

## **The Emotion Potential of Words and Passages in Reading Harry Potter - An fMRI Study**

Chun-Ting Hsu<sup>1,2</sup>, Arthur M. Jacobs<sup>1,2,3</sup>, Francesca M.M. Citron<sup>2,4</sup>, Markus Conrad<sup>2,5</sup>

<sup>1</sup>Department of Education and Psychology, Freie Universität Berlin, D-14195 Berlin, Germany

<sup>2</sup>Languages of Emotion, Freie Universität Berlin, D-14195 Berlin, Germany

<sup>3</sup>Dahlem Institute for Neuroimaging of Emotion (D.I.N.E.), Freie Universität Berlin, D-14195 Berlin, Germany

<sup>4</sup>Humanities Council, Princeton University, Princeton, NJ, 08544 USA

<sup>5</sup>Department of Cognitive, Social and Organizational Psychology, Universidad de La Laguna, 38205 Spain

**Email addresses of authors:** Chun-Ting Hsu: [hsuchunting@gmail.com](mailto:hsuchunting@gmail.com)  
Arthur M. Jacobs: [ajacobs@zedat.fu-berlin.de](mailto:ajacobs@zedat.fu-berlin.de)  
Francesca M.M. Citron: [fm.citron@gmail.com](mailto:fm.citron@gmail.com)  
Markus Conrad: [maconrad@ull.es](mailto:maconrad@ull.es)

**Corresponding author:** Chun-Ting Hsu  
Email: [hsuchunting@gmail.com](mailto:hsuchunting@gmail.com)  
Telephone number: +49 30 838 50388  
Telefacsimile numbers: +49 30 838 4 50388  
  
Postal address:  
Freie Universität Berlin  
Excellence Cluster "Languages of Emotion",  
Habelschwerdter Allee 45, D-14195 Berlin, Germany

**Conflict of Interest:** The authors declare no competing financial interests.

## **Abstract**

Previous studies suggested that the emotional connotation of single words automatically recruits attention. We investigated the potential of words to induce emotional engagement when reading texts. In an fMRI experiment, we presented 120 text passages from the Harry Potter book series. Results showed significant correlations between affective word (lexical) ratings and passage ratings. Furthermore, affective lexical ratings correlated with activity in regions associated with emotion, situation model building, multi-modal semantic integration, and Theory of Mind. We distinguished differential influences of affective lexical, inter-lexical, and supra-lexical variables: differential effects of lexical valence were significant in the left amygdala, while effects of arousal-span (the dynamic range of arousal across a passage) were significant in the left amygdala and insula. However, we found no differential effect of passage ratings in emotion-associated regions. Our results support the hypothesis that the emotion potential of short texts can be predicted by lexical and inter-lexical affective variables.

**Keywords:** emotion-laden words, reading, valence, arousal, arousal-span, fMRI

## 1. Introduction

Literary reading of novels or poems brings pleasures, including feelings of suspense, vicarious joy (or fear), or beauty, which are unique and important to human beings (Brewer & Lichtenstein, 1982; Nell, 1988). However, the neurocognitive processes underlying these feelings are poorly understood (Kringelbach, Vuust, & Geake, 2008; Schrott & Jacobs, 2011; Wolf, 2007). A recent neurocognitive poetics model of literary reading (Jacobs, 2011, 2014) has attempted to link the findings from the few existing neurocognitive studies on literary reading with results from cognitive linguistics, poetics, and aesthetics, by formulating hypotheses about which text elements evoke cognitive, emotional, or aesthetic processes. According to the model, literary reading can be viewed as a process of constructive content simulation (Mar & Oatley, 2008), closely linked to perspective taking and relational inferences associated at the neural level with the activation of the extended language network (ELN, Ferstl, Neumann, Bogler, & von Cramon, 2008), the Theory of Mind (ToM) network (Mason & Just, 2009), brain regions associated with affective or mood empathy (Altmann, Bohrn, Lubrich, Menninghaus, & Jacobs, 2012, 2014; Frith & Frith, 2003; Lüdtke, Meyer-Sickendiek, & Jacobs, 2014; Mar & Oatley, 2008), the creation of “event gestalts” (Speer, Zacks, & Reynolds, 2007), and with reward and aesthetic pleasure (Bohrn, Altmann, Lubrich, Menninghaus, & Jacobs, 2013). These processes converge to fulfill the goal of literary reading, i.e., meaning construction and the closure of *meaning gestalts* which depends on many factors, including the affective meaning of single words and passages (Iser, 1976; Jacobs, 2014).

At the cognitive level, discourse comprehension involves at least three different processing steps (Schmalhofer & Glavanov, 1986; Van Dijk & Kintsch, 1983): 1) the construction of the *surface structure* of a text, which is a mental representation of the exact text read; 2) a *text-base representation*, which contains idea units explicitly stated in the text, including bridging inferences that help connecting consecutive clauses; and 3) a *situation model* of the text, in which

the current linguistic input (i.e., the linguistic meaning of the sentence or paragraph being read) is integrated with both general world knowledge and the prior discourse context (Graesser, Millis, & Zwaan, 1997; Zwaan & Radvansky, 1998).

Affective aspects of discourse comprehension seem to involve embodied representations evoked through empathy, the simulation of emotion-related behavior, and autobiographical emotions related to memories of similar events. Empirical studies investigating emotional inferences during text comprehension (Gillioz, Gygas, & Tapiero, 2012; Gygas, Tapiero, & Carruzzo, 2007) have shown that the protagonist's behavioral reactions to an event are likely to be an important component of emotional inferences in discourse comprehension, and are highly associated with the emotional arousal level of the protagonist's behavior. Neuroscientific evidence that the affective aspect of meaning construction is possibly based on perspective simulation comes from an fMRI study of Moseley, Carota, Hauk, Mohr, and Pulvermuller (2012), who demonstrated that emotion-laden words indeed activated precentral cortex, including body-part-specific areas that are somatotopically activated by face- or arm-related words.

All these views on discourse processing suggest that affective processes during text reading may require, unlike in the respective minimalistic proposal of Jacobs' (2014) model, more complex processes than those elicited by the emotion potential of single words (note that Jacobs' model also elaborates on many higher level processes like immersion, situation model update, etc., but nevertheless posits some degree of texts' emotion potential to directly arise from the single word level).

Regarding the emotion potential of literary texts at the neuronal level, Ferstl, Rinck, and von Cramon (2005) were the first to show that listening to emotion-laden text passages indeed activates affect-related brain areas like the ventromedial prefrontal cortex (vmPFC), the left amygdala, and the pons. By auditorily presenting a near complete recording of "*The Ugly Duckling*" to participants, Wallentin et al. (2011) reported the neural correlates of intensity

ratings of each line of the text in the bilateral temporal, inferior frontal and premotor cortices, the thalamus, and the right amygdala. However, further neuroscientific evidence is still scarce (e.g., Bohrn et al., 2013; Hsu, Conrad, & Jacobs, 2014; Hsu, Jacobs, & Conrad, 2015; Lehne, Engel, Menninghaus, Jacobs, & Koelsch, in revision) compared to the more substantial body of research on the relationship between language and emotion using words in isolation (see Citron, 2012, for a review).

The emotion potential of words is usually operationalized in terms of valence and arousal ratings: the former refers to how positive or negative a word is, whereas the latter refers to its physiological intensity (Bradley & Lang, 1999; Osgood, 1969). These properties are normally distributed across all words, which can be categorized into negative, neutral, and positive classes (see Jacobs et al., in prep.; Schmidtke, Schröder, Jacobs, & Conrad, 2014; Vö et al., 2009; Vö, Jacobs, & Conrad, 2006, for normative German lexical databases), and have significant effects at the three relevant psychological levels: the experiential (e.g., subjective ratings, self-reports; Vö et al., 2009; Vö et al., 2006), the behavioral (e.g., response times, oculo- and pupillometric responses; Briesemeister, Kuchinke, & Jacobs, 2011a, 2011b; Vö et al., 2006; Vö et al., 2008), and the neuronal level, using both electrophysiological (e.g., Briesemeister, Kuchinke, & Jacobs, 2014; Conrad, Recio, & Jacobs, 2011; Fischler & Bradley, 2006; Hofmann, Kuchinke, Tamm, Vö, & Jacobs, 2009; Recio, Conrad, Hansen, & Jacobs, 2014) and functional neuroimaging methods (e.g., Citron, Gray, Critchley, Weekes, & Ferstl, 2014; Kuchinke et al., 2005). In particular, neuroscientific research on emotion-laden word processing using event-related potentials (ERP) has shown that such words capture attention more strongly than neutral words since early processing stages (e.g., Citron, Weekes, & Ferstl, 2013; Fritsch & Kuchinke, 2013; Hofmann et al., 2009; Kissler, Assadollahi, & Herbert, 2006; Kissler, Herbert, Peyk, & Junghofer, 2007; Schacht, Adler, Chen, Guo, & Sommer, 2012; see Citron, 2012, for a review). The attention capture seems to be automatic, i.e., it occurs even when participants attend to non-emotional

features of the words (e.g., Scott, O'Donnell, Leuthold, & Sereno, 2009), and it can also occur in the absence of conscious perception (Straube, Sauer, & Miltner, 2011).

In the present study, we addressed the question to what extent the emotion potential of supra-lexical units like narrative text passages is a function of the affective values of their constituting words, as proposed by Bestgen (1994) and by Whissell (2003; see also Jacobs, 2014).

Certainly, one might wonder to what extent findings for isolated words could generalize to the processing of larger text units and literary reading. A text is more than a list of words, and the way these words are combined, or the context they are embedded in, clearly matter. For example, basic phenomena such as negation, or more sophisticated rhetorical elements like metaphor or irony, can turn the emotional information of a text unit into the opposite of what single words' emotional content might suggest (Nagels et al., 2013; Regel, Gunter, & Friederici, 2011).

ERP studies have addressed the issue of how emotionally relevant sentence context influences the way in which the brain processes otherwise identical lexical material. Van Berkum, Holleman, Nieuwland, Otten, and Murre (2009) reported increased N400 amplitudes for target words whose combination with the preceding sentence context resulted in a moral statement that was inconsistent with the reader's political attitude. Schauenburg, Conrad, Ambrasat, Von Scheve, and Schröder (2013) extended this finding to the inconsistency concerning the general emotional content of single words across sentences describing social interactions. Their findings suggested that supra-lexical phenomena – like the relation between words within a sentence – affect emotion processing during sentence reading. In contrast, other ERP studies have shown that emotional features of single words can also reduce general context effects like the N400 component, which is typically interpreted to reflect increased difficulty of semantic integration into context. In the study of Delaney-Busch and Kuperberg (2013), incongruent context only modulated N400 amplitudes for neutral, but not for emotion-laden words, whereas N400 amplitudes were generally decreased for the latter. This suggests that emotional salience ensures

a partly context-independent processing of single lexical units, which “pop out” of context because of their general emotional relevance. In another study, Wang, Bastiaansen, Yang, and Hagoort (2013) presented question-answer pairs. Each answer contained a critical word (either positive, negative or neutral) that could be either the focus of the question or not. In both studies, the authors reported significant main effects of emotion and of congruency/focus, as well as a significant interaction between them, in the N400 component. Post-hoc analyses showed that the effect of congruency was only significant for neutral words, but not for both positive and negative words. One possible interpretation of the latter findings, proposed by Delaney-Busch and Kuperberg (2013), is that the emotional salience of such words gives rise to their prioritized processing from the earliest stages of meaning extraction, and leads readers to bypass a possible semantic incongruency with context. In this regard, both studies reviewed above suggest that the context-independent, highly automatic attention-capture potential of emotion-laden words can sometimes override integration processes or context effects during discourse comprehension. In the present study, we attempted to disentangle emotional effects during reading that arise at the lexical level from those that arise at the supra-lexical level. Following Jacobs’ (2014) proposal, in the absence of irony or similar stylistic devices, the emotion potential of supra-lexical units such as narrative passages can be hypothesized to be a direct function of the emotion potential of the constituting words. On the other hand, the principle of the whole being more than (or different from) the sum of its parts would predict emotional effects to be a function of more holistic processing of larger units, and thus emotional effects of a passage could not be fully accounted for by the affective values of its constituting words.

In his pioneering study, Bestgen (1994) acquired valence ratings from 120 participants at three different processing levels: 1) the textual level: participants read the text in the context of narrative reading and rated all sentences sequentially; 2) The sentential level: sentences from all four stories were divided into three sets, each including one third of all sentences from each of the

four stories, but two sentences from the same story were not presented one after another; 3) The lexical level: words extracted from the four texts, excluding articles, pronouns, prepositions, proper names, and the verbs “to be” and “to have”. Bestgen (1994) computed the correlations between the rated valence of the sentences at the textual, sentential, and lexical level (by calculating mean valence values of words composing the sentences), in each of the four texts. The linear correlations between rated valence at the textual and lexical level, and between textual and sentential level were large and significant (between .55 and .84, depending on the text). A similar approach was adopted by Whissell using the *Dictionary of Affect in Language* (Whissell & Dewson, 1986), which contains valence and arousal ratings of 4500 words as an estimate of the affective tone of existing passages of literature (Whissell, 2003, 2010, 2011). With principle component analysis, Whissell (2003) showed that ratings of 20 excerpts of romantic poetry from 68 participants on dimensions of pleasantness, activation, romanticism, and preoccupation with nature were consistent with estimations based on the *Dictionary of Affect*. By comparing two parts of Byron's *Child Harold's Pilgrimage*, written before and after an interruption of several turbulent years in England, Whissell (2010) showed that the later part employed fewer extreme emotional words and more abstract words than the pre-interruption part. Whissell (2011) found evidence for different pleasantness levels in four categories of speech (condescension, control, self-definition, and courtship of good opinion) given by Shakespeare's character Henry V. We adopted this general approach of estimating a text's emotion potential as a function of affective values of the constituting words and investigated it at the neural level.

### **The present study**

We used text passages from the Harry Potter book series as stimuli, featuring concepts, events and descriptions that display a sufficiently wide range of lexical valence and arousal values. We acquired functional magnetic resonance imaging (fMRI) data to investigate the neural correlates



of the emotional potential of narratives, operationalized via: individual emotion ratings of entire passages (termed henceforth “subjective passage ratings”); and normative affective values of their constituting words (termed henceforth “affective lexical variables”). The lexical variables were derived from a recent, large-scale, normative database comprising affective ratings for 6,600 words (Conrad, Schmidtke, Vo, & Jacobs, in prep.), which extends the Berlin Affective Word List (BAWL-R, Võ et al., 2009) and the Affective Norms for German Sentiment Terms (ANGST, Schmidtke et al., 2014). In particular, we use both mean and spread measures of affective lexical variables across a given piece of text because they might represent different specific aspects of the text’s emotion potential at the level of lexical surface features: While the mean of lexical valence and arousal values across a text may best represent its emotion potential as a function of the appearance of emotionally rather consistent concepts, spread measures of arousal or valence may, in turn, represent a proxy of dynamic changes or contrasts in the affective reactions a reader experiences. We consider “arousal-span” (i.e., the range of arousal values of single words across a text; see Jacobs, 2014, for a theoretical proposal concerning effects of this measure) and, accordingly, “valence-span” (the range of respective valence values) as the most appropriate lexical spread measures of a text’s emotion potential. Using data from a recent study by Lehne et al. (in revision), Jacobs (2014) showed that arousal-span could account for about 25% of the variance in suspense ratings from readers of E.T.A. Hoffmann’s black-romantic story „*The Sandman*“.

We used these affective lexical variables and subjective passage ratings as parametric predictors of brain activation, assuming that the former would contribute significantly to the emotional reading experience. At the neuronal level, using parametric fMRI analyses, we expected both the affective lexical and supra-lexical variables to be correlated with BOLD signal intensity in the temporal, inferior frontal and premotor cortices, vmPFC, thalamus, amygdala, and the pons, which have all been reported to be specifically associated with emotional discourse

comprehension (Ferstl et al., 2005; Wallentin et al., 2011). In addition, our aim was to disentangle potential differential effects of the following variables that reflect a text's emotion potential at three different levels (see also examples given in Table 2):

- 1) The elementary level of general emotion content of lexical units or concepts – represented by the means of lexical valence and arousal values of single words, appearing in a given text. Note that this operationalization assumes that respective values for single elements (words or concepts) would condense into a homogenous affective impression. An illustrative example from our stimulus material is the sentence “‘You disgusting little Squib, you filthy little blood traitor!’ roared Gaunt, losing control” (Rowling, 2005), whose lexical units are consistently negative in valence and high in arousal (“disgusting”, “filthy”, “blood”, “traitor”, “roar”, and “losing”);
- 2) The inter-lexical level, between single words and holistic supra-lexical processing – operationalized by lexical valence- and arousal-span. Note that, in contrast to the mean measures described above, these lexical span measures should especially account for dynamic changes and salient contrasts characterizing the relation between single lexical units and concepts in a text (Jacobs, 2014). These aspects may not be captured by overall lexical mean measures. For example, in the sentence “‘And then a silence fell over the crowd, from the front first, so that a chill seemed to spread down the corridor” (Rowling, 1999), a high lexical arousal-span is produced by the contrast between the low arousal of “silence” and the high arousal of “chill”, whereas the mean lexical arousal of the whole sentence would be rather moderate;
- 3) The supra-lexical level, represented by subjective valence and arousal ratings for whole passages. Specific supra-lexical emotion potential may go beyond what lexical values alone would predict: e.g., the passage “Ginny glanced round, grinning, winked

at Harry, then quickly faced the front again. Harry's mind wandered a long way from the marquee, back to afternoons spent alone with Ginny in lonely parts of the school grounds.” (Rowling, 2007) was rated as positive, while its mean lexical valence was neutral. The emotional impact probably results from the drift of Harry’s mind into the past remembering his relationship with Ginny, that the reader is rather invited to imagine than actually being told about.

As a main hypothesis, we expect lexical variables to significantly predict subjective ratings and BOLD responses in emotion-related brain areas, therefore corroborating the approach of Bestgen (1994) and Whissell (2003), i.e., the prediction of the affective impact of text units from mean lexical valence and arousal values. Furthermore, we expect lexical variables’ effects on BOLD responses to be, at first glance, largely comparable to those of subjective ratings of the whole text passages. However, we also hypothesize that our data might reveal discernable effects of emotion potential at lexical and supra-lexical levels: In the usage of language, we can employ different words to express the same thing, and the specific choice of our words has a huge impact on how our message will be received at the affective level. Thus, it makes a difference whether we say: “why don’t you clean up your *mess* vs. *thing* behind you?”, or whether we call a person living on the street a “*homeless*” or a “*bum*”. In this case, lexical affective variables could produce emotional effects beyond the ones obtained for subjective ratings of entire text units, because the latter ones might lose focus on fine-grained emotional connotations at the lexical level when trying to evaluate complex parts of text as a whole.

In addition, the automatic attention-capture potential of emotion-laden words can override integration processes involved in discourse comprehension by ensuring their processing prioritization relative to the context. Therefore, we expected to find significant lexical effects

after partialling out the variance accounted for by passage ratings, in particular in the “salience network” (Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012; Seeley et al., 2007).

Furthermore, specific lexical effects may not be limited to single words’ salience alone: all words we encounter embedded in a sentence have already been seen or heard previously from us in many other and potentially very different contexts. All of these contribute to the complex pattern of emotional connotations a single word can have in our memory, which has the potential to “project” affective processes triggered by single words via associative links far beyond the given context of a sentence – which is another reason to expect lexical effects to be at least somewhat independent from more general text evaluations.

Finally, after partialling out the variance accounted for by affective lexical variables first, we might expect rating values to correlate with activation in areas associated with emotional conceptualization, evaluation (Lindquist et al., 2012), and the ToM network (Mar, 2011), which are supposed to represent the specific emotion potential of supra-lexical units (e.g., Bohrn, Altmann, & Jacobs, 2012).

## **2. Materials and Methods**

### ***2.1 Participants***

Twenty-four right-handed native German speakers (16 women) gave written consent to take part in the experiment, which was approved of by the ethics committee of the Freie Universität Berlin. Their age ranged from 18 to 31 years (mean  $\pm$  SD = 23.71  $\pm$  3.67). All participants had read at least one Harry Potter book, and were therefore familiar with its context enough to understand the novel-specific contents. They all had normal or corrected-to-normal vision, and reported no neurological or psychiatric disorders. Participants were compensated properly monetarily or with course credits for their participation.

## ***2.2 Stimuli***

To prepare our stimulus material we screened all seven HP novels (Rowling, 1997, 1998, 1999, 2000, 2003, 2005; 2007, German translations by Klaus Fritz, Carlsen Verlag, Hamburg) for text passages featuring either particularly emotional or neutral moments or events. We finally selected 120 passages, each of which was 4 lines long. The passages provide a wide range of valence and arousal as demonstrated in two previous studies (Hsu et al., 2014; Hsu et al., 2015) based on the same set of fMRI data. Passage selection further ensured that: 1) comprehension of the passages did not require a high level of familiarity with Harry Potter novels; 2) the emotional connotations of the passages clearly emerged at their very beginning; 3) emotional contents were unambiguous and generally consistent throughout the passage.

## ***2.3 Design***

The 120 text passages were divided into two subsets of 60. During the experiment, each participant read one subset in German and the other one in English. Each subset was presented to 12 participants in German and English respectively. Only the data for reading in German was used for analyses in this study (see Hsu et al., 2015, for analyses of the complete data following a bilingualism research question).

## ***2.4 Procedure***

The experiment consisted of four runs, each containing 15 German and 15 English passages. The order of presentation was pseudo-randomized so that the distribution of language switch positions was balanced in each run and across all participants. Similar to the design of a previously successful fMRI experiment on text-reading (Altmann et al., 2012), each passage was presented for 14 s in the MR scanner, distributed on 4 lines (shown consecutively for 3.5 s each), and then

followed by 14 s of fixation cross. The visual input was presented on a computer screen and was reflected to the participants' eyes via a mirror.

To make sure that participants were attentive and actively comprehended the passages, they had been informed that several (leaving open how many) randomly selected passages in each run were immediately followed by a context-specific yes/no question, unrelated to emotion (e.g., 'Was Harry in a train station?' 'Was the alarm clock broken again?'), to which participants responded via button press.

### ***2.5 fMRI data acquisition***

Functional data were acquired on a Siemens Tim Trio 3T MR scanner. Four runs of 440 volumes were measured using a  $T_2^*$ -weighted echo-planar sequence [slice thickness: 3 mm, no gap, 37 slices, repetition time (TR): 2s, echo time (TE): 30ms, flip angle:  $70^\circ$ , matrix:  $64 \times 64$ , field of view (FOV): 192 mm, voxel size:  $3.0 \text{ mm} \times 3.0 \text{ mm} \times 3.0 \text{ mm}$ ] and individual high-resolution T1-weighted anatomical data (MPRAGE sequence) were acquired (TR: 1.9, TE: 2.52, FOV: 256, matrix:  $256 \times 256$ , sagittal plane, slice thickness: 1 mm, 176 slices, resolution:  $1.0 \text{ mm} \times 1.0 \text{ mm} \times 1.0 \text{ mm}$ ).

### ***2.6 Post-scan ratings, affective lexical variables, and correlational analyses***

Following the experiment in the MR scanner, participants rated all 120 passages in the language version they had read inside the scanner on valence, scaled from -3 (very negative) to +3 (very positive, see Vö et al., 2006), and arousal, scaled from 1 (very calming) to 5 (very arousing, see Bradley & Lang, 1999; Schmidtke et al., 2014; Vö et al., 2009; Vö et al., 2006). For each passage, we calculated affective lexical variables based on valence and arousal values provided by a large scale German normative database for lemma forms of single words contained in the passages (Conrad et al., in prep., see also BAWL-R, Vö et al., 2009, and ANGST, Schmidtke et al., 2014).

Note that this database containing normative affective rating values for over 6,600 German lemmata provided matches for overall 54% of content words of our stimulus material. Based on these values we computed the following affective lexical variables for all of our text passages stimuli: 1) mean lexical valence, 2) mean lexical arousal, 3) lexical valence-span and 4) lexical arousal-span. Bivariate and partial correlations between affective lexical variables and mean values of valence and arousal ratings for each passage are given in Table 1. Note that, although mean rating values correlated with lexical values, respective correlations were not too large ( $<0.6$  in all cases between ratings on the one hand and lexical values on the other, except .66 between arousal ratings and lexical arousal span; higher correlations only occurred in some cases between variables at one and the same level – lexical vs. ratings); this leaves open the possibility to obtain differential effects for ratings by partialling out shared variance with lexical values, or vice versa, on brain activation.

### ***2.7 fMRI preprocessing***

The fMRI data were preprocessed and analyzed using the software package SPM8 ([www.fil.ion.ucl.ac.uk/spm](http://www.fil.ion.ucl.ac.uk/spm)). Preprocessing consisted of slice-timing correction, realignment for motion correction, and sequential coregistration. Structural images were segmented into grey matter, white matter, cerebrospinal fluid, bone, soft tissue, and air/background with the ‘New Segment’ module (Ashburner & Friston, 2005). A group anatomical template was created with DARTEL (Diffeomorphic Anatomical Registration using Exponentiated Lie algebra, Ashburner, 2007) toolbox from the segmented grey and white matter images. Transformation parameters for structural images were then applied to functional images to normalize them to the brain template of the Montreal Neurological Institute (MNI) supplied with SPM. For the univariate analysis, functional images were resampled to a resolution of  $1.5 \times 1.5 \times 1.5$  mm, and spatially smoothed with a kernel of 6 mm full-width-at-half-maximum during normalization.

## ***2.8 fMRI Analyses***

We calculated statistical parametric maps by multiple regressions of the data onto a model of the hemodynamic response (Friston et al., 1995). In this study, we tried to disentangle specific or potentially differential influences of emotion potential at lexical and supra-lexical levels on brain activation using three different models of parametric modulation (see Fig. 1): all models contained two regressors for German and English conditions, and each passage lasted 14 seconds. Six parametric modulators and their polynomial expansions (detailed below for each model separately) were added after the German condition. The context-specific questions were modeled as the third condition, and each question lasted four seconds. The six realignment parameters were modeled as six additional regressors. Regressors were convolved with the canonical hemodynamic response function in SPM8.

In all models, parametric modulators (predictors) and their second-order derivatives were serially orthogonalized from the first to the last one, in a Gram-Schmidt process (Büchel, Holmes, Rees, & Friston, 1998; Büchel, Wise, Mummery, Poline, & Friston, 1996; Golub & Van Loan, 1996) by SPM8. Note that, despite orthogonalization, first order effects (i.e., effects of the first predictors entered in the model), potentially include shared variance that cannot be uniquely attributed to any of the lexical, inter-lexical, or supra-lexical variables. On the other hand, effects of the last predictors entered in the model represent differential effects specific to those predictors, i.e., variance that could be uniquely attributed to a specific level, since all shared variance has already been partialled out by previously entered predictors. The current study, thus, uses conservative inferences concerning differential effects specific to each level of text processing. In Model 1, six predictors were entered in the following order (see Fig. 1): 1) linear lexical arousal mean; 2) linear and quadratic lexical valence mean; 3) linear arousal-span; 4) linear valence-span; 5) linear arousal ratings; 6) linear and quadratic valence ratings. This model reveals



first order effects of affective lexical variables and differential, or unique, effects of subjective passage ratings – after partialling out variance shared with affective lexical variables.

In Model 2, the order of entry was: passage ratings (predictor 1 & 2), arousal- and valence-span (predictor 3 & 4), and lexical means (predictor 5 & 6). This model focused on the differential effects of lexical means after partialling out variance due to subjective passage ratings and lexical arousal- and valence-span.

In Model 3, arousal- and valence-span (predictor 5 & 6) were entered after lexical means (predictor 3 & 4) and passage ratings (predictor 1 & 2). Therefore, this model focused in particular on differential effects specific to lexical arousal- and valence-span.

To explore the effect of reading vs. fixation at the group level, beta-images of the contrast [German reading > fixation] from each participant were entered into a random effect, one-sample t-test. For each parametric effect at the group level, beta-images of each polynomial expansion of each parametric modulator were taken from each participant and entered into random effect, one-sample t-tests.

We calculated correlations with both first and second order polynomial expansions (Büchel et al., 1998; Büchel et al., 1996) for the valence dimension because 1) valence is a bipolar variable; and 2) a meta-analysis on emotion studies suggested that some emotion-related structures, like amygdala and pregenual anterior cingulate cortex (ACC), would respond to bipolar emotional valence of the passages, possibly in a quadratic fashion (e.g., Lewis, Critchley, Rotshtein, & Dolan, 2007); 3) according to Lewis et al. (2007), both arousal and quadratic valence code for a generalized form of “salience”, by directing attention toward behaviorally important goals (Cunningham, Raye, & Johnson, 2004; Winston, O'Doherty, & Dolan, 2003). Results from Recio et al. (2014) indicate that valence and arousal contribute independently to early attentional stages of word processing. In line with these findings, while the second order of valence ratings for our stimuli strongly correlates with arousal ratings ( $r = 0.54$ ), at the lexical level, quadratic mean

lexical valence is quite independent of mean lexical arousal ( $r = -0.18$ ). All this justifies using both first and second order valence and arousal values as predictors for our models in order to explore their potentially independent contributions to brain activations. We always entered arousal before valence variables into our models in order not to exclude potential arousal effects, given that valence already enters in both linear and quadratic fashions. In this way, variance attributed to quadratic valence is not attributable to arousal.

None of the affective lexical variables or subjective passage ratings showed significant correlations with passage length (all  $ps > 0.1$ ), operationalized in numbers of letters and words, passage-wise average word imageability taken from the BAWL-R (Vö et al., 2009) and the ANGST (Schmidtke et al., 2014) databases, or passage-wise average log frequency of words given in the Leipzig Wortschatz Lexicon (available at <http://wortschatz.uni-leipzig.de/>).

Therefore, our material involves no confound of affective dimensions with length, frequency, or imageability, thus we did not use these variables as additional predictors in order not to inflate our models.

Unlike Wallentin et al. (2011), who factored out “physiological nuisance effects” including cardiac and respiratory oscillation, as well as six motion parameters, we did not attempt to partial out respective effects of physiological regressors from emotion related fMRI effects, though this may represent an interesting matter of debate:

Emotion theorists and empirical evidence have long associated affective processing and behaviors with the autonomic (sympathetic and parasympathetic) nervous system (ANS) activity, e.g., Damasio’s “somatic marker hypothesis” (Damasio, 1998; Damasio & Carvalho, 2013), Critchley’s emphasis of internal bodily feedback of autonomic arousal influencing subjective emotional states (Critchley, 2005, 2009; Critchley, Wiens, Rotshtein, Ohman, & Dolan, 2004), see also Barrett, Mesquita, Ochsner, and Gross (2007) and Nummenmaa, Glerean, Hari, and Hietanen (2014). Respiratory and cardiac oscillations are regulated by the ANS. Hence,

physiologically correlated BOLD signals could include variance due to ANS activity, as well as variance due to noise of cardiac pulse and respiratory movements. Among fMRI studies, physiological recordings are often used as regressors of no interest (Lund, Madsen, Sidaros, Luo, & Nichols, 2006). Dagle, Ingeholm, and Haxby (1999) have found that cardiac cycle related BOLD variability could effectively hide or mask changes of the BOLD signal due to neural activity on the order of 1 to 2% in regions, thus reduce sensitivity in the detection of cognitive activity in regions near major arterial vessels: anterior insula, anterior temporal lobe, and medial frontal cortex. On the other hand, Gray et al. (2009) stressed the value of physiological recordings in fMRI analysis to investigate the brain-body interaction for emotional/cognitive neuroscience and psychosomatic medicine. In line with the theoretical proposals of Damasio and Critchley, Beissner, Meissner, Bar, and Napadow (2013) performed a meta-analysis based on 43 neuroimaging studies that used simultaneous autonomic measures in the fMRI analysis as regressors of interest to identify the central autonomic network, which included left amygdala, right anterior and left posterior insula, and mid-cingulate cortex, most of which are important parts of the salience network (Seeley et al., 2007). Considering the close neuroanatomical and functional overlapping between the salience network and the central autonomic network, we expect respiratory and cardiac oscillations to co-vary with the neural activity in the central autonomic network during reading of emotion-laden texts. Therefore, we did not use physiological recordings as regressors of no interest to avoid possible type II errors (false negative) due to factoring out activity in the central autonomic network that co-varies with autonomic oscillations. On the other hand, the brain-body interaction is not the focus of the current study. Therefore, physiological recordings were also not treated as regressors of interest. Moreover, by using a block design lasting 14 seconds in each block, the current study seems less prone to the type II errors associated with cardiac cycle related BOLD variability (Dagle et al., 1999) by averaging out BOLD signals across multiple cardiac cycles. However, readers are

advised to note that results of the current study may contain type I errors (false positive) due to physiological noise, which could not be disentangled from ANS activity associated BOLD signal due to the correlational nature of the current study. For further discussion on the balance of type I and II error in neuroimaging studies in affective neuroscience in a different context, please see Lieberman and Cunningham (2009).

The fMRI analyses were conducted at the whole brain level. Furthermore, because current meta-analyses strongly suggested the amygdala to be involved in emotion processing (Costafreda, Brammer, David, & Fu, 2008; Lindquist et al., 2012; Murphy, Nimmo-Smith, & Lawrence, 2003; Phan, Wager, Taylor, & Liberzon, 2002), especially in emotional discourse comprehension (Ferstl et al., 2005), we performed small volume correction (SVC) with a bilateral amygdala mask for contrasts showing parametric emotion effects. The bilateral amygdala mask in the MNI template was defined by the WFU Pickatlas Tool (Maldjian, Laurienti, Kraft, & Burdette, 2003). For the one-sample t-test showing the effect of reading vs. fixation, we used the voxel-level family-wise error (FWE) corrected  $p < 0.05$  and cluster threshold of 5 to report only very strong activation differences. For the parametric analyses at the whole brain level, we used an initial voxel-level threshold of uncorrected  $p < 0.005$ , then a cluster-level threshold of false discovery rate (FDR) corrected  $p < 0.05$  for the entire image volume, as suggested by Liebermann and Cunningham (2009) for studies in cognitive, social and affective neuroscience. For the SVC analyses of the amygdala, we used initial voxel-level threshold of uncorrected  $p < 0.005$  for the entire image volume, then the threshold of voxel-level FWE corrected  $p < 0.05$  after applying the SVC with a bilateral amygdala mask. The labels reported were taken from the ‘TD Labels’ (Lancaster et al., 1997; Lancaster et al., 2000) or ‘aal’ labels in the WFU Pickatlas Tool. The Brodmann areas (BA) were further checked with the Talairach Client using nearest grey matter search after coordinate transformation with the WFU Pickatlas Tool.

### **3. Results and discussion**

As our results involve a large amount of data for different statistical models, we will briefly discuss results of each of the following subsections before moving to the next set of results to enhance comprehensibility of the entire report.

#### ***3.1 Behavioral data***

##### ***3.1.1. Task performance***

We randomly inserted four context-specific questions in each run, i.e., 16 questions throughout the experiment. All participants correctly responded to questions in the scanner above chance ( $\geq 62.5\%$ ) with overall mean accuracy of  $81.47\% \pm 13.16\%$  suggesting good comprehension of the presented text passages.

##### ***3.1.2. Correlations between affective lexical variables and ratings***

Correlations and partial correlations between ratings and lexical parameters are listed in Table 1, indicating that mean lexical valence values (taken from Conrad et al., in prep.) accounted for 28% of the variance in the valence ratings of text passages, while the mean lexical arousal values accounted for about 35% of the variance of arousal ratings. Surprisingly, while lexical arousal-span showed a significant positive correlation with mean lexical arousal values and lexical valence-span, it accounted for even more of the variance of the arousal ratings ( $r^2 = 0.44$ ) than did mean lexical arousal values ( $r^2 = 0.35$ ) and valence-span ( $r^2 = 0.27$ ). Because of the high intercorrelations of the affective lexical variables, we calculated partial correlations to remove joint variance. These showed significant correlations between mean lexical valence and valence ratings ( $r = 0.34, p < 0.001$ ), between mean lexical arousal and arousal ratings ( $r = 0.36, p < 0.001$ ), and between lexical arousal-span and arousal ratings ( $r = 0.40, p < 0.001$ ).

Taken together, these results partially support the hypothesis that a lexical measure of single words' mean emotion potential – as estimated by standard databases – significantly contributes to the subjective emotional reading experience (cf. Bestgen, 1994; Jacobs, 2014). Importantly, our data also show that, at the behavioral level, lexical mean values alone are not sufficient to predict all aspects of affective impact at the text level: lexical means neglect the emotion potential of dynamic changes of lexical values across a text – as assessed by our valence- and arousal-span measures. More specifically, the latter ones explained additional variance of passage arousal ratings as hypothesized by the neurocognitive poetics model of literary reading (Jacobs, 2014).

## ***3.2. fMRI data***

### ***3.2.1 reading (German) vs. fixation***

Results are summarized in Table A.1 and shown in Figure 2. Clusters more active in the reading condition included bilateral inferior prefrontal gyrus, dorsolateral prefrontal cortex (dlPFC), vmPFC, medial supplementary motor area (SMA), visual cortex, precuneus, superior temporal sulcus (STS), temporo-parietal junction (TPJ), thalamus, amygdala, pons, and cerebellum.

The simple contrast between reading German passages vs. fixation showed extensive activation of the ELN (Ferstl et al., 2008), in line with previous studies of text comprehension. Bilateral STS, IFG, and medial supplementary motor area are associated with language processing (Price, 2012); dlPFC, TPJ, anterior temporal lobe (aTL), precuneus, and amygdala are associated with ToM or affective empathy processing (Mar, 2011; Walter, 2012), whereas aTL and vmPFC are associated with multimodal (semantic) integration and emotional conceptualization (Binder & Desai, 2011; Binder, Desai, Graves, & Conant, 2009; Lindquist et al., 2012). Finally, significant activation peaks in bilateral amygdala, vmPFC, and pons are in line with previous findings from Ferstl et al. (2005), during presentation of emotion-laden texts. The results clearly suggest that our participants read the texts by connecting sentences, producing inferences and keeping

previously-read information in the working memory, rather than just picking up emotion-laden words, as if those were presented in isolation. Note that such elaborate, comprehension processing is further supported by different analyses of our data showing activation differences for passages describing events that violate participants' world knowledge (Hsu, Jacobs, Altmann, & Conrad, accepted).

### ***3.2.2. Overall (unspecific) effects at the lexical, intra-lexical and supra-lexical levels: first order effects in different models***

#### *3.2.2.1. Model 1: unspecific effects of mean lexical arousal*

When lexical arousal was entered as the first predictor into the model, hemodynamic responses in the following neural substrates increased significantly when reading passages with higher mean lexical arousal (Fig. A.1, panel A-C in red color; Table A.2): bilateral middle superior temporal sulcus (STS, BA 21, 22 & 41), bilateral caudate tail, left IPL and supramarginal gyrus (SMG, BA 40), left primary visual cortex (BA 17) and PCC (BA 30), right cerebellum, and after SVC, bilateral amygdala (Fig. A.1, panel D).

Hemodynamic responses in the following neural substrates decreased significantly when reading passages with higher mean lexical arousal (Fig. A.1, panel A-C in green color; Table A.2): bilateral parahippocampal cortex (PHC, BA 28 & 36, including left hippocampus), bilateral anterior cingulate cortex (ACC, BA 32 & 25) and vmPFC (BA 10 & 11), bilateral dorsolateral prefrontal cortex (dlPFC, BA 8), bilateral middle superior temporal gyrus (STG), right anterior temporal lobe (aTL, BA 20 & 21), right posterior cingulate cortex (PCC, BA 23, 30 & 31), right medial premotor cortex (BA 6), right medial frontopolar cortex (BA 10), right TPJ (BA 19 & 39), left precuneus (BA 19), and left cerebellum.

#### *3.2.2.2. Model 1: unspecific effects of mean lexical valence*

After partialling out variance due to mean lexical arousal, brain activity in the following regions significantly increased with more positive mean lexical valence of passages (Fig. A.1, panel E & F in red color; Table A.2): bilateral aTL (BA 21, 22 & 38), left premotor cortex (BA 6), left dlPFC (BA 8 & 9), left TPJ (BA 39), left PHC (BA 30), left PCC & precuneus (BA 31 & 7), right extrastriate visual cortex (BA 18 & 30), and right cerebellum.

Hemodynamic responses in left posterior middle temporal gyrus (MTG, BA 37) and left IPL (BA 40) decreased significantly with more positive mean lexical valence (Fig. A.1, panel E & F in green color; Table A.2).

BOLD responses in bilateral caudate tail and right superior parietal lobule increased quadratically when reading passages with either positive or negative mean lexical valence (Fig. A.1, panel G in red color; Table A.2), while in the following regions the BOLD responses decreased quadratically with mean lexical valence (Fig. A.1, panel G & H in green color; Table A.2): right middle STS (BA 21 & 22), left posterior temporal cortex (BA 20, 22 & 38), left temporo-occipital junction (TOJ, one cluster in BA 39 & 18, another cluster in BA 36, 20 & 19), bilateral PCC and precuneus (BA 30 & 31), and left frontopolar cortex (BA 9 & 10).

### *3.2.2.3. Model 1: unspecific effects of lexical arousal-span*

After partialling out variance due to mean lexical arousal and valence, we found significant positive correlation between BOLD responses and arousal-span in the following neural substrates (Fig. A.1, panel Q to T in red color; Table A.3): left inferior frontal gyrus (IFG) pars triangularis including anterior insula (BA 13, 45 & 46), left IFG pars opercularis including anterior insula (BA 9, 13 & 44), right IFG pars triangularis and orbitalis (BA 46 & 47), bilateral pre- and postcentral gyrus (BA 2, 4 & 6), bilateral posterior MTG and inferior temporal gyrus (ITG, BA 37), bilateral extrastriate cortex (BA 18, 19 & 30), left premotor cortex (two clusters, BA 6), left



MTG and fusiform gyrus (BA 20 & 37), left middle cingulate gyrus (BA 24), and right middle STS (BA 21, 22 & 38).

On the other hand, there was a significant negative correlation between arousal-span and BOLD responses in bilateral PCC (BA 29 & 31) and right SMG (BA 39 & 40; Fig. A.1, panel Q to T in green color; Table A.3).

#### *3.2.2.4. Model 1: unspecific effects of lexical valence-span*

After partialling out variance due to affective lexical means and lexical arousal-span, hemodynamic responses in left amygdala and globus pallidus and right cerebellum showed significant positive correlation with valence-span (Fig. A.1, panel U; Table A.3). No neural substrate correlated negatively with valence-span.

#### *3.2.2.5. Model 2: unspecific effects of arousal ratings.*

Right IFG including anterior insula (BA 47 & 13), left posterior MTG (BA 21, 39 & 37), and left IPL (BA 40) showed positive correlation with arousal ratings (Fig. A.1, panel I to K in red color; Table A.4).

Bilateral aTL (BA 21), dlPFC, dorsal ACC (BA 32), vmPFC (BA 11), TPJ (BA 39 & 40), pericentral cortex (BA 3, 4 & 6), and left PCC (BA 23 & 31) showed negative correlation with arousal ratings (Fig. A.1, panel I to K in green color; Table A.4).

#### *3.2.2.6. Model 2: unspecific effects of valence ratings*

Right aTL (BA 21), right TPJ (BA 40 & 22), right ACC (BA 32), vmPFC (BA 11), and left PCC (BA 23, 29 & 30) showed linear positive correlation with valence ratings (Fig. A.1, panel M & N; Table A.4).

Left IFG including anterior insula (BA 9 & 13), bilateral precuneus (BA 31 & 7), visual cortex (BA 18 & 19; Fig. A.1, panel O & P; Table A.4) and after SVC, bilateral amygdala (Fig. A.1, panel L) showed quadratic positive correlation with valence ratings.

No region showed negative linear or quadratic correlation with valence ratings.

### ***3.2.3. Discussion of overall, unspecific effects at lexical, intra- and supra-lexical levels***

#### ***3.2.3.1. Unspecific effects of affective mean lexical variables and inter-lexical variables.***

We obtained robust correlations of the affective lexical variables with the BOLD signal intensities (Table A.2 & A.3). Most remarkably, in Model 1, we found a significant positive linear correlation in both amygdalae with lexical arousal means as well as with valence-span. The amygdala was proposed to be an integral part of the emotion network in general (Costafreda et al., 2008; Lindquist et al., 2012; Murphy et al., 2003; Phan et al., 2002), and the salience network in particular (Seeley et al., 2007), and our result is in line with a previous study on the comprehension of texts containing emotional information (Ferstl et al., 2005). Correlated activity in ACC, PCC, vmPFC, and PHC is associated with emotion conceptualization (Lindquist et al., 2012).

Among other neural correlates associated with lexical means and spans (Table A.1 & A.2), IFG, aTL, and TPJ are all associated with the ELN (Ferstl et al., 2008), with ToM or affective empathy processing (Altmann et al., 2012, 2014; Mar, 2011; Walter, 2012), and with multi-modal semantic integration (Binder & Desai, 2011; Binder et al., 2009). In addition, aTL, vmPFC, and PCC have been associated with emotional semantic processing (Binder & Desai, 2011; Bohrn, Altmann, Lubrich, Menninghaus, & Jacobs, 2012).

The fact that affective lexical values correlated with activity in many emotion-related brain areas provides strong evidence for the assumption that lexical surface features play an important role for the overall affective reading experience (see Bestgen, 1994; Jacobs, 2014; Whissell, 2003).

The comparison of these effects with those of subjective passage ratings may further inform us about how adequate the estimation of a text's emotion potential from lexical affective surface features is at the neuronal level.

#### *3.2.3.2. Unspecific effects of subjective passage ratings*

Among first order effects for subjective passage ratings, it is most noteworthy that the hemodynamic responses in bilateral amygdala correlated significantly in a quadratic fashion with passages valence ratings. We also found significant correlations among neural substrates associated with the ELN (Ferstl et al., 2008), ToM or affective empathy processing (Altmann et al., 2012, 2014; Mar, 2011; Walter, 2012), and multi-modal semantic integration (Binder & Desai, 2011; Binder et al., 2009) like IFG, dlPFC, aTL, TPJ, PCC, dorsal ACC, vmPFC as discussed in the previous section.

#### *3.2.3.3. Shared variance between unspecific effects of affective lexical variables and subjective passage ratings.*

All results presented so far cannot be attributed exclusively to either of the relevant text levels: they potentially represent unspecific effects of variance shared by passage ratings and affective lexical variables. Indeed, we found several neural substrates for which correlations between neural activity and emotional variables are consistent concerning both affective lexical variables and subjective passage ratings. Most importantly, amygdala activity correlated with valence ratings in a quadratic fashion, and with mean lexical arousal values and valence-span in a linear fashion. This corresponds well with the classical U-shaped distribution of valence and arousal (Bradley & Lang, 1999; Lewis et al., 2007; Võ et al., 2009). Moreover, in the two-dimensional affective space described by valence and arousal (Russell, 1980, 2003), both arousal and quadratic valence are measures of emotional salience (Lewis et al., 2007). The linear correlation

of activation with lexical arousal and the quadratic correlation with valence ratings support the notion that it is emotional intensity and attention to (emotionally) salient stimuli – rather than a specific type of emotion, i.e., positive vs. negative – that is crucial for amygdala activation (Anderson & Sobel, 2003; Costafreda et al., 2008; Small et al., 2003; Wallentin et al., 2011).

Other neural substrates showing consistent correlations with BOLD responses for both affective lexical variables and subjective passage ratings include ACC (BA 32), PCC (BA 31), vmPFC (BA 11), PHC (BA 36 & 28), bilateral precuneus (BA 7), left IPL including SMG (BA 40 & 39), right middle temporal cortex (BA 21), and right dlPFC (BA 8).

Despite the problem of unambiguously attributing unspecific effects to the lexical or supralexical level in particular – due to shared variance between the two – the fact that affective lexical variables and passage ratings could account in parallel for such a big range of effects concerning activity in emotion-related brain area is additional strong evidence for the general assumption that the emotional impact of texts can be adequately predicted by merely averaging affective values for constituting words at the lexical surface level. Unlike in Bestgen (1994), our computation of lexical affective means was not even based on an exhaustive sample of values for all words, but rather on valence and arousal values for 54% of content words as provided by a recent large scale normative German lexical database featuring valence and arousal ratings for 6,600 words.

However, these matches probably involve a very high percentage of emotionally relevant words comprised in the texts, because attempting to include as many emotionally relevant words as possible was a main construction principle of this database. Our data thus show for the first time that emotional brain responses to complex texts can be directly predicted by text surface features at lexical level as provided by large scale normative databases – in much the same way as using individual affective evaluations of the complete passages.

### ***3.2.4. Results of differential effects, specific to either lexical, inter-lexical, or supra-lexical levels in fMRI analyses***

#### *3.2.4.1. Model 1: differential effects of arousal ratings*

After partialling out the variance due to affective lexical mean and span variables, the hemodynamic response in left temporo-occipital junction (BA 19 & 39) showed a significant negative correlation with passage arousal ratings (Table 3).

#### *3.2.4.2. Model 1: differential effects of valence ratings*

We found the hemodynamic response in left posterior MTG (BA 21) to be negatively correlated with valence ratings in a quadratic fashion (Table 3), i.e. this region gets more active when the passage is more neutral.

#### *3.2.4.3. Model 2: differential effects of mean lexical arousal*

We found no neural substrate in which the hemodynamic response is specifically correlated with mean lexical arousal values.

#### *3.2.4.4. Model 2: differential effects of mean lexical valence*

After partialling out the variance due to subjective passage ratings, lexical arousal- and valence-span, and mean lexical arousal, hemodynamic responses in right lower TOJ, including PHC, fusiform and lingual gyrus (BA 19), right STS (BA 22), and left amygdala after SVC, correlated positively with mean lexical valence (Table 3, Fig. 3A). No neural substrate showed significant negative linear correlation or quadratic correlation with mean lexical valence.

#### *3.2.4.5. Model 3: differential effects of lexical arousal-span*

After partialling out the variance due to the subjective passage ratings and mean lexical valence and arousal values, the hemodynamic response of the following neural substrates correlated positively with arousal-span (Fig. 3, panel B to D): left IFG pars triangularis including anterior insula (Fig. 3C; BA 13, 45 & 46), left premotor cortex (BA 6), bilateral posterior MTG and ITG (BA 37), left MTG and fusiform gyrus (BA 20 & 37), left middle cingulate gyrus (BA 24), right occipital pole (BA 17, 19 & 30), right globus pallidus and thalamus (ventral lateral nucleus and pulvinar; Fig. 3D, Table 4) and, after SVC, the left amygdala.

The hemodynamic response of the right PCC (BA 23 & 31) and precuneus (BA 7; Table 4) showed a significant negative correlation with arousal-span.

#### *3.2.4.6. Model 3: differential effects of lexical valence-span*

After partialling out the variance due to the subjective passage ratings, lexical means, and lexical arousal-span, the hemodynamic response in right occipital base (BA 19) and two clusters in right cerebellum correlated positively with valence-span (Table 4). No neural substrate correlated negatively with valence-span.

### ***3.2.5. Discussion of differential effects of affective lexical variables on neural activity***

#### *3.2.5.1. Differential effects of mean lexical valence values*

Our second parametric model revealed neural correlates representing effects of the emotion potential of words that go beyond 1) the holistic reading experience (as consciously expressed in passage ratings) and 2) effects due to inter-lexical affective variables (i.e., span). Our present finding of increasing lexical valence triggering amygdala activation after partialling out rating and arousal effects seems well in line with a recent EEG study that showed that early attention shifts to emotional words are restricted to the positive valence domain, once arousal is held constant (Recio et al., 2014). While amygdala activity has been often correlated with arousal,

recent evidence has demonstrated that also valence manipulations can engage this substrate of salience detection (Garvert, Friston, Dolan, & Garrido, 2014; Redondo et al., 2014; Stillman, Van Bavel, & Cunningham, 2014) – emerging as additional explanation of variance due to mean lexical valence in our statistical approach.

#### *3.2.5.2. Differential effects of lexical arousal-span*

We found significant correlations between lexical arousal-span and hemodynamic responses in many neural substrates, representing the emotion potential of dynamics and contrasts of arousal across the words constituting the text – beyond passage ratings and mean lexical values, effects of which had been partialled out. The neural correlates included striate and extrastriate visual cortex, amygdala, left anterior insula (extending from IFG), and thalamus. The latter three belong to the core affect regions in the psychological constructionist hypothesis of emotion (Lindquist et al., 2012), and the salience network (Seeley et al., 2007).

The anterior insula (BA 13), in which the activity is correlated with arousal-span, has been associated with awareness of bodily sensations and affective feelings (Craig, 2002, 2003, 2009), and the integration of autonomic and visceral information with emotional and motivational functions (Jones, Ward, & Critchley, 2010). Insula activation has also been shown to reflect interactive effects of emotional valence and arousal on lexical processing of emotional words (Citron et al., 2014), and to be responsive to discrete emotion information in words, in particular disgust (Ponz et al., 2013).

The activity in the striate and extrastriate visual cortices correlated with arousal-span probably represents enhanced visual processing via an attention gain control mechanism exerted by the amygdala through the direct connection with the visual cortex (Herbert et al., 2009), and the indirect modulation through a fronto-parietal network as proposed in the Multiple Attention Gain Control (MAGiC) Model (Pourtois, Schettino, & Vuilleumier, 2013).

Furthermore, arousal-span correlated positively with activation in bilateral premotor cortex (BA 6) and right middle globus pallidus. Respective activity in neural substrates associated with motor function, especially the premotor cortex, could reflect elicitation of action tendencies or preparation for action – implicitly associated with arousal in the component process model of emotion (Scherer, 2005). Motion-associated neural substrates have also been shown to be associated with embodied processing of abstract emotional meaning (Moseley et al., 2012). Lesions in globus pallidus have been associated with poor motivation, poor reward sensitivity, and apathy (Adam et al., 2013; RoCHAT et al., 2013; Vijayaraghavan, Vaidya, Humphreys, Beglinger, & Paradiso, 2008), while functional connectivity between right nucleus accumbens and right globus pallidus seems to be weaker in apathetic depressed patients than healthy controls (Alexopoulos et al., 2013). The meta-analysis of Hattingh et al. (2012) showed that patients with social anxiety disorder have stronger activation in right globus pallidus when perceiving socially emotive cues than healthy controls. Thus, apart from the association with motoric embodied emotion processing, activity in right globus pallidus may possibly be associated with the processing of reward or fear/anxiety – correlated with the increase of arousal-span in our data. The robust correlation between lexical arousal-span of texts and emotion-related neural correlates supports the idea that this variable is a promising predictor of emotional experience related to suspense and immersion in reading (Jacobs, 2011, 2014).

#### *3.2.5.3. Differential effects of subjective passage ratings*

After partialling out variance accounted for by affective lexical variables, we found no emotion-associated neural activation to be correlated in particular with valence or arousal ratings. Instead, we found negative correlation with arousal ratings and the quadratic term of valence ratings in left posterior temporal cortex and higher-level visual cortex. These negative correlations may best be explained assuming that the default network (Buckner, Andrews-Hanna, & Schacter, 2008;



Sprengh, Mar, & Kim, 2009) activity decreases with increasing perceived emotionality at the supra-lexical level. This seems reasonable assuming that supralexical effects may be very heterogeneous in nature and, accordingly, not relate to single brain regions but rather cause wide spread activity that becomes evident in our data as deactivation of the default network.

Furthermore, note that “Null effects” concerning residual variance for ratings also suggest that when reading our stimulus material (i.e., selected Harry Potter passages with rather unambiguous and passage-wise consistent specific emotional contents), the emotional salience and connotations of the constituting words and the dynamics/contrasts among them can – alone – account for the overall pattern of activity in emotion-related neural networks.

#### **4. General discussion, limitation and conclusion**

In this study, we investigated to what extent the affective impact of larger text units on the reader is a function of the emotional values of the words constituting them. We therefore considered the intercorrelations between three sources of data: 1) affective lexical variables provided by existing affective word databases (Conrad et al., in prep., see also BAWL-R and ANGST), representing lexical surface features; 2) subjective passage ratings of valence and arousal, representing the ‘holistic’ evaluation of the emotional reading experience by the readers; and 3) BOLD responses representing neural activity during reading

Intuitively, emotions we experience are generally considered a most individual, intimate and private issue – especially when, for instance, reading a book. Therefore, one can assume that participants’ individual affective evaluations should be the best predictors of their brain activity in emotion-related areas when reading a text. Besides robust parametric (unspecific) effects for such ratings given in our data, the results show that brain activity can be predicted with at least comparable efficiency via normative affective values of lexical surface features, as provided by large scale databases providing affective ratings of words – that is via quasi objective,

quantifiable patterns of basic elements, completely visible to everybody at the surface of a text. Our study, thus, builds upon previous behavioral approaches (Bestgen, 1994; Whissell, 2003) in trying to predict the emotional impact of text via the mean affective values of constituting words; the considerable overlap between parametric effects of individual passage ratings on the one hand and those for lexical valence and arousal means on the other – concerning their correlations with brain activity – in principle corroborates this view. In addition, our data also go a step further in providing evidence for genuine lexical effects beyond those of whole text ratings: mean lexical valence of words in a text still correlated significantly with amygdala activation after all alternative predictors' effects had been partialled out. We propose that these effects might reflect the special power of words to evoke affect in the following ways: Emotion-laden words have the potential to capture attention (Kissler et al., 2006) – and they do so in at least a partially context-independent manner (Delaney-Busch & Kuperberg, 2013; Wang et al., 2013); in addition, when we read a text, specific words reverberate in our minds beyond the more complex message conveyed by the text; and finally, the art of choosing the *right* words with the appropriate affective impact is part of what defines the skill of good writers or speakers. Moreover, while a text is clearly more than (or different from) the sum or average of its constituting words, single words also have a meaning beyond the specific context we find them embedded in: we have encountered these words in many other contexts before and our semantic representation of these words potentially contains traces of all these different contexts – giving words potentially complex emotional connotations. Superficially, these may not have much to do with the specific text we read, but rather the automatic activation of complex emotional connotations in our memory may add the lexical grain of affective salt that flavors text processing, therefore enabling us to joyfully read between the lines.

Furthermore, our data revealed the obvious limitations of the “lexical mean approach” by providing evidence for the importance of dynamic shifts concerning emotional content of words

across a text for affective reading experience: These shifts and their relevance cannot be captured via lexical means, because increasing shifts would always make means drift towards the neutral mean of the respective scale.

Accordingly, after partialling out the influence of alternative predictors including lexical means, the span of lexical arousal values of words encountered in a text passage proved to be the most fruitful differential predictor of activation in emotion-related brain area in our study. Respective findings involved the salience and emotion processing network and it seems plausible that sharp affective contrasts that we encounter between different elements of a text are particularly salient and trigger a range of emotional responses – in particular those involving preparation of immediate reactions of the organism to unexpected emotionally relevant events (Lang, Bradley, & Cuthbert, 1998; Scherer, 2005).

In sum, our data show that

- 1) affective evaluation of texts is strongly associated with mean and spread of lexical affective values,
- 2) means of lexical affective values and mean affective text evaluations both predict a widely shared pattern of brain activation in emotion-related brain area.
- 3) arousal-span represents a fruitful operationalization for the prediction of the emotion potential of texts at the inter-lexical level.

The absence of differential effects specific to the individual affective ratings of passages remains a particular finding that cannot necessarily be generalized. We would certainly not claim that the emotional impact of all literary productions could always be predicted in similar, sufficient ways via lexical and inter-lexical surface features as in our study. Rather, some aspects of our approach (concerning stimulus length and selection, questions of literary genre and stylistics) may have particularly favored effects specific to lexical features to arise; other types of stimuli such as

entire poems (Lüdtke et al., 2014) may instead be more appropriate to show that the whole is more than the sum of its parts when it comes to neural emotional correlates of literary reading. For instance, one may argue that short text samples of only four lines (like our stimuli) may simply not provide enough context to evoke specific supra-lexical effects. On one hand, the respective probability of detecting unique supra-lexical effects may indeed increase with text length. But notice that in the present case of a parametric approach, a rather large number of stimuli involving a sufficient spread of manipulated variables are necessary – which constrains possible individual stimulus length for reasons of maximum experiment duration. On the other hand, however, many different factors may account for supra-lexical emotional effects, e.g., figurativeness, context, event/situational change, affective empathy, or suspense, and not all of them necessarily require particularly long texts. In fact, metaphors, idioms, and irony/sarcasm can evoke emotional effects already at the *sentential* level (Bohn, Altmann, & Jacobs, 2012; Citron & Goldberg, 2014).

Additional evidence comes from Gernsbacher et al. (Gernsbacher, Goldsmith, & Robertson, 1992; Gernsbacher, Hallada, & Robertson, 1998; Gernsbacher & Robertson, 1992), who used short stories of 5 to 6 lines that led participants to correctly identify the emotional status of the protagonist, without explicitly using emotion-laden words. Furthermore, the current study follows the design of Altmann et al. (2012), who presented stories of five lines, line by line, in the scanner. The authors showed the effective connectivity from the medial prefrontal cortex to other regions involving affective empathy and ToM processing, including left amygdala and bilateral insula, when reading negatively vs. neutrally-valenced short stories. Specifically, the medial prefrontal cortex is associated with moral judgement when negative stories are liked by the reader. All these results clearly seem to reflect supra-lexical effects. In the current study, all our participants were sufficiently familiar with the novel-specific contents, hence contextual

knowledge, which is necessary for supra-lexical emotional effects to arise, was present even if participants read only short passages.

However, stylistic elements usually relax or untie the relation between affective features of lexical elements and the possible affective impact of a text. Therefore, a possible explanation of why our study failed to obtain specific supra-lexical effects – beyond an apparent deactivation of the default network by increasing emotion intensity expressed in supra-lexical ratings – may lie in the specific choice of literature. Harry Potter books are aimed to reach a broad population of readers, and their content might be relatively straightforward to comprehend in comparison with other literary materials, e.g., works by Proust, Chekhov, or Tolstoy. For example, in other literary works, authors may deliberately provide ambiguous information for the closure of *meaning gestalts* (Iser, 1976), so that readers would constantly consider new interpretations and possible outcomes upon reflection (Jacobs, 2011). In contrast, in our selected passages emotional content was rather straightforward, lacking more sophisticated stylistic elements that may have specifically contributed to the supra-lexical emotional reading experience (Jacobs, 2014). Finally, methodological constraints of fMRI, in particular the low temporal resolution of the BOLD response, let us choose text passages in which the emotional content evolved over the entire text passage in rather consistent ways. Clearly, such (consistent) stimulus characteristics are at odds with specific literary phenomena of sharply disrupting previously cheerfully nourished expectations that presumably evoke strong supralexical emotional effects as, for example, in the poems of Heinrich Heine (see deVega, Diaz, & Leon, 1997; deVega, Leon, & Diaz, 1996; Speer, Reynolds, Swallow, & Zacks, 2009; Speer et al., 2007 for psychological accounts of pertinent (rapid) updates of situation models).

To conclude, our data make a strong case for the prevailing importance of basic features concerning the emotional impact of overall very complex samples of human communication: the mean and spread of affective features of single words encountered in texts. Future studies will

hopefully extend and differentiate these findings to different literary genres and styles tapping into more specific peculiarities of literary reading at the supra-lexical level.

## **Acknowledgements**

This research was supported by two grants from Deutsche Forschungsgemeinschaft (DFG) to Markus Conrad: “Bilingualism and Affectivity in Reading”, Project 201 and “Sound-physiognomy in Language Organization, Processing and Production”, Project 410 from the Research Excellence Cluster “Languages of Emotion” at the Freie Universität Berlin.

## References

- Adam, R., Leff, A., Sinha, N., Turner, C., Bays, P., Draganski, B., et al. (2013). Dopamine reverses reward insensitivity in apathy following globus pallidus lesions. *Cortex*, 49(5), 1292-1303. doi: 10.1016/j.cortex.2012.04.013
- Alexopoulos, G. S., Hoptman, M. J., Yuen, G., Kanellopoulos, D., Seirup, J. K., Lim, K. O., et al. (2013). Functional connectivity in apathy of late-life depression: a preliminary study. *J Affect Disord*, 149(1-3), 398-405. doi: 10.1016/j.jad.2012.11.023
- Altmann, U., Bohrn, I. C., Lubrich, O., Menninghaus, W., & Jacobs, A. M. (2012). The power of emotional valence—from cognitive to affective processes in reading. *Front Hum Neurosci*, 6, 192. doi: 10.3389/fnhum.2012.00192
- Altmann, U., Bohrn, I. C., Lubrich, O., Menninghaus, W., & Jacobs, A. M. (2014). Fact vs fiction—how paratextual information shapes our reading processes. *Soc Cogn Affect Neurosci*, 9(1), 22-29. doi: 10.1093/scan/nss098
- Anderson, A. K., & Sobel, N. (2003). Dissociating intensity from valence as sensory inputs to emotion. *Neuron*, 39(4), 581-583. doi: 10.1016/S0896-6273(03)00504-X
- Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1), 95-113. doi: 10.1016/j.neuroimage.2007.07.007
- Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, 26(3), 839-851. doi: 10.1016/j.neuroimage.2005.02.018
- Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. *Annu Rev Psychol*, 58, 373-403. doi: 10.1146/annurev.psych.58.110405.085709
- Beissner, F., Meissner, K., Bar, K. J., & Napadow, V. (2013). The autonomic brain: an activation likelihood estimation meta-analysis for central processing of autonomic function. *J Neurosci*, 33(25), 10503-10511. doi: 10.1523/JNEUROSCI.1103-13.2013
- Bestgen, Y. (1994). Can emotional valence in stories be determined from words? *Cognition & Emotion*, 8(1), 21-36. doi: 10.1080/02699939408408926
- Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends Cogn Sci*, 15(11), 527-536. doi: 10.1016/j.tics.2011.10.001



Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb Cortex*, 19(12), 2767-2796. doi: 10.1093/cercor/bhp055

Bohrn, I. C., Altmann, U., & Jacobs, A. M. (2012). Looking at the brains behind figurative language-A quantitative meta-analysis of neuroimaging studies on metaphor, idiom, and irony processing. *Neuropsychologia*, 50(11), 2669-2683. doi: 10.1016/J.Neuropsychologia.2012.07.021

Bohrn, I. C., Altmann, U., Lubrich, O., Menninghaus, W., & Jacobs, A. M. (2012). Old proverbs in new skins - an fMRI study on defamiliarization. *Front Psychol*, 3, 204. doi: 10.3389/fpsyg.2012.00204

Bohrn, I. C., Altmann, U., Lubrich, O., Menninghaus, W., & Jacobs, A. M. (2013). When we like what we know--a parametric fMRI analysis of beauty and familiarity. *Brain Lang*, 124(1), 1-8. doi: 10.1016/j.bandl.2012.10.003

Bradley, M. M., & Lang, P. J. (1999). *Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings*. Gainesville, FL: University of Florida; NIMH Center for the Study of Emotion and Attention.

Brewer, W. F., & Lichtenstein, E. H. (1982). Stories Are to Entertain - a Structural-Affect Theory of Stories. *Journal of Pragmatics*, 6(5-6), 473-486. doi: 10.1016/0378-2166(82)90021-2

Briesemeister, B. B., Kuchinke, L., & Jacobs, A. M. (2011a). Discrete emotion effects on lexical decision response times. *PLoS One*, 6(8), e23743. doi: 10.1371/journal.pone.0023743

Briesemeister, B. B., Kuchinke, L., & Jacobs, A. M. (2011b). Discrete emotion norms for nouns: Berlin affective word list (DENN-BAWL). *Behav Res Methods*, 43(2), 441-448. doi: 10.3758/s13428-011-0059-y

Briesemeister, B. B., Kuchinke, L., & Jacobs, A. M. (2014). Emotion word recognition: Discrete information effects first, continuous later? *Brain Res*, 1564, 62-71. doi: 10.1016/j.brainres.2014.03.045

Büchel, C., Holmes, A. P., Rees, G., & Friston, K. J. (1998). Characterizing Stimulus-Response Functions Using Nonlinear Regressors in Parametric fMRI Experiments. *NeuroImage*, 8(2), 140-148. doi: 10.1006/nimg.1998.0351

Büchel, C., Wise, R. J., Mummery, C. J., Poline, J. B., & Friston, K. J. (1996). Nonlinear regression in parametric activation studies. *NeuroImage*, 4(1), 60-66. doi: 10.1006/nimg.1996.0029

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Ann N Y Acad Sci*, 1124, 1-38. doi: 10.1196/annals.1440.011

Citron, F. M. (2012). Neural correlates of written emotion word processing: a review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain Lang*, 122(3), 211-226. doi: 10.1016/j.bandl.2011.12.007

Citron, F. M., & Goldberg, A. E. (2014). Metaphorical sentences are more emotionally engaging than their literal counterparts. *J Cogn Neurosci*, 26(11), 2585-2595. doi: 10.1162/jocn\_a\_00654

Citron, F. M., Gray, M. A., Critchley, H. D., Weekes, B. S., & Ferstl, E. C. (2014). Emotional valence and arousal affect reading in an interactive way: Neuroimaging evidence for an approach-withdrawal framework. *Neuropsychologia*, 56(0), 79-89. doi: 10.1016/j.neuropsychologia.2014.01.002

Citron, F. M., Weekes, B. S., & Ferstl, E. C. (2013). Effects of valence and arousal on written word recognition: time course and ERP correlates. *Neurosci Lett*, 533, 90-95. doi: 10.1016/j.neulet.2012.10.054

Conrad, M., Recio, G., & Jacobs, A. M. (2011). The Time Course of Emotion Effects in First and Second Language Processing: A Cross Cultural ERP Study with German-Spanish Bilinguals. *Front Psychol*, 2, 351. doi: 10.3389/fpsyg.2011.00351

Costafreda, S. G., Brammer, M. J., David, A. S., & Fu, C. H. (2008). Predictors of amygdala activation during the processing of emotional stimuli: a meta-analysis of 385 PET and fMRI studies. *Brain Res Rev*, 58(1), 57-70. doi: 10.1016/j.brainresrev.2007.10.012

Craig, A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nat Rev Neurosci*, 3(8), 655-666. doi: 10.1038/nrn894

Craig, A. D. (2003). Interoception: the sense of the physiological condition of the body. *Curr Opin Neurobiol*, 13(4), 500-505. doi: 10.1016/S0959-4388(03)00090-4

Craig, A. D. (2009). How do you feel--now? The anterior insula and human awareness. *Nat Rev Neurosci*, 10(1), 59-70. doi: 10.1038/nrn2555

Critchley, H. D. (2005). Neural mechanisms of autonomic, affective, and cognitive integration. *J Comp Neurol*, 493(1), 154-166. doi: 10.1002/cne.20749

Critchley, H. D. (2009). Psychophysiology of neural, cognitive and affective integration: fMRI and autonomic indicants. *Int J Psychophysiol*, 73(2), 88-94. doi: 10.1016/j.ijpsycho.2009.01.012

Critchley, H. D., Wiens, S., Rotshtein, P., Ohman, A., & Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nat Neurosci*, 7(2), 189-195. doi: 10.1038/nn1176

Cunningham, W. A., Raye, C. L., & Johnson, M. K. (2004). Implicit and explicit evaluation: fMRI correlates of valence, emotional intensity, and control in the processing of attitudes. *J Cogn Neurosci*, 16(10), 1717-1729. doi: 10.1162/0898929042947919

Dagli, M. S., Ingeholm, J. E., & Haxby, J. V. (1999). Localization of cardiac-induced signal change in fMRI. *NeuroImage*, 9(4), 407-415. doi: 10.1006/nimg.1998.0424

Damasio, A. R. (1998). Emotion in the perspective of an integrated nervous system. *Brain Res Brain Res Rev*, 26(2-3), 83-86.

Damasio, A. R., & Carvalho, G. B. (2013). The nature of feelings: evolutionary and neurobiological origins. *Nat Rev Neurosci*, 14(2), 143-152. doi: 10.1038/nrn3403

Delaney-Busch, N., & Kuperberg, G. (2013). Friendly drug-dealers and terrifying puppies: affective primacy can attenuate the N400 effect in emotional discourse contexts. *Cogn Affect Behav Neurosci*, 13(3), 473-490. doi: 10.3758/s13415-013-0159-5

deVega, M., Diaz, J. M., & Leon, I. (1997). To know or not to know: Comprehending protagonists' beliefs and their emotional consequences. *Discourse Processes*, 23(2), 169-192. doi: 10.1080/01638537709544988

deVega, M., Leon, I., & Diaz, J. M. (1996). The representation of changing emotions in reading comprehension. *Cognition & Emotion*, 10(3), 303-321. doi: 10.1080/026999396380268

Ferstl, E. C., Neumann, J., Bogler, C., & von Cramon, D. Y. (2008). The extended language network: a meta-analysis of neuroimaging studies on text comprehension. *Hum Brain Mapp*, 29(5), 581-593. doi: 10.1002/hbm.20422

Ferstl, E. C., Rinck, M., & von Cramon, D. Y. (2005). Emotional and temporal aspects of situation model processing during text comprehension: an event-related fMRI study. *J Cogn Neurosci*, 17(5), 724-739. doi: 10.1162/0898929053747658

Fischler, I., & Bradley, M. (2006). Event-related potential studies of language and emotion: words, phrases, and task effects. *Prog Brain Res*, 156, 185-203. doi: 10.1016/S0079-6123(06)56009-1

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C., Frackowiak, R. S., et al. (1995). Analysis of fMRI time-series revisited. *NeuroImage*, 2(1), 45-53. doi: 10.1006/nimg.1995.1007

- Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philos Trans R Soc Lond B Biol Sci*, 358(1431), 459-473. doi: 10.1098/rstb.2002.1218
- Fritsch, N., & Kuchinke, L. (2013). Acquired affective associations induce emotion effects in word recognition: an ERP study. *Brain Lang*, 124(1), 75-83. doi: 10.1016/j.bandl.2012.12.001
- Garvert, M. M., Friston, K. J., Dolan, R. J., & Garrido, M. I. (2014). Subcortical amygdala pathways enable rapid face processing. *NeuroImage*, 102 Pt 2, 309-316. doi: 10.1016/j.neuroimage.2014.07.047
- Gernsbacher, M. A., Goldsmith, H. H., & Robertson, R. R. W. (1992). Do readers mentally represent characters' emotional states? *Cognition and Emotion*, 6(2), 89-111. doi: 10.1080/02699939208411061
- Gernsbacher, M. A., Hallada, B. M., & Robertson, R. R. W. (1998). How Automatically Do Readers Infer Fictional Characters' Emotional States? *Scientific Studies of Reading*, 2(3), 271-300. doi: 10.1207/s1532799xssr0203\_5
- Gernsbacher, M. A., & Robertson, R. R. W. (1992). Knowledge Activation Versus Sentence Mapping When Representing Fictional Characters Emotional States. *Language and Cognitive Processes*, 7(3-4), 353-371. doi: Doi 10.1080/01690969208409391
- Gillioz, C., Gygax, P., & Tapiero, I. (2012). Individual differences and emotional inferences during reading comprehension. *Can J Exp Psychol*, 66(4), 239-250. doi: 10.1037/a0028625
- Golub, G. H., & Van Loan, C. F. (1996). *Matrix Computations* (3 ed.). Baltimore, MD: Johns Hopkins University Press.
- Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annu Rev Psychol*, 48, 163-189. doi: 10.1146/annurev.psych.48.1.163
- Gray, M. A., Minati, L., Harrison, N. A., Gianaros, P. J., Napadow, V., & Critchley, H. D. (2009). Physiological recordings: basic concepts and implementation during functional magnetic resonance imaging. *NeuroImage*, 47(3), 1105-1115. doi: S1053-8119(09)00525-4 [pii] 10.1016/j.neuroimage.2009.05.033
- Gygax, P., Tapiero, I., & Carruzzo, E. (2007). Emotion Inferences During Reading Comprehension: What Evidence Can the Self-Pace Reading Paradigm Provide? *Discourse Processes*, 44(1), 33-50. doi: 10.1080/01638530701285564
- Hattingh, C. J., Ipser, J., Tromp, S. A., Syal, S., Lochner, C., Brooks, S. J., et al. (2012). Functional magnetic resonance imaging during emotion recognition in social anxiety disorder: an

activation likelihood meta-analysis. *Front Hum Neurosci*, 6, 347. doi: 10.3389/fnhum.2012.00347

Herbert, C., Ethofer, T., Anders, S., Junghofer, M., Wildgruber, D., Grodd, W., et al. (2009). Amygdala activation during reading of emotional adjectives--an advantage for pleasant content. *Soc Cogn Affect Neurosci*, 4(1), 35-49. doi: 10.1093/scan/nsn027

Hofmann, M. J., Kuchinke, L., Tamm, S., Võ, M. L., & Jacobs, A. M. (2009). Affective processing within 1/10th of a second: High arousal is necessary for early facilitative processing of negative but not positive words. *Cogn Affect Behav Neurosci* 9(4), 389-397. doi: 10.3758/9.4.389

Hsu, C.-T., Conrad, M., & Jacobs, A. M. (2014). Fiction feelings in Harry Potter: haemodynamic response in the mid-cingulate cortex correlates with immersive reading experience. *NeuroReport*, 25(17), 1356-1361. doi: 10.1097/WNR.0000000000000272

Hsu, C.-T., Jacobs, A. M., Altmann, U., & Conrad, M. (accepted). The magical activation of left amygdala when reading Harry Potter: An fMRI study on how descriptions of supra-natural events entertain and enchant. *PLoS ONE*.

Hsu, C.-T., Jacobs, A. M., & Conrad, M. (2015). Can Harry Potter still put a spell on us in a second language? An fMRI study on reading emotion-laden literature in late bilinguals. *Cortex*, 63, 282-295. doi: 10.1016/j.cortex.2014.09.002

Iser, W. (1976). *Der Akt des Lesens: Theorie ästhetischer Wirkung*. Munich, Germany: Fink Verlag.

Jacobs, A. M. (2011). Neurokognitive Poetik: Elemente eines neurokognitiven Modells des literarischen Lesens [Neurocognitive poetics: Elements of a neurocognitive model of literary reading]. In R. Schrott & A. M. Jacobs (Eds.), *Gehirn und Gedicht: Wie wir unsere Wirklichkeiten konstruieren* (pp. 492–524). Munich, Germany: Carl Hanser Verlag.

Jacobs, A. M. (2014). Towards a Neurocognitive Poetics Model of literary reading. In R. M. Willems (Ed.), *Towards a Cognitive Neuroscience of Natural Language Use* (pp. 135-159). Cambridge: Cambridge University Press.

Jacobs, A. M., Braun, M., Briesemeister, B., Conrad, M., Hofmann, M., Kuchinke, L., et al. (in prep.). *10 years of BAWLing into affective and aesthetic processes in reading*. Unpublished manuscript.

Jones, C. L., Ward, J., & Critchley, H. D. (2010). The neuropsychological impact of insular cortex lesions. *J Neurol Neurosurg Psychiatry*, 81(6), 611-618. doi: 10.1136/jnnp.2009.193672

Kissler, J., Assadollahi, R., & Herbert, C. (2006). Emotional and semantic networks in visual word processing: insights from ERP studies. *Prog Brain Res*, 156, 147-183. doi: 10.1016/S0079-6123(06)56008-X

Kissler, J., Herbert, C., Peyk, P., & Junghofer, M. (2007). Buzzwords: early cortical responses to emotional words during reading. *Psychological Science*, 18(6), 475-480. doi: 10.1111/j.1467-9280.2007.01924.x

Kringelbach, M. L., Vuust, P., & Geake, J. (2008). The pleasure of reading. *Interdiscipl Sci Rev*, 33(4), 321-335. doi: 10.1179/174327908x392889

Kuchinke, L., Jacobs, A. M., Grubich, C., Vo, M. L., Conrad, M., & Herrmann, M. (2005). Incidental effects of emotional valence in single word processing: an fMRI study. *NeuroImage*, 28(4), 1022-1032. doi: 10.1016/j.neuroimage.2005.06.050

Lancaster, J. L., Rainey, L. H., Summerlin, J. L., Freitas, C. S., Fox, P. T., Evans, A. C., et al. (1997). Automated labeling of the human brain: a preliminary report on the development and evaluation of a forward-transform method. *Hum Brain Mapp*, 5(4), 238-242. doi: 10.1002/(SICI)1097-0193(1997)5:4<238::AID-HBM6>3.0.CO;2-4

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., et al. (2000). Automated Talairach atlas labels for functional brain mapping. *Hum Brain Mapp*, 10(3), 120-131. doi: 10.1002/1097-0193(200007)10:3<120::AID-HBM30>3.0.CO;2-8

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1998). Emotion, motivation, and anxiety: brain mechanisms and psychophysiology. *Biol Psychiatry*, 44(12), 1248-1263. doi: 10.1016/S0006-3223(98)00275-3

Lehne, M., Engel, P., Menninghaus, W., Jacobs, A. M., & Koelsch, S. (in revision). *On first looking into Hoffmann's "The Sandman": Neural correlates of suspense during the first reading of a literary text*. Unpublished manuscript.

Lewis, P. A., Critchley, H. D., Rotshtein, P., & Dolan, R. J. (2007). Neural correlates of processing valence and arousal in affective words. *Cereb Cortex*, 17(3), 742-748. doi: 10.1093/cercor/bhk024

Lieberman, M. D., & Cunningham, W. A. (2009). Type I and Type II error concerns in fMRI research: re-balancing the scale. *Soc Cogn Affect Neurosci*, 4(4), 423-428. doi: 10.1093/scan/nsp052

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: a meta-analytic review. *Behav Brain Sci*, 35(3), 121-143. doi: 10.1017/S0140525X11000446

Lüdtke, J., Meyer-Sickendiek, B., & Jacobs, A. M. (2014). Immersing in the Stillness of an Early Morning: Testing the Mood Empathy Hypothesis in Poems. *Psychology of Aesthetics, Creativity, and the Arts*, 8, 363-377. doi: 10.1037/a0036826

Lund, T. E., Madsen, K. H., Sidaros, K., Luo, W. L., & Nichols, T. E. (2006). Non-white noise in fMRI: Does modelling have an impact? *NeuroImage*, 29(1), 54-66. doi: Doi 10.1016/J.Neuroimage.2005.07.005

Maldjian, J. A., Laurienti, P. J., Kraft, R. A., & Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage*, 19(3), 1233-1239. doi: 10.1016/S1053-8119(03)00169-1

Mar, R. A. (2011). The neural bases of social cognition and story comprehension. *Annu Rev Psychol*, 62, 103-134. doi: 10.1146/annurev-psych-120709-145406

Mar, R. A., & Oatley, K. (2008). The function of fiction is the abstraction and simulation of social experience. *Perspect Psychol Sci* 3(3), 173-192. doi: 10.1111/J.1745-6924.2008.00073.X

Mason, R. A., & Just, M. A. (2009). The Role of the Theory-of-Mind Cortical Network in the Comprehension of Narratives. *Lang Linguist Compass*, 3(1), 157-174. doi: 10.1111/j.1749-818X.2008.00122.x

Moseley, R., Carota, F., Hauk, O., Mohr, B., & Pulvermuller, F. (2012). A role for the motor system in binding abstract emotional meaning. *Cereb Cortex*, 22(7), 1634-1647. doi: 10.1093/cercor/bhr238

Murphy, F. C., Nimmo-Smith, I., & Lawrence, A. D. (2003). Functional neuroanatomy of emotions: a meta-analysis. *Cogn Affect Behav Neurosci*, 3(3), 207-233. doi: 10.3758/CABN.3.3.207

Nagels, A., Kauschke, C., Schrauf, J., Whitney, C., Straube, B., & Kircher, T. (2013). Neural substrates of figurative language during natural speech perception: an fMRI study. *Front Behav Neurosci*, 7, 121. doi: 10.3389/fnbeh.2013.00121

Nell, V. (1988). *Lost in a Book: The Psychology of Reading for Pleasure*. New Haven/London: Yale University Press.

Nummenmaa, L., Glerean, E., Hari, R., & Hietanen, J. K. (2014). Bodily maps of emotions. *Proc Natl Acad Sci U S A*, 111(2), 646-651. doi: 10.1073/pnas.1321664111

Osgood, C. E. (1969). On the whys and wherefores of E, P, and A. *Journal of Personality and Social Psychology*, 12(3), 194-199. doi: 10.1037/h0027715

Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *NeuroImage*, 16(2), 331-348. doi: 10.1006/nimg.2002.1087

Ponz, A., Montant, M., Liegeois-Chauvel, C., Silva, C., Braun, M., Jacobs, A. M., et al. (2013). Emotion processing in words: a test of the neural re-use hypothesis using surface and intracranial EEG. *Soc Cogn Affect Neurosci*. doi: 10.1093/scan/nst034

Pourtois, G., Schettino, A., & Vuilleumier, P. (2013). Brain mechanisms for emotional influences on perception and attention: what is magic and what is not. *Biol Psychol*, 92(3), 492-512. doi: 10.1016/j.biopsycho.2012.02.007

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, 62(2), 816-847. doi: 10.1016/j.neuroimage.2012.04.062

Recio, G., Conrad, M., Hansen, L. B., & Jacobs, A. M. (2014). On Pleasure and Thrill: The Interplay between Arousal and Valence during Visual Word Recognition. *Brain Lang*, 134, 34-43. doi: 10.1016/j.bandl.2014.03.009

Redondo, R. L., Kim, J., Arons, A. L., Ramirez, S., Liu, X., & Tonegawa, S. (2014). Bidirectional switch of the valence associated with a hippocampal contextual memory engram. *Nature*, 513(7518), 426-430. doi: 10.1038/nature13725

Regel, S., Gunter, T. C., & Friederici, A. D. (2011). Isn't it ironic? An electrophysiological exploration of figurative language processing. *J Cogn Neurosci*, 23(2), 277-293. doi: 10.1162/jocn.2010.21411

Rochat, L., Van der Linden, M., Renaud, O., Epiney, J. B., Michel, P., Sztajzel, R., et al. (2013). Poor reward sensitivity and apathy after stroke: implication of basal ganglia. *Neurology*, 81(19), 1674-1680. doi: 10.1212/01.wnl.0000435290.49598.1d

Rowling, J. K. (1997). *Harry Potter and the Philosopher's Stone*. London: Bloomsbury.

Rowling, J. K. (1998). *Harry Potter and the Chamber of Secrets*. London: Bloomsbury.

Rowling, J. K. (1999). *Harry Potter and the Prisoner of Azkaban*. London: Bloomsbury.

Rowling, J. K. (2000). *Harry Potter and the Goblet of Fire*. London: Bloomsbury.

Rowling, J. K. (2003). *Harry Potter and the Order of the Phoenix*. London: Bloomsbury.



- Rowling, J. K. (2005). *Harry Potter and the Half Blood Prince*. London: Bloomsbury.
- Rowling, J. K. (2007). *Harry Potter and the Deathly Hallows*. London: Bloomsbury.
- Russell, J. A. (1980). A circumplex model of affect. *J Pers Soc Psychol*, 39, 1161-1178. doi: 10.1037/h0077714
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychol Rev*, 110(1), 145-172. doi: 10.1037/0033-295X.110.1.145
- Schacht, A., Adler, N., Chen, P., Guo, T., & Sommer, W. (2012). Association with positive outcome induces early effects in event-related brain potentials. *Biol Psychol*, 89(1), 130-136. doi: 10.1016/j.biopsycho.2011.10.001
- Schauenburg, G., Conrad, M., Ambrasat, J., Von Scheve, C., & Schröder, T. (2013). EEG correlates of affective coherence in visually presented sentences, *11th Symposium of Psycholinguistics*. Spain.
- Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44(4), 695-729. doi: 10.1177/0539018405058216
- Schmalhofer, F., & Glavanov, D. (1986). Three components of understanding a programmer's manual: Verbatim, propositional, and situational representations. *J Mem Lang* 25(3), 279-294. doi: 10.1016/0749-596X(86)90002-1
- Schmidtke, D. S., Schröder, T., Jacobs, A. M., & Conrad, M. (2014). ANGST: Affective Norms for German Sentiment Terms derived from the Affective Norms for English Words. *Behav Res Methods*, 46(4), 1108-1118. doi: 10.3758/s13428-013-0426-y
- Schrott, R., & Jacobs, A. M. (2011). *Gehirn und Gedicht: Wie wir unsere Wirklichkeiten konstruieren (Brain and Poetry: How We Construct Our Realities)*. München, Germany: Hanser.
- Scott, G. G., O'Donnell, P. J., Leuthold, H., & Sereno, S. C. (2009). Early emotion word processing: evidence from event-related potentials. *Biol Psychol*, 80(1), 95-104. doi: 10.1016/j.biopsycho.2008.03.010
- Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., et al. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *J Neurosci*, 27(9), 2349-2356. doi: 10.1523/JNEUROSCI.5587-06.2007

- Small, D. M., Gregory, M. D., Mak, Y. E., Gitelman, D., Mesulam, M. M., & Parrish, T. (2003). Dissociation of neural representation of intensity and affective valuation in human gustation. *Neuron*, 39(4), 701-711.
- Speer, N. K., Reynolds, J. R., Swallow, K. M., & Zacks, J. M. (2009). Reading stories activates neural representations of visual and motor experiences. *Psychol Sci*, 20(8), 989-999. doi: 10.1111/j.1467-9280.2009.02397.x
- Speer, N. K., Zacks, J. M., & Reynolds, J. R. (2007). Human brain activity time-locked to narrative event boundaries. *Psychol Sci*, 18(5), 449-455. doi: 10.1111/j.1467-9280.2007.01920.x
- Spreng, R. N., Mar, R. A., & Kim, A. S. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *J Cogn Neurosci*, 21(3), 489-510. doi: 10.1162/jocn.2008.21029
- Stillman, P. E., Van Bavel, J. J., & Cunningham, W. A. (2014). Valence Asymmetries in the Human Amygdala: Task Relevance Modulates Amygdala Responses to Positive More than Negative Affective Cues. *J Cogn Neurosci*, 1-10. doi: 10.1162/jocn\_a\_00756
- Straube, T., Sauer, A., & Miltner, W. H. (2011). Brain activation during direct and indirect processing of positive and negative words. *Behav Brain Res*, 222(1), 66-72. doi: 10.1016/j.bbr.2011.03.037
- Van Berkum, J. J., Holleman, B., Nieuwland, M., Otten, M., & Murre, J. (2009). Right or wrong? The brain's fast response to morally objectionable statements. *Psychol Sci*, 20(9), 1092-1099. doi: 10.1111/j.1467-9280.2009.02411.x
- Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Vijayaraghavan, L., Vaidya, J. G., Humphreys, C. T., Beglinger, L. J., & Paradiso, S. (2008). Emotional and motivational changes after bilateral lesions of the globus pallidus. *Neuropsychology*, 22(3), 412-418. doi: 10.1037/0894-4105.22.3.412
- Võ, M. L., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behav Res Methods*, 41(2), 534-538. doi: 10.3758/BRM.41.2.534
- Võ, M. L., Jacobs, A. M., & Conrad, M. (2006). Cross-validating the Berlin Affective Word List. *Behav Res Methods*, 38(4), 606-609. doi: 10.3758/bf03193892

Võ, M. L., Jacobs, A. M., Kuchinke, L., Hofmann, M., Conrad, M., Schacht, A., et al. (2008). The coupling of emotion and cognition in the eye: introducing the pupil old/new effect. *Psychophysiology*, 45(1), 130-140. doi: 10.1111/j.1469-8986.2007.00606.x

Wallentin, M., Nielsen, A. H., Vuust, P., Dohn, A., Roepstorff, A., & Lund, T. E. (2011). Amygdala and heart rate variability responses from listening to emotionally intense parts of a story. *NeuroImage*, 58(3), 963-973. doi: 10.1016/j.neuroimage.2011.06.077

Walter, H. (2012). Social Cognitive Neuroscience of Empathy: Concepts, Circuits, and Genes. *Emotion Review*, 4(1), 9-17. doi: 10.1177/1754073911421379

Wang, L., Bastiaansen, M., Yang, Y., & Hagoort, P. (2013). ERP evidence on the interaction between information structure and emotional salience of words. *Cogn Affect Behav Neurosci* 13(2), 297-310. doi: 10.3758/s13415-012-0146-2

Whissell, C. M. (2003). Readers' opinions of romantic poetry are consistent with emotional measures based on the Dictionary of Affect in Language. *Percept Mot Skills*, 96(3 Pt 1), 990-992. doi: 10.2466/pms.2003.96.3.990

Whissell, C. M. (2010). Poet interrupted: differences in the emotionality and imagery of Byron's poetry associated with his turbulent mid-career years in England. *Psychol Rep*, 107(1), 321-328. doi: 10.2466/10.21.28.PR0.107.4.321-328

Whissell, C. M. (2011). Explaining inconsistencies in Shakespeare's character Henry V on the basis of the emotional undertones of his speeches. *Psychol Rep*, 108(3), 843-855. doi: 10.2466/28.PR0.108.3.843-855

Whissell, C. M., & Dewson, M. R. J. (1986). A Dictionary of Affect In Language: III. Analysis of Two Biblical and Two Secular Passages. *Perceptual and Motor Skills*, 62(1), 127-132. doi: 10.2466/pms.1986.62.1.127

Winston, J. S., O'Doherty, J., & Dolan, R. J. (2003). Common and distinct neural responses during direct and incidental processing of multiple facial emotions. *NeuroImage*, 20(1), 84-97. doi: 10.1016/S1053-8119(03)00303-3

Wolf, M. (2007). *Proust and the Squid – The Story and Science of the Reading Brain*. New York: Harper Perennial.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychol Bull*, 123(2), 162-185. doi: 10.1037/0033-2909.123.2.162

### **Figure 1. The Design Matrices of Three Models of Parametric Analyses**

Each model contains the regressors of the German Reading condition, six parametric modulators and their polynomial expansions, English Reading, and Question conditions, followed by six motion parameters. All regressors are serially orthogonalized. Three models differ in their sequence of entry of the parametric modulators. L. Arousal = lexical arousal mean; L. Val. 1<sup>st</sup> = first order lexical valence mean; L. Val 2<sup>nd</sup> = second order lexical valence mean; Aro.-span = arousal-span; Val.-span = valence-span; Arousal R. = arousal ratings; Val. R. 1<sup>st</sup> = first order valence ratings; Val. R. 2<sup>nd</sup> = second order valence ratings.

### **Figure 2. Results of Reading (German) vs. Fixation**

Regions showing significantly stronger BOLD response in the German reading condition than fixation. The voxel-level threshold is FWE corrected  $p < 0.05$ , the size threshold 5.

A: left hemisphere render image.

B: right hemisphere render image.

C: sagittal section at  $x = 0$  showing activation in medial supplementary motor cortex, visual cortex, and pons.

D: transverse section at  $z = -18$ . The crosshair highlights ventromedial prefrontal cortex (0 48 - 18). The section also shows activation in bilateral anterior temporal lobe and parahippocampal cortex including amygdala.

### **Figure 3. Specific Effects of Affective Lexical Variables**

A: transverse section at  $z = -15$  showing significant linear positive correlation between BOLD responses in the left amygdala (crosshair, -28 -4 -15, after SVC) and lexical valence means after partialling out variance due to ratings, spans, and lexical arousal means in Model 2.

B to D: results of significant positive correlation between BOLD responses and arousal-span after partialling out variance due to ratings, lexical means in Model 3.

B: left hemisphere render showing significant clusters in left IFG including insula, premotor area, and temporo-occipital junction.

C: sagittal section at  $x = -40$  highlighting the significant cluster in left IFG and insula (crosshair, -40 29 12).

D: sagittal section at  $x = 16$  highlighting the significant cluster in right globus pallidus and thalamus (crosshair, 16 -7 3).

The initial voxel-level threshold is uncorrected  $p < 0.005$ , the size threshold is according to the cluster-level FDR correction in each analysis.

### **Figure A.1 Unspecific Effects of Affective Lexical Variables**

Render images were produced with *xjview*, in which red color indicates significant positive correlations; green indicate significant negative correlations.

Panel A to D show unspecific effects of *mean lexical arousal* in Model 1. Panel D shows significant positive correlation in bilateral amygdala (crosshair -26 -6 -23) after SVC.

Panel E and F show unspecific linear effects, and panel G and H show unspecific quadratic effects of *mean lexical valence* in Model 1.

Panel I to K show unspecific effects of *arousal ratings* in Model 1.

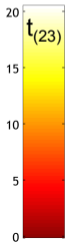
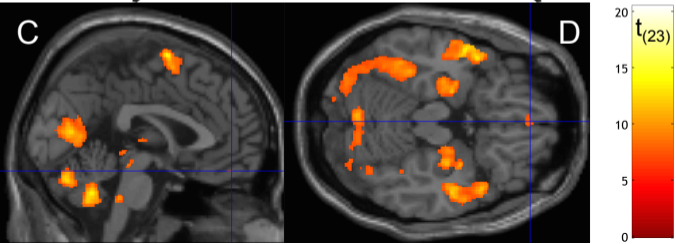
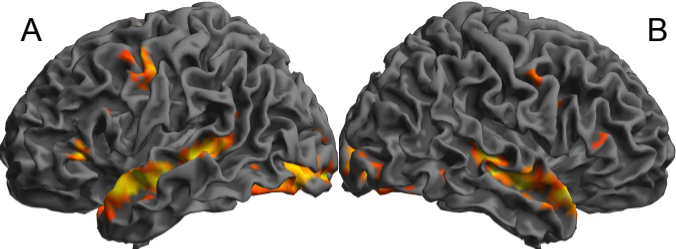
Panel M and N show unspecific linear effects, and panel L, O and P show unspecific quadratic effects of *valence ratings* in Model 2. Panel L shows significant positive quadratic correlation with *valence ratings* in bilateral amygdala (crosshair 27 -6 -14) after SVC.

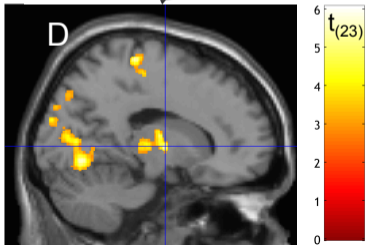
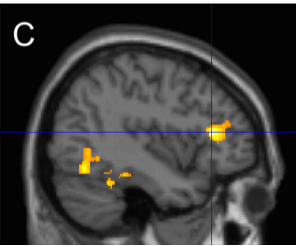
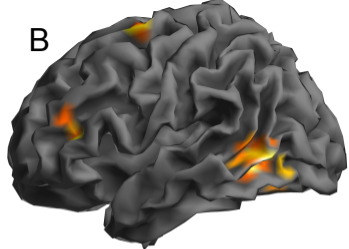
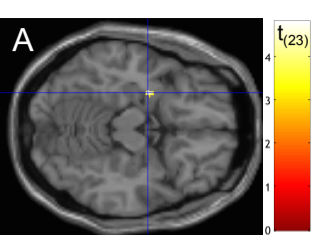
Panel Q to T: unspecific effects of *arousal-span* in Model 1.

Panel U: unspecific effects *valence-span* in Model 1 highlighting significant positive correlation in the left amygdala (crosshair -24 -7 -17).

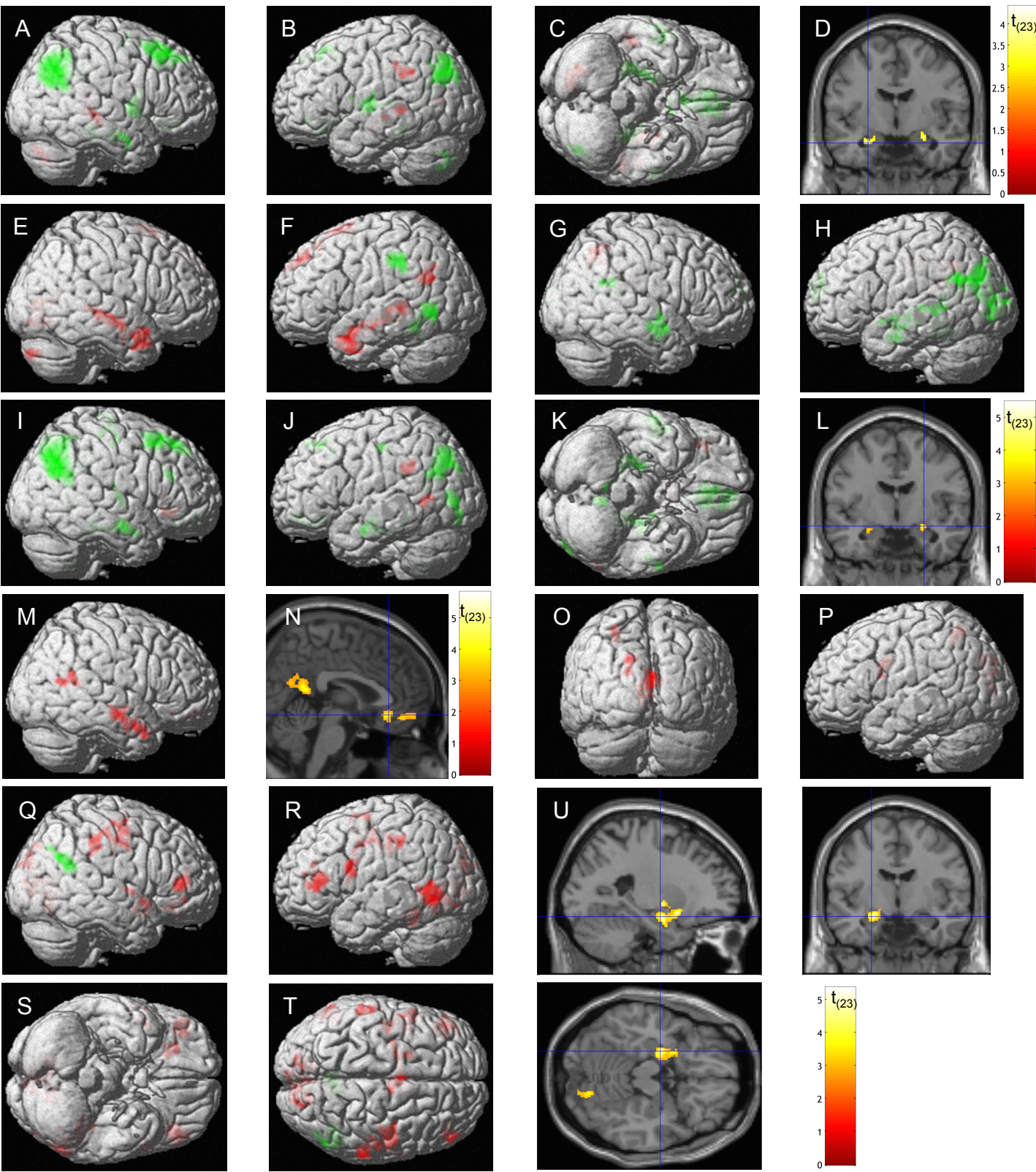
The initial voxel-level threshold is uncorrected  $p < 0.005$ , the size threshold is according to the cluster-level FDR correction in each analysis.

Reg.	Model 1	Model 2	Model 3	Serial Orthogonalization ↓
1	German Reading Condition			
2	L. Arousal	Arousal R.	Arousal R.	
3	L. Val. 1 <sup>st</sup>	Val. R. 1 <sup>st</sup>	Val. R. 1 <sup>st</sup>	
4	L. Val. 2 <sup>nd</sup>	Val. R. 2 <sup>nd</sup>	Val. R. 2 <sup>nd</sup>	
5	Aro.-span	Aro.-span	L. Arousal	
6	Val.-span	Val.-span	L. Val. 1 <sup>st</sup>	
7	Arousal R.	L. Arousal	L. Val. 2 <sup>nd</sup>	
8	Val. R. 1 <sup>st</sup>	L. Val. 1 <sup>st</sup>	Aro.-span	
9	Val. R. 2 <sup>nd</sup>	L. Val. 2 <sup>nd</sup>	Val.-span	
10	English Reading Condition			
11	Questions			
12-17	Six motion parameters			









**Table 2. Examples of Passages with High Mean Lexical Valence/Arousal and Valence-  
/Arousal-span values**

<b>Passages</b>	<b>Lexical Valence</b>	<b>Lexical Arousal</b>	<b>Valence- span</b>	<b>Arousal- span</b>
<p><i><b>Highly positive mean lexical valence</b></i></p> <p>But James merely laughed, permitted his mother to kiss him, gave his father a fleeting hug, then leapt on to the rapidly filling train. They saw him wave, then sprint away up the corridor to find his friends. (Book 7)</p>	1.74	2.67	2.30	2.07
<p><i><b>Highly negative mean lexical valence</b></i></p> <p>'You disgusting little Squib, you filthy little blood traitor!' roared Gaunt, losing control, and his hands closed around his daughter's throat. Both Harry and Ogden yelled 'No!' at the same time. (Book 6)</p>	-0.81	3.13	3.82	2.54
<p><i><b>High mean lexical arousal</b></i></p> <p>Wormtail screamed, screamed as though every nerve in his body was on fire, the screaming filled Harry's ears as the scar on his forehead seared with pain; he was yelling, too. (Book 4)</p>	-0.69	3.57	3.64	2.23

<p><b><i>Large valence-span</i></b></p> <p>When the Dementors approached him, he heard the last moments of his mother's life, her attempts to protect him, Harry, from Lord Voldemort, and Voldemort's laughter before he murdered her.... Harry dozed fitfully, sinking into dreams full of clammy, rotted hands and petrified pleading. (Book 3)</p>	0.11	2.95	4.60	2.35
<p><b><i>Large arousal-span</i></b></p> <p>And then a silence fell over the crowd, from the front first, so that a chill seemed to spread down the corridor. The Fat Lady had vanished from the portrait, which had been slashed so viciously that strips of canvas littered the floor; great chunks of it had been torn away completely. (Book 3)</p>	-0.015	2.76	3.40	2.83

**Table 3. Differential Effects of Affective Lexical Variables and Subjective Passage Ratings**

<b>H</b>	<b>Regions</b>	<b>Cluster size</b>	<b>p*</b>	<b>T</b>	<b>B.A.</b>	<b>[x, y, z]</b>
<i>Model 1: Negative linear Correlation with Arousal Ratings</i>						
L	TOJ (MOG, MTG & IOG)	642	0.007	4.49	19/39	-45 -84 0
<i>Model 1: Negative quadratic Correlation with Valence Ratings</i>						
L	Posterior MTG	1723	0.000	5.71	21	-63 -48 3
<i>Model 2: Positive linear Correlation with mean Lexical Valence</i>						
R	PHC, fusiform and lingual gyrus	933	0.000	5.32	19	27 -57 -8
R	STS (STG & MTG)	519	0.010	4.36	22	46 -27 -2
L	Amygdala (SVC)	19	0.007	4.79		-28 -4 -15
<i>Model 3: Positive linear Correlation with Arousal-span</i>						
L	IFG pars triangularis, insula & MFG	459	0.027	6.04	13/45/46	-40 29 12
R	Occipital pole (lingual & cuneus)	2543	0.000	5.88	19/30/17	15 -69 -8
L	Posterior MTG & ITG	1923	0.000	5.65	37/19	-57 -64 -3
R	Middle globus pallidus, thalamus (ventral lateral nucleus & pulvinar)	500	0.024	5.48		16 -7 3
L	Cerebellum, MTG, fusiform gyrus	362	0.042	5.19	20/37	-38 -45 -27
R	TOJ (MTG, MOG & ITG)	443	0.027	5.10	37	51 -61 3
L	Premotor cortex (MFG, SFG, med. FG)	392	0.036	4.72	6	-26 -4 64
L+R	Cingulate gyrus	387	0.036	4.70	24	-4 -1 46
L	Amygdala (SVC)	21	0.021	4.24		-22 -1 -26
<i>Model 3: Negative linear Correlation with Arousal-span</i>						
R	PCC & Precuneus	780	0.001	4.79	31/23/7	0 -57 24
<i>Model 3: Positive linear Correlation with Valence-span</i>						
R	Lingual gyrus & Cerebellum	622	0.009	4.70	18	16 -85 -15

R Cerebellum

562 0.009 3.91

39 -76 -35

\* Cluster level false-discovery-rate-corrected p-values

Abbreviations: FG = frontal gyrus; IFG = inferior frontal gyrus; IOG = inferior occipital gyrus; ITG = inferior temporal gyrus; MFG = middle frontal gyrus; MOG = middle occipital gyrus; MTG = middle temporal gyrus; PCC = posterior cingulate cortex; PHC = parahippocampal cortex; SFG = superior frontal gyrus; STG = superior temporal gyrus; STS = superior temporal sulcus; SVC = small volume correction; TOJ = temporo-occipital junction; H = hemisphere; L = left; R = right; p = p-value; T = T-value; B.A. = Brodmann area; x, y, z = MNI coordinates

**Table A.1 Main Effects of Reading (German) vs. Fixation**

<b>H</b>	<b>Regions</b>	<b>Cluster size</b>	<b>p*</b>	<b>T</b>	<b>B.A.</b>	<b>[x, y, z]</b>
<i>Reading (German) &gt; Fixation</i>						
L+R	Thalamus, PHC incl. L Amygdala	3447	0.000	20.39		22 -27 -0
			0.000	14.91		-21 -30 -0
			0.000	11.93		-28 -6 -21
L	STS incl. aTL (STG & MTG)	3992	0.000	17.07	38	-51 11 -14
			0.000	15.00	22	-57 -3 -9
			0.000	14.01	22	-54 -39 4
L+R	Cerebellum & Lingual gyrus	13379	0.000	16.83		30 -64 -50
			0.000	15.15	17	-12 -97 -6
			0.000	14.90	17	16 -87 -3
R	STS incl. aTL (STG & MTG)	2714	0.000	16.46	41	51 -21 -8
			0.000	14.58	22	50 -27 -0
			0.000	14.31	21	54 -9 -14
L	Medial SMA (SFG)	676	0.000	14.99	6	-3 2 67
			0.000	10.33	6	0 3 60
			0.002	8.68	6	-4 9 63
L	Cerebellar Nodule & Tonsil	1083	0.000	13.77		-2 -55 -35
			0.000	12.57		-20 -40 -44
			0.001	8.92		-9 -49 -41
R	PHC incl. hippocampus & Amygdala	380	0.000	12.69		33 -1 -20
			0.000	12.11		30 -10 -20
			0.000	9.49	34	21 -10 -18
L	Precentral gyrus	822	0.000	12.31	6	-51 -1 48
			0.000	10.68	4	-50 -12 43

			0.002	8.82	6	-46	-6	37
R	Cerebellar Declive	316	0.000	12.27		42	-66	-24
			0.000	10.97		33	-61	-26
L	IFG pars triangularis	277	0.000	11.81	47	-50	27	-2
R	Cerebellar Tonsil	107	0.000	10.58		21	-43	-44
R	Precentral & MFG	168	0.000	10.02	6	60	-7	42
			0.003	8.35	6	51	5	49
L+R	Pons	71	0.000	9.94		0	-36	-39
L	Pons	45	0.000	9.49		-6	-25	-38
L	Putamen	195	0.001	9.15		-20	5	7
R	Postcentral gyrus	18	0.001	9.10	3	51	-18	55
R	IFG pars triangularis & MFG	202	0.001	8.86	46	56	33	9
			0.012	7.66	45	57	27	3
L	IFG pars triangularis & MFG	191	0.002	8.80	46	-42	17	22
R	Paracentral lobule	45	0.002	8.76	6	9	-33	58
R	Precentral gyrus	27	0.004	8.31	6	30	-19	66
L	Paracentral lobule	24	0.004	8.23	6	-8	-34	57
R	dIPFC (MFG)	129	0.005	8.14	9	41	9	30
			0.007	7.90	46	41	20	25
R	Precentral gyrus	11	0.005	8.09	6	28	-19	55
R	Caudate tail	7	0.006	7.99		22	-31	21
L+R	vmPFC (Orbital gyrus)	22	0.006	7.98	11	0	48	-18
L	Cuneus	59	0.007	7.95	18	-21	-81	19
R	Cingulate	18	0.007	7.94	31	12	-24	48
R	ITG	13	0.008	7.85	37	50	-49	-12
R	TPJ (MTG)	79	0.008	7.84	39	32	-67	28

L	TPJ (IPL)	37	0.010	7.75	40	-26 -49 52
R	Precentral gyrus	8	0.011	7.70	6	36 -10 63
R	Precuneus	11	0.011	7.70	7	30 -55 51
R	Caudate body	15	0.014	7.54		8 9 10
R	Fusiform	8	0.018	7.40	37	38 -43 -17
L	Fusiform	9	0.021	7.33	20	-39 -12 -30

\* Voxel level FWE-corrected

Abbreviations: aTL = anterior temporal lobe; dlPFC = dorsolateral prefrontal cortex; IFG = inferior frontal gyrus; ITG = inferior temporal gyrus;

MFG = middle frontal gyrus; MTG = middle temporal gyrus; PHC = parahippocampal cortex; SMA = supplementary motor area; STG = superior

temporal gyrus; STS = superior temporal sulcus; TPJ = temporo-parietal junction; vmPFC = ventromedial prefrontal cortex; H = hemisphere; L =

left; R = right; p = p-value; T = T-value; B.A. = Brodmann area; x, y, z = MNI coordinates



**Table A.2 Unspecific Effects of Affective Lexical Variables in Model 1**

<b>H</b>	<b>Regions</b>	<b>Cluster size</b>	<b>p*</b>	<b>T</b>	<b>B.A.</b>	<b>[x, y, z]</b>
<i>Positive linear Correlation with mean lexical Arousal</i>						
L	Cuneus & PCC	337	0.033	5.76	17/30	-18 -76 9
R	STS (STG & MTG)	526	0.009	5.34	22	50 -30 -3
R	Cerebellum	614	0.009	5.30		15 -78 -30
L	IPL, supramarginal & postcentral gyrus	474	0.011	5.06	40/2	-52 -37 34
L	Caudate tail, hippocampus & putamen	521	0.009	4.97		-33 -24 -6
L	MTG & STG	520	0.009	4.96	22/41/21	-51 -36 4
R	Caudate tail	372	0.025	4.81		22 -28 21
R	Amygdala (SVC)	35	0.015	4.41		27 -9 -17
L	Amygdala (SVC)	35	0.028	4.09		-26 -6 -23
<i>Negative linear Correlation with mean lexical Arousal</i>						
L	Precuneus, SPL & cerebellum	1827	0.000	7.60	19/7	-42 -79 34
R+L	ACC, vmPFC (subcallosal gyrus & med FG)	1873	0.000	6.94 4.99	32 25/11	8 38 -9 -6 24 -12
R	TPJ (IPL, MTG & AG)	3405	0.000	6.44	19/39	42 -73 37
R	PHC & hippocampus	1000	0.000	6.37	28	30 -34 -12
R	dIPFC (MFG & SFG)	2371	0.000	6.16	8	39 27 49
L	PHC	714	0.001	6.11	36/28	-24 -39 -15
R	PCC	1586	0.000	6.02	23/30	6 -60 18
R	PCC	1575	0.000	5.88	31	9 -42 39
L	STG & postcentral gyrus	562	0.002	4.82	22/42/41	-57 -9 3
R	aTL (MTG & ITG)	477	0.005	4.73	20/21	50 -6 -23

L	dIPFC (MFG)	342	0.019	4.57	8	-28 32 51
R	STG & precentral gyrus	507	0.004	4.41	22/44	53 -1 1
R+L	Medial premotor cortex (medial FG)	327	0.021	4.37	6	15 -3 60
	& cingulate gyrus			3.35	24	-6 -6 52
R+L	Medial frontopolar cortex	252	0.039	4.15	10	12 59 1
	(SFG & medial FG)			3.69	10	-6 65 7
L	Cerebellum	272	0.039	4.15		-50 -67 -35

***Positive linear Correlation with mean lexical Valence***

R	aTL (MTG & STG)	2088	0.000	5.86	38/22/21	51 10 -18
L	aTL (MTG & STG)	2207	0.000	5.66	21/22	-50 6 -26
L	Premotor cortex (SFG)	423	0.009	5.66	6	-8 6 67
R	Cerebellum	394	0.011	5.23		24 -85 -38
R	Extrastriate visual cortex	1317	0.000	5.06	18/30	21 -85 -9
	(lingual gyrus & cuneus)					
L	dIPFC (SFG)	499	0.004	4.87	9/8	-20 42 43
L	TPJ (MTG, IPL &	767	0.000	4.68	39/13	-44 -60 24
	supramarginal gyrus)					
L	PHC & lingual gyrus	278	0.043	4.49	30/19	-21 -55 1
L	PCC & precuneus	717	0.000	4.14	31/7	-3 -63 24

***Negative linear Correlation with mean lexical Valence***

L	MTG	1163	0.000	6.27	37	-54 -57 -2
L	Postcentral gyrus & IPL	661	0.001	5.41	2/40	-60 -33 42

***Positive quadratic Correlation with mean lexical Valence***

R	SPL & caudate tail	994	0.000	4.67	7	34 -57 51
L	Caudate tail	804	0.000	4.58		-14 -39 18

***Negative quadratic Correlation with mean lexical Valence***

R	STS (MTG & STG)	1131	0.000	5.71	21/22	51 -1 -15
L	STG, MTG & ITG	1369	0.000	5.59	22/20/38	-63 -25 3
L	TOJ (MTG & cuneus)	4912	0.000	5.53	39/18	-48 -66 25
L	TOJ (PHC, fusiform, lingual gyrus)	863	0.000	5.03	36/20/19	-26 -43 -11
L+R	PCC & precuneus	963	0.000	4.86	30	-12 -63 13
				4.15	31	8 -63 18
L	Frontopolar cortex (SFG, med. FG)	588	0.003	3.94	10/9	-12 63 10

\* Cluster level FDR-corrected for the whole brain; voxel-level FWE-corrected for SVC

Abbreviations: ACC = anterior cingulate cortex; AG = angular gyrus; aTL = anterior temporal lobe; dlPFC = dorsolateral prefrontal cortex; FG = frontal gyrus; IFG = inferior frontal gyrus; IPL = inferior parietal lobule; ITG = inferior temporal gyrus; MFG = middle frontal gyrus; MTG = middle temporal gyrus; PCC = posterior cingulate cortex; PHC = parahippocampal cortex; SFG = superior frontal gyrus; SPL = superior parietal lobule; STG = superior temporal gyrus; STS = superior temporal sulcus; SVC = small volume correction; TOJ = temporo-occipital junction; TPJ = temporo-parietal junction; vmPFC = ventromedial prefrontal cortex; H = hemisphere; L = left; R = right; p = p-value; T = T-value; B.A. = Brodmann area; x, y, z = MNI coordinates

**Table A.3 Unspecific Effects of Arousal- and Valence-span in Model 1**

<b>H</b>	<b>Regions</b>	<b>Cluster size</b>	<b>p*</b>	<b>T</b>	<b>B.A.</b>	<b>[x, y, z]</b>
<i>Positive linear Correlation with Arousal-span</i>						
L	IFG pars triangularis, insula & MFG	1034	0.000	6.97	13/45/46	-40 29 10
L	Posterior MTG & ITG	2175	0.000	5.70	37/19	-56 -63 -3
R	IFG pars triangularis and orbitalis	765	0.001	5.70	46/47	47 36 6
L+R	Cingulate, medial FG	786	0.001	5.58	24	-2 -4 49
				4.78	6/32	9 -9 60
L+R	Extrastriate visual cortex (Cuneus & Lingual gyrus)	4209	0.000	5.32	18/19	10 -69 -3
				4.77	30	-2 -73 7
R	Middle STS (MTG & STG)	430	0.013	5.11	21/22/38	51 2 -14
L	IFG pars opercularis & insula	821	0.001	4.97	44/9/13	-52 8 16
R	Precentral & postcentral gyrus	1265	0.000	4.63	4/6/2	44 -15 60
L	IPL & postcentral gyrus	374	0.020	4.50	40/2	-56 -34 43
L	Cuneus, MOG & precuneus	464	0.010	4.43	18/19	-18 -87 24
L	Premotor cortex (precentral gyrus, MFG & IFG)	379	0.020	4.29	6/9	-39 -12 51
L	Premotor cortex (SFG, MFG, med. FG)	329	0.032	4.24	6	-20 8 60
R	IFG pars orbitalis	310	0.038	4.13	47	21 29 -12
R+L	Extrastriate visual cortex (cuneus & precuneus)	659	0.002	4.04	19	16 -91 24
				3.86	19	-6 -97 22
<i>Negative linear Correlation with Arousal-span</i>						
L+R	PCC	1597	0.000	5.60	31/29	-2 -61 24
				5.27	31	9 -57 21
R	Supramarginal gyrus	1089	0.000	4.54	40/39	51 -52 22

***Positive linear Correlation with Valence-span***

L	Amygdala, subcallosal gyrus & lateral globus pallidus	868	0.000	5.35	-24 -7 -17
R	Cerebellum	1296	0.000	4.91	28 -78 -21

\*Cluster level FDR-corrected for the whole brain

Abbreviations: FG – frontal gyrus; IFG – inferior frontal gyrus; IPL – inferior parietal lobule; ITG – inferior temporal gyrus; MFG – middle

frontal gyrus; MOG – middle occipital gyrus; MTG – middle temporal gyrus; PCC – posterior cingulate cortex; STG – superior temporal gyrus;

STS – superior temporal sulcus; H = hemisphere; L = left; R = right; p = p-value; T = T-value; B.A. = Brodmann area; x, y, z = MNI coordinates

**Table A.4 Unspecific Neural Correlates of Subjective Ratings**

<b>H</b>	<b>Regions</b>	<b>Cluster size</b>	<b>p*</b>	<b>T</b>	<b>B.A.</b>	<b>[x, y, z]</b>
<i>Positive linear Correlation with Arousal Ratings</i>						
L+R	Caudate body	654	0.002	5.48		0 6 18
				4.10		-9 -16 22
R	IFG (orbitalis & triangularis) & AI	292	0.039	5.12	47/13	39 33 -8
L	IPL & PI	441	0.011	4.74	40/13	-57 -45 30
L	Posterior MTG	304	0.039	4.05	21/39/37	-60 -54 0
<i>Negative linear Correlation with Arousal Ratings</i>						
L	PCC	4906	0.000	8.20	23/31	-8 -60 16
R	TPJ (AG & IPL)	4592	0.000	7.84	39/40/7	45 -67 31
R	PHC & Cerebellum	946	0.000	7.71	36/28	27 -33 -15
R	dIPFC (SFG & MFG)	2134	0.000	7.08	9/6/8	21 42 45
R+L	ACC & vmPFC (medial FG)	1688	0.000	6.59	11	3 41 -9
				5.31	11	-10 47 -14
L	Precuneus, AG & IPL	2339	0.000	6.22	19/39/7	-40 -78 37
L	TOJ (MOG, IOG & MTG)	725	0.000	5.91	19/39	-48 -82 4
L	dACC & dIPFC (MFG & SFG)	386	0.008	5.51	32/8	-16 20 42
R	aTL (MTG)	673	0.000	5.43	21	54 0 -23
L+R	Cerebellum	368	0.008	5.12		-6 -52 -50
				4.09		6 -49 -47
L	aTL (MTG)	376	0.008	5.07	21	-50 -15 -21
L	Cerebellum & PHC	965	0.000	4.97	28	-22 -30 -24
R	Precentral gyrus	353	0.009	4.94	4/6	36 -28 58
L	Postcentral gyrus	223	0.049	4.94	3/4	-48 -18 46
R	Premotor (SFG & MFG) &	244	0.040	4.53	6/4	15 -19 70

	postcentral gyrus					
R	Precentral, postcentral gyrus & insula	443	0.004	4.21	43	54 -12 9
R	dIPFC (MFG)	240	0.040	3.91	46/9	44 29 24

***Positive linear Correlation with Valence Ratings***

R	aTL (MTG & ITG)	1184	0.000	5.81	21	57 2 -23
R	TPJ (STG & PI)	536	0.007	5.25	40/13/22	54 -51 19
L	PCC	1544	0.000	4.73	29/23/30	0 -46 13
R	ACC & vmPFC (medial FG)	716	0.001	4.40	32/11	3 32 -11

***Positive quadratic Correlation with Valence Ratings***

L	IFG (opercularis), AI & precentral	396	0.026	5.65	9/13/6	-40 3 30
L+R	Precuneus & cuneus	1333	0.000	4.98	31/17	0 -70 15
L	SPL & precuneus	386	0.026	4.77	7	-28 -58 60
L	Precuneus & cuneus	483	0.017	4.28	7/18/19	-14 -75 34
L	Amygdala (SVC)	64	0.010	4.61		-27 0 -21
R	Amygdala (SVC)	31	0.023	4.22		27 -6 -14

\* Cluster level FDR-corrected for the whole brain; voxel-level FWE-corrected for SVC

Abbreviations: ACC = anterior cingulate cortex; AG = angular gyrus; AI = anterior insula; aTL = anterior temporal lobe; dIPFC = dorsolateral prefrontal cortex; FG = frontal gyrus; IFG = inferior frontal gyrus; IOG = inferior occipital gyrus; IPL = inferior parietal lobule; ITG = inferior temporal gyrus; MFG = middle frontal gyrus; MOG = middle occipital gyrus; MTG = middle temporal gyrus; PCC = posterior cingulate cortex; PI = posterior insula; PHC = parahippocampal cortex; SFG = superior frontal gyrus; SPL = superior parietal lobule; STG = superior temporal gyrus; SVC = small volume correction; TOJ = temporo-occipital junction; TPJ = temporo-parietal junction; vmPFC = ventromedial prefrontal cortex; H = hemisphere; L = left; R = right; p = p-value; T = T-value; B.A. = Brodmann area; x, y, z = MNI coordinates