# Cover letter for JPA submission JPA-2019-057: "Measuring Implicit Motives With the Picture Story Exercise (PSE): Databases of Expert-Coded German Stories, Pictures, and Updated Picture Norms."

Dear Dr. Cain,

We were happy to hear that you think our work would be of interest to journal readers, and are grateful for the opportunity to submit a revised version.

During the revision, we managed to include two additional studies to the database, raising the number of stories from 21,941 to 26,389. As one of the new studies included 5 of the new pictures, we can include them now in the main descriptive table. Furthermore, we managed to add participant-level data on age and gender, which allowed us to compute a meta-analysis on the gender-effect on nAff, including a comparison of the different options of correcting for story length.

We first reply to the points that you highlighted as most important for a successful revision, and then reply to each additional comment of the reviewers.

### The editor’s points:

E1: Like Reviewer 1, I wondered if you could give a bit more rationale for the proposed deviations from the established coding rules. This section wasn’t particularly clear to me and I think you could provide additional arguments for these deviations.

*One change concerned the application of the “second sentence rule”. The original coding systems for the TAT - which are the basis for Winter’s coding system - have no such rule. Winter introduced this rule mainly for the coding of historical text material and speeches; hence, we think that there is rather a need for justification why such a rule should be employed in PSEs, which are close to the original TAT procedure. An omission of the second sentence rule brings the coding system closer to the original intention of coding fantasy stories. Furthermore, omitting codings might lead to a loss of relevant diagnostic information. Finally, coding every sentence has psychometric advantages, as the high prevalence of “null”-codings (which is amplified by the second sentence rule) complicates the estimation of person parameters in psychometric models. Potential interdependencies between codings should be handled in psychometric models, not by removing the data in the first place.*

*Another change concerns the possibility of having a second coding of the same motive if another motive is coded in between. This happens very rarely (in about 0.5% of all sentences in the current data base). We opted to use the simplified coding scheme in order to harmonize the slightly different coding systems in the data base.*

*In the revised manuscript, we now present these arguments in a more elaborate way.*

E2: Like Reviewer 1, I think it would strengthen your contribution if you could provide additional analysis on the relationship between motive scores derived by different ways to control for story length and psychological/behavioral outcomes or the effects of age, gender, education on pull of picture cards.

*As suggested by R2, we used the well-established gender difference in the implicit intimacy/affiliation motive as showcase to demonstrate the different ways to control for story length. For this purpose, we used the newly added information about participants’ gender to replicate a published meta-analysis on the nAff gender effect. A comparison of three different ways to control for story length revealed that the recommended procedure (with robust regression residuals) showed the strongest effect size, and that density scores showed the weakest effect (see new subsection “Effect of correction type on the affiliation gender effect.” in the revised manuscript).*

E3: Reviewer 2 also highlighted important aspects of the data sets that would be helpful to include, such as details about age, gender, word count, etc. It may not be possible to include all of this information, but more information about the data sets will strengthen your overall contribution.

*As suggested, we added the word count to the summary table of the picture descriptives. Furthermore, we managed to associate individual-level information on age and gender wherever possible. For two (archival) data sets, it was not possible to match the associated demographics, but for 24 of the 26 data sets we now provide this information.*

*Additionally, we now provide additional information on the study-level: Date and location of data collection, Administration method (handwritten/ computer-written online study / computer-written in lab), Setting of data collection (individual / group testing), Population (students / mostly students / mostly non-students / non-students) (see Table 2 in the revised manuscript).*

E4: At JPA we require authors to report statistical results using standard effect size measures (e.g., r, Cohen's d) in addition to statistical significance. Please include effect sizes that correspond to all of your statistical findings when you submit a revised manuscript.

*In the revised manuscript, effect sizes are reported wherever feasible. Specifically, Table 7 reports marginal R² for the predictor sets as an effect size measure. The new Table 8 for the gender effect meta-analysis includes Hedge’s g.*

*For the analysis of the decline effect, all of our inferential tests involve mixed effect models, due to the nested data structure. To our knowledge, there is no agreed-upon method of computing standardized effect sizes in these models, as multiple ways of standardizing are possible. For these reasons several scholars recommend to report unstandardized coefficients (REF TODO!), as we did, because they allow much clearer interpretations. As another approach, Rights and Sterba (2019) recently proposed 12 different R² measures (for the full model, not for single coefficients), each carrying another meaning. We don’t think that this is relevant or helpful for guiding the interpretation in the current analyses. However, if the journal has a preferred way of computing standardized estimates for these models, we certainly can provide them in another revision.*

### Reviewer 1

R1.1: [...] a number of recommendations are given that mostly mirror procedures typically implemented by researchers in the field. Given this widespread agreement, I wonder whether colleagues in the field really check different ways of controlling for story length to gain favorable results (see page 10).

*We toned that part down and now write about a “potential researcher’s degree of freedom”, without implying that this indeed has been employed in the field.*

R1.2: Given the coding rules recommended by Winter (1994), the authors propose a slightly different approach in motive coding. They recommend scoring a) an indicator for a given motive only once per sentence and b) every single sentence. As there are clear differences in verbal fluency between participants, I wonder whether such an approach does not lay too much stress on motive imagery in relatively short sentences. It also seems that basically two labs have contributed to the promoted databases (see Table 2). However, even within those two labs different coding rules were applied (e.g., sentence rule). The authors might provide additional arguments for the proposed deviations from established coding rules.

*Please, see our response to E1.*

R1.3: In general, the significance of the paper could be strengthened if the authors would provide additional analyses on, for example, the relationship between motive scores derived by different ways to control for story length and psychological/behavioral outcomes or effects of age, gender, and education on pull of picture cards.

*This is an excellent idea, which we implemented in an analysis on the well-known gender effect on nAff (see also our response to E2).*

R1.4: The authors repeatedly state that their databases include 54 picture cues. However, 30 (new) cards were used in a single study with only 53 participants that produced a total sum of 276 stories. Even if the paper lists no information on details of this study, I wonder whether these pictures cues are ready to be included into the database (e.g., approx. 10 stories / new card).

*We could add another study to the database, which provides 192 stories each for 5 of the new pictures. Therefore we include these pictures now in the descriptive table. The reviewer is absolutely correct that the descriptive information about the other new pictures with ~10/picture is not sufficient to make a clear recommendation about the inclusion of these pictures in a new study. Nonetheless we think it makes sense to include the pictures in the database in order to provide researchers a larger pool of pictures to select from, and to be able to conduct additional validation studies on this standardized set of pictures.*

Minor issues:

In Table 8, details on 23 picture cards are listed.

*This changed in the revised version; the Table (which is now Table 9) now lists information about 28 pictures.*

The construct activity inhibition is not introduced.

*We added a short explanation and a reference pointing to more detailed information about the construct.*

### Reviewer 2:

R2.1: [...] Beyond these data sets, the paper presents some technical principles for preparing and scoring narrative data that likewise generalize beyond PSE motive measures. Lack of a citeable source for these principles has impeded forward movement in this area of our field. The Abstract should end with mention of these broader applications and the generalizable principles found on pp. 6 and 10-13.

R2.2: The paper’s purpose statement at the top of p. 4 could beneficially mention them as well.

*We added this aspect to the Abstract and the paper’s purpose statement.*

R2.3. Because historical and cultural differences in motives are known to exist (e.g., Skolnick, 1966; Veroff, Reuman, & Feld, 1984), thorough documentation of each data set regarding location and year(s) of data collection facilitates further analyses of this type. Thus, sample descriptive information (p. 4 l. 50, Tables 1 & 2) should be pursued energetically whenever possible. Including administration method (group handwritten, individual computerized, individual dictated (clinical) would also be useful for method comparisons.

*We added study-level descriptive data for all data sets on the following variables: Date and location of data collection, Administration method (handwritten/ computer-written online study / computer-written in lab), Setting of data collection (individual / group testing), Population (students / mostly students / mostly non-students / non-students), % female*

*(see also our response to E3).*

R2.4. Regarding individual-level data, because gender effects are also known (e.g., Skolnick, 1966; Stewart & Chester, 1982; Winter, 1988), gender, age, and culture (race/ethnicity/immigration status) data at the individual level would also be valuable for as many data sets as feasible, and the absence of these would be disappointing.

*We associated individual-level information on age and gender wherever possible. For two (archival) data sets, it was not possible to match the associating demographics, but for 24 of the 26 data sets we now provide this information. Race/ethnicity was not assessed in the vast majority of the samples. But we now provide the location of data collection in the study-level descriptives, which can be seen as a proxy.*

R2.5. The preprocessing noted on p. 6 will be much appreciated by future users.

*Thank you.*

R2.6. The picture-level data in Table 8 would be even more valuable for picture selection if number of words could be included as well.

*We added word count to what is nowTable 9*

R2.7. The important recommendations on pp. 10-13 set a citeable standard for quantitative narrative research. The authors could beneficially comment on their generalizability to other picture-story scoring systems including but not limited to those that may be used clinically, and to other narrative material such as Pennebaker-type expressive writing essays.

*Thank you for this excellent idea. We added a paragraph with our assessment on how these recommendations might or might not generalize to other coding systems.*

R2.8. Similarly with generalizability in mind, might it be possible to include code for other packages such as SPSS and SAS that are used more widely in the U.S.?

*We added information about how to do this approach in SPSS and SAS.*

R2.9. The analyses of picture position are of particular interest for followers of the Atkinson and Birch line of motive research, but here too the question of fatigue effects generalizes to other stimuli.

*We agree, and added a note to the section on the fatigue section.*

R2.10. Given widespread dissatisfaction among clinicians with the outdated Morgan-Murray TAT pictures, the picture database is especially welcome.

*Thank you - we are happy that the database is useful.*

Minor corrections

P. 18 l. 56—What is meant by “less motive images”? Lower motive scores?

*Changed to “less motive scores”.*

P. 19 l. 15—“satisfactory processes” is ambiguous in this context. Consider “satisfaction processes” or even “satiation processes”.

*Changed to “satiation processes”.*