In poetry, if meter has to help memory, it takes its time

[version 1; peer review: awaiting peer review]

Sara Andreetta\textsuperscript{1}, Oleksandra Soldatkina\textsuperscript{1}, Vezha Boboeva\textsuperscript{1,2}, Alessandro Treves\textsuperscript{1}

\textsuperscript{1}Cognitive Neuroscience, SISSA, Trieste, 34136, Italy
\textsuperscript{2}Bioengineering, Imperial College London, London, SW7 2AZ, UK

Abstract

To test the idea that poetic meter emerged as a cognitive schema to aid verbal memory, we focused on classical Italian poetry and on three components of meter: rhyme, accent, and verse length. Meaningless poems were generated by introducing prosody-invariant non-words into passages from Dante's Divina Commedia and Ariosto's Orlando Furioso. We then ablated rhymes, modified accent patterns, or altered the number of syllables. The resulting versions of each non-poem were presented to Italian native speakers, who were then asked to retrieve three target non-words. Surprisingly, we found that the integrity of Dante's meter has no significant effect on memory performance. With Ariosto, instead, removing each component downgrades memory proportionally to its contribution to perceived metric plausibility. Counterintuitively, the fully metric versions required longer reaction times, implying that activating metric schemata involves a cognitive cost. Within schema theories, this finding provides evidence for high-level interactions between procedural and episodic memory.

Keywords

Schema theory, Memory retrieval, Hendecasyllables, Sequence replay, Dynamical attractors.

This article is included in the Excellent Science gateway.
Corresponding author: Alessandro Treves (ale@sissa.it)

Author roles: Andreetta S: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Writing – Original Draft Preparation; Soldatkina O: Formal Analysis, Resources, Software, Writing – Review & Editing; Boboeva V: Software, Visualization, Writing – Review & Editing; Treves A: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Methodology, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No [765549], (Memory research: Ground-breaking, Applied, and Technological Exchanges [M-GATE]) and by the Human Frontier Science Program grant RGP0057/2016.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 Andreetta S et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Andreetta S, Soldatkina O, Boboeva V and Treves A. In poetry, if meter has to help memory, it takes its time [version 1; peer review: awaiting peer review] Open Research Europe 2021, 1:59 https://doi.org/10.12688/openreseurope.13663.1

First published: 28 May 2021, 1:59 https://doi.org/10.12688/openreseurope.13663.1
Plain language summary

We have tested the common idea that poetic meter has emerged in order to help verbal memory. We have focused on classical Italian poetry and on three components of meter: rhyme, accent and verse length. We selected four passages from Dante’s Divina Commedia and four from Ariosto’s Orlando Furioso, and from them derived meaningless poems by replacing key words with non-words, in such a way as to retain the original prosody. Alongside these eight “original” non-poems, we created three metrically defective variants of each, by either ablating rhymes or modifying accent patterns or altering the number of syllables. One of the resulting four versions of each non-poem was presented repeatedly to Italian native speakers, who were then asked, a day later, to remember some of the non-words. Surprisingly, we found that the integrity of Dante’s meter has no significant effect on memory performance. With Ariosto, instead, removing each component downgrades memory proportionally to its contribution to perceived metric plausibility, with rhymes the most important one. Counterintuitively, the fully metric versions required longer reaction times, implying that activating a given metric pattern, an example of a recurring schema, is helpful but involves a cognitive cost. Within schema theories, this finding provides evidence for high-level interactions between procedural and episodic memory.

Introduction

Poems, nursery rhymes, traditional songs: they are found in every culture, and they have been around for ages, well before the advent of writing systems. Sometimes, they have or had the crucial mission of carrying an important message for the listeners: a list to know by heart, an event happening every year, a warning of a potential danger. What do these texts have in common? At least one aspect: they adopt a variety of devices that help hold verbal material in memory.

Human memory can, in fact, fail spectacularly at times. Writing systems have helped safely store verbal information, in a format relatively difficult to tamper with; before, when our ancestors had to rely on their fallible memory, a number of linguistic devices crystallized to help them remember words and verbal material. Cultural transmission, then, has depended for ages on these devices, which in poetry we can broadly refer to as “meter”. These devices may range from the use of repeated metaphors: “rosy-fingered dawn” in Homer (Reece, 2011), to the ring composition in the Zoroastrian Yasna (Hintze, 2002) to semantic repetition as in Biblical poetry: “In the way of righteousness is life; and in the pathway thereof there is no death.” in Prov. 12:28 (King James Version).

In several Western literary traditions, including the Italian one, the local structure of poetry revolves around the verse, and includes a constant number of syllables, a limited variability in the pattern of accents, and a specific organization of rhymes. These components of meter have gradually lost their centrality or at least their perceived necessity over the course of several centuries, but they were in full sway at least between the 13th and 16th centuries, from the emergence of modern Italian (so called “volgare”, the language of the people) as an acceptable literary language to the diffusion of the printing press. The Divina Commedia by Dante Alighieri and the Orlando Furioso by Ludovico Ariosto are two lengthy masterpieces towards the beginning and, respectively, the end of this golden age. With 14,233 verses in the Commedia and 38,736 in the Orlando Furioso, neither of which contains material which is absolutely necessary to remember in order to carry on with one’s life, it may be asked whether their metric structure still retained a primary memory function, or is already a purely esthetic ornament for cultured readers (Rubin, 1995).

Can the role of metrical features be explained from a neuro-cognitive point of view, with respect to memory? Recently, in memory literature the notion of schemata, long seen as important (see e.g., (Brewer & Dupree, 1983), has been discussed again (Ghosh & Gilboa, 2014), stimulated also by the analysis of its neurobiological basis in rodents (Tse et al., 2007). A schema, whether directly functional like those involved in preparing coffee (Norman & Shallice, 1986) or social/ornamental, like rituals of salutations (Taylor & Crocker, 1981), can be considered as a set of regularities that help organize and retrieve information (Van Kesteren et al., 2012). In this context, we consider the components of meter as schemata which, by encouraging regularities, facilitate the recall of verses. They operate as schemata insofar as they help us recruit, and possibly produce, the next element of a sequence stored in our memory.

In facilitating verbal sequence replay, metrical features appear to be effective with extended “trajectories”, lasting even several verses. These are extended relative to the short trajectories thought to be produced by the phonological loop of Baddeley’s model of working memory, which are presumed to last only a couple of seconds, precisely because of the lack of specific devices that extend their range (Baddeley, 1992). To the best of our knowledge, though, the effectiveness of these features has never been quantified. In this study, we aim at measuring the strength of some metric devices. Specifically, we focus on the three main characteristics of classical Italian meter: rhyme, pattern of accents, and verse length.

Methods

We extracted passages from two masterpieces of Italian literature: the Divina Commedia by Dante Alighieri (1265–1321), and the Orlando Furioso by Ludovico Ariosto (1474–1533). From the latter we chose ottave (octaves, stanzas of eight verses) from canti XIII, XV, XIX and XXX, and one from canto I to train subjects, while from the former we selected sequences of three consecutive terzine (hence nine verses) from two canti from Inferno (XXIV for the experiment, V for training), two from Purgatorio (VI and XVI) and one from Paradiso (XXVII). All passages had only proper (Italian) hendecasyllables with an accented 10th syllable, and were, to our arbitrary judgement, devoid of explicit or easily reconstructed memorable content or references.
Poem manipulation

The original texts were manipulated in a number of different ways. Firstly, most content words were converted into non-words in order to eliminate discernible semantic content, hence semantic effects on memory; an effort was made to change phonemes with similar ones, while maintaining the original prosody. Function words were not modified. This applied equally to all texts and resulted in “original non-poems” (ONPs).

The second stage of manipulations focused on metrical patterns. We created three conditions:

1) a condition where we eliminated rhymes (“NPR” – Non-Poem without Rhyme)

2) a condition where the accent patterns of four-five verses per passage were replaced with less standard ones (“NPA” – Non-Poem with modified Accents). To validate proper original accents, we consulted with an expert scholar for Orlando Furioso, whereas for Divina Commedia we referred to the “Archivio Metrico Italiano”, a database collecting masterpieces of Italian literature with their accents annotated (www.maldura.unipd.it).

3) a condition where the number of syllables per verse, which in the ONPs were strictly 11 throughout (regular hendecasyllables) was altered again in four-five verses, to nine, 10, 12 or 13 (“NPS” – Non-Poem with wrong numbers of Syllables). Note that by adding or subtracting one or two syllables, also the pattern of accents was perforce altered, but we attempted to make the alteration less noticeable than the number change, in contrast to the NPA condition in which, while there were strictly 11 syllables/verse, the accents followed more unusual patterns.

These manipulations were applied to all texts. All texts were then recited by a professional actor and audio recorded.

For the experiment, every subject was administered four texts in total, by the same (original) author: one per canto and one per condition. Therefore, twenty-four combinations were created.

An example of an NPS we used, from the Commedia, is presented in Figure 1 together with its original spectrogram, and all non-poems are available as Extended data (Andreetta et al., 2021).

Ranking

We conducted an online survey about how these manipulated poems were perceived by a group of Italian native

![Figure 1. NPS example derived from Purgatorio, canto VI. For each terzina (vv.127–135) the NPS text, shown below the sound wave by the professional actor, maintains rhymes, in color, and accents, in boldface, as in the ONP version; whereas overall three syllables have been added and two taken away, in gray. The underlined non-words were the targets of the memory test; underlined blanks denote synizesis (when two syllables are pronounced as one).](image-url)
speakers. Participants were asked to listen to the four conditions and give a ranking of preference, from the one that sounded the most “poetically plausible” to them, to the one they perceived as the strangest.

**Consent statement.** Written informed consent for participation was obtained in advance from all participants.

**Subjects.** 62 people participated in the online survey for Ariosto (F=32, M=30, mean age = 29.06, sd= 8.13) and 65 people for Dante (F=35, M=30, mean age = 26.48, sd = 6.26). Part of either cohort were the participants in the main experiment below, but tested with the other author, and they were asked to complete this survey after the end of the second session of the main experiment. Another group of participants was recruited through the online platform Prolific (www.prolific.co). This last group was compensated with five euros. We had aimed for 72 rankers in each cohort, but had to exclude some a posteriori, who failed to complete the ranking in full.

**Experimental design.** The online survey was designed with the open-source toolkit Psytoolkit (Stoet, 2017). After an example, presented as training also in the main experiment below, they listened to the four poems one at a time. Every poem was associated with a name, in order to help participants refer to that specific condition. If they wanted, they were allowed to listen again and again to the same poem before proceeding to the next.

At the end they were asked to rank the four poems: from the one they perceived as the best, to the one that sounded worse to them. From the rankings by all participating subjects we extracted an average index of metric plausibility by assigning a value 0.6 to the first –ranked condition (e.g., NPA), 0.3, to the second, 0.1 to the third, and 0 to the fourth. The logic behind this assignment is that subjects occasionally reported being unsure as to which passage sounded the strangest. The rankings were collapsed across canti, with the relatively large number of participants ensuring approximately even sampling (each canto was presented originally 18 times per condition, which came down to 16+/2 after the exclusions). As a result, the average metric plausibility of each condition could in principle range from 0 to 0.6, but in practice was much more restricted, particularly with passages from Dante, to values around the average of 0.25 (see Figure 2).

**Memory experiment**

Ethical approval for this study was granted by the SISSA Ethics Committee at the Scuola Internazionale Superiore di Studi Avanzati with deliberation 2018/16/ib on Nov. 5th, 2018 transmitted by act prot. 15534-III/13.

**Subjects.** 48 native Italian speakers who had been exposed to Italian literature through one of the national high school curricula were recruited for the main experiment. Half of them were administered material from Ariosto (F= 15, M=9, mean age= 26.34, sd=4.02), the other half from Dante (F=15, M=9, mean age= 26.12, sd=3.61).

Participants were recruited through the SISSA recruiting platform and social media. None of them had a previous history of psychiatric or neurological illness, learning disabilities, nor hearing or visual loss. They were asked to participate in a study on memory and poetry, which would have involved them for two consecutive days, for about 30 minutes the first and about 10 minutes the second. Due to the pandemic situation, they were asked to be connected remotely with their own devices.

![Figure 2. Relative metric plausibility.](image)

The different versions of the same four passages from the *Divina Commedia* (red) and *Orlando Furioso* (blue) were ranked in the same order, but the plausibility index (see Methods) is more spread out for Ariosto.
**Experimental design.** The experiment was designed with Psychopy Builder (Peirce et al., 2019). It included a study phase of about 30 minutes the first day, and the test of about 10 minutes the second day.

We aimed at an almost exclusively auditory experiment, in order to assess how memory relies on meter if listening is the only available channel to learn from (Najme et al., 2020). Indeed, the material included audio files only, with the sole exception of written material when a fill in the gap task appeared.

Besides the four passages, verses from two other cantì were used for training, as indicated above. However, these verses were presented only in the ONP condition, leaving meter intact.

Every passage, including the training, was associated with an image, taken from among Gustave Doré’s illustrations of the *Divina Commedia*. The images were consistent for the same canto in different conditions and were intended to help engage memory without, at the same time, biasing the linguistic material (see Extended data).

Every poem, including the training, was presented divided into three consecutive portions.

**Notice on the use of the Zoom platform.** The pandemic of 2020 forced us to find new methods to administer our experiment to subjects in remote mode.

After evaluating several options, we decided that, for this study, it was important to keep a degree of dialogue with participants. Also, we wanted to make sure that they were focused on the task and that they did the second part at the same time the following day.

For these reasons, among others, we thought that a good option was to have them on streaming in an open source platform. We chose Zoom, for which we had an institutional account.

Unfortunately, this meant that lab conditions could not be fully guaranteed. To overcome potential biases, we gave participants specific instructions:

- be connected with a computer or a tablet. Smartphones were not allowed, because of the small size of their screens and because they could potentially be distracting in case of notifications during the experiment
- be in a silent room with no disturbances
- be in the same room during the experiment for both days

Moreover, we used the accompanying images mentioned in the Experimental design section of the main text, in order to focus the visual attention of the participants away from distracting visual stimuli in the rooms they were in. The images, shown in Extended Data (Andreetta et al., 2021), were taken from among Gustave Doré’s illustrations of the *Divina Commedia*, but accompanied passages derived from cantì in the *Commedia* unrelated to those the artist referred to, as well as the *a priori* unrelated passages derived from the *Orlando Furioso*, and they were further enriched with graded color hues. Each image was consistently paired to each passage, whichever version, ONP, NPS, NPA or NPR, was presented to the subject. The variability across images, enhanced also by leaving each image in its original non-commensurate format, was thus adding an independent component to the natural variability across passages, but not contributing to the metric plausibility effect; and possibly helped reduce the variability, across participants, due to their heterogeneous testing-at-home conditions.

The dialogue over zoom was always conducted by the first author, but the experiment was self-paced by the participants using the Psychopy platform, with the images displayed on the shared screen and the passages played auditorily from the prior recordings by actor Sara Alzetta. The recordings are available upon request. When tested with the muted non-word, participants would read aloud the left, central or right alternative appearing on the screen, and the experimenter would press the corresponding key (leftward, downward or rightward) on her own keyboard.

**Study phase.** The study phase started with the training ONP. First, participants listened once to all verses. Then they listened to the first part (three verses) repeated five times, with a three second pause between repetitions. Afterwards, another repetition of the same verses followed, but this time a non-word was muted. The task for the participant was then to retrieve the correct non-word. Three written alternatives appeared on the screen and the participant had to read aloud his/her choice.

After this training, each of the experimental passages was played, in a separate block, in the one of the four versions that had been assigned to that subject in the design. Then, after listening once to the entire poem (nine verses in Dante, eight verses in Ariosto), participants listened to the first part (three verses for both authors) five times. Next, they had to complete the task with the muted non-word.

The same happened with the second part of the poem, repeated four times. In this case the audio started from silence with the first part ramping up linearly in intensity, until it continued smoothly into the second part at the standard volume. Therefore, this allowed them to have feedback about the test they just completed.

For the third part they just listened to the repetitions (three verses for Dante, two verses for Ariosto; repeated three times), starting now with an acoustically smoothed version of the second part, but there was no test.

The alternatives to the correct non-word were chosen by maintaining the same number of syllables, and the same accent. Typically, stems and intermediate vowels or consonants changed. Target non-words were generally in the third verse,
aiming not to overload working memory from the moment they listened to the silent word until the test time. In the cases where this was not possible (because there were no appropriate non-words in the third verse, for instance) a non-word in the second verse was chosen, if possible towards the end. Notably, for every text we chose options which were consistent among all conditions, allowing a fair comparison in the results.

**Test phase.** The following day participants were connected at the same time, so 24 hours had passed. They listened directly to the test parts of the text, so the verses with a muted non-word, and they were asked to identify the muted target in all three parts, including then the third one. For the first and second part, tested also the previous day, target non-words were different. After completing the three tests per text, they listened to the entire poem, to receive feedback.

**Data analysis.** The outcome of interest is essentially the presence of a significant correlation between the dependent variables measured in the ranking (the plausibility index, see Figure 2) and in the memory experiment (correct responses, and reaction times, see Figure 4), that would indicate a joint dependence on the type of passage manipulation. Correlations were considered significant at $p < 0.05$. Additional methodological considerations and controls are detailed below.

**Entropy in the accent distribution.** Two simple entropy measures were used to quantify the variability in the pattern of accents in the eight passages from the *Divina Commedia* and *Orlando Furioso* from which we derived the non-poems used in the experiment. First, the pattern of accents in each verse (from a total of 36 verses from the *Commedia* and 32 from the *Orlando*) was codified, based on the consensus in the literature, as a binary string of length 11, where each syllable was assigned a 1 if accented and a 0 if not. Since all 68 verses were regular hendecasyllables with the 10th syllable accented and the 11th not, we focused on the first nine digits in each string.

The first measure is based on the simplifying assumption of independent accents on neighboring syllables and calculates, for each author, the sum of the binary entropies for the syllables in each ordinal position, a sum which can range from 0 to nine bits.

The second measure is the entropy of the distribution, for each author, of distinct binary strings, and it ranges from 0 to $\log_2(36)$ for Dante and from 0 to 5 bits for Ariosto.

**Word frequency.** The targets were derived from words of widely different frequency, covering the entire spectrum from $5 \times 10^{-8}$ to $1.5 \times 10^{-3}$ in SUBTLEX-IT, the corpus of Italian Subtitle-based Word Frequency Estimates, containing 517,564 entries (poster presented at the Annual Meeting of the Italian Ass Expl Psychol, Rovereto, Sept. 2015: Crepaldi, Amenta, Mandera, Keuleers and Brysbaert, SUBTLEX-IT. Subtitle-based word frequency estimates for Italian. Available online: https://turlac.sissa.it/publications/frequency-estimates-different-registers-explain-different-aspects-visual-word).

**Results**

Two separate cohorts or rankers, for Dante- and Ariosto-derived non-poems, were presented with a combination of the four passages from the same author, one in each of the ONP, NPS, NPS and NPR versions, and were asked to rank them in order of metric plausibility. The fully balanced design allowed us to extract a passage-independent plausibility score.

**The contribution of distinct components to metric plausibility.**

Both when derived from passages by Dante and Ariosto, non-sense poems were found most plausible in their fully metric ONP versions, somewhat less when the number of syllables was manipulated (NPS), even less when the pattern of accents was altered in the NPA renditions, and the least when rhymes were removed, NPR. Remarkably, however, differences in the plausibility index are shown in Figure 2 to be quite limited, making the fully balanced design essential. The variance was particularly limited in passages from the *Commedia*, which may be due to Dante’s taking more liberties with the meter he had adopted (the same hendecasyllables as Ariosto, but in terzine rather than ottave). To quantify this perception, at least in relation to accent patterns, which are more accessible to analysis, we applied two independent measures of accent variability to the four original passages by each author.

Dante appears to be slightly more variable in his accent patterns relative to Ariosto, but the main observation that can be gleaned off Figure 3 is that both poets are far from using a fixed pattern, utilizing over half of the maximum entropy they had available in terms of accenting those passages.

**Meter can facilitate memory.** Does such a loose structure help remember individual words? Table 1 and Figure 4 (upper) show that it does, only for the non-poems derived from

![Figure 3. Variability in the pattern of accented syllables in the eight original passages by the two authors.](image-url)
Memory and reaction times both increase with metric plausibility. (Upper) Overall correct responses (out of 72) for each condition, ordered in terms of their metric plausibility, as in Figure 2, for passages from Dante (red) and Ariosto (blue). (Lower) Reaction times (in seconds) for correct (circles) and wrong responses (dots) are regressed against plausibility for each author, with a single slope parameter. The slope is significant and similar to that characterizing the Ariosto data alone, whereas it is denoted with a dashed line for the Dante data, because the latter would not produce a significant correlation on its own.

There were three such targets in each non-poem. While in the case of those derived from passages in the *Divina Commedia* the overall number of correct responses per condition was unrelated to its metric plausibility ($r^2=0.04$), seemingly fluctuating as much as the correct responses to the first, second, and third query taken alone (Table 1), for the passages from *Orlando Furioso* the correlation with metric plausibility was remarkable ($r^2=0.98$) and highly significant ($p<0.01$). Interestingly, the total score of the two cohorts was nearly identical, 147 for Ariosto and 148 for Dante, out of a total of 288 (24×4×3).

### Meter helps, but not for free.

The analysis of reaction times helps interpret the above results. As shown in Figure 4 (lower), overall it took longer for participants to pick a wrong answer over the correct answer (on average, 733ms more), and it took longer for participants tested with Ariosto, relative to those tested with Dante, to respond (on average, 547ms more). Most importantly, in each of the four types of trials above, the more metrically plausible the passage, the longer the reaction time. However, the trend is significant only with Ariosto, if data from the two authors are analyzed separately, and it is significant overall ($p<0.004$) with a slope mainly determined by the Ariosto data, as shown in Figure 4. The slope for the Dante data alone would be higher, but not significant, likely because of the limited plausibility range spanned.

The overall distribution of reaction times is reported in Figure 5. Note that to avoid biasing RT results with the occasional outliers, only RTs < 10s where included in the averages in Figure 4, leaving out 14 trials for Dante and 20 for Ariosto, each out of 288. Including them (or alternatively excluding the three trials with RT< 2.8s), does not change the results, in fact it widens the RT gap between Dante and Ariosto.

These findings suggest that processing meter in order to help retrieve a non-word heard the day before has a cognitive cost, and takes the order of hundreds of ms extra time, depending on exactly how much meter is “used up” in the process. For passages derived from Dante, it appears that although outwardly the metric structure is essentially the same (with the slight qualification reported in Figure 3, and the note that a passage is a sequence of three *terzine* rather than a single *ottava*), meter is used less, and the very same memory performance is attained on average in less time.

Word frequency does not have major effects. While targets derived from more frequent words tended to be remembered marginally better, the same trend was observed for both authors (Figure 6), and each target appeared by design in all four conditions.

A strong bias makes subject favor the left alternative, among the three non-word options, but mainly in their wrong responses and the extent of the bias does not correlate with metric plausibility (Figure 7).

---

Table 1. Correct responses out of a total of 24 participants for the first, second, and third query.

<table>
<thead>
<tr>
<th></th>
<th>Dante</th>
<th></th>
<th>Ariosto</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NPR</td>
<td>NPA</td>
<td>NPS</td>
<td>ONP</td>
<td>NPR</td>
<td>NPA</td>
<td>NPS</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>15</td>
<td>13</td>
<td>10</td>
<td>11</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>13</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>17</td>
<td>16</td>
<td>12</td>
<td>11</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>All</td>
<td>33</td>
<td>45</td>
<td>38</td>
<td>32</td>
<td>31</td>
<td>34</td>
<td>40</td>
</tr>
</tbody>
</table>
**Figure 5. Distribution of reaction times.** As explained in the results section, only RTs < 10s were included, leaving out 14 trials for Dante (red) and 20 for Ariosto (blue).

**Figure 6. Frequencies.** Word frequencies (log of occurrences per million) in SUBTLEX-IT for the 24 (8×3) target words used in non-poems derived from Dante (red) and Ariosto (blue). On the y-axis, the memory score is the number of times each target word has been correctly selected, by 24 participants.

**Figure 7. Left Bias.** In full color (red for Dante, blue for Ariosto) the correct responses on the left, central and right non-words. The left alternative, among the three non-word options, was often chosen when the correct non-word was central or on the right (left-tilted striped segments). Blank segments are responses in the center, when the correct non-word was left or right. Right-tilted striped responses are on the right, when the correct non-word was left or center.
Discussion

The connection between meter and memory is not new to cognitive science: in a seminal book Rubin described oral traditions and the linguistic devices they use, highlighting in particular their role in memory as limiting the choice (one could say the entropy, (Shannon & Weaver, 1949) of larger units: by indicating a specific word ending, for instance, choices will be limited to those words which have the same ending, if a rhyme is expected (Rubin, 1995).

In music, Schulkind has investigated memory mechanisms by having participants listen to well-known and novel songs which were altered in their rhythm. Results showed that unaltered versions were identified significantly better than the altered ones, and this applied to both known and novel songs (Schulkind, 1999).

Analogously, Sachs investigated the retention of semantic and syntactic information in discourse by having participants listen to short prose stories. By selectively manipulating the meaning or the syntactic form of a target sentence, she could show that meaning is remembered, in prose, better and for longer that meaning-irrelevant sentence form (Sachs, 1967).

In a similar design, Tillman and colleagues have tested short term memory in prose and poetry. Also in this case, a sentence, considered the target, was changed in its form or in meaning. While with prose memory for surface characteristics declined over time, as expected, the same did not happen with poetry, for which form, in addition to content, was efficiently retrieved (Tillmann & Dowling, 2007).

With this study, we had hoped to be able to quantify, in rather absolute terms, the contribution of different aspects of meter to memory retention, using “material” from the classical period of Italian literature, before the advent of the printing press diminished the perceived value of memorability per se, and promoted the further ritualization of the written verse into a primarily esthetic construct. The results belied our naïve expectation, in that meter seems to be ‘perceived’ much more (in terms of our plausibility index) in passages derived from Ariosto than in those from Dante, and to contribute to memory in the former but not in the latter. Yet the meter employed by the two authors is nearly identical, with a discrete difference in the concatenation of hendecasyllables (in terzine in the Divina Commedia, in ottave in the Orlando Furioso) and a presumably small quantitative difference in the variability with which the common meter is used (Figure 3). Therefore, one would expect that the listener, or the reader, activates the very same cognitive schema, at least locally, within the few verses of a single ottava or three terzine.

The time the subjects from the two statistically indistinguishable cohorts needed to react to the memory tests suggest an account of the main finding: the metric schema is the same, but it is activated to a different extent. Somewhat counterintuitively, it appears to be activated less with Dante, an author with whom most people who have been in high school in Italy are rather familiar, than with Ariosto, who has been relegated, especially in the last few decades, to a marginal niche in standard Italian curricula. This appears to discount a possible interpretation of this difference, i.e., that we are seeing two competing effects, whereby both congruence and incongruence with established schemata can enhance memory, the latter a novelty effect (Bonasia et al., 2018): novelty presupposes the activation of the schema it contradicts. An alternative interpretation is that Dante’s verses are just more interesting and tend to focus one’s attention to other aspects than the components of meter. Even if this interpretation were to be shown to be correct, it is quite surprising that it would apply, in our paradigm, to verses that have been deprived of their meaning. Moreover, our replacing several of the key words chosen by Dante with our untalented choice of non-words would have been expected to remove other potentially useful devices from the poet’s bag-of-tricks, like alliteration, onomatopoeia, use of liquids, of newly crafted words, etc. (Robey, 1985).

Can the hypothesis of differential schema activation be tested experimentally? In principle, yes, and one approach would be by looking at evoked response potentials (ERPs), which have been widely used to reveal brain signals that reflect violations of expectation, whether (in the language domain) syntactic, semantic, or just phonological (Brown & Hagoort, 1993; Hagoort, 2003). With poetry, there have been ERP studies of aesthetic appreciation and ease of processing (Obermeier et al., 2016) and of brain activity during poem composition (Liu et al., 2015). For an experimental design like ours, however, one challenge is how to obtain the large number of trials per condition needed in order to obtain valid ERP measures. Another one is to what extent one can rely on single ERPs to characterize a heterogeneous variety of metric components. It is possible that addressing both challenges will require a change in perspective from whole brain dynamics to one which articulates the cortex into a plurality of interacting local networks, as embodied e.g., by the Potts model (Naim et al., 2018). Distinct processes, among the many that concur to the overall perception, appreciation and memory of a poem, including the components of meter, are likely reflected differentially in the dynamics of distinct cortical networks, just like other, better studied types of memory such as episodic and spatial memories (Robin & Moscovitch, 2017), which have stimulated theories about the interactions between the medial temporal lobe and medial prefrontal cortex (Van Kesteren et al., 2012). Meter, with its multiple components, indicates the need to go beyond the somewhat coarse distinctions available to neuropsychology, what is at the moment accessible only through mathematical models (Spalla et al., 2020), with which one can study forms of partial coherence among multiple local networks, reminiscent of that of systems unable to attain long-range order (Stella et al., 2020).

While partial coherence might seem to detract from the wholeness attributed to conscious processing (Dehaene et al., 1998), it is entirely consistent with the idea of a mixture of automatic and controlled components concurring to memory.
encoding and retrieval (Wang & Morris, 2010). With meter, the notion that different schemata might be activated only optionally, at times, and then partially and incoherently with others, and when activated might offer only an incremental contribution, suggests a more nuanced take on high-level cognition in general. Using many filters to interpret reality, and to a variable degree, implies check-and-balances and minimal recourse to prevailing or dominant schemata, those that often reflect biases or prejudice.

Data availability
Underlying data
Repository: Meter and Memory. DOI: 10.17605/OSF.IO/A825X (Andreetta et al., 2021).

The project contains the following underlying data:

- all Arioisto.xlsx (Responses to the non-poems derived from the Orlando Furioso).
- all Dante.xlsx (Responses to the non-poems derived from the Divina Commedia).

Extended data
Repository: Meter and Memory. DOI: 10.17605/OSF.IO/A825X (Andreetta et al., 2021).

The project contains the following extended data:

- Andreetta_ExtendedData.pdf (the non-poems and the associated images)

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Acknowledgements
An earlier version of this article can be found on bioRxiv (https://doi.org/10.1101/2021.03.14.435310)

We are grateful to the HFSP collaboration on Analog computations underlying language mechanisms, in particular to Yair Lakretz, with whom the original idea for this study was discussed, and to Elisa Ciaramelli, Rodolfo Zucco, Sergio Bozola, Andrea Tabarroni and Sara Alzetta, who offered different perspectives.

References


Publisher Full Text

Publisher Full Text

Publisher Full Text

Publisher Full Text

Publisher Full Text

Publisher Full Text

Publisher Full Text

Publisher Full Text

Publisher Full Text

Publisher Full Text

Publisher Full Text

Reference Source

PubMed Abstract | Publisher Full Text

PubMed Abstract | Publisher Full Text | Free Full Text

Reference Source

Publisher Full Text

Publisher Full Text

Reference Source

Publisher Full Text

PubMed Abstract | Publisher Full Text

Reference Source
Spalla D, Cornachia IM, Treves A: Continuous attractors for dynamic...

Publisher Full Text

PubMed Abstract | Publisher Full Text

Publisher Full Text

Reference Source

PubMed Abstract | Publisher Full Text

PubMed Abstract | Publisher Full Text

PubMed Abstract | Publisher Full Text

PubMed Abstract | Publisher Full Text
Comments on this article

Version 1

Reader Comment 16 Jul 2021

Johann-Mattis List, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Thanks for this very interesting study. As a linguist interested in the evolution of poetry across languages and cultures, the work you are doing is of crucial importance. What I miss from the current study, however, are more explicit explanations on the data which you have shared (detailed description of column names, which information is used where in the article, etc.), and also that you share more detailed information on the software that was used for plotting. For example, you mention the use of the SUBTLEX-IT data for assessing word frequencies, but I had to look quite a bit when I was trying to find where in the data files you had this information provided. In order to avoid that readers interested in the details of your methods have to second-guess what part of the data relates to what part of the article, it is always recommended to be very verbose about the data, ideally providing a README file that provides all necessary information, specifically explaining what one can find in which column. As a scientist who has been struggling a lot with studies in which code is not being shared fully, I'd also recommend to share your plotting code for the individual data plots, also in order to allow young scholars to learn from your expertise. Thanks again for this very interesting study. I am curious to see the reviews.

Competing Interests: No competing interests were disclosed.