Measuring Implicit Motives With the Picture Story Exercise (PSE): Databases of Expert Coded German Stories, Pictures, and Updated Picture Norms

Felix D. Schönbrodt
Ludwig-Maximilians-University, Munich

Birk Hagemeyer[a], Veronika Brandstätter[b], Thomas Czikmantori[b], Peter Gröpel[c], Marie Hennecke[b], Kevin Janson[d], Nina Kemper[e], Martin G. Köllner[d], Philipp Kopp[f], Andreas Mojzisch[g], Raphael Müller-Hotop[f], Johanna Prüfer[h], Markus Quirin[f], Bettina Scheidemann[i], Lena Schiestel[j], Larissa Sust[j], Caroline Zygar[j], Oliver C. Schultheiss[d]

[a]Friedrich Schiller Universität Jena, [b]University of Zurich, [c]University of Vienna, [d]Friedrich-Alexander-Universität Erlangen, [e]Frankfurt University, [f]Technical University Munich, [g]University of Hildesheim, [h]University of Göttingen, [i]Osnabrück University, [j]Ludwig-Maximilians-Universität München

We present two freely accessible databases related to the assessment of implicit motives using Picture Story Exercises (PSEs): (a) A database of 20,039 German PSE stories that have been scored by experts using Winters (1994) scoring system for the implicit affiliation/intimacy, achievement, and power motive, and (b) a database of 53 classic and new pictures which have been used as stimuli in the PSE. Updated picture norms are provided which can be used to select appropriate pictures for PSE applications. Based on an analysis of the relation between raw motive scores, word count, and sentence count, we give recommendations on how to control motive scores for story length. Several potential use cases for the databases are discussed, including (un)supervised machine learning of text content, psychometrics, and better reproducibility of PSE research.

Keywords: picture story exercise, database, pictures, manual coding, machine learning

Implicit motives are nonconscious motivational needs that orient, select, and energize behavior (McClelland, 1987). A common approach to measuring implicit motives, such as the affiliation, power, or achievement motive, is the Picture Story Exercise (PSE; Schultheiss & Pang, 2007, Smith, Atkinson, McClelland, & Veroff, 1992), which is a modern, experimentally validated version of the classic Thematic Apperception Test (Morgan & Murray, 1935). In this task, several ambiguous pictures are presented to participants who are asked to write an imaginative story in response to each picture. These stories then are scored by trained coders using empirically derived and validated content coding systems, which quantify the amount of motive imagery in each story. Motive-related imagery then is taken as an indicator for the strength of the implicit motive.

"Picture Story Exercise" is a rather generic term, as instructions, pictures, and coding systems can vary between applications. However, some standardization has taken place in the recent years. For example, a standard set of six pictures has been suggested, which provides a roughly balanced motivational pull for each of the achievement, affiliation, and power motives (Schultheiss & Pang, 2007). The existence of such a standard picture set, however, does not mean that other pictures are discouraged. In contrast, it has been recommended to use tailored picture sets that focus on the focal motive, or depict situations that are close to the relevant criterion variables that are to be predicted by PSE scores (Schultheiss & Pang, 2007). Furthermore, multiple coding systems exists for several motives (for an overview, see Schultheiss & Brunstein, 2010, or Smith et al., 1992). Many coding systems focus on one single motive, but one prominent exception is David Winter's (1994) *Manual for scoring motive imagery in running text*. This integrated scoring system allows to assess three implicit motives simultaneously

(Winter, 1991): the need for achievement (*nAch*), power (*nPow*), and affiliation/intimacy (*nAff*).[1] This system is the most commonly employed system in the last two decades for PSE stories, and will be the focus of this publication.

The current paper has three goals: (1) Present a large database of stories that have been coded for implicit motives using the Winter scoring system, (2) provide a systematic database of 53 classic and new picture stimuli that have been used in PSEs, and (3) provide updated norms for picture pulls (i.e. the propensity of a picture to elicit a certain kind of motive imagery).

A Database of Scored PSE Stories

Several labs contributed existing datasets for building a large database of scored PSE stories in German language. The inclusion criteria were (a) the stories were coded using the Winter scoring system, (b) all coders were trained by experts, had extensive coding experience ($\gg$ 1000 stories), and achieved good convergence with training material scored by experts (such as ICC $\geq$ .85, category agreement $\geq$ .85), and (c) the stories were scored sentence-wise. The included datasets come from a diverse range of studies, including lab and online administration of the PSE tasks, differing numbers and types of pictures, and diverse samples. Some of the datasets come from published work (*TODO: cite all papers that are based on (parts) of this dataset*), others are from undocumented archival datasets. For some of these archival datasets no sample descriptives could be recovered.

One of the primary purposes of collecting these expert-scored text data is to provide an extensive, annotated training dataset for automatic text analysis. Therefore the focus of this database is the text data itself and its relation to the codings according to the scoring system. As we aimed to compile a database that is as large as possible, we decided to include all available datasets without providing sample characteristics such as age and gender. Table 1 explains all variables of the database and their meaning.

Winter's (1994) Scoring System

All stories were coded according to the Winter (1994) scoring system, which defines rules when to score a motive imagery for each motive category. A motive imagery is defined as "an action (past, present, future or hypothetical), a wish or concern, or some other internal state" (p. 4) which is attributed to any character in a PSE story. Four to six specific content categories are defined for each motive (see Table 2).

The unit of scoring is the sentence. Each sentence can be independently scored for the presence of any of the three motive categories *ach*, *aff*, or *pow*. The manual defines an exception to this rule: If a certain motive has been scored (e.g., *aff*), then another motive imagery is present (e.g., *pow*) and then the first motive category *aff* is present again in the same sentence, it can be scored twice in a sentence. The current database, however, does not incorporate such double scorings of a motive in a single sentence and only scores the dichotomous presence (= 1) or absence (= 0) of a motive imagery for each sentence. Following from the combinations of the three motive categories, a sentence can belong to no category (*null*), a single category (e.g., *ach*), or multiple categories (e.g., *achaff* or *achaffpow*).

A second deviation from the manual concerns the "2nd-sentence-rule". This scoring convention states that a motive of a certain category cannot be scored in two consecutive sentences. For example, if *ach* imagery is present in three consecutive sentences, it is only scored in the first and the third sentence. However, the same motive can be scored in both of two consecutive sentences if the two categories are separated by imagery for another motive. Several labs, however, abandoned this 2nd-sentence rule, as it unnecessarily increases the frequency of the *null* category and distorts analyses for psychometric models. The majority of all stories (89%) was coded without applying the 2nd-sentence-rule. Hence, in this variation each sentence is scored independently of the scores of the previous sentence.

Finally, some of the stories of the included studies were scored by multiple coders. In some cases, differences were resolved via discussion and coders agreed on a final scoring. In other cases, however, the diverging scores were averaged, which could lead to fractional scores, such as 0.5 *aff*. As

Felix D. Schönbrodt, Department of Psychology, Ludwig-Maximilians-Universität München, Germany. We embrace the values of openness and transparency in science (http://www.researchtransparency.org/). We therefore publish all data necessary to reproduce the reported results and provide reproducible scripts for all data analyses reported in this paper (XXX).

Acknowledgements. We want to thank the following additional coders: Kira Bleck, Julia Boppel, Franziska Dombaj, Julia Fenkl, Miriam Frisch, Franziska Herbig, Sinja Hondong, Jennifer Jochheim, Tobias Kaiser, Eva Margraf, Anna Ossmann, Dominik Özbe, Lea T. Riegl, Anna Röltgen, Anne Schnabel, Maria Schultheiss, Lukas de Wall. We thank Claudio Lauper for transforming many stories in to a sentence-wise format.

Parts of this research were funded by the German Research Foundation (DFG SCHO 1334/1-1, Felix Schönbrodt and Birk Hagemeyer; SCHU 1210/3-1, Oliver Schultheiss; 254142454 / GRK 2070, Stefan Schulz-Hardt and Andreas Mojzisch) and the Schweizer Nationalfond (SNF 100019_156516, Marie Hennecke and Veronika Brandstätter).

Correspondence concerning this article should be addressed to Felix Schönbrodt, Leopoldstr. 13, 80802 München, Germany. Email: felix.schoenbrodt@psy.lmu.de.

[1] Affiliation/intimacy is a fusion of originally separate coding systems for affiliation and intimacy. Here we use the abbreviation *nAff* for the combined affiliation/intimacy category.

Table 1
Codebook for the PSE Story Database.

Variable name	Data type	Comment	Values
study_id	factor	Identifier for the original study/data set	
codingLab	factor	Lab where the coders were trained	Munich, Erlangen, Osnabrueck, Trier
scoringType	factor	Second sentence rule applied?	eachSentence, 2nd_sentence_rule
PID	factor	Unique person identifier	
USID	factor	Unique story identifier	
UTID	factor	Unique text identifier (each sentence is one 'text')	
pic	factor	Unique picture identifier	See https://osf.io/pqckn/
pic.position	numeric	Position of picture in PSE task	
pic.order	factor	Picture order in PSE task fixed for all participants?	fixed, variable
unit	numeric	Sentence number within each story	
wc	numeric	Word count (at text level)	
sc	numeric	Sentence count (at story level)	
pow	numeric	Presence of power imagery	0 (absent) or 1 (present)
ach	numeric	Presence of achievement imagery	0 (absent) or 1 (present)
aff	numeric	Presence of affiliation/intimacy imagery	0 (absent) or 1 (present)
motclass	factor	Multiclass combination of aff, ach, and pow codings. All mixed codings are collapsed into the category 'mixed'.	none, ach, aff, pow, mixed
motclassFull	factor	Multiclass combination of aff, ach, and pow codings with all possible combinations.	none, ach, aff, pow, achpow, affach, affpow, affachpow
text	character	The text of the sentence.	

Table 2
Categories for Scoring Motive Imagery (Winter, 1991; Winter, 1994)

Motive	Categories
Affiliation/Intimacy	aff1: Positive, friendly, or intimate feelings towards others aff2: Negative feeling about separation aff3: Affiliative, companionate activities aff4: Friendly nurturant acts
Achievement	ach1: Adjectives that positively evaluate performance/outcomes ach2: Descriptions of goals/performances that suggest positive evaluation ach3: Winning or competing with others ach4: Negative feelings about failure, doing badly, lack of excellence ach5: Unique accomplishment
Power	pow1: Strong, forceful actions which inherently have impact on other people pow2: Control or regulation pow3: Attempts to convince, persuade, influence, argue, make a point, etc. pow4: Giving help, support, or advice that is not explicitly solicited pow5: Impressing others, concern about fame, prestige, reputation pow6: Strong emotional reactions in one person to actions of another person

one main purpose of the database is to provide training data for automatic text analysis, which requires unambiguous assignments of sentences to categories, we decided to enter only distinct scores of 0 or 1. In case that multiple coders scored the same stories and did not agree, we relied on the coder who demonstrated the better performance, measured by agreement with expert-scored material.

Stories were minimally preprocessed by automatically splitting them into sentences, converting all words to lower case, and by removing trailing and leading whitespace. Furthermore we put in some effort to correct spelling errors. However, given the size of the database and that no fully automatic correction is possible, there still might be a considerable number of typos. Table 3 shows some rows of the dataset, and how sentences are scored.

Descriptive Statistics

The database combines coded PSE stories from 22 studies. Overall, 53 different pictures have been used, although 29 of them ("newpic") were just recently added and have only very few coded stories (see also below). Only 264 (=1.3%) of all stories were written in response to these new pictures, therefore all picture-related descriptive statistics below have been computed on all other pictures, which had at least 81 stories each.

Overall, the database consists of 134,602 sentences coded with the Winter system, which are nested in 20,039 stories provided by 3,614 participants. Most participants wrote stories to 5 pictures (36.9%) or 6 pictures (37.2%) during their PSE task. A story had on average 6.7 (SD = 3.3) sentences and 91.9 (SD = 34.9) words. These counts were roughly comparable for all pictures, ranging from an average sentence count of 5.1 for picture *neymar* to 7.5 for *soccer_duel*. The average word count was between 81 and 106, except for picture *neymar* which had only 59 words on average.

Table 4 shows the frequency of scorings for each of the three motives. These proportions were only computed on studies that did not apply the 2nd-sentence-rule. Most sentences (59.4%) did not receive any motive category, and only few sentences scored simultaneously on two or even all three motive categories. The power category is slightly overrepresented compared to typical PSE datasets, as one of the included studies ("*JP*") was tailored for assessing the power motive and used only pictures with a strong pull for power imagery.

The Relation of Story Length and Motive Raw Scores

It is well-known that motive counts have a strong correlation with the length of the story, indicated by either word count or sentence count (Pang, 2010; Schultheiss & Pang, 2007). This phenomenon can have multiple causes: (A) To some extent, it follows from the structure of the coding system. As the unit of scoring in the Winter system is the sentence, longer stories with more sentences can (potentially) accumulate more motive scores. More specifically, the coding system imposes an upper limit on scorable motive scores. For example, a story with four sentences cannot have more than four motive scores for each of the three motives, if the 2nd-sentence rule is not applied. Other coding systems impose other constraints. For example, the Partner-Related Agency and Communion Test (PACT; Hagemeyer & Neyer, 2012), which is a relationship-specific variant of PSEs, defines seven content categories for the partner-related need for communion, and each category can be scored maximally once per story. In this case, the number of content categories, but not the number of sentences, limits the sum of motives scores per story. (B) A confounding with unrelated variables, such as verbal fluency, typing speed, creativity, or general vividness of fantasy can cause the relationship. From this perspective, persons who have more experience in typing on a computer keyboard have longer stories and are consequently ascribed stronger motives in the absence of a control for story length. (C) The length of the story can also contain an actual signal related to implicit motives. Persons with a strong implicit motive are assumed to have a dense associative network which connects autobiographical experiences, situational cues, emotional experiences, and behavioral strategies around a motivational theme (Schultheiss, 2001; McClelland, 1987). It is plausible that such a dense associative network makes it easier to generate rapidly available motive-related imagery, which results in more elaborate stories. From this perspective, an increased number of scores due to longer stories is a valid indicator of motive strength.

In practice, every PSE dataset contains probably a mixture of all factors. The challenge is that motive researchers typically want to control for A and B, but not for C. But any attempt to control for one factor probably has unwanted side-effects on factors that contain a true signal ("overcontrolling'). Consequently, there is no easy solution to this problem. Typically two methods have been employed to deal with these confounds (Schultheiss & Pang, 2007): Either (linearly) residualizing motive scores for word count, or computing density scores (i.e., motive scores per 1000 words). Since the unit of scoring is the sentence, however, both residuals and density scores could arguably be computed with sentence count instead of word count.[2]

For an empirical analysis of the word/sentence count and

[2]Given that the modeled outcome variable (i.e., raw motive scores) represents strictly non-negative count data, more specific regression approaches would be appropriate. The distributions of raw motive scores of all three motives follow very closely a negative binomial distribution, which suggests a corresponding generalized linear model for count data. However, the main focus is not the hypothesis test, and the residuals from a Gaussian linear regression correlate > .93 with residuals from a negative binomial regression. Therefore, for increasing the simplicity in practical application of the correction, we focus on the traditionally applied Gaussian linear

Table 3
Exemplary Sentences and Their Scores for Motive Imagery.

Text (original)	Text (translation)	ach	aff	pow	motclassFull
der reporter im bild versucht sich einen eindruck vom leben der beschäftigten der schifffahrt in der vergangenheit zu machen.	The reporter in this picture is trying to get an impression of the life of shipboard employees in the past.	0	0	0	null
als er erfährt, dass dieser kapitän bei einem unwetter über 100 leben gerettet hat, beginnt er aufgeregt der sache auf den grund zu gehen.	When he finds out that this captain saved more than 100 lives during a storm, he excitedly begins to investigate the matter.	0	0	1	pow
immerhin könnte das die geschichte sein, auf die er seit langem wartet.	After all, this could be the story he has been waiting for a long time.	0	0	0	null
zwei freundinnen treffen sich um eine party vorzubereiten.	Two friends meet up to prepare a party.	0	1	0	aff
dazu sitzen auf der terasse in einem restaurant und sammeln ideen für ein motto.	For this purpose, they are sitting on the terrace of a restaurant collecting ideas for the party's theme.	0	1	0	aff
außerdem wollen kurz aufteilen wer welche aufgaben bei der vorbereitung übernimmt.	Furthermore, they briefly want to distribute the preparation tasks among themselves.	0	0	0	null
hinzu kommt ein weiterer freund, der die beiden erkannt hat.	Another friend, who has recognized them, joins.	0	1	0	aff
er möchte kurz eine minute aufmerksamkeit der beiden haben um hallo zu sagen.	He wants to get the girls' attention for one minute to say hello.	0	1	0	aff
die beiden sind so vertieft in ihre arbeit, dass sie ihn gar nicht erst wahrnehmen.	Both girls are so absorbed in their work they do not even notice him.	0	0	0	null
da er scheinbar schon länger steht ist er bereits etwas genervt.	As apparently he has been standing there for a while now, he is already a little annoyed.	0	0	0	null
wir befinden uns im zirkus rogalli.	We are at circus Rogalli.	0	0	0	null
die zwei akrobaten im bild sind bekannt für ihre gefährlichen kunststücke am trapez.	The two acrobats in the picture are famous for their dangerous feats on the trapeze.	1	0	1	achpow
mit ihrer neuen nummer gehen sie noch ein stück weiter.	They go another step further with their new stunt.	1	0	0	ach

motive count relation, we reduced the dataset to 17,340 stories which had $\leq$ 17 sentences (99.2% of all stories)[3] and did not apply the 2nd-sentence rule. Table 5 shows bivariate correlations between key variables. As will be shown below, however, the relationship is partly non-linear. Hence, these linear correlations should be treated with caution.

The joint impact of sentence and word count on overall motive scores. Concerning potential indicators for story length, sentence count sets an upper limit of attainable motive scores. But word count could have an incremental contribution, as longer sentences might have a higher chance of getting a motive score. Therefore, we analyze the impact of both indicators of story length. Furthermore, slopes for word and sentence count might vary between pictures, and also between studies. To allow and account for such variations, we computed mixed effects models with z-standardized sentence and word count as predictors, and random intercepts and slopes for the grouping variables "picture" and "study ID". Finally, we explored the incremental contribution of squared sentence and word count. We added squared predictors as fixed effects, but did not add random slopes for the squared terms due to convergence problems.

Table 6 summarises the explained variance of the fixed effects (marginal R^2, Nakagawa & Schielzeth, 2013, Johnson, 2014) and the random variance of the linear slopes.

models and acknowledge the model misspecification.

[3]We excluded stories with more than 17 sentences as we did not want to model the influence of such rare outliers.

Table 4
Frequency of motive scores and their combinations.

Motive category	Frequency
null	59.4%
aff	14.0%
pow	13.4%
ach	8.9%
affpow	2.2%
achpow	1.5%
affach	0.4%
affachpow	0.2%

For all three motives, models including the squared terms showed a better fit then models without (ΔAIC > 40, all LR test *ps* < .001). However, given the only small increase in R^2, for parsimony and simplicity we decided to focus on models with linear main effects only for further analyses and application in practice.

To disentangle the shared and unique contributions of sentence and word count, we performed a commonality analysis (Nimon, Lewis, Kane, & Haynes, 2008). This analysis allows to partition the explained variance into parts that are unique to certain predictor variables or common to the shared variance of predictors. Table 6 shows how much of the explained variance in each motive raw score could be attributed to the shared variance of sentence and word count, or uniquely to either word or sentence count. The largest explanatory power could be attributed to the common variance of both length indicators, but both indicators had also unique contributions in predicting raw motive counts.

Recommendation: How to control for story length. Having multiple ways of controlling for story length is a potential researcher's degree of freedom (John, Loewenstein, & Prelec, 2012) which allows tweaking a data analysis towards more favorable results by trying out multiple alternative analytical pipelines, and choosing the one that "works best". We see three steps to ensure result-independent preprocessing of data, which in turn reduces false-positive results in the literature and increases generalizability and robustness of analyses.

First, the specific way of how to control for story length can be preregistered before data collection. Such a preregistration can also contain if-then contingencies, such as "If the correlation exceeds .15, we residualize for sentence count". Second, as such analytical pipelines presumably do not change between studies of a lab, each lab can develop standard operating procedures (SOPs) that define a standard workflow which is routinely applied in all similar studies (Lin & Green, 2016). Deviations from this lab-internal standard are of course possible, but have to be justified. Third, such SOPs ideally are harmonized across labs towards a field-wide standard. Below, we suggest such a general approach.

A potential goal of the current analysis was to recommend a fixed, "global" linear correction that can be applied across all pictures and for all studies, using the same regression coefficients. Such an approach would have had the advantage of having comparable corrected motive scores on the same scales across studies. However, as the mixed effects models have shown a huge between-picture and between-study variability in these slopes, we recommend to correct on the sample level, but always to provide the raw data as open data, so that alternative ways of correcting can be applied.

Hence, based on the present most extensive available analysis, we suggest some general recommendations and a specific procedure how to control for story length:

1. Do not use density scores. Although previous publications suggested to use density scores (e.g., Winter, 1991), we recommend *not* to use them. First, they do not really control for word and sentence count (see Table 5), but, in contrast, even reverse the relationship. Second, they overemphasize very short stories and punish long stories. A single-sentence story with a motive score receives the maximally attainable density of 100%, while a long, elaborate story that has many scores in most, but not all sentences, has a lower density. This directly contradicts assumption (C) which states that dense implicit motive networks are supposed to lead to longer stories.

2. Control for both sentence count and word count. Both indicators of story length have unique and substantial contributions in predicting raw motive count, and consequently should be controlled for. Controlling for the linear effect is sufficient for practical purposes.

3. Control on picture level. Typically, motive scores across all administered picture stimuli have been controlled for overall word count across all stories. As slopes substantially vary between pictures, we suggest to control on picture level (i.e., predicting raw motive scores for a picture by the sentence count and word count of that picture). This results in motive score residuals for each picture, which then can be averaged for a total motive residual score.

4. Control for each motive separately.

5. Use a robust regression approach. A potential drawback of controlling on sample level, with potentially small sample sizes, is the susceptability to single outliers and the risk of overfitting. Therefore we suggest to use a robust regression approach that automatically takes care of outliers and is robust non-normality, such as MM-estimators implemented in the *lmrob* function of the *R* package *robustbase* (Maechler et al., 2018; for an overview, see Yu, Yao, & Bai, 2014).

The recommended procedure. For each picture and each motive separately, predict raw motive scores using a robust regression model. Enter *z*-transformed sentence count and word count as predictors. Extract the residuals, which then

Table 5
Descriptive Statistics and Correlations for Overall Motive, Word, and Sentence Count.

	Mean	SD	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Overall motive score	2.99	1.99	-	.90	.92	.75	.58	.40	.43
(2) Overall motive score, word count resid.	-0.00	1.80		-	.93	.90	.64	.12	-.00
(3) Overall motive score, sentence count resid.	0.00	1.83			-	.81	.77	-.00	.18
(4) Overall motive density score (per 100 words)	3.40	2.15				-	.68	.02	-.14
(5) Overall motive density score (per sentence)	0.50	0.35					-	-.30	.01
(6) Sentence count per story	6.57	3.00						-	.67
(7) Word count per story	91.01	34.29							-

Note. Analyses are based on 17,340 stories which had ≤ 17 sentences and did not apply the 2nd-sentence rule.

Table 6
Mixed Effects Models for Predicting Raw Motive Scores by Story Length.

	Model / predictor	aff	ach	pow
marginal R^2	$sc + wc$	5.6%	1.7%	7.7%
	$sc + wc + sc^2 + wc^2$	5.8%	2.1%[b]	7.8%
Commonality analysis: How much of the explained variance can be attributed to unique and common parts of predictors?	Common to $sc + wc$	59.9%	51.9%	49.6%
	Unique to sc	8.6%	21.1%	14.7%
	Unique to wc	31.6%	26.9%	35.7%
Fixed effects (SE) (all predictors standardized, linear main effects only)	sc	0.14 (0.039)	0.08 (0.032)	0.14 (0.034)
	wc	0.26 (0.049)	0.09 (0.018)	0.25 (0.034)
Random slope variances (*SD*s) based on “picture”[a]	sc	0.02 (0.15)	0.02 (0.13)	0.02 (0.12)
	wc	0.03 (0.16)	0.00 (0.07)	0.01 (0.10)
Random slope variances (*SD*s) based on “study ID”[a]	sc	0.01 (0.08)	0.00 (0.05)	0.00 (0.07)
	wc	0.02 (0.13)	NA (NA)	0.01 (0.09)

Note. *sc* = sentence count, *wc* = word count. [a]The random variances are based on the models including only linear terms as fixed and random effects. [b]Model with squared terms for *ach* excluded the random slope for word count due to convergence problems.

are used in subsequent analyses. This can be accomplished with the following *R* code:

```
# install required package (only has to be
    done once):
install.packages("robustbase")
library(robustbase) # The setting =
    "KS2014" is strongly recommended

# Do the following analysis for each
    picture and each motive:
rlm.aff.pic1 <- lmrob(formula =
    aff.raw.pic1 ~ sc.pic1 + wc.pic1, data =
    dat, setting = "KS2014")
aff.residual.pic1 <- resid(rlm.aff.pic1)
```

No Decline Effect for Later Pictures in the PSE Task

Writing imaginative stories can be exhausting, and existing literature (TODO CITATION NEEDED) speculated that pictures which are administered later during the PSE task are shorter and generate less motive scores. This hypothesis, however, could not be confirmed in a subset of the database that administered the pictures at random position (i.e., not with a fixed order; n = 4365 stories; see Figure 1).

For a formal test, we conducted mixed effect models with picture position as predictor, overall motive score, sentence count, and word count as dependent variable, and a random intercept and a random slope for “picture”. If anything, analyses revealed the positive trend that later picture positions had more overall motive scores (b = 0.05, p = .002), higher sentence count (b = 0.06, p = .051), and higher word count (b = 0.96, p < .001).

A Database of Pictures Used in PSEs

All 53 pictures which have been used in the PSE database are provided in an OSF project (https://osf.io/pqckn/). This project also includes a table that shows the license and the provenance of each picture, as far as these informations could be reconstructed.

This collection of pictures includes some classic pictures (such as the “standard six” set and some TAT pictures), but

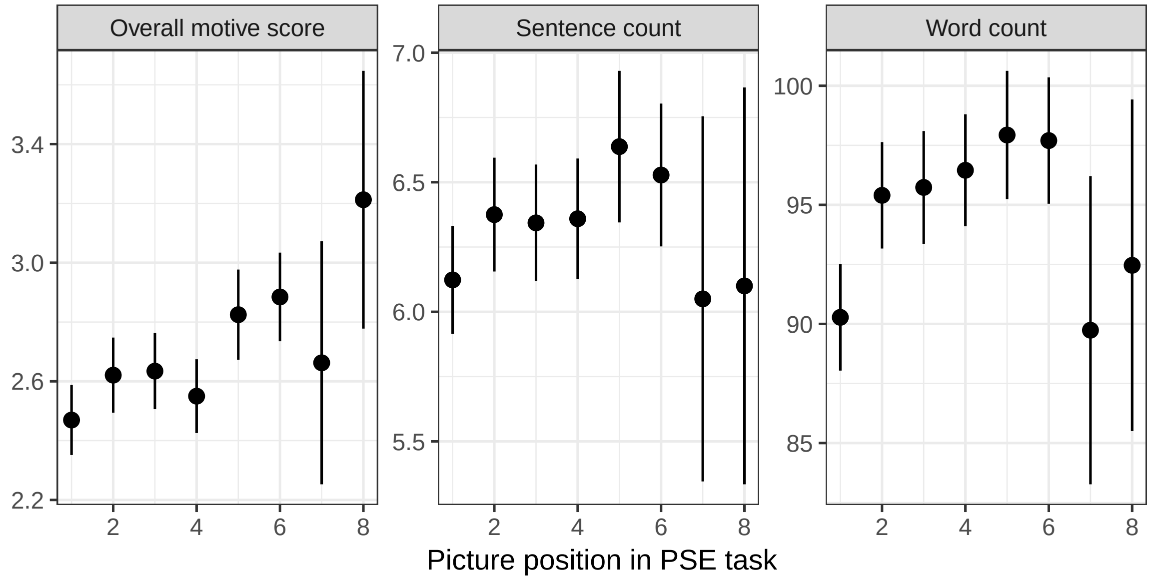

Figure 1. Descriptive overall motive score, sentence count, and word count for each picture position. Error bars are 95% confidence intervals. Figure available at https://osf.io/pqckn/, under a CC-BY4.0 license.

also modern pictures that have been used in PSE research. As the license of some pictures is not clear, four experts (Birk Hagemeyer, Felix Schönbrodt, Lena Schiestel, and Larissa Sust) searched for 29 new pictures, all of which promised to have a considerable motive pull. All of these new pictures (starting with the label "newpic") have an open license (CC0, CC-BY, or CC-BY-SA) and therefore can be safely reused for research and other purposes. Figure 2 exemplarily shows six of these new pictures which had a strong overall motive pull in a preliminary dataset.

Updated Picture Norms

Norms of picture pulls have been published for the six standard pictures by Schultheiss and Brunstein (2001) (n = 424, German stories), Pang and Schultheiss (2005) (n = 320, English stories), and Schultheiss, Yankova, Dirlikov, and Schad (2009) (n = 190, English stories). All of these norms employed the 2nd-sentence rule.

Here we present updated norms for German PSE stories, which are based on larger samples and sentence-wise coding without the 2nd-sentence rule. Sample sizes vary between pictures, depending on how often a picture has been used in the studies included in the database. Table 7 shows descriptive statistics for all pictures that had at least 50 stories, ordered by overall motive pull.[4]

Pictures differ in their ability to pull multiple types of motive imagery. Some pictures are mostly monothematic, such as *couple_bridge* or *cafe* which almost exclusively elicit affiliation imagery. Other pictures elicit imagery from two motives (e.g., *women_laboratory* for *ach* and *pow*), or even all three motives to some extent (e.g., *applause* or *trapeze*). The ability of a picture to elicit imagery from multiple motives has also been termed *cue ambiguity* (Pang, 2010). Figure 3 shows a ternary plot (Hamilton, 2017) that visualizes whether pictures are rather monothematic (located at the corners of the triangle), pull for two motives (around the midpoint of each side of the triangle), or pull for multiple motives (in the middle of the triangle).

Availability of the databases

Both the database on scored PSE stories and the picture database can be downloaded and reused freely from the Open Science Framework (https://osf.io/pqckn/) under a CC-BY 4.0 license. Please cite this publication if you use either database in your work.

Discussion

In this paper, we presented two databases: (a) A database of 134,602 sentences, nested in 20,039 stories, scored by experts using the Winter (1994) *Manual for scoring motive imagery in running text*, and (b) a database of 53 classic and new pictures which have been used in PSE research. Furthermore, we provided descriptive statistics on typical sentence and word counts, analyses and recommendations for how to correct motive scores for story length, and updated norm values for picture pulls.

We see several potential use cases for these databases. The primary intention for creating the PSE story database was to

[4] Descriptive statistics for all pictures, including the new pictures, are in the online supplementary material.

newpic18

newpic12

newpic10

newpic9

newpic7

newpic2

Figure 2. Examples of new pictures with an open license. Credits: *newpic12*: Pete Lewis / Department for International Development; TODO ADD OTHER LICENSES.

provide a large training dataset for automatic text analysis. We want to emphasize that these expert-scored sentences go beyond a simple sentiment analysis (e.g., positive vs. negative product reviews) that can quite easily be implemented using dictionaries (e.g., Feldman, 2013). In contrast, scoring implicit motives requires deep semantic processing, evaluating nuances in meaning, differentiating negations, hypothetical from actual actions, and much more. To what extent mathematical text models or machine learning algorithms are able to replicate human scorings in the Winter scoring system is an open question (see, however, Schultheiss, 2013 for a potential approach to automatic coding). In addition to supervised learning that tries to approximate human codings, the dataset can also be used to infer structures using unsupervised learning methods, such as topic models and latent dirichlet allocation (Blei, Ng, & Jordan, 2003).

Another use case lies in psychometric modeling. It has been argued that measurement models based on classical test theory violate assumed underlying processes in PSEs and therefore are not applicable (Atkinson, 1981; Hibbard, 2003; Schultheiss, Liening, & Schad, 2008). This large database allows testing and developing alternative measurement models that might provide more appropriate estimates of reliability and shed light on the response processes during a PSE task (see, for example, Tuerlinckx, De Boeck, & Lens, 2002, Lang, 2014, Runge et al., 2016). We present the first large dataset that provides motive scores on sentence level, thus allowing to investigate within-story dynamics separately from between-story dynamics. This allows testing decade-old theories with high statistical power.

Finally, now a systematic investigation of differences between labs in coding style is possible. Although all labs employed the same manual, effects such as coder drift (Schultheiss & Pang, 2007) can lead to an evolution of implicit coding rules that lets labs drift apart.

The updated picture norms allow to select appropriate sets of pictures for a PSE. For example, it has been recommended to select pictures with a high motivational pull for the tar-

Table 7
Means and Standard Deviations of Raw Scores Across Coding Categories and Pictures.

Number	Picture ID	Aff	Ach	Pow	Overall	*n*
1	applause	**1.84** (1.46)	0.82 (1.13)	**1.82** (1.52)	4.48	1004
2	sorrow	**2.46** (2.00)	0.16 (0.61)	**1.65** (1.58)	4.27	141
3	beachcombers	0.62 (1.10)	0.14 (0.50)	**3.40** (1.87)	4.17	797
4	three_people	**2.95** (1.84)	0.02 (0.22)	0.64 (0.94)	3.62	81
5	*nightclub	**2.33** (1.60)	0.15 (0.40)	1.00 (1.20)	3.48	2259
6	burglar	**2.03** (1.72)	0.15 (0.46)	**1.25** (1.48)	3.43	141
7	woman	**1.55** (1.62)	0.16 (0.70)	**1.70** (1.52)	3.41	120
8	kennedy_nixon	0.10 (0.38)	**1.30** (1.32)	**1.82** (1.42)	3.22	799
9	architect	**2.21** (1.66)	0.48 (0.80)	0.49 (0.84)	3.18	409
10	*women_laboratory	0.33 (0.76)	**1.51** (1.25)	**1.33** (1.34)	3.17	1778
11	*couple_bridge	**2.55** (1.79)	0.04 (0.25)	0.36 (0.69)	2.95	1802
12	*boxer	0.32 (0.79)	**1.76** (1.37)	0.81 (1.08)	2.89	1173
13	violin	0.97 (1.12)	0.76 (1.03)	**1.04** (1.15)	2.77	144
14	bicycle	0.18 (0.50)	**1.74** (1.34)	0.72 (0.74)	2.64	585
15	cafe	**2.26** (1.07)	0.04 (0.19)	0.34 (0.60)	2.64	502
16	*trapeze	0.68 (1.08)	**1.06** (1.10)	0.88 (0.99)	2.62	2127
17	neymar	0.35 (0.79)	**1.20** (1.13)	0.86 (1.10)	2.40	354
18	lacrosse_duel	0.19 (0.47)	**1.86** (1.28)	0.30 (0.52)	2.35	98
19	*ship_captain	0.48 (0.90)	0.22 (0.56)	**1.60** (1.42)	2.31	2062
20	soccer_duel	0.16 (0.49)	**1.45** (1.27)	0.65 (0.95)	2.27	643
21	group	0.84 (1.23)	0.20 (0.55)	**1.12** (1.15)	2.16	126
22	window	0.94 (1.41)	0.11 (0.34)	0.72 (1.14)	1.77	124
23	canyon	0.67 (0.96)	0.18 (0.43)	0.88 (1.24)	1.72	112
24	men_on_ship	0.08 (0.34)	0.27 (0.49)	0.76 (0.77)	1.11	97

Note. Overall is the sum of all three motive categories. Pictures are ordered along their overall motive pull. Average raw motive scores ≥ 1 are printed in bold. Pictures of the "standard six" set are marked with an asterisk. The actual pictures are provided in an OSF project (https://osf.io/pqckn/).

geted motives (Schultheiss & Pang, 2007). With the large collection of available pictures and their updated norms, such choices can be empirically informed. Finally, we encourage researchers who use a PSE to refer to the unique IDs of the picture database in their methods section, and to contact the first author if they want to add a new picture to the database. (Preferably pictures with a permissive license that allows reuse.) A common and standardized catalogue of PSE pictures enhances the replicability of studies, but also the reusability and interoperability of research results, as such clear identifiers allow the aggregation and reanalysis of datasets.

To conclude, we hope that these two public databases are a helpful resource both for PSE researchers and more generally for researchers interested in text content analysis, and that they refuel a renewed interest in methodological and psychometric research about measuring implicit motives with Picture Story Exercises.

Author contributions

Felix Schönbrodt: Aggregated, cleaned, preprocessed, and analyzed data; provided multiple datasets; scored stories; searched for new pictures; wrote manuscript. *Birk Hagemeyer*: Provided dataset; searched for new pictures; wrote manuscript. *Veronika Brandstätter*: Provided multiple datasets. *Thomas Czikmantori*: Provided multiple datasets. *Peter Gröpel*: Provided dataset. *Marie Hennecke*: Provided multiple datasets. *Kevin Janson*: Provided dataset; scored stories. *Nina Kemper*: Provided dataset; scored stories. *Martin Köllner*: Provided multiple datasets. *Philipp Kopp*: Provided dataset. *Andreas Mojzisch*: Provided dataset. *Raphael Müller-Hotop*: Provided dataset. *Johanna Prüfer*: Provided dataset. *Markus Quirin*: Provided dataset. *Bettina Scheidemann*: Provided multiple datasets; scored stories. *Lena Schiestel*: Provided dataset; scored stories; searched for new pictures. *Stefan Schulz-Hardt*: Provided dataset; scored stories. *Larissa Sust*: Provided dataset; scored stories; searched for new pictures. *Caroline Zygar*: Provided dataset; scored stories. *Oliver Schultheiss*: Provided multiple datasets; scored stories.

All coauthors critically reviewed, commented, and approved the final manuscript.

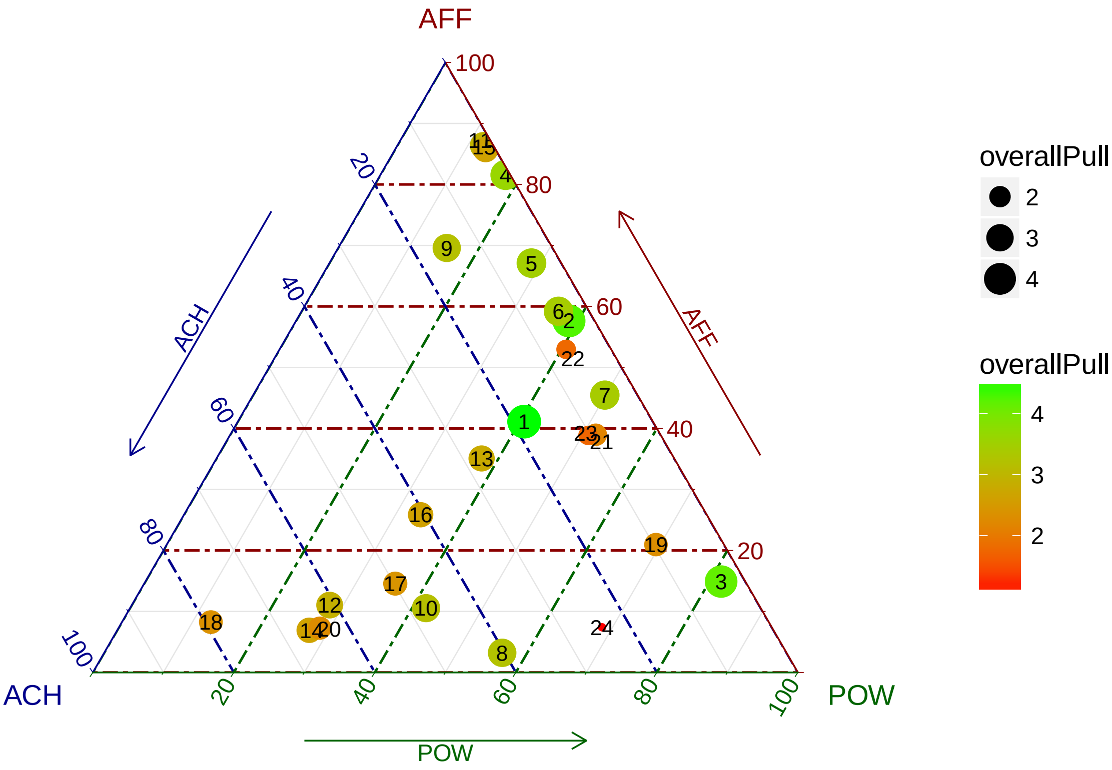

Figure 3. Relative motive pull of pictures. Numbers correspond to picture numbers in Table XXX. Figure available at https://osf.io/pqckn/, under a CC-BY4.0 license.

References

Atkinson, J. W. (1981). Studying personality in the context of an advanced motivational psychology. *American Psychologist*, *36*, 117–128. doi:10.1037/0003-066X.36.2.117

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *J. Mach. Learn. Res. 3*, 993–1022. Retrieved August 9, 2018, from http://dl.acm.org/citation.cfm?id=944919.944937

Feldman, R. (2013). Techniques and Applications For Sentiment Analysis. *Communications of the ACM*, *56*, 82–89. Retrieved August 13, 2018, from https://cacm.acm.org/magazines/2013/4/162501-techniques-and-applications-for-sentiment-analysis/abstract

Hagemeyer, B. & Neyer, F. J. (2012). Assessing implicit motivational orientations in couple relationships: The Partner-Related Agency and Communion Test (PACT). *Psychological Assessment*, *24*, 114–128. doi:10.1037/a0024822. pmid: 21787092

Hamilton, N. (2017). *Ggtern: An Extension to 'ggplot2', for the Creation of Ternary Diagrams*. R package version 2.2.1. Retrieved from https://CRAN.R-project.org/package=ggtern

Hibbard, S. (2003). A critique of Lilienfeld et al.'s (2000) 'The scientific status of projective techniques'. *Journal of Personality Assessment*, *80*, 260–271. doi:10.1207/S15327752JPA8003_05

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*, 524–532. doi:10.1177/0956797611430953

Johnson, P. C. (2014). Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods in Ecology and Evolution*, *5*, 944–946. doi:10.1111/2041-210X.12225. pmid: 25810896

Lang, J. (2014). A dynamic Thurstonian item response theory of motive expression in the picture story exercise : Solving the internal consistency paradox of the PSE. *Psychological Review*, *121*, 481–500. doi:http://dx.doi.org/10.1037/a0037011

Lin, W. & Green, D. P. (2016). Standard Operating Procedures: A Safety Net for Pre-Analysis Plans. *PS: Political Science & Politics*, *49*, 495–500. doi:10.1017/S1049096516000810

Maechler, M., Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., … di Palma, M. A. (2018). *Robustbase: Basic Robust Statistics*. Retrieved from http://robustbase.r-forge.r-project.org/

McClelland, D. C. (1987). *Human motivation*. New York: Cambridge University Press.

Morgan, C. D. & Murray, H. A. (1935). A method for examining fantasies: The Thematic Apperception Test. *Archives of Neurology and Psychiatry*, *34*, 289–306.

Nakagawa, S. & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*, 133–142. doi:10.1111/j.2041-210x.2012.00261.x

Nimon, K., Lewis, M., Kane, R., & Haynes, R. (2008). An R package to compute commonality coefficients in the multiple regression case: An introduction to the package and a practical example. *Behavior Research Methods*, *40*, 457–466. doi:10.3758/BRM.40.2.457

Pang, J. S. & Schultheiss, O. C. (2005). Assessing implicit motives in US college students: Effects of picture type and position, gender and ethnicity, and cross-cultural comparisons. *Journal of Personality Assessment*, *85*, 280–294.

Pang, J. S. (2010). Content Coding Methods in Implicit Motive Assessment: Standards of Measurement and Best Practices for the Picture Story Exercise. In O. C. Schultheiss & J. C. Brunstein (Eds.), *Content Coding Methods in Implicit Motive Assessment: Standards of Measurement and Best Practices for the Picture Story Exercise* (pp. 119–150). Oxford University Press. Retrieved August 15, 2018, from http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195335156.001.0001/acprof-9780195335156-chapter-5

Runge, J. M., Lang, J. W. B., Engeser, S., Schüler, J., den Hartog, S. C., & Zettler, I. (2016). Modeling motive activation in the Operant Motive Test: A psychometric analysis using dynamic Thurstonian item response theory. *Motivation Science*, *2*, 268–286. doi:10.1037/mot0000041

Schultheiss, O. C. (2001). An information processing account of implicit motive arousal. In M. L. Maehr & P. Pintrich (Eds.), *Advances in motivation and achievement (Vol. 12: New directions in measures and methods)* (pp. 1–41). Greenwich: JAI Press.

Schultheiss, O. C. (2013). Are implicit motives revealed in mere words? Testing the marker-word hypothesis with computer-based text analysis. *Frontiers in Psychology*, *4*. doi:10.3389/fpsyg.2013.00748. pmid: 24137149

Schultheiss, O. C. & Brunstein, J. (2010). *Implicit Motives*. Oxford University Press.

Schultheiss, O. C. & Brunstein, J. C. (2001). Assessment of implicit motives with a research version of the TAT: Picture profiles, gender differences, and relations to other personality measures. *Journal of Personality Assessment*, *77*, 71–86.

Schultheiss, O. C., Liening, S. H., & Schad, D. J. (2008). The reliability of a Picture Story Exercise measure of implicit motives: Estimates of internal consistency, retest reliability, and ipsative stability. *Journal of Research in Personality*, *42*, 1560–1571.

Schultheiss, O. C. & Pang, J. S. (2007). Measuring implicit motives. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 322–344). Guilford Press.

Schultheiss, O. C., Yankova, D., Dirlikov, B., & Schad, D. J. (2009). Are implicit and explicit motive measures statistically independent? A fair and balanced test using the Picture Story Exercise and a cue-and response-matched questionnaire measure. *Journal of P0ersonality Assessment*, *91*, 72–81.

Smith, C. P., Atkinson, J. W., McClelland, D. C., & Veroff, J. (1992). *Motivation and personality: Handbook of thematic content analysis* (C. P. Smith, J. W. Atkinson, D. C. McClelland, & J. Veroff, Eds.). New York, NY US: Cambridge University Press.

Tuerlinckx, F., De Boeck, P., & Lens, W. (2002). Measuring needs with the Thematic Apperception Test: A psychometric study. *Journal of Personality and Social Psychology*, *82*, 448–461.

Winter, D. G. (1994). *Manual for scoring motive imagery in running text (4th edition)*. Ann Arbor: University of Michigan.

Winter, D. G. (1991). Measuring personality at a distance: Development of an integrated system for scoring motives in running text. In *Perspectives in personality, Vol. 3: Part A: Self and emotion; Part B: Approaches to understanding lives* (pp. 59–89). London, England: Jessica Kingsley Publishers.

Yu, C., Yao, W., & Bai, X. (2014). Robust Linear Regression: A Review and Comparison. arXiv: 1404.6274 [stat]. Retrieved August 22, 2018, from http://arxiv.org/abs/1404.6274

Appendix A: picture norms for all pictures in the database

Table 8
Means and Standard Deviations of Raw Scores Across Coding Categories and Pictures.

Picture ID	Aff	Ach	Pow	Overall	*n*	Sentence count	Word count
newpic12	**1.33** (1.03)	**1.67** (1.37)	**1.83** (1.47)	4.83	6	6.8 (2.9)	78 (26)
newpic5	**1.36** (1.12)	**1.73** (1.56)	**1.55** (0.69)	4.64	11	6.0 (2.6)	66 (26)
newpic18	**3.25** (1.71)	0.00 (0.00)	**1.25** (1.50)	4.50	4	10.5 (3.7)	118 (31)
applause	**1.84** (1.46)	0.82 (1.13)	**1.82** (1.52)	4.48	1004	6.5 (2.8)	92 (33)
sorrow	**2.46** (2.00)	0.16 (0.61)	**1.65** (1.58)	4.27	141	7.5 (3.4)	90 (33)
beachcombers	0.62 (1.10)	0.14 (0.50)	**3.40** (1.87)	4.17	797	6.8 (2.9)	97 (34)
newpic6	**2.38** (1.77)	0.38 (0.74)	**1.25** (1.16)	4.00	8	5.1 (2.5)	70 (32)
newpic8	0.89 (0.93)	**1.56** (1.42)	**1.44** (1.51)	3.89	9	7.1 (3.4)	73 (30)
newpic29	**2.75** (1.49)	0.00 (0.00)	**1.12** (1.64)	3.88	8	7.8 (4.1)	106 (57)
newpic17	0.25 (0.62)	**2.83** (1.99)	0.67 (0.65)	3.75	12	7.8 (2.1)	75 (27)
three_people	**2.95** (1.84)	0.02 (0.22)	0.64 (0.94)	3.62	81	6.2 (3.2)	96 (28)
newpic32	**1.57** (0.79)	0.14 (0.38)	**1.86** (2.34)	3.57	7	10.6 (7.4)	119 (73)
newpic31	**2.45** (1.21)	0.55 (0.82)	0.55 (0.82)	3.55	11	5.9 (4.3)	77 (42)
*nightclub	**2.33** (1.60)	0.15 (0.40)	1.00 (1.20)	3.48	2259	6.9 (3.3)	93 (36)
newpic25	**2.18** (2.23)	0.09 (0.30)	**1.18** (1.40)	3.45	11	7.4 (2.6)	92 (41)
newpic9	0.44 (0.73)	**1.78** (1.92)	**1.22** (1.09)	3.44	9	5.7 (4.3)	75 (41)
burglar	**2.03** (1.72)	0.15 (0.46)	**1.25** (1.48)	3.43	141	7.3 (3.2)	92 (31)
woman	**1.55** (1.62)	0.16 (0.70)	**1.70** (1.52)	3.41	120	7.0 (3.9)	94 (41)
newpic30	0.60 (0.83)	**1.33** (1.95)	**1.47** (1.51)	3.40	15	7.1 (4.0)	88 (44)
newpic27	**2.33** (1.15)	0.67 (0.58)	0.33 (0.58)	3.33	3	10.0 (4.6)	113 (56)
newpic33	**1.89** (0.78)	0.33 (0.71)	**1.11** (1.05)	3.33	9	8.1 (3.8)	88 (37)
newpic14	**1.88** (1.13)	0.00 (0.00)	**1.38** (1.30)	3.25	8	9.8 (4.2)	100 (45)
kennedy_nixon	0.10 (0.38)	**1.30** (1.32)	**1.82** (1.42)	3.22	799	5.9 (2.7)	85 (31)
architect	**2.21** (1.66)	0.48 (0.80)	0.49 (0.84)	3.18	409	7.1 (3.3)	106 (34)
*women_laboratory	0.33 (0.76)	**1.51** (1.25)	**1.33** (1.34)	3.17	1778	6.4 (3.0)	91 (33)
newpic24	0.77 (0.93)	0.69 (1.18)	**1.69** (2.25)	3.15	13	9.5 (4.0)	104 (40)
*couple_bridge	**2.55** (1.79)	0.04 (0.25)	0.36 (0.69)	2.95	1802	6.8 (3.4)	95 (37)
newpic22	0.70 (1.06)	**1.20** (1.32)	**1.00** (1.25)	2.90	10	8.2 (2.6)	105 (53)
*boxer	0.32 (0.79)	**1.76** (1.37)	0.81 (1.08)	2.89	1173	6.7 (3.6)	87 (36)
newpic11	0.40 (0.70)	0.10 (0.32)	**2.30** (1.57)	2.80	10	6.8 (3.9)	87 (52)
violin	0.97 (1.12)	0.76 (1.03)	**1.04** (1.15)	2.77	144	7.1 (3.9)	99 (47)
newpic16	**1.29** (1.11)	0.00 (0.00)	**1.43** (0.79)	2.71	7	6.3 (2.8)	84 (31)
newpic10	**1.00** (1.10)	0.83 (1.60)	0.83 (0.41)	2.67	6	6.2 (1.9)	88 (28)
bicycle	0.18 (0.50)	**1.74** (1.34)	0.72 (0.74)	2.64	585	7.4 (3.4)	95 (37)
cafe	**2.26** (1.07)	0.04 (0.19)	0.34 (0.60)	2.64	502	6.8 (2.8)	94 (31)
*trapeze	0.68 (1.08)	**1.06** (1.10)	0.88 (0.99)	2.62	2127	6.9 (3.4)	91 (35)
newpic21	**1.89** (1.45)	0.22 (0.44)	0.44 (0.53)	2.56	9	8.2 (5.7)	92 (35)
newpic20	**1.75** (0.46)	0.00 (0.00)	0.75 (1.75)	2.50	8	6.2 (4.5)	73 (44)
newpic7	0.80 (0.86)	0.87 (1.36)	0.80 (0.77)	2.47	15	5.6 (3.2)	69 (36)
neymar	0.35 (0.79)	**1.20** (1.13)	0.86 (1.10)	2.40	354	5.1 (2.4)	59 (25)
newpic23	0.62 (0.74)	0.25 (0.46)	**1.50** (1.93)	2.38	8	7.5 (4.8)	85 (39)
lacrosse_duel	0.19 (0.47)	**1.86** (1.28)	0.30 (0.52)	2.35	98	6.4 (3.2)	92 (36)
*ship_captain	0.48 (0.90)	0.22 (0.56)	**1.60** (1.42)	2.31	2062	6.3 (3.1)	94 (34)
newpic3	**2.10** (1.29)	0.10 (0.32)	0.10 (0.32)	2.30	10	9.6 (6.4)	98 (49)
soccer_duel	0.16 (0.49)	**1.45** (1.27)	0.65 (0.95)	2.27	643	7.5 (3.4)	91 (33)
newpic13	**1.18** (1.40)	0.18 (0.40)	0.82 (0.98)	2.18	11	6.2 (3.5)	75 (33)
group	0.84 (1.23)	0.20 (0.55)	**1.12** (1.15)	2.16	126	6.8 (3.9)	88 (41)
newpic4	0.25 (0.71)	**1.12** (0.99)	0.62 (0.74)	2.00	8	7.8 (6.0)	93 (44)
newpic28	**1.20** (1.32)	0.13 (0.35)	0.60 (1.06)	1.93	15	6.9 (3.6)	87 (40)
window	0.94 (1.41)	0.11 (0.34)	0.72 (1.14)	1.77	124	7.4 (5.0)	89 (43)
canyon	0.67 (0.96)	0.18 (0.43)	0.88 (1.24)	1.72	112	6.6 (4.8)	81 (44)
men_on_ship	0.08 (0.34)	0.27 (0.49)	0.76 (0.77)	1.11	97	5.5 (2.7)	84 (30)
newpic26	0.67 (0.58)	0.00 (0.00)	0.33 (0.58)	1.00	3	6.3 (3.5)	83 (59)

Note. Overall is the sum of all three motive categories. Pictures are ordered along their overall motive pull. The actual pictures are provided in an OSF project (https://osf.io/pqckn/).