme	Keyword Source	Keywords
Air	Bulkeley 2.0 Template	Air, Wind, Breath, Blow, Cloud, Sky, Tornado
Anger	Bulkeley 2.0 Template	Angry, Irritated, Disgusted, Anger, Mad, Furious, Annoy, Rage, Argument, Pissed, "Fuck You"
Animals	Bulkeley 2.0 Template	Dog, Animal, Snake, Cat, Horses, Insects, Bird, Duck, Kitten, Fish, Moose, Deer, Bug, Bull, Dolphin, Boar, Crow, Gorilla, Wolf, Pet, Bear
Architecture	Bulkeley 2.0 Template	House, Room, Home, Door, Floor, Castle, Restaurant, Bathroom, "Air Duct", "Dining Hall", "Dining Room", Wall, Roof, Ceiling, "Living Room", Hospital, Garage, Temple, Hallway, Window
Art	Bulkeley 2.0 Template	Movie, Dance, Music, Dancing, Museum, Drawing, Sing, Actress, Actor, Paint, Theatre, "Marching Band", Instrument
Causation	Cognitive Processing	Explanation, Motive, Origin, Purpose, Root, Source, Motivate, Motivation, Cause, Causation
Clothing	Bulkeley 2.0 Template	Clothes, Dress, Wearing, Trousers, Shirt, Boot, "High Heel", Jean, Shorts, Khaki, Jacket, Hoodie, Sweatshirt, Socks, Underwear, Bra, Panties, Swimsuit, Shoe, Pajama, Jewelry, Necklace, Bracelet, Robe, Outfit, Attire
Color	Bulkeley 2.0 Template	Red, White, Blue, Black, Green, Yellow, Purple, Orange, Beige, Grey, Glow, Rainbow, Color, Light, Dark, Crimson, Burgundy
Death	Bulkeley 2.0 Template	Dead, Death, Die, "Pass Away", Dying, Murder, Killed
Discrepancy	Cognitive Processing	Difference, Different, Distinct, Variation, Divergent, Divergence, Disparity, Inconsistency, Discrepant, Discrepancy, Disagreement, Unexpected, Dissimilarity
Earth	Bulkeley 2.0 Template	Hill, Mountain, Dirty, Land, Stone, Dirt, Mud, Rock, Valley, Ground, Grass, Sand, Plants, Quarry, Garden, Volcano, Tree
Exclusion	Cognitive Processing	Exclude, Eliminate, Elimination, Eviction, Prohibit, Refuse, Reject, Remove, Removal, Segregate, Separate, Seperation, Suspend, Ban, Block, Eject, Exception, "Keep Out"
Falling	Bulkeley 2.0 Template	Fell, Fall, Drop, Landed
Family	Bulkeley 2.0 Template	Brother, Father, Mother, Wife, Sister, Married, Nephew, Aunt, Family, Baby, Dad, Mom, Fiance, Husband, Son, Daughter, Grandpa, Grandma, Grandparent, Child, Cousin, Uncle, Families, Parents, Granny, Niece
Fantastic Beings	Bulkeley 2.0 Template	Monter, Witch, Spirit, Fairy, Alien, "Santa Claus", Supernatural, Wizard, Vampire, Magic, Knight, King, Queen, Prince, Zombie, Dwarf, Werewolves
Fear	Bulkeley 2.0 Template	Afriad, Fear, Frightened, Anxious, Worried, Upset, Embarrassed, Apologetic, Trapped, Terrified, Terrify, "Freaked Out", Scare, Scary
Female References	Bulkeley 2.0 Template	Her, She, Girl, Mother, Lady, Woman, Mum, Lesbian, Nanny, Mom, Sister, Wife, Female, Daughter, Niece
Fire	Bulkeley 2.0 Template	Fire, Sun, Star, Burn, Star, Lava, Flame, " Heat", Fiery
Flying	Bulkeley 2.0 Template	Fly, Float, Flew, Flies, Jump
Food & Drink	Bulkeley 2.0 Template	Drink, Food, Dinner, Eat, Lunch, Coffee, Drunk, Dine, Café, Candy, Cake, Sweets, Tea, Hunger, Hungry, Noodles, Sushi, Fruit, Vegetables, Wine, Beer, Alcohol, Chocolate, Restaurant, Breakfast, Bread, Snack, Pizza, Toast, "Ice Cream", Syrup, Buffet, Beverage, Soup, Cook, Vodka, Sauce
Friendliness	Bulkeley 2.0 Template	Friend, Married, Party, Offer, Save, Help, Love, Classmate
Happiness	Bulkeley 2.0 Template	Happy, Relieved, Glad, Pleased, Amused, Relaxed, Cheerfully, Bliss, Joy, Wonderful, "Great Time", "Good Time", Relief, Relieved
Hearing	Bulkeley 2.0 Template	Hear, Listen, Loud, Noise, Sound, Noisy, " Hear", " Loud", "Tells Us", "Tells Me", "Tell you", "Sound Like", "Sounds Like", Ruckus
Inclusion	Cognitive Processing	Include, Including, Inclusion, Admit, Incorporate, Involve, Form, Embody, Embodiment, Embrace, Encompass, Insertion, "Let in", Welcomed
Inhibition	Cognitive Processing	Inhibit, Barrier, Wall, Restrict, Hinder, Shy, Block, Prohibit, Reserve, Restraint, Suppress, Obstacle
Insight	Cognitive Processing	Insight, Judgement, Observe, Understand, Vision, Wisdom, Perception, Perceive, Intuition, Experience
Male References	Bulkeley 2.0 Template	He, Him, His, Man, Boy, Father, Brother, Dad, Guy, Husband, Masculine, " Male"
Money & Work	Bulkeley 2.0 Template	Work, Money, Office, Business, Cents, Rich, Wealthy, Gold, Shop, Dollar, Paycheck, Paid
Physical Aggression	Bulkeley 2.0 Template	" Hit", Shot, Killed, Enemy, Fight, Kill, Throw, Threw, Push, Hurt, Banging, " Bit", Biting, Rape, Defend, Attack, Stab, Kidnap, "Slam"
Reading/Writing	Bulkeley 2.0 Template	Book, Letter, Written, Read, Writing, Write, Note, Poem, Card, Journal, Diary, Spelled
Religion	Bulkeley 2.0 Template	Church, Christmas, Priest, Altar, Religious, Spirit, Hell, Bible, Salvation, God, Christian, Catholic, Jew, Muslim, Islam, Judaism, " Sin", " Cult", Apocalypse, Anglican
Sadness	Bulkeley 2.0 Template	Disappointed, Sad, Distress, Sadly, Lonely, Unhappy, Miserable, Balling, Sobbing, Cry, Cries

Theme	Keyword Source	Keywords
School	Bulkeley 2.0 Template	School, Teacher, College, Student, Library, Test, University, Semester, Class
Sexuality	Bulkeley 2.0 Template	Intercourse, "Make Love", Kiss, Sex, Naked, "Fooled Around", "Fooling Around", "Hooked Up", "Hooking Up", Rape, Flirt, Tease, Vibrator, "Making Love", "Made Love", Sensual, Orgy, Breast, Penis, Boob, Butt, Fondling, Virgin
Smell & Taste	Bulkeley 2.0 Template	Nose, Smell, Sweet, Taste, Delicious, Odor, Tongues, Stink, Whiff, Sniff, Scent
Speaking	Bulkeley 2.0 Template	Said, Call, Say, Talk, Answer, Ask, Scream, Converse, Tell, Mocked, Holler, Whisper, Explain, Shout, Argument
Sports	Bulkeley 2.0 Template	Football, Baseball, Gym, Basketball, Tennis, Golf, Athletic, Sport, Soccer, Game, Hockey, "Track and Field", Boxing
Technology & Science	Bulkeley 2.0 Template	Machine, Phone, Radio, Television, Telephone, Biology, Engine, Biological, Engineer, Video, YouTube, Computer, Laptop, Xbox, Playstation, Headphone, "Video Game", Movie
Tentativeness	Cognitive Processing	Unsure, Uncertain, Negative, Hesitant, Delay, Doubt, Indecision, Skeptic, Reluctant, Reluctance, Timid
Thinking	Bulkeley 2.0 Template	Thought, Think, Notice, Realize, Decide, Aware
Touch	Bulkeley 2.0 Template	Hand, Held, Hold, Hug, "Pick It Up", "Pick Up", "Picked it Up", "Picks Up", Grab, Throw, Threw, Touch, Pull, Felt, Push, Lift
Transportation	Bulkeley 2.0 Template	Car, Stairs, Boat, Cars, Street, Road, Wheelchair, Train, Freeway, Railroad, Highway, Trail, Drove, Drive, Rode, Ride, Plane, Jet, Truck, Rollercoaster
Vision	Bulkeley 2.0 Template	Saw, See, Sight, View, Watch, Eyes, Vision, Observe
Walking/Running	Bulkeley 2.0 Template	Ran, Walk, Step, Run, Jog, Explore, "Head into", Jump, Wander
Water	Bulkeley 2.0 Template	Water, Snow, Lake, Rain, River, Wet, Ocean, Glacier, Sea, Hurricane, Tsunami, Pool, Swim, Drown, Wave, Flood, Boat, Ice, Pond, Swam
Weapons	Bulkeley 2.0 Template	Gun, Knife, Rifle, Bomb, Spear, Bullet, Sword, Explosion, Grenade, Pistol, Hammer, Flamethrower, Ammo, Taser, Mercenary, Chainsaw, "Power Drill", " Axe", Hatchet, Scissor, Trigger, "Buck Shot"
Wonder/Confusion	Bulkeley 2.0 Template	Sudden, Surprise, Confused, Confusing, Wonder, Shock, Disoriented