



ELSEVIER

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Poetics

journal homepage: [www.elsevier.com/locate/poetic](http://www.elsevier.com/locate/poetic)

# What the eyes reveal about (reading) poetry

Winfried Menninghaus<sup>a,1</sup>, Sebastian Wallot<sup>b,1,\*</sup>

<sup>a</sup> Max Planck Institute for Empirical Aesthetics, Grüneburgweg 14, 60322 Frankfurt am Main, Germany

<sup>b</sup> Institute of Psychology, Leuphana University of Lüneburg, Lüneburg, Germany

## ARTICLE INFO

### Keywords:

Eye tracking  
Poetry  
Rhyme  
Meter  
Aesthetic evaluation

## ABSTRACT

This study investigated how rhyme and meter affect eye movements and subjective aesthetic evaluations while reading poems. Earlier findings suggest that the effects might include prosodic predictability-driven cognitive and affective rewards from increased processing fluency (Blohm, Wagner, Schlesewsky and Menninghaus, 2018, McGlone and Tofiqbakhsh, 2000), but also semantic and syntactic disfluency, as rhyme and meter are often implemented at the expense of unusual word forms and word order (Menninghaus et al., 2015, Wallot and Menninghaus, 2018). This study set out to investigate the extent to which eye movements might reveal not only distinct effects of fluency and disfluency at the same time, but potentially also hedonic responses that are associated with longer rather than shorter self-motivated exposure, in line with the hypothesis of “savoring” (Frijda and Sundararajan, 2007). The results reveal several fluency-enhancing effects of rhyme and meter on reading times for more global features of the poems, but also increased disfluency effects on gaze durations for some more local features of the poems. Moreover, some of the latter effects are readily interpretable in terms of the savoring hypothesis. Eye movement characteristics that were predictive of higher aesthetic evaluation—irrespective of the presence or absence of rhyme and meter—similarly resulted in increased fluency, disfluency, and savoring effects. Our study thus reveals, for the first time, a complex picture of effects that co-occur while reading poetic prosody, based on analyzing different dimensions of a single psychophysiological variable, namely, eye movements.

## 1. Introduction

Poetic language is a very old form of human language use, likely predating written language (Finnegan, 2017). Prototypically, it differs from non-poetic language through a pronounced focus on patterns of recurrence across multiple levels, i.e., phonology, morphology, syntax, and semantics. These patterns of recurrence have all come to be subsumed under the concept of poetic parallelism (Fabb, 2015; Jakobson, 1960). These patterns can be found far beyond poetry, for instance, in proverbs, advertisements, and even quotidian conversation (Jakobson, 1960), and they have received substantial attention ranging from ancient poetics to recent humanist scholarship. Moreover, parallelistic patterns are linguistic analogues of repetition and variation in music (Lerdahl and Jackendoff; 1977, Margulis, 2014) and also of symmetries in visual aesthetics (Berlyne, 1974; Palmer and Hemenway, 1978). They are hence key factors of aesthetic appeal across domains. Many parallelistic features of poetry are fairly well circumscribed, thus allowing for selective experimental modification for the purpose of testing their distinct effects on language processing.

\* Corresponding author.

E-mail address: [sebastian.wallot@ae.mpg.de](mailto:sebastian.wallot@ae.mpg.de) (S. Wallot).

<sup>1</sup> Both authors contributed equally to preparation of the manuscript and hence share first authorship.

<https://doi.org/10.1016/j.poetic.2020.101526>

Received 5 August 2019; Received in revised form 23 November 2020; Accepted 25 November 2020

Available online 6 January 2021

0304-422X/© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

The present study focused on two preeminent parallelistic features of poetic diction: meter and rhyme. Meter and rhyme are ongoing repetitive structures that recruit our capacities for predictive coding and allow for rapid entrainment (Cutler and Foss, 1977; Essens and Povel, 1985; Obermeier et al., 2016; Obermeier et al., 2013; Rothermich, Schmidt-Kassow, Schwartze and Kotz, 2010; Teng et al., 2020). As rhyme and meter are used in many communication domains, an understanding of how they are processed and what they can do also holds importance beyond poetry.

Experimental studies on the effects of the presence vs. absence of rhyme and meter can replace neither close readings of entire poems nor their interpretation in historical contexts. Rather, such empirical studies pursue complementary desiderata, with an exclusive focus on the aesthetic and emotional effects of the special treatment of language in poetry. Whereas neuroscientific methods have been used in several studies on reading poetry (Obermeier et al., 2016; Thierry et al., 2008; Vaughan-Evans et al., 2016; Wassiliwizky et al., 2017), reading research that uses eye tracking has largely focused on nonliterary prose and artificial texts. Moreover, it has primarily focused on basic components of the reading process, such as word recognition and syntactic parsing (for reviews, see Clifton and Duffy, 2001; Graesser, Millis and Zwaan, 1997; Rayner, 1998).

Experimental studies on proverbs, humoristic verses, and poems have shown that meter and rhyme render proverbs more beautiful, succinct, and persuasive and render poems more beautiful, moving, and joyful, as well as sadder (Menninghaus et al., 2014; Menninghaus et al., 2015; Menninghaus et al., 2017; Obermeier et al., 2016; Obermeier et al., 2013). In addition, parallelistic features enhance verbatim recall and memorability (Hanauer, 1996; Hanauer, 1998; Lea et al., 2008; Menninghaus et al., 2014; Tillmann and Dowling, 2007)

Previous eye-tracking studies on poems have investigated local variations of meter-mediated reader expectations in limericks (Breen and Clifton, 2011) and in haikus; however, the latter do not feature rhyme and ongoing meter (Mueller et al., 2017). Another study has addressed the effects of properties of lines on eye movement characteristics (Roberts, Stabler, Fischer and Otty, 2013), but without specifically targeting or experimentally modifying rhyme and meter.

The present study is the first to investigate the eye movement signature when reading poems with a special focus on the role of meter and rhyme. We presented participants with original poems and modified variants of these from which either rhyme or meter or both had been removed (for details, see the “Stimuli” section). Our research questions were: 1. Does the presence vs. absence of rhyme and meter influence eye movements while reading poetry? 2. Does their presence vs. absence reflect readers’ subjective aesthetic evaluations of poems?

These questions take two basic approaches in empirical aesthetics as their point of departure. On the one hand, specific properties of objects have been systematically investigated for their effects on subjectively experienced aesthetic evaluation (see (Berlyne, 1974)). On the other hand, research on tears, chills, and goosebumps in response to artworks has shown that not only do physiological responses provide important additional insight into the nature of subjective aesthetic evaluation (Blood and Zatorre, 2001; Panksepp, 1995; Salimpoor et al., 2009); they can also guide an analysis of particularly relevant textual features of poems that prime these responses (Wassiliwizky et al., 2017).

## 2. General mechanisms and indicators of aesthetic evaluation and reward

Rhyme and meter enhance the ease of perceptual/prosodic processing by making upcoming words, and specifically the closure of lines and stanzas, more predictable (McGlone and Tofighbakhsh, 2000; Menninghaus et al., 2015). This increased processing fluency is reflected in higher reading speeds (Lea et al., 2008) and distinct EEG signatures (lower deflections for rhymed and metered poems) in the N400 and P600 components (Obermeier et al., 2016). Increased prosodic fluency is often experienced as inherently rewarding; as with other types of cognitive fluency, it tends to enhance aesthetic liking (Menninghaus et al., 2017; Reber, Schwarz and Winkielman, 2004). It might even lead readers to find the respective sentences more truthful and persuasive (McGlone and Tofighbakhsh, 2000; Menninghaus et al., 2015).

However, meter and rhyme are also associated with increased difficulty of semantic/conceptual and/or syntactic comprehension and, consequently, decreased reading speed, as they tend to impose significant constraints on the most natural choice and order of words (Wallot and Menninghaus, 2018). Existing data suggest that the fluency effects of parallelistic patterning are often not cancelled out by the concomitant disfluency effects, at least in the case of prototypical poems of the rhymed and metered tradition (Menninghaus et al., 2017; Obermeier et al., 2016) and proverbs (Menninghaus et al., 2015). The fact that many successful commercial ads are both highly parallelistic in prosody and highly elliptic in syntax and/or semantics suggests that the same holds for them. Accordingly, Wallot and Menninghaus (Wallot and Menninghaus, 2018) have proposed a two-factor model that entails provisions for both the fluency and the disfluency effects of poetic language.

Finally, the reception of artworks and other objects of aesthetic appeal has dimensions other than fluency or disfluency of cognitive processing. In particular, higher aesthetic liking has a motivational effect not only in terms of seeking repeated exposure to a preferred stimulus, but also in terms of extending the exposure time where possible (Menninghaus et al., 2019; Mitschke, Goller and Leder, 2017; Tschacher et al., 2012). Frijda and Sundararajan (2007) have called the latter phenomenon a “savoring” effect. The design of our study allowed us to test whether the self-paced reading of poems also shows such liking-dependent savoring effects.

For a single dimension of poetry processing, the hypotheses of fluency vs. disfluency are mutually exclusive, that is, longer reading times can only be interpreted as increased disfluency and shorter reading times only as increased fluency. The hypothesis of combined fluency and disfluency renders the two otherwise antagonistic hypotheses compatible by identifying increased fluency and increased disfluency with respect to different processing dimensions. Thus, higher prosodic fluency can co-occur with higher disfluency in semantic or syntactic processing.

However, the inclusion of potential savoring effects in our study gives rise to conflicting interpretations: Longer reading times can

be indicative of either reduced aesthetic reward qua disfluency or increased aesthetic reward qua savoring. In this case, only the larger picture of the respective results can lead to a conclusion about which interpretation is more likely to be correct for the observed eye-movement parameter.

### 3. Eye-movement parameters that are potentially distinctive of cognitive fluency/disfluency and/or savoring while reading poems

Studies on textual features other than rhyme and meter have identified various parameters that influence reading fluency. For example, low-frequency words (Schroeder, 2011; Wallot, Hollis and van Rooij, 2013), complex grammatical constructions (Frazier and Rayner, 1982; Meseguer, Carreiras and Clifton, 2002), and unexpected turns in a narrative (Blanchard and Iran-Nejad, 1987; Booth et al., 2016; Hoeken and van Vliet, 2000) decrease situational reading fluency. These features can also be mapped onto specific eye movement characteristics: Low-frequency words elicit longer durations of first fixations, whereas syntactic and semantic violations can lead to an increase in regressive eye movements (Clifton, Staub and Rayner, 2007; Kliegl, Grabner, Rolfs and Engbert, 2004). Moreover, low-frequency words elicit longer fixation durations (Rayner and Duffy, 1986), and garden-path sentences elicit regressive eye movements (Meseguer, Carreiras and Clifton, 2002).

For rhyme- and meter-enhanced ease of processing (*fluency*), we expected shorter fixation durations, fewer regressive eye movements, and faster overall or localized reading times. Regarding the hypothesis of combined fluency and disfluency, two outcomes could be expected: First, we might see a combination of effects in which some eye movement features show increased fluency (e.g., shorter fixation durations), whereas others show decreased fluency (e.g., higher numbers of regressive eye movements). In such cases, the antagonistic effects can well coexist. However, when affecting the same measure, increases in (prosodic) fluency and (semantic) disfluency might also cancel each other out.

As to a potential *savoring* effect of rhyme and meter, this effect would by definition require longer reading times for one or several of our measures, making it difficult to distinguish between potential disfluency and savoring effects. (For a list of all eye movement characteristics tested in the study, see the “Data Processing and Reduction” section).

### 4. Eye movement characteristics that are potentially predictive of subjective aesthetic evaluation

Regarding our second research question about the relation between eye movement characteristics and subjective aesthetic evaluations, our study cannot rely on previous results since it is the first to investigate this issue. Still, our study is analogous to research aimed at predicting text comprehension from reading process data, using process variables as predictors and performance on comprehension questionnaires as outcome variables (LeVasseur, Macaruso and Shankweiler, 2008; Schroeder, 2011; Wallot, O'Brien, Coey and Kely-Stephen, 2015; Wallot et al., 2014).

In our study, we analyzed the eye movement characteristics measured during the reading of poetry as potential predictors of subjective aesthetic evaluation—specifically, predictors of ratings for how *beautiful*, *moving*, and *melodious* the poems were perceived to be. These rating items have been shown to capture a significant part of the emotional and aesthetic effects of poetry (Knoop, Wagner, Jacobsen and Menninghaus, 2016; Menninghaus et al., 2017). The cognitive fluency hypothesis suggests a positive correlation between high ratings on these scales and eye-movement features that are indicative of fluent reading. The hypothesis of combined fluency and disfluency suggests that some eye-movement features (e.g., few regressive eye movements) might be of a facilitating nature and enhance the behavioral ratings, whereas others (e.g., longer fixation durations) might indicate a handicapping effect and have an adverse effect on the aesthetically evaluative ratings. The hypothesis of savoring suggests that longer reading times are associated with higher aesthetic evaluation.

We investigated our research questions with a focus on both the overall eye-movement patterns observed throughout the entire duration of reading the poem variants and on specific regions of interest within the poems. For the latter, we specifically focused on two features. Guided by the great importance that poetics attributes to the closure (clausula, cadence) of both entire texts and important sub-units (Smith, 1968), we measured the duration of gazes on the final words of each verse, expecting that the words that close rhymes and that conclude the poems should show particularly pronounced closure effects. We also measured the reading times for the individual lines of the poems, as these might be indicative of a temporal dynamics of potential fluency, disfluency, or savoring effects.

## 5. Method

### 5.1. Apparatus

Participants' eye movements were recorded using an EyeLink 1000 eye tracker (SR Research Ltd 2009) with a sampling rate of 1000 Hz and a maximum spatial resolution of 0.01° visual angle. A computer screen placed directly above the eye tracker (24 in., 1920 × 1200 pixels) displayed the poems. During data collection, participants' heads were stabilized with a head rest, situating participants' eyes at a distance of approximately 65 cm from the screen. The height of the chair remained at 55 cm above floor level. Eye tracking was performed by tracking the participants' right eyes (monocular eye tracking). However, due to an experimenter mistake, the left eyes of 10 participants (out of 40) were also tracked.

**Table 1**  
Illustration of the Poem Modifications Using the First Stanza of William Blake's "Ah Sun-flower!"

Variant	Example	Glossary
A Original	Ah Sun-flower! weary of time. Who countest the steps of the Sun: Seeking after that sweet golden clime Where the travellers journey is done.	Four verse lines, anapaestic trimeter (with verse-initial variation), cross-rhymed (ABAB)
B Meter, no rhyme	Ah Sun-flower! weary of time. Who countest the steps of the Sun: Seeking after that sweet golden place Where the travellers journey is through.	Rhyme removed by replacing the verse-final words in lines 3 and 4
C Rhyme, no meter	Ah Sun-flower! weary of the time. Who countest all the steps of the Sun: You are seeking after that sweet golden clime Where the journey of the traveller is done.	Meter removed by adding syllables to break up the pattern
D No rhyme, no meter	Ah Sun-flower! weary of the time. Who countest all the steps of the Sun: You are seeking after that sweet golden place Where the journey of the traveller is through.	Integration of modifications B and C

Note. The glossary column provides more detailed explanations of the specific modifications.

## 5.2. Participants

All procedures employed in the present study were approved by the Ethics Council of the Max Planck Society, and all participants provided written informed consent prior to data recording. In total, 44 native German speakers were tested as paid participants at the Max Planck Institute for Empirical Aesthetics. None of them had a diagnosed weakness in reading or writing. All had normal or corrected-to-normal vision. The data for four participants were excluded due to technical problems with the eye tracker or problems with the experimental software. The remaining 40 participants (32 females), aged between 19 and 40 years (mean age = 24.83 years,  $SD = 4.63$ ), were included in the analysis.

## 5.3. Materials

The original poems and modified variants of these poems were adopted from a larger corpus of 40 poems (Menninghaus et al., 2017). From this corpus, we selected 10 poems for which the reference study had reported significant differences in aesthetic evaluations of the rhymed and metered original poems compared to modified variants that featured neither rhyme nor meter. Moreover, all variant of the selected poems had the same number of lines, regardless of differences in their rhyme pattern, and were similar in length (i.e., the number of words). We used these latter criteria in order to make the eye-movement characteristics as comparable as possible in terms of the average number of return sweeps and the resulting overshooting (Hofmeister, Heller and Radach, 1999). All 10 poems were rated as both highly emotionally moving and highly beautiful. In addition, five of the 10 poems were rated as high in sadness and low in joy, whereas the reverse applied to the other five poems (Menninghaus et al., 2017).

The corpus includes unfamiliar poems by famous authors as well as poems by relatively unknown authors from the later 18th century through the mid-20th century. However, none of the poems contain historical wording that has become unfamiliar or that was used with a meaning that is no longer accessible, nor do they contain wording that was designed to be unusually difficult to understand. Beyond ongoing meter and rhyme, the poems also feature additional parallelistic features, such as local assonances and alliterations. These features of nonsystematic parallelism were kept constant across the experimental variant used in the present study.

Poems can have a great number of rhyme schemes and types of meter. Our study was not aimed at revealing subtle differences between distinct types of rhyme or meter. Rather, in the absence of any previous knowledge about the potential eye-tracking signatures of rhyme and meter, we drew on a mix of trochaic and iambic poems (which are, in this order, the two most widely used metrical patterns in German poetry). Besides the predominant ABAB rhyme scheme, the 10 poems also included four other rhyme schemes and hence substantial variation. The reference study showed that the presence vs. absence of all of these variants of rhyme and meter yielded convergent effects on behavioral ratings for emotional affection and aesthetic evaluation. This led us to conclude that the subtle differences between the subtypes of meter and rhyme were less potent than the categorical difference between the presence vs. absence of meter and rhyme altogether.

In order to dissolve the higher-order gestalt effect of rhyme, one of the two words of each rhyme pair needed to be replaced by a synonym. For the purpose of also changing the sustained metrical patterning of an entire quatrain into more ordinary prosody, at least two further words had to be replaced by synonyms, unless the same effect could be arrived at by merely changing the order of words.

On top of the intended modification of the prosody, some minor differences in meaning could also result from these changes. However, these were likely to be rather unsystematic and only of local relevance, whereas the modification of rhyme and meter was systematic and sustained. Across a variety of poems and a greater number of readers, it is therefore likely that the obtained effects would indeed mostly be driven by the intentionally implemented differences in the formal target features. We also performed several tests designed to determine differences in the verbal material between the original poems and their three modified variants.

**Table 2**  
Statistical stimulus characteristics.

Poem modification	Poem length (#words)	Word length (#letters)	Word freq. (log)
Original (R+M+)	78.1 (13.6)	4.65 (0.38)	2.29 (0.24)
Rhyme only (R+M-)	80.9 (13.3)	4.67 (0.35)	2.38 (0.25)
Meter only (R-M+)	77.7 (13.7)	4.68 (0.36)	2.26 (0.21)
None (R-M-)	77.8 (13.4)	4.70 (0.29)	2.39 (0.20)

*Note.* The table reports average text length, word length, and word frequency for the original poems and for each of their three modifications. Standard deviations are presented in parentheses.

The five joyful poems were: August Graf von Platen, “Laß tief in dir mich lesen,” 1819; Ernst Moritz Arndt, “Frühling im Alter,” 1841; Hans-Eckardt Wenzel, “August,” 1998; Johanna Ambrosius, “Frei,” 1894; and Johannes R. Becher, “Das lachende Herz,” 1949. The five sad poems were: Joseph von Eichendorff, “Wehmut,” 1807; Klaus Johann Groth, “Heimweh III,” 1893; Friedrich Hebbel, “Nachtgefühl,” 1836; Heinrich Heine, “Wasserfahrt,” 1817; and Rainer Maria Rilke, “Abschied,” 1907.

To illustrate the manipulations, in [Table 1](#) we present an English analogue to our four variants based on the poem “Ah Sunflower!” by William Blake.

In the selective elimination of rhyme and meter, great care was taken to replace as few of the original words as possible and, where words were replaced, to keep the semantic content of the modified poems as identical as possible (see [\(Menninghaus et al., 2017\)](#)).

To evaluate the similarity between the original poems and their modifications, we conducted a Bayesian repeated measures ANOVA (using a relatively uninformative prior, setting the  $r$  scale to 0.5 for the Cauchy prior distribution for fixed effects) on the variables *poem length*, *word length*, and *word frequency* using JASP (version 0.8.6; [\(JASP Team 2018\)](#)). We chose a Bayesian repeated measures ANOVA because this approach provides a proper test of the null hypothesis that our poem modifications do not differ from the original poems with regard to poem length, word length, or word frequency. Instead of a  $p$ -value, this analysis provides a Bayes factor value ( $BF_{01}$ ) indicating how likely the null hypothesis (i.e., no difference between the poems) is relative to the alternative hypothesis (i.e., the poems differ in terms of poem length, word length, and word frequency; see [\(Marsman and Wagenmakers, 2017\)](#), for an introduction to JASP). The descriptive statistics for the different poem variants are reported in [Table 2](#).

The resulting Bayes factors were:  $BF_{01} = 5.00$  for poem length,  $BF_{01} = 5.07$  for word length, and  $BF_{01} = 0.01$  for word frequency. Because the Bayes factors for poem length and word length report the likelihood of the null hypothesis over the alternative hypothesis (01), and because their values are greater than one for poem length and word length, we can conclude that we have a moderate degree of evidence that all poem modifications are similar with regard to these features. Specifically, the data obtained are five times more likely under the null hypothesis compared to the alternative hypothesis. However, the low  $BF_{01}$  value for word frequency suggests that the poems differ systematically in terms of word frequency (to be exact, the corresponding Bayes factor suggests that the observed word frequencies are 214 times more likely under the alternative hypothesis compared to the null hypothesis). We therefore decided to include word frequency as a random effect in the analysis of the potential effects of the rhyme and meter manipulation on aesthetic evaluation.

#### 5.4. Design

All original and modified poems were presented to all participants. Experimental research on ease of processing has often resorted to within-participant designs, as interindividual differences can easily override the effects obtainable by the subtle experimental modifications that are typically employed in this type of research (for a discussion of this issue, see [\(Forster, Leder and Ansoorge, 2013\)](#)). However, as is well known, a within-participant design carries a danger of unwanted downsides. Reading four different variants of a poem is likely to enhance familiarity across the four encounters, and familiarity is known to increase aesthetic liking even independently of any differences in textual features ([\(Bohrn et al., 2012; Hideyuki and Menninghaus, 2018\)](#)). As a result, the stepwise increases in familiarity are prone to confound the effects driven by differences in the target textual features.

In order to control for this potential confound, we decided to present the original poems as well as their three modifications to half of the participants in ascending order and to the other half in descending order regarding the presence of the target variables *meter* and *rhyme*. Our underlying reasoning was as follows: In the ascending presentation order, the increases in familiarity with the poems throughout the trials might push the aesthetic evaluation driven by the textual variables *rhyme* and *meter* to even higher levels. In the descending presentation order, in contrast, the stepwise increased familiarity might still equally enhance aesthetic evaluation, but the elimination of meter and rhyme should simultaneously reduce the evaluation and hence affect overall aesthetic evaluation in the opposite direction.

This asymmetry has a methodically interesting implication in our specific case: if familiarity does play an important role, it should prevent the two presentation orders from yielding a by-and-large inverse effect pattern. That is, only in the case of the ascending order, which starts with the poem variants from which both meter and rhyme were removed and which stepwise proceeds towards the original poem featuring both meter and rhyme, should the effects of increased familiarity and increases in parallelistic features (i.e., rhyme and meter) be additive. In the descending order, by contrast, familiarity-driven fluency should equally increase step by step, but the aesthetic evaluation effects driven by rhyme and meter should decrease. Here, the two effects go in opposite directions and might even cancel each other out. Hence, if familiarity were to affect the results, the overall effects of the textual variants on aesthetic evaluation ratings should *not* be mirror symmetrical. Inversely, if we were still to find a roughly mirror-symmetrical pattern for the two

presentation orders, then familiarity would likely not be significantly affecting the results.

A second critical issue regarding our choice of design is the principle of contrast, which predicts asymmetrical effects of the presentation order (Fechner, 1876; Parker, Bascom, Rabinovitz and Zellner, 2008; Thoussignant and Bodner, 2014): If a more appealing stimulus precedes a less appealing one, the less appealing stimulus will fare worse than in the absence of the preceding contrast. Inversely, if a less appealing stimulus precedes a more appealing one, the more appealing stimulus profits from this comparison *ex negativo*. The study from which we adopted the poems (Menninghaus et al., 2017) reported such effects. However, that study compared only two variants that were even further apart from one another than the original poem and the variant with neither rhyme nor meter in the present study: not only were meter and rhyme removed from the modified poems used in the previous study, but also alliterations, assonances, and other features of phonological parallelism.

By contrast, our stepwise modifications, along with the continually descending or ascending presentation order, are far more subtle and hence far less likely to elicit marked contrast effects between subsequent variants of a poem. Across the two presentation orders, each variant of all poems presented was read and rated exactly as often before and after a more and a less parallelistic variant (and hence a potentially more or less aesthetically liked variants). Therefore, we expected the two inverse contrast effects to balance each other out in our design across the two groups of participants. In any event, if we were to obtain a roughly mirror-symmetrical effect of the experimental modifications for both presentation orders, this would speak in favor of the absence of any significant effect of asymmetrical contrasts.

Regardless of these considerations concerning the design, the two reading conditions (i.e., ascending vs. descending presentation order) were in any event controlled for using linear-mixed-effects regression models (see the “Data Analysis” section).

In sum, we employed a mixed  $4 \times 2$  design, with the within-participant factor poem modification (4 levels: A [original poem], B [rhyme only], C [meter only], or D [neither rhyme nor meter]) and the between-participants factor presentation order (2 levels: increasing vs. decreasing number of parallelistic features).

### 5.5. Procedure

First, participants were provided with a written introduction to the study and given the opportunity to ask questions. After all questions were answered, participants signed the written consent form. Then they were asked to sit comfortably on a chair in a sound-proof cabin. The chair was placed directly in front of a chin rest on which participants placed their chin while leaning their head against a forehead rest. Depending on the participants' height, the height of the table was adjusted so that participants could see the center of the screen directly ahead. A nine-dot calibration and validation followed, and written instructions for the study were presented.

All stimuli were presented to the participants using MatLab 2015b (Mathworks 2015), with a custom PTB-MatLab script (Brainard, 1997). The study started with a familiarization trial. First, a screen with the title of a test poem and the year of its publication was displayed. This was followed by a blank screen with a cross that was presented for 400 ms in the upper left corner, marking the optimal viewing position for the first word of the poem (see (McConkie et al., 1989)). Subsequently, the first variant of the test poem was presented. Participants pressed a button to indicate that they had finished reading the poem and were ready for the next step. Then they were asked to rate the poem, using the numeric keypad, on several scales ranging from 1 (*not at all*) to 7 (*very much*); each scale was presented on a separate screen. After all questions were answered, the second and remaining variants of the test poem were presented in an identical fashion.

The experimental trials were presented in exactly the same way as the test trials. Between exposure to the different variants of a poem, readers had the opportunity to take a short break. After the participants had read and rated half of the poems, they were offered the opportunity to take a longer break if they wanted. Before starting with the second half of the stimuli, the eye tracker was recalibrated. After reading all poems, participants were asked to answer some basic demographic questions as well as questions regarding their reading habits and preferences. They were also asked to fill out the Multidimensional Aesthetic Responsiveness Scale (AREa) questionnaire (Schlotz et al., 2020), which measures individual differences in responsiveness to aesthetically appealing stimuli. Due to differences in individual reading speeds and the length of the pauses taken, the sessions took between 50 and 90 min.

### 5.6. Data availability

The data are available online from Figshare: Doi: [10.6084/m9.figshare.13387475](https://doi.org/10.6084/m9.figshare.13387475)

### 5.7. Data preprocessing and reduction

Trials for which more than 10% of the data were improperly recorded (due to blinks, movements, and other sources that led to loss of the corneal reflection during the reading of a poem) were discarded from the analysis; this led to 3.8% of the trials being discarded. Individual missing data points within a trial in the remaining 96.2% of the trials were removed from the record before the extraction of fixations and saccades. Subsequently, fixations and saccades were extracted from the raw position data using Engbert and Mergenthaler's (Engbert and Mergenthaler, 2006) saccade detection algorithm.

For cognitive processing, we analyzed fixation durations, the number of fixations, saccade lengths, regressive eye movements, and total reading times. All of these parameters have been shown to be informative about how fluently or disfluently readers comprehend single sentences as well as short texts similar in length to our poems (Rayner, Chace, Slattery and Ashby, 2006). However, rather than using average fixation durations (as in (Rayner, Chace, Slattery and Ashby, 2006)), we calculated first fixation durations (i.e., the fixation duration for a word when fixated-upon for the first time) and total gaze duration (i.e., the sum of fixation durations for each

**Table 3**  
Correlations between the three ratings for aesthetic evaluation.

	Beautiful	Moving	Melodious
Beautiful	–	0.71	0.63
Moving		–	0.54
Melodious			–

fixated-upon word), because these two measures, albeit related, show differential sensitivity to lexical characteristics, probably reflecting differences in processing depth (Kliegl, Grabner, Rolfs and Engbert, 2004).

In order to capture affective processing, which is of great importance for the aesthetic perception of poetry (Menninghaus et al., 2017; Obermeier et al., 2013), we added measures of pupil dilations and blink rates to our set of eye-movement characteristics. Pupil dilation has been related to cognitive load (Staners, Coulter, Sweet and Murphy, 1979), but also to the arousing quality and affective valence of visual stimuli (Granholm and Steinhauer, 2004). Violations of rhyme have been reported to lead to a delayed increase in subjects' pupil dilation when listening to limericks (Scheepers, Mohr, Fischer and Roberts, 2013).

Before the average pupil dilation was calculated, distortions of the pupil size as a function of position were corrected (Hayes and Petrov, 2016). Pupil dilation time series were low-pass filtered using a third-order Butterworth filter (0.05–4.0 Hz; (de Gee, Knapen and Donner, 2014)). Blink rate has been associated with sustained attention during cognitively demanding episodes (Siegle, Ichikawa and Steinhauer, 2008). Insofar as rhyme and meter might influence ease of processing (Menninghaus et al., 2015), the blink rate might be informative about facilitating or handicapping effects of rhyme and meter during poetry reading.

In order to capture the effects of rhyme and meter on eye movements, we defined two regions of interest in our texts. The first was the *final word of each verse*. Verse-final words are of interest, because they are, at least in the original poem, the rhyme words and hence particularly susceptible to the rhyme manipulation. To capture the effects of rhyme and meter on eye movements related to the final word of each verse, we defined the surface coordinates of each of these words on the screen, and summed all fixations falling into these regions to yield a measure of gaze duration for each participant and poem variant. The resulting total gaze durations were divided by the word length. (As meter is a feature of syllabic units, eye movements do not allow locally capturing the processing of meter.)

The *individual lines of the poems* were a second region of interest, because the sequence of the total reading times for these lines might capture a differential effect of the presence vs. absence of meter: the more both the rhyme pattern and the metrical scheme are repeated throughout the reading trajectory, the stronger the effect could become. Gaze durations for the individual lines were analyzed analogously to the way the verse-final words only were analyzed (see the preceding paragraph).

Finally, we factorized subjective ratings that capture dimensions of aesthetic evaluation along with emotional responses after examining their correlation matrices (Table 3). In the experiment, we asked the participants to rate the poems on scales regarding how *beautiful*, *emotionally moving*, and *melodious* they perceived them to be (for the particular salience of these three items for poetry, see (Knoop, Wagner, Jacobsen and Menninghaus, 2016) and (Menninghaus et al., 2017)).

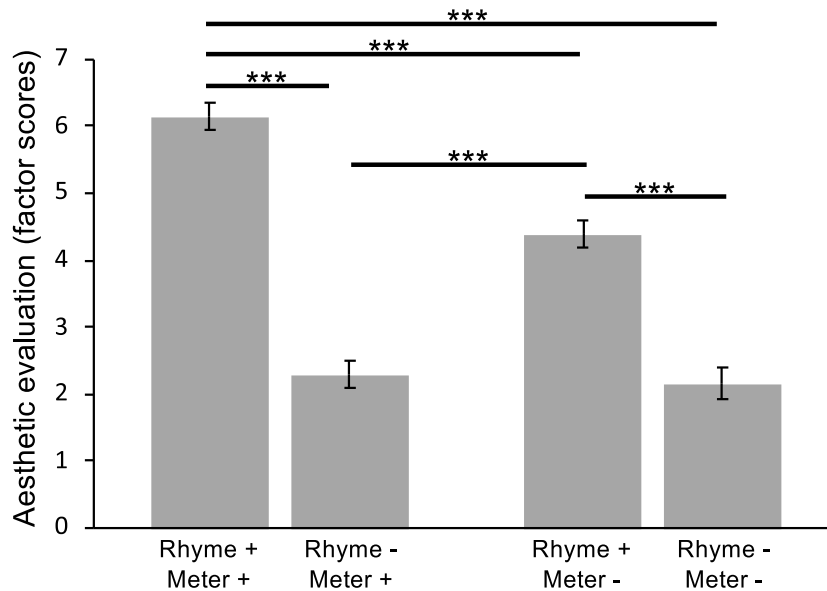
We subjected the data for these three ratings to an exploratory maximum likelihood (ML) factor analysis. Our examination of the scree plot, as well as a parallel analysis (Horn, 1965), suggested extracting a single factor that captured 75.33% of the covariance between the three items. This is reasonable, as ratings for being moved reflect not only a content-driven emotional response, but also tend to be directly predictive of ratings for beauty and other dimensions of aesthetic evaluation in art reception contexts (Hanich et al., 2014; Menninghaus et al., 2020; Menninghaus et al., 2017; Wassiliwizky et al., 2017). Consequently, we named the entire factor the “aesthetic evaluation” factor.

In addition to the three rating items constituting the aesthetic evaluation factor that were of prime interest for our research question, we also collected subjective ratings for *joyful*, *sad*, *negative affect*, *positive affect*, and *comprehensibility*. These variables were not further analysed for the purposes of the present study.

### 5.8. Data analysis

We performed the data analysis with R 3.4.0 (R Core Team 2017). We used the psych package to conduct a factor analysis of the three ratings (*beautiful*, *moving*, *melodious*), the lavaan package for path analysis, and the lme4 package for inferential statistics (linear mixed modelling). All linear mixed models contained crossed participant and poem-specific intercepts as random factors (Baayen, Davidson and Bates, 2008). We chose only random intercepts because inclusion of random slopes led to problems with model convergence due to sample size restrictions.

Additionally, all linear mixed models predicting aesthetic evaluation contained the values of the AREA total scores as a person-specific random covariate capturing sensitivity to aesthetic experiences (Schlotz et al., 2020). Linear mixed models relating the poem manipulation to aesthetic experience contained random intercepts for word frequency (average of logarithmized word frequency for each poem), as this variable was found to differ significantly between the original poem and its modifications. After the model estimation, we computed all pairwise within-participant contrasts using Tukey's HSD test.



**Fig. 1.** Effects of poem modification on subjective aesthetic evaluation. The figure displays the observed means bracketed by standard errors of the mean across participants. Dark lines and asterisks indicate within-participant differences between poem modifications. See Table S2 in the Supplemental Material for model parameter estimates. \*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

## 6. Results and discussion

### 6.1. Effects on subjective aesthetic evaluation

Our analysis of aesthetic evaluation (i.e., the factor scores derived from participants' ratings of how beautiful, moving, and melodious they found the original poems and their modifications to be) revealed a pronounced effect of the textual target variables: the rhymed and metered original poems were clearly rated highest. Overall, ratings for the other variants show a decreasing pattern. However, the differences were not significant for the ratings of variants featuring neither rhyme nor meter compared to the variants featuring no rhyme but meter (see Fig. 1). Ratings for the latter variants may have failed to profit from the presence of meter because of their near-total lack of familiarity and prototypicality. German poetry has many free-verse poems that feature neither rhyme nor meter; however, poems featuring meter but not rhyme are extremely rare outliers. (18<sup>th</sup>-century odes written in strophic meters adopted from ancient Greek and Latin poetry are the only such cases, and these are typically completely unknown to nonexperts.)

Importantly, the results for the two reading conditions (ascending vs. descending order; presence vs. absence of rhyme and meter) showed the same pattern. As explained in the "Methods" section, this suggests that the potential effects of stepwise increased familiarity as well as the asymmetrical contrasts of the two presentation orders were marginal relative to the effects of poem modification. (Details for the two reading conditions are reported in the Supplemental Material, Tables S1 and S2.)

### 6.2. Effects of meter and/or rhyme on overall eye-movement characteristics

All potentially relevant overall eye-movement characteristics were subjected to the same model as the scores from the aesthetic evaluation factor. However, none of these characteristics showed significant differences as a function of poem modification. (See Tables S3–S10 in the Supplemental Material for model parameter estimates.)

### 6.3. Effects of meter and/or rhyme on gaze durations for three types of verse-final words

We analyzed potential effects of meter and/or rhyme on gaze durations for individual words, the predictors being (1) all verse-final words and, more specifically, (2) the verse-final words that open a rhyme pair vs. those that close it, (3) as well as the respective final words of the entire poems. To be sure, in rhymed and metered poems, poem-final words also close a rhyme pair. However, they have the additional function of *concluding the entire poem* and therefore are of particular importance in the theory of *poetic closure* (Smith, 1968). For this reason, we also treated the poem-final words as a separate category.

All analyses of these three types of verse-final words included the predictors poem variant (coding for the presence vs. absence of rhyme and meter), the beginning and closing word of a rhyme pair (in a stanza with an ABAB rhyme scheme, the last words of the first and second verses open a rhyme, and the last words of the third and fourth verses, respectively, close it), as well as the very last word of each poem, which might capture known wrap-up effects in reading or, alternatively, effects of savoring. Moreover, we added interaction terms between the poem variant and beginning/closing rhymes, as well as between the poem variant and the last word, because



**Table 4**

Effects of poem modification, rhyme, and poemfinalword on reading times for verse-final words.

Predictor	B	SE	t	p
(Intercept)	0.065	0.004	15.29	< .001
R+M-	-0.002	0.002	-1.29	.196
R-M+	-0.001	0.002	-0.66	.509
R-M-	-0.004	0.002	-2.42	.015
BeginCloseRhyme	-0.004	0.002	-2.10	.036
PoemFinalWord	0.031	0.003	8.92	< .001
R+M-:BeginCloseRhyme	0.003	0.003	0.99	.334
R-M+:BeginCloseRhyme	0.003	0.003	1.04	.299
R-M-:BeginCloseRhyme	0.008	0.003	2.98	.003
R+M-:PoemFinalWord	-0.016	0.005	-3.21	.001
R-M+:PoemFinalWord	-0.020	0.005	-4.08	< .001
R-M-:PoemFinalWord	-0.016	0.005	-3.21	.001

Note. The predictors R and M code for the presence (+) or absence (-) of the modified features, namely, rhyme (R) and meter (M).

**Table 5**

Effects of poem modification, verse number, and stanza-final verse on total reading times for verses.

Predictor	B	SE	t	p
(Intercept)	0.054	0.002	24.77	< .001
R+M-	-0.005	0.001	-4.45	< .001
R-M+	-0.002	0.001	-1.75	.080
R-M-	-0.003	0.001	-2.76	.006
VerseNr	-0.000	0.000	-4.15	< .001
StanzaFinalVerse	-0.028	0.002	-13.42	< .001
R+M-:VerseNr	0.000	0.000	0.08	.932
R-M+:VerseNr	-0.000	0.000	-1.98	.047
R-M-:VerseNr	-0.000	0.000	-0.62	.534
R+M-:StanzaFinalVerse	0.014	0.003	4.61	< .001
R-M+:StanzaFinalVerse	0.004	0.003	1.46	.145
R-M-:StanzaFinalVerse	0.010	0.003	3.31	< .001
VerseNr:StanzaFinalVerse e	0.003	0.000	12.51	< .001
R+M-:VerseNr:StanzaFinalVerse	-0.001	0.000	-4.77	< .001
R-M+:VerseNr:StanzaFinalVerse	-0.000	0.000	-1.49	.137
R-M-:VerseNr:StanzaFinalVerse	-0.001	0.000	-3.14	.002

Note. The predictors R and M code for the presence (+) or absence (-) of the modified features, namely, rhyme (R) and meter (M). The predictor VerseNr was entered as continuous variable into the model. The predictor StanzaFinalVerse was entered as a discrete variable into the model.

we were specifically interested in how the presence of rhyme and meter influences these effects. The results are summarized in Table 4.

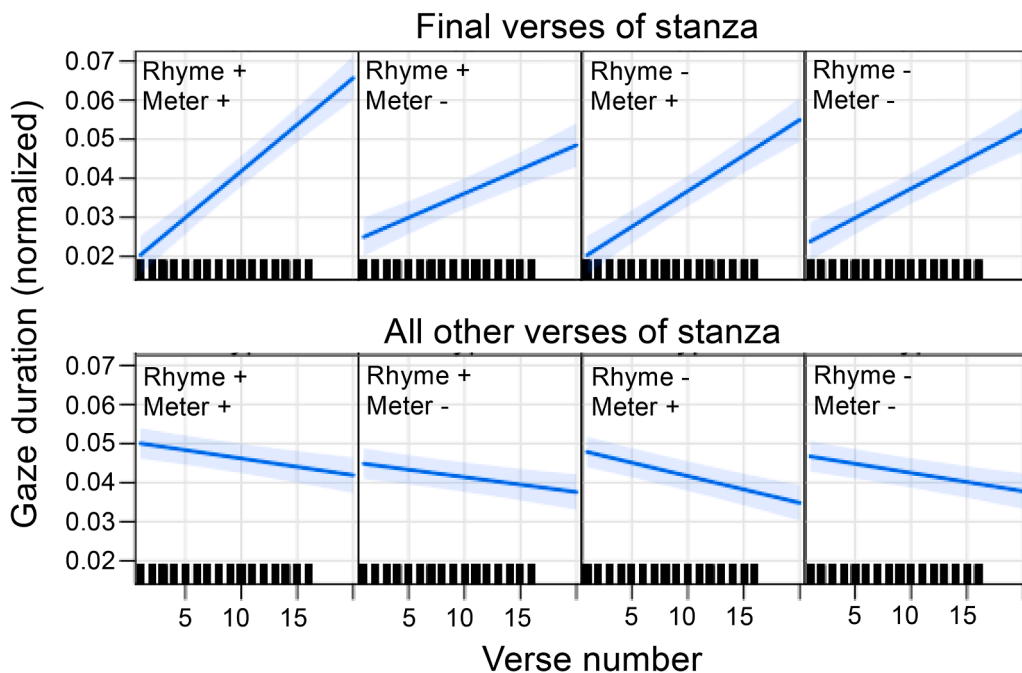
Across all verse-final words, the total gaze durations were longer for the three poem variants featuring rhyme only, rhyme plus meter, or meter only compared to those featuring neither rhyme nor meter (R-M-,  $t = -2.42$ ,  $p = .015$ , see Table 4). Moreover, the original poems featuring rhyme and meter show the most pronounced closure effect: gaze durations are longer for the poem-final words compared to the preceding verse-final words (for interaction effects of the presence of rhyme and meter with LastWord, see Table 4). This effect is by no means at odds with the shorter reading times, and hence the fluency effect, observed for entire verses (with the exception of the stanza-final verses). Rather, it can readily be interpreted as a local disfluency, or savoring effect within a global fluency effect of rhyme and meter on reading times.

Words that close a rhyme pair are read faster than words that open this pair (BeginCloseRhyme,  $t = -2.10$ ,  $p = .036$ , see Table 4). Because readers can have a fairly precise phonological anticipation only with regard to the second word of a rhyme pair, this result is likewise fully in accord with the assumption of increased processing ease. The additional presence of meter enhances this effect, as in the nonrhyming, nonmetered poem variants, verse-final words in the position of the words that close rhymes in the original poem variants showed no facilitating effect (R-M-,  $t = 2.98$ ,  $p = .002$ , see Table 4).

#### 6.4. Effects of meter and/or rhyme on total reading times for each verse

In addition to the overall analysis of eye movement characteristics, we performed analyses regarding particular regions of interest, specifically, total gaze durations for the final words of each verse and total reading durations for each verse of a poem.

The analysis of the final words of each verse yielded effects of the predictors (1) poem variant, (2) verse number, and (3) final verse of each stanza vs. the other verses within a stanza for the reading times of each verse. The effects are moderated by a three-way-interaction between these predictors (see Table 5). In general, reading times decreased as the stanzas proceeded from verse to verse (VerseNr,  $t = -4.15$ ,  $p < .001$ , see Table 5). However, the final lines show a reverse effect: here, reading times increased compared to the preceding lines (see Fig. 2 and the three-way-interactions between the presence of rhyme and meter, VerseNr, and StanzaFinalVerse in Table 5). Both of these effects are most pronounced in the poem variants featuring meter plus rhyme or meter only.



**Fig. 2.** Interaction effect of verse number and stanza-final verses on gaze durations for verses. As can be seen, overall verse gaze duration increases with the verse number if they are the final verses of a stanza, while gaze duration tends to decrease with the verse number if they are not the final verses of a stanza. This effect is particularly pronounced for the original poems, as well as for poem variants with meter but not rhyme. Note that gaze durations were normalized by the length of a verse to avoid confounds due to mere spatial length of a verse.

**Table 6**  
Prediction of aesthetic evaluation by global eye-movement characteristics.

Predictor	B	SE	t	p
(Intercept)	-1.91	0.77	-2.48	0.013
First fixation duration (ms)	-0.78	3.04	-0.26	0.799
Total gaze duration (ms)	6.68	3.36	1.99	0.047
Number of fixations	0.02	0.01	2.34	0.019
Saccade length (px)	0.01	0.01	0.60	0.548
Regressions (%)	-3.08	1.32	-2.34	0.019
Pupil dilation (arbitrary)	-0.00	0.00	-0.54	0.587
Blinks (%)	-0.10	0.06	-1.66	0.098
Total reading time (sec)	-0.06	0.03	-2.13	0.033
AReA total score	0.24	0.12	1.92	0.055

The general decrease in reading times is readily interpretable along the lines of increased processing ease: the more a metered stanza proceeds towards its conclusion, the more predictable its formal patterning becomes, and hence the more fluent its processing. The slower reading times for the final lines of each stanza suggest a closure effect. This could either be the well-known wrap-up effect of semantic processing (e.g., (Hirotoni, Frazier and Rayner, 2006; Rayner, Kambe and Duffy, 2000)) or a savoring of the closure of the stanza’s formal gestalt, or a combination of these two effects.

Notably, just as with the longer reading times observed for poem-final verses, the longer gaze durations observed for verse- and stanza-final words can also be interpreted not only as closure, but also as savoring effects, as rhyming words are generally better liked (e.g., (Hayes, Chemelski and Palmer, 1982; Obermeier et al., 2013)). Verse-final words and stanza-final verses share, at different levels, a critical role in the prosody of poems. For the final words of each line, as for the final lines of each stanza, all preceding words, or lines, drive predictions regarding the upcoming concluding units to increasingly precise and narrow limits, more so for rhymed and metered poems than free verse poems. In the case of successful completion, the increased attentional focus on prosodic closure can be associated with pleasure and a concomitant potential savoring effect.

Using an analogous stimulus set—i.e., traditional rhymed and metered German poems along with variants of these poems from which rhyme and/or meter were selectively and cautiously removed—a neuroscientific study reported a rhyme- and meter-driven facilitation effect in both the N 400 and the P 600 components of the EEG precisely for the processing of all stanza-final words for which we reported longer reading times (Obermeier et al., 2016). Importantly, this facilitation effect correlated positively with subjective aesthetic evaluations. This strongly supports the assumption that the longer reading times observed in our study need not,

and perhaps even should not, be interpreted as evidence for an adverse effect on cognitive processing, let alone as evidence for reduced hedonic processing.

### 6.5. Relation between overall eye movements and subjective aesthetic evaluations

Some eye-movement variables that characterize the overall reading of all poem variants turned out to be predictive of the factor of subjective aesthetic evaluation. In particular, *during the reading of poems that received higher aesthetic evaluation scores, individual words were read longer*. At the same time, *progression during reading was more tightly paced, i.e., showed more fixations, and reading proceeded in a more pronounced feed-forward manner, i.e., showed fewer regressions* (see Table 6). Together, these two effects of aesthetic evaluation on eye tracking yielded an interesting pattern for the poems that were highly aesthetically liked compared to the less liked ones: the overall total reading times were shorter, even though fixation durations for individual words were longer. This implies that readers proceeded more carefully word by word through the text, yet still sped up the overall reading process by virtue of not spending extra time rereading and/or moving forward from word to word at a faster pace.

### 6.6. Relation between effects of final word and verse reading times on aesthetic evaluation

Just as with the overall measures of the reading process, we used the reading times for the final words of each line, the reading time difference between the opening and closing words of a poem, and the general trend in reading duration across verses to predict aesthetic evaluation. Among these predictors, only the decrease in total reading times across verses (from which the final lines of stanzas are notably exempt) turned out to be predictive of aesthetic evaluation: *The less steep the decrease of verse reading times across the poems, the higher the aesthetic evaluation scores*. That a general decrease of verse reading times is aesthetically preferred clearly speaks in favor of the fluency hypothesis. At the same time, the less pronounced this decrease, the more a poem is liked. This suggests that the rewards of fluency should best be checked by counteracting trends limiting these effects, be these increased semantic and/or syntactic deviance and hence higher complexity, or tendencies to “savor” the reading, i.e., to prolong the reading times for hedonic reasons, if only within the confines of an overall fluency effect.

## 7. General discussion

In the present study, we investigated the effects of rhyme and meter on eye movements and subjective aesthetic evaluation. Our theoretical points of departure were the hypothesis of cognitive fluency (McGlone and Tofiqbakhsh, 2000; Menninghaus et al., 2015; Reber, Schwarz and Winkielman, 2004), according to which higher perceptual/prosodic fluency of processing is positively correlated with perceived aesthetic appeal, the hypothesis of combined fluency and disfluency (Jakobson, 1960; Wallot and Menninghaus, 2018), according to which the parallelistic patterns of poetic language use involve increased (semantic and syntactic) processing difficulty, but at the same time greater (prosodic) fluency, and finally the hypothesis of savoring, according to which people spend more time in processing aesthetically preferred artworks that allow for self-paced exposure, such as paintings and silently read poems.

To start with, we replicated earlier experimental findings according to which the presence vs. absence of rhyme and meter in traditional German poems enhances aesthetic evaluation, specifically, ratings of how beautiful, melodious, and emotionally moving the poems are perceived to be.

Regarding the core focus of the present study—i.e., what do eye-movement characteristics reveal about reading poems?—, overall measures of global eye movements turned out to be not sensitive to the objective textual differences resulting from the experimental modifications of rhyme and meter. At the same time, the following results provide clear evidence for the importance of increased rhyme and meter-driven (prosodic) processing ease as measured by reading times for several constituents of the verses:

- 1 In general, verse reading times decrease as the stanzas proceed from line to line. This effect is most pronounced in the metered poem variants.
- 2 Words that close a rhyme pair are read faster than words opening it in rhymed and metered poems.

In contrast, the following results reveal increased gaze durations for constituents of rhymed and metered verses:

- 1 For all verse-final words, total gaze durations are longer for the three poem variants featuring rhyme only, rhyme plus meter, or meter only compared to those that feature neither rhyme nor meter.
- 2 For the very concluding words of entire poems, the original poems featuring both rhyme and meter show a more pronounced effect of longer gaze durations than the three other variants do.
- 3 For the final lines of each stanza of the poems, the general decrease in reading times from the first through the subsequent lines is inverted. Here, reading times increase compared to the preceding lines, in the rhymed and/or metered poem variants more so than in the other variants.

Importantly, the fluency and the disfluency effects of the presence vs. absence of rhyme and meter reported here are readily compatible. The latter affect only local constituents of verses, whereas the former also include a more global effect, i.e., a general decrease of reading times throughout all stanzas, if only with the notable exception of the poem-final lines.

All longer gaze duration effects are also compatible with the hypothesis of increased savoring. A neuroscientific study (Obermeier

**Table 7**  
Prediction of aesthetic evaluation by text-based eye-movement characteristics.

Predictor	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>
(Intercept)	−0.70	0.15	−4.48	<.001
GazeSlopeLineNr	1061.97	244.10	4.35	<.001
GazeMean	−26.83	22.84	−1.17	.240
GazeDiff	−21.52	18.07	−1.19	.234
AReA total score	0.23	0.05	4.62	<.001

*Note.* GazeSlopeLineNr is the rate of change for gaze durations across verses of a poem variant. GazeMean is the average gaze duration for the final word of a poem variant, and GazeDiff is the difference in total gaze duration between the opening and closing words of a rhyme pair. Including the interaction effect of line number with the verse end (Table 5) did not improve the the prediction of aesthetic liking scores.

et al., 2016) that did not measure reading times, but rather event-related potentials in responses to a fully analogous stimulus set, supports this interpretation: it reported a facilitation effect precisely for the stanza-final words for which we observed longer reading times. In fact, in the absence of both an unambiguous and pronounced disfluency effect on our global measures, the hypothesis of savoring is even more readily compatible with our findings of prolonged gaze durations than is the hypothesis of cognitive disfluency. Specifically, the majority of the longer rhyme- and meter-driven gaze duration effects, as well as reading times in relation to subjective aesthetic reward, support the assumption that the closures of individual lines, of entire stanzas, and most notably of entire poems are more hedonically processed in the original rhymed and metered poems. This is not surprising, as rhymed and metered prosody builds stronger predictions regarding the total verbal gestalt. Under this premise, living up to more pronounced predictions in a final resolution of the verses' and the poems' trajectories can be experienced as more self-rewarding.

Independently of rhyme and meter, the correlations we observed between more global and more local eye movement measures and aesthetic evaluation likewise show a pattern that can be interpreted as combining fluency and disfluency/savoring effects:

- 1 During the reading of poems that received higher aesthetically evaluative ratings, gaze durations for *individual words* were longer. At the same time, *progression during reading* was more tightly paced, i.e., showed *more fixations*, and reading proceeded in a more pronounced feed-forward manner, i.e., showed *fewer regressions*. As a result, the *overall total reading times were shorter*, even though the total gaze durations for individual words were longer.
- 2 Higher subjective aesthetic evaluation is also associated with decreased verse reading times across stanzas. Moreover, this decrease is most liked if it is not steep and hence balanced out by counteracting processes moderating the trend towards increased fluency.

Measures of blink rates and pupil dilations were not informative in our study. As a potential role of blinks in higher-level text processing is not yet known, it is difficult to interpret this result. One would expect pupil dilations to show effects of the subjective experience during poetry reading relating to emotional responses, such as being moved. However, pupil dilations are not only affected by emotional factors, but also by cognitive factors, and this might be one explanation of why there is no substantial correlation between the aesthetic reward ratings and pupil dilation. Alternatively, it might also be the case that the poems presented here did not affect the participants' emotions strongly enough in order to reliably elicit effects on pupil dilation.

Clearly, none of the three theoretical hypotheses guiding our study can fully explain all of the observed effects. Rather, our results suggest that an integrative approach is called for, one that does not consider these hypotheses as mutually exclusive alternatives. The nuanced blend of effects that we report here using eye-tracking measures should encourage further studies using other methods and/or combining a variety of methods. In any event, our results strongly suggest that longer processing times do not invariably indicate adverse effects on aesthetic liking and hedonic processing, and that the hypothesis of savoring is worth being considered as an additional interpretive option.

### 7.1. Limitations and outlook

Aesthetic evaluations were collected only after the poems had been read, and participants were likely to have started with an eye-movement pattern—or a processing strategy—that could at best have been partially reflective of their final and overall aesthetic evaluation. Hence, additional research is needed that is ideally just as sensitive to the time course of reading with regard to aesthetic evaluation as it is with regard to the continuously collected eye-movement data. Moreover, progress needs to be made towards a more controlled dissociation of the contributions that we attribute in our interpretation of the results to the processes stipulated under the three hypotheses guiding our study. Specifically, the findings that, compared to nonrhymed and nonmetered variants, the stanza-final words of rhymed and metered poems showed longer gaze durations in our study, yet enhanced rather than reduced processing ease in event-related potentials in an earlier study using a fully analogous stimulus set (Obermeier et al., 2016), suggest that the interpretation of eye-tracking data may best be informed by objective data obtained through other methods. Finally, even though our study investigated the role of aesthetic evaluations in poetry reading, we conjecture that our findings are—in principle—not limited to poetry, but may apply more generally to the reading of texts that use specific linguistic features designed to enhance aesthetic evaluation. The degree to which this might be the case, however, needs to be investigated in future studies. (Table 7)

## Acknowledgements

We thank Christine Knoop for the exemplification of the relevant poem modifications with an English poem, as well as Claudia Lehr and Freya Materne for their help with the data collection. Sebastian Wallot acknowledges funding from the Heisenberg programme of the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), grant number 442405852.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.poetic.2020.101526.

## References

- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.
- Berlyne, D. E. (1974). *Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic evaluation*. Sons: John Wiley.
- Blanchard, H. E., & Iran-Nejad, A. (1987). Comprehension processes and eye movement patterns in the reading of surprise-ending stories. *Discourse Processes*, *10*, 127–138.
- Blohm, S., Wagner, V., Schlesewsky, M., & Menninghaus, W. (2018). Sentence judgments and the grammar of poetry: Linking linguistic structure and poetic effect. *Poetics*, *69*, 41–56.
- Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 11818–11823.
- Bohrn, I. C., Altmann, U., Lubrich, O., Menninghaus, W., & Jacobs, A. M. (2012). Old proverbs in new skins: An fMRI study on defamiliarization. *Frontiers in Psychology*, *3*, 204.
- Booth, C. R., Brown, H. L., Eason, E. G., Wallot, S., & Kelly-Stephen, D. G. (2016). Expectations on hierarchical scales of discourse: Multifractality predicts both short- and long-range effects of violating gender expectations in text reading. *Discourse Processes*, *55*, 12–30.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.
- Breen, M., & Clifton, C., Jr. (2011). Stress matters: Effects of anticipated lexical stress on silent reading. *Journal of Memory and Language*, *64*, 153–170.
- Clifton, C., Jr., & Duffy, S. A. (2001). Sentence and text comprehension: Roles of linguistic structure. *Annual Review of Psychology*, *52*, 167–196.
- Clifton, C., Jr., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 341–371). North Holland: Elsevier.
- Cutler, A., & Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Language and Speech*, *20*, 1–10.
- de Gee, J. W., Knapen, T., & Donner, T. H. (2014). Decision-related pupil dilation reflects upcoming choice and individual bias. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, E618–E625.
- Engbert, R., & Mergenthaler, K. (2006). Microsaccades are triggered by low retinal image slip. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 7192–7197.
- Essens, P. J., & Povel, D.-J. (1985). Metrical and nonmetrical representations of temporal patterns. *Perception & Psychophysics*, *37*, 1–7.
- Fabb, N. (2015). *What is poetry?: Language and memory in the poems of the world*. Press: Cambridge University.
- Fechner, G. T. (1876). *Vorschule der Aesthetik*. Härtel: Breitkopf &.
- Finnegan, R. (2017). *Oral literature in Africa*. Publishers: Open Book.
- Forster, M., Leder, H., & Ansorge, U. (2013). It felt fluent, and I liked it: Subjective feeling of fluency rather than objective fluency determines liking. *Emotion*, *13*, 280–289.
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*, 178–210.
- Frijda, N. H., & Sundararajan, L. (2007). Emotion refinement: A theory inspired by Chinese poetics. *Perspectives on Psychological Science*, *2*, 227–241.
- Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, *48*, 163–189.
- Granholm, E., & Steinhauer, S. R. (2004). Pupillometric measures of cognitive and emotional processes. *International Journal of Psychophysiology*, *52*, 1–6.
- Hanauer, D. (1996). Integration of phonetic and graphic features in poetic text categorization judgements. *Poetics*, *23*, 363–380.
- Hanauer, D. (1998). Reading poetry: An empirical investigation of formalist, stylistic, and conventionalist claims. *Poetics Today*, *19*, 565–580.
- Hanich, J., Wagner, V., Shah, M., Jacobsen, T., & Menninghaus, W. (2014). Why we like to watch sad films: The pleasure of being moved in aesthetic experiences. *Psychology of Aesthetics, Creativity, and the Arts*, *8*, 130–143.
- Hayes, D. S., Chemelski, B. E., & Palmer, M. (1982). Nursery rhymes and prose passages: Preschoolers' liking and short-term retention of story events. *Developmental Psychology*, *18*, 49–56.
- Hayes, T. R., & Petrov, A. A. (2016). Mapping and correcting the influence of gaze position on pupil size measurements. *Behavior Research Methods*, *48*, 510–527.
- Hideyuki, H., & Menninghaus, W. (2018). The eye tracks the aesthetic appeal of sentences. *Journal of Vision*, *18*, 1–22.
- Hirofani, M., Frazier, L., & Rayner, K. (2006). Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language*, *54*, 425–443.
- Hoeken, H., & van Vliet, M. (2000). Suspense, curiosity, and surprise: How discourse structure influences the affective and cognitive processing of a story. *Poetics*, *27*, 277–286.
- Hofmeister, J., Heller, D., & Radach, R. (1999). The return sweep in reading. In W. Becker, H. Deubel, & T. Mergner (Eds.), *Current oculomotor research* (pp. 349–357). pp: Springer.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179–185.
- Jakobson, R. (1960). Linguistics and poetics. In T. A. Sebeok (Ed.), *Style in language* (pp. 350–377). pp: Wiley.
- JASP Team. (2018). JASP (Version 0.8.6) [Computer software]. Retrieved from <https://jasp-stats.org/>.
- Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*, 262–284.
- Knoop, C. A., Wagner, V., Jacobsen, T., & Menninghaus, W. (2016). Mapping the aesthetic space of literature from “below. *Poetics*, *56*, 35–49.
- Lea, R. B., Rapp, D. N., Elfenbein, A., Mitchel, A. D., & Romine, R. S. (2008). Sweet silent thought: Alliteration and resonance in poetry comprehension. *Psychological Science*, *19*, 709–716.
- Lerdahl, F., & Jackendoff, R. (1977). Toward a formal theory of tonal music. *Journal of Music Theory*, *21*, 111–171.
- LeVasseur, V. M., Macaruso, P., & Shankweiler, D. (2008). Promotion gains in reading fluency: A comparison of three approaches. *Reading and Writing*, *21*, 205–230.
- Margulis, E. H. (2014). *On repeat: How music plays the mind*. Press: Oxford University.
- Marsman, M., & Wagenmakers, E. J. (2017). Bayesian benefits with JASP. *European Journal of Developmental Psychology*, *14*, 545–555.
- Mathworks (2015). *Matlab 2015b user manual*. Retrieved from <http://de.mathworks.com/help/>.

- McConkie, G. W., Kerr, P. W., Reddix, M. D., Zola, D., & Jacobs, A. M. (1989). Eye movement control during reading: II. Frequency of refixating a word. *Attention, Perception, & Psychophysics*, *46*, 245–253.
- McGlone, M. S., & Tofighbakhsh, J. (2000). Birds of a feather flock conjointly (?): Rhyme as reason in aphorisms. *Psychological Science*, *11*, 424–428.
- Menninghaus, W., Bohrn, I. C., Altmann, U., Lubrich, O., & Jacobs, A. M. (2014). Sounds funny? Humor effects of phonological and prosodic figures of speech. *Psychology of Aesthetics, Creativity, and the Arts*, *8*, 71–76. <https://doi.org/10.1037/a0035309>.
- Menninghaus, W., Bohrn, I. C., Knoop, C. A., Kotz, S. A., Schlotz, W., & Jacobs, A. M. (2015). Rhetorical features facilitate prosodic processing while handicapping ease of semantic comprehension. *Cognition*, *143*, 48–60.
- Menninghaus, W., Schindler, L., Wagner, V., Wassiliwizky, E., Hanich, J., Jacobsen, T., & Koelsch, S. (2020). Aesthetic emotions are a key factor in aesthetic evaluation: Reply to Skov and Nadal. *Psychological Review*, *127*, 650–654.
- Menninghaus, W., Wagner, V., Hanich, J., Wassiliwizky, E., Jacobson, T., & Koelsch, S. (2017). The Distancing-Embracing model of the enjoyment of negative emotions in art reception. *Behavioral and Brain Sciences*. <https://doi.org/10.1017/S0140525X17000309>. Advance online publication.
- Menninghaus, W., Wagner, V., Wassiliwizky, E., Jacobsen, T., & Knoop, C. A. (2017). The emotional and aesthetic powers of parallelistic diction. *Poetics*, *63*, 47–59.
- Menninghaus, W., Wagner, V., Wassiliwizky, E., Schindler, L., Hanich, J., Jacobsen, T., & Koelsch, S. (2019). What are aesthetic emotions. *Psychological Review*, *126*, 171–195.
- Mesuger, E., Carreiras, M., & Clifton, C. (2002). Overt reanalysis strategies and eye movements during the reading of mild garden path sentences. *Memory & Cognition*, *30*, 551–561.
- Mitschke, V., Goller, J., & Leder, H. (2017). Exploring everyday encounters with street art using a multimethod design. *Psychology of Aesthetics, Creativity, and the Arts*, *11*, 276–283.
- Mueller, H. J., Geyer, T., Günther, F., Kacian, J., & Pierides, S. (2017). Reading English-language haiku: Processes of meaning construction revealed by eye movements. *Journal of Eye Movement Research*, *10*, 1–33.
- Obermeier, C., Kotz, S. A., Jessen, S., Raettig, T., von Koppenfels, M., & Menninghaus, W. (2016). Aesthetic evaluation of poetry correlates with ease of processing in event-related potentials. *Cognitive, Affective, & Behavioral Neuroscience*, *16*, 362–373.
- Obermeier, C., Menninghaus, W., von Koppenfels, M., Raettig, T., Schmidt-Kassow, M., Otterbein, S., & Kotz, S. A. E. (2013). Aesthetic and emotional effects of meter and rhyme in poetry. *Frontiers in Psychology*, *4*, 10.
- Palmer, S. E., & Hemenway, K. (1978). Orientation and symmetry: Effects of multiple, rotational, and near symmetries. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 691–702.
- Panksepp, J. (1995). The emotional sources of “chills” induced by music. *Music Perception*, *13*, 171–207.
- Parker, S., Bascom, J., Rabinovitz, B., & Zellner, D. (2008). Positive and negative hedonic contrast with musical stimuli. *Psychology of Aesthetics, Creativity, and the Arts*, *2*, 171–174.
- R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from <https://www.R-project.org/>.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.
- Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, *10*, 241–255.
- Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*, 191–201.
- Rayner, K., Kambe, G., & Duffy, S. A. (2000). The effect of clause wrap-up on eye movements during reading. *The Quarterly Journal of Experimental Psychology: Section A*, *53*, 1061–1080.
- Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver’s processing experience? *Personality and Social Psychology Review*, *8*, 364–382.
- Roberts, A. M., Stabler, J., Fischer, M. H., & Otty, L. (2013). Space and pattern in linear and postlinear poetry: Empirical and theoretical approaches. *European Journal of English Studies*, *17*, 23–40.
- Rothermich, K., Schmidt-Kassow, M., Schwartze, M., & Kotz, S. A. (2010). Event-related potential responses to metric violations: Rules versus meaning. *NeuroReport*, *21*, 580–584.
- Salimpoor, V. N., Benovoy, M., Longo, G., Cooperstock, J. R., & Zatorre, R. J. (2009). The rewarding aspects of music listening are related to degree of emotional arousal. *PLoS One*, *4*, e7487.
- Scheepers, C., Mohr, S., Fischer, M. H., & Roberts, A. M. (2013). Listening to limericks: a pupillometry investigation of perceivers’ expectancy. *PLoS One*, *8*, e74986.
- Schroeder, S. (2011). What readers have and do: Effects of students’ verbal ability and reading time components on comprehension with and without text availability. *Journal of Educational Psychology*, *103*, 877–896.
- Siegle, G. J., Ichikawa, N., & Steinhauer, S. (2008). Blink before and after you think: Blinks occur prior to and following cognitive load indexed by pupillary responses. *Psychophysiology*, *45*, 679–687.
- Smith, B. H. (1968). *Poetic closure: A study of how poems end*. Press: University of Chicago.
- SR Research Ltd. (2009). EyeLink® 1000 user manual. *SR Research*.
- Staners, R. F., Coulter, M., Sweet, A. W., & Murphy, P. (1979). The pupillary response as an indicator of arousal and cognition. *Motivation and Emotion*, *3*, 319–340.
- Teng, X., Ma, M., Yang, J., Blohm, S., Cai, Q., & Tian, X. (2020). Constrained structure of ancient Chinese poetry facilitates speech content grouping. *Current Biology*. <https://doi.org/10.1016/j.cub.2020.01.059>.
- Thierry, G., Martin, C. D., Gonzalez-Diaz, V., Rezaie, R., Roberts, N., & Davis, P. M. (2008). Event-related potential characterisation of the Shakespearean functional shift in narrative sentence structure. *NeuroImage*, *40*, 923–931.
- Thoussignant, C., & Bodner, G. E. (2014). Context effects on beauty ratings of photos: Building contrast effects that erode but cannot be knocked down. *Psychology of Aesthetics, Creativity, and the Arts*, *8*, 81–86.
- Tillmann, B., & Dowling, W. J. (2007). Memory decreases for prose, but not for poetry. *Memory & Cognition*, *35*, 628–639.
- Tschacher, W., Greenwood, S., Kirchberg, V., Wintzerith, S., van den Berg, K., & Tröndle, M. (2012). Physiological correlates of aesthetic perception of artworks in a museum. *Psychology of Aesthetics, Creativity, and the Arts*, *6*, 96–103.
- Vaughan-Evans, A., Trefor, R., Jones, L., Lynch, P., Jones, M. W., & Thierry, G. (2016). Implicit detection of poetic harmony by the naïve brain. *Frontiers in Psychology*, *7*, 1859.
- Wallot, S., Hollis, G., & van Rooij, M. (2013). Connected text reading and differences in text reading fluency in adult readers. *PLoS One*, *8*, e71914.
- Wallot, S., & Menninghaus, W. (2018). Ambiguity effects of rhyme and meter. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. <https://doi.org/10.1037/xlm0000557>. Advance online publication.
- Wallot, S., O’Brien, B., Coey, C. A., & Keltz-Stephen, D. (2015). Power-law fluctuations in eye movements predict text comprehension during connected text reading. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th annual meeting of the cognitive science society* (pp. 2583–2588). Society: Cognitive Science.
- Wallot, S., O’Brien, B. A., Haussmann, A., Kloos, H., & Lyby, M. S. (2014). The role of reading time complexity and reading speed in text comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1745–1765.
- Schlotz, W., Wallot, S., Omigie, D., Masucci, M. D., Hoelzmann, S. C., & Vessel, E. A. (2020). The Aesthetic Responsiveness Assessment (AREA): A screening tool to assess individual differences in responsiveness to art in English and German. *Psychology of Aesthetics, Creativity, and the Arts*. <https://doi.org/10.1037/aca0000348>.
- Wassiliwizky, E., Koelsch, S., Wagner, V., Jacobsen, T., & Menninghaus, W. (2017). The emotional power of poetry: Neural circuitry, psychophysiology and compositional principles. *Social Cognitive and Affective Neuroscience*, *12*, 1229–1240.

Winfried Menninghaus is Director of the Department of Language and Literature at the Max Planck Institute for Empirical Aesthetics (Frankfurt am Main). He previously held professorships at Freie Universität Berlin, the universities of Jerusalem, Berkeley, Yale, Princeton, and Rice. His fields of research are rhetoric and poetics, philosophical and empirical aesthetics, and literature from 1750 until present. Publications include scientific articles on poetic and rhetorical language and its cognitive, emotional and aesthetic effects, and monographs on the aesthetics of disgust, Kant and Ludwig Tieck, and Friedrich Hölderlin. He is a member of the Berlin-Brandenburg Academy of Sciences.

Sebastian Wallot is a senior researcher at the Department of Language and Literature at the Max Planck Institute for Empirical Aesthetics (Frankfurt am Main). He got his Ph.D. in experimental psychology from the University of Cincinnati, OH, USA and worked as postdoctoral research at the Interacting Minds Centre, Aarhus University, Denmark. His research focusses on reading, particularly the relation between reading process characteristics and comprehension, joint action, and the development of and application of nonlinear time-series analyses.