Distribution Agreement

Signature:

______________________________ _______________

Aoxiang Xu Date

Autobiographical Retrieval Revealed in the Eyes of the Narrator: Dynamic Changes in Pupillometry

By

Aoxiang Xu
Master of Arts

Psychology

Patricia J. Bauer, Ph.D.
Advisor

Robyn Fivush, Ph.D.
Committee Member

Stephan Hamann, Ph.D.
Committee Member

Joseph Manns, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Autobiographical Retrieval Revealed in the Eyes of the Narrator: Dynamic Changes in Pupillometry

By

Aoxiang Xu
M.S., Beijing Normal University, 2011
B.S., Jilin University, 2006

Advisor: Patricia J. Bauer, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Arts
in Psychology
2016

Abstract

Autobiographical Retrieval Revealed in the Eyes of the Narrator: Dynamic Changes in Pupillometry

By Aoxiang Xu

Autobiographical memory (AM) refers to memory for personally significant events and experiences. Though the retrieved outcome of autobiographical memory (e.g., narratives) have been well researched, the dynamic processes involved in search and access and subsequent elaboration of autobiographical memories are only beginning to be explored. Moreover, it is unclear how factors such as gender and emotion influence the AM retrieval process. In the present study, we developed a novel application of pupillometry to study effects of gender and emotion as participants retrieved and typed narratives of emotional autobiographical events. Participants produced more words and had a longer duration of typing for emotional narratives than non-emotional narratives; females used more words and typed for longer relative to males. A longitudinal analysis applied to the pupillary dilation revealed a curvilinear trend of pupil dilation. The effects of gender emerged over time. For males only, pupil size was negatively correlated to the number of emotional words in their narratives. This study confirmed the promising application of pupillometry to the study of AM retrieval, and further explored the temporal process of the effects of gender and emotion on AM retrieval.

Autobiographical Retrieval Revealed in the Eyes of the Narrator: Dynamic Changes in Pupillometry

By

Aoxiang Xu
M.S., Beijing Normal University, 2011
B.S., Jilin University, 2006

Advisor: Patricia J. Bauer, Ph.D.

A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Arts
in Psychology
2016

TABLE OF CONTENTS

As the recollection of personal experiences, autobiographical memory (AM) is characterized by its complex, multi-faceted representation (Rubin, 2005). Consider the example of recalling a wedding one have recently attended, the details of which may include the buildings, people, conversations, music, feelings, and so forth. The involvement of multiple sensory modalities, compared to other types of episodic memories based on unisensory modality (e.g., vision or hearing), makes AM difficult to be manipulated in experiments in laboratory environments. Moreover, retrieval of AMs involves processes that unfold over time, including search for and access of the memory, followed by reactivation of the entire pattern of activity associated with initial encoding, and not infrequently, reconstructive processes that "fill in" some of the bits and pieces that have gone missing as a result of incomplete encoding and consolidation. Consequently, although the dynamic cognitive process of AM is widely recognized, it has not been well researched, due to a shortage of effective methods. In the present study, we addressed this issue by developing a real-time pupillometry measurement that we employed as participants retrieved AMs and typed narrative descriptions of them. We used changes in pupillometry as a "window" on the ebb and flow of effort throughout the retrieval process.

Like other types of memories, retrieval of AM takes time. According to the Self-Memory System (SMS) model presented by Conway and Pleydell-Pearce (2000), autobiographical information is stored in a network of interconnected memory nodes at different levels of specificity. The activation of memory nodes will spread upwards or horizontally as AMs are retrieved. fMRI studies confirmed that AM retrieval is a temporally extended process. Cabeza and St Jacques (2007) summarized the data on AM retrieval, noting the activation of left lateral prefrontal cortex (PFC) in constructive process of AM, and the continuous activation of amygdala, hippocampus, and visual cortex during the retrieval of AM. For example, Daselaar and colleagues (2008) dissociated the time course of cue presentation, memory access, elaboration, and rating of memory during an fMRI session of AM retrieval based on participants' button presses. They observed greater activation of hippocampus, retrosplenial cortex, and right lateral PFC during the access of memories. The activation of visual cortex, precuneus, and left lateral PFC ascended as the elaboration phase started.

In the present research, we examined relations between the temporal dynamics of AM retrieval and behavioral performance as AM were retrieved and elaborated. Specifically, we used pupillometry to examine changes in cognitive processing as individuals created typed narratives about autobiographical experiences. Pupillometry refers to the study of changes in the diameter of the pupil, as a function of cognitive processing (Sirois & Brisson, 2014). The change of the diameter of the pupil has been observed to reflect the "intensity" of mental activity and the changes in mental states (Laeng, Sirois, & Gredebäck, 2012). For example, dilated pupillary diameter has been observed in a short-term memory task (Kahneman & Beatty, 1966). Participants were instructed to remember and report digit strings. The pupil dilated while the stimulus was presented and constricted during the report. The rates of change of both dilation and constriction were positively correlated with the span of the string of digits to be remembered. These findings establish the link between dilation of the pupil and cognitive load.

Cognitive neuroscience studies have provided neural evidence of the relation between pupillary dilation and the activation of locus coeruleus (LC) which plays a role in regulation of attention. LC is the key brain region evoking the sole source of the neuro-transmitter norepinephrine (NE) to the cortex, cerebellum, and hippocampus (see Aston-Jones & Cohen, 2005 for a review). The activation of LC leads to the secretion of NE through the hypothalamic-pituitary-adrenal axis, and alters the activation of PFC. Thus the LC-NE system is considered to mediate shifts of attention (e.g., Sara, 2009; Sterpenich et al., 2006). This relation has been confirmed through single-cell recording studies in monkeys, which have revealed a tight correlation between the changes in pupillary diameter and the changes in activity in LC neurons (e.g. Rajkowski, Kubiak, &Aston-Jones, 1993; Rajkowski et. al., 2004). In summary, pupillary response is considered to reflect the activity of LC-NE system and the change of cognitive effort.

Historically, pupillometry has been considered as an index of cognitive process (Hess & Polt, 1960). For decades, pupillometry has been utilized to study cognitive process across a broad range of psychological domains, from perception (e.g. Rayner & Pollatsek, 1992), to language (e.g. Rayner, 1998), to short term memory (e.g. Heitz, Schrock, Payne, & Engle, 2008). As the role of Locus Coeruleus in memory retrieval has been confirmed (Sterpenich et al., 2006), the task-evoked pupillary reflexes (TEPRs) presented and implicated a role of memory processes. A small but growing trend of investigating long-term memory using pupillometry is evident in the literature, from recognition to retrieval of episodic memory (e.g. Goldinger & Papesh, 2012; Papesh & Goldinger, 2011; Papesh, Goldinger, & Hout, 2012). To date, there has been no use of the pupillary response to study AM. Since pupillometry has long been proven as an effective index of cognitive effort and workload (e.g. Marshall, 2002; Piquado, Isaacowitz, & Wingfield, 2010), the recording and measurement of pupillary response during the AM retrieval could precisely reflect the change of the cognitive effort spent across the whole session.

Pupillometry also holds promise for studying the temporal processes of the influence of factors like gender and emotion on AM retrieval. Gender difference has been observed in AM research. That is, several studies have reported that females report more vivid memories, relative to males, and they include more details in their reports (see Grysman & Hudson, 2013 for a review). What is unknown is when the gender difference emerges in the process of AM retrieval. Is the gender difference consistent, decreasing, or increasing throughout the retrieval session? These questions are ripe for address by comparing the change of pupillary response across females and males.

The temporal dynamic of emotion's effect on AM is also unclear. Analyses of reports of personal experience have found that more perceptual details are included in descriptions of emotional experiences than non-emotional experiences (e.g. Comblain, D'Argembeau, & Van der Linden, 2005; Conway et al., 2009). However, it is unclear how this effect works during the process of AM retrieval. With the aid of pupillometry, we expected to locate the timestamp where the emotion's effect started, and to track the potential difference as time flows.

Another advantage of the use of pupillary dilation is that it can be linked to gender differences and emotional expression in narratives. For example, to understand the difference in the affective qualities in autobiographical memories, a main method is to code the reports for mention of internal states terms, like emotions, cognitions, perceptions, and physiological states (e.g. Bauer, Stennes, & Height, 2003). By correlating the pupillary dilation to the number of specified coded terms, the source of the cognitive effort might be traced. A similar strategy can be followed for examining gender differences.

Because highly correlated to the mental activity, other mental-state related factors, like fatigue, excitement, and sleepiness, could also influence the pupillary dilation (e.g. see Goldwater, 1972 for a review). The effect of fatigue is vital for a study using pupillometry to explore the temporal dynamics of cognitive process. The fatigue effect reflects the fluctuating activation of the autonomic nervous system, which governs the change of pupillary response. For instance, a visual fatigue study found that after 24 hours of sleep deprivation, people's average pupil size was negatively correlated with the change of level of sleepiness (Morad, Lemberg, Yofe,& Dagan, 2000). In the present research, we conducted analyses to examine the possible influence of fatigue on performance.

In the present study, we utilized pupillometry as a measure of the cognitive effort during AM retrieval process. We were also interested in the availability of pupillary response as an index to the effect of gender and emotion. Accordingly, we developed a narrative production paradigm in which participants typed descriptions of their personal experiences based on given emotional cue word (i.e., POSITIVE, NEGATIVE, OR NEUTRAL). Because searching for and accessing a memory happens at the beginning of memory retrieval, and the information about the retrieved event that is "mentally" held in mind will be gradually decreased as it is committed to writing (which is the "typing" behavior in the experiment), constricting pupil size, reflecting decreasing cognitive effort throughout the AM retrieval, was expected. Based on previous studies about gender difference and emotional AM, we also expected more dilated pupil in females than males, and in emotional memories than non-emotional ones.

Methods

Participants

Twenty-two native English-speaking undergraduate students at Emory University (11 females, 11 males) participated in the study (Mean age = 19.6 years old, SD = 0.98). All participants had normal or corrected-to-normal vision. Participants were required to be "touch typists"—they could type without frequent looks to the keyboard to locate the desired keys. Participants read and signed an informed consent form before the start of the experiment. After the experiment, participants received course credit as compensation. The protocol was approved by the Emory University Institutional Review Board. Five additional participants were recruited but excluded because they did not follow task instructions ($N = 3$) or the recorded pupillometry samples were below the threshold (lower than 70%; see below; $N = 2$).

Procedure

Participants were tested individually in a laboratory setting. They were instructed to type a total of six narratives about personal experiences, each in a separate, blank Microsoft Word® document. Three words (positive, negative, and neutral) were presented as cues for the emotional valence of the experience the participants were to describe. Each cue word was shown twice, for a total of six narrative reports. We developed three pseudo-random orders in which the narratives were to be produced, constrained such that cue words for the same emotional valence did not appear in immediate succession. Each order was used approximately equally often across participants. The instruction was as follows:

"In this experiment, we would like for you to type your very deepest thoughts and/or feelings about a specified experience that occurred within the last year. There will be 6 experiences presented in total. For each experience, you will first see a word specifying the type of experience you should write about. When you are ready to provide your memory, you will then press the space bar and type your experience within the word document that appears. In your writing, please describe your experience in detail, and explore your very deepest emotions and thoughts. Do not worry about spelling, sentence structure, or grammar. On the keyboard, you can use any letter, number, symbol, and arrow keys as well as the space bar, and the enter key. However, the mouse and function keys like Home, End, and Page Up/Page Down should not be used. If you need to make corrections, please use the arrow keys and backspace instead of the mouse. Please limit your writing to no more than one page. All of your writing will be completely confidential. Please press the Esc key when you have finished your writing."

The computer was set to hide the Microsoft Office® menu, word count and grammar correction. At the start of each trial, the cue word (positive, negative, or neutral) was shown on the middle of screen, indicating what kind of experience to be described. The cue was visible until the participant pressed the space bar on the keyboard and began to type.

We monitored participants' keystrokes and the diameter of their pupils throughout production of each narrative. We also measured word typing speed and the total word count for each narrative. We did not measure fixations or saccades because participants made frequent visual references to what they had just typed. As a result, we considered these behaviors uninformative regarding narrative production.

Eye tracking

The pupillary dilation was measured using a Tobii T120 eye tracker with a sampling rate of 60 Hz. Calibrations were done before the data collection of each narrative production task. To avoid the influence of light reflex on pupillary response, the data record of the first second was excluded from these calculations and from analyses. The threshold of the ratio of recorded samples during the whole narrative production was set to 70%. Narratives with lower than 70% were excluded.

Coding

Internal states terms in the narratives were coded into one of three, mutually exclusive categories: (1) emotion (e.g., happy, sad), (2) mental (e.g., think, guess), and (3) evaluation (e.g., almost, special). Internal states reflecting perceptual and physiological experiences were not coded. The terms were categorized according to their meaning in the narrative. Examples from each category are provided in the Appendix A. All of the 122 narratives were coded by one coder, and 27(22%) narratives were independently coded by another coder. Cohen's Kappa results were .87 for the mental category, .78 for emotion category, and .73 for evaluation category, indicating that reliability was established.

Results

Narrative Production Performance

Descriptive statistics for the mean number of words and mean duration of narrative production (i.e., typing) are provided in Table 1. To test for differences in the mean number of words included in the narrative and the mean duration of typing in each condition, we conducted 2(gender: female &male) x 3(emotional valence: neutral, positive, and negative) mixed analyses of variance (ANOVA) for each dependent variable. Tukey's HSD multiple pairwise comparison was used to make all pair-wise comparisons. Significant main effects of gender were found for both the number of words included in the narratives, $F(1,61) = 10.75$, $p = .002$, $\eta^2 = .13$, and the mean duration of typing, $F(1,61) = 4.07$, $p = .05$, $\eta^2 = .06$. Across valence conditions, females included more words in their narratives and spent more time typing relative to males (*p*s < .05). There also was a main effect of valence for the number of words included in the narratives, $F(2,59) = 6.24$, $p = .003$, $\eta^2 = .15$. Across gender groups, participants featured more words in their narratives about both positive and negative events than neutral events (*p*s < .05). The number of words in narratives about positive and negative events did not differ ($p = 1.00$). The was not a significant main effect of valence for the duration of typing and there were no interactions between the variables, for either dependent measure.

Descriptive statistics for the use of internal states terms in narratives are provided in Table 2. The total number of words in the narratives was correlated with the total number of internal states terms, as well as with the number of mental terms, emotion terms, and evaluation terms, $r(20) = .85, .75, .65, .71$; *p*s <.001. Consistent correlations were found in males, r(9) = .93, .81, .78, .87; *p*s < .01. For females, the total number of words in the narratives was correlated with the total number of internal states terms and the number of mental terms, r(9) = .69, .75; *p*s < .05, but not correlated with the number of emotion terms and evaluation terms, r(9) = .46, .25; *p*s = .15, .46. The correlation results revealed that participants who produced longer narratives also tended to include more internal states terms. In light of these relations, we conducted analyses of use of internal states terms as a function of gender and emotional valence of the events first without controlling for narrative length and then with narrative length controlled, to determine whether the narratives were differentially saturated with internal states terms. For the analysis without narrative length controlled, we conducted 2(gender: female and male) x 3(emotional valence: neutral, positive, and negative) ANOVAs for the total number of internal states terms used and for each category of internal states terms separately. For the analysis with narrative length controlled, we conducted a parallel ANCOVA.

Total internal states terms. The Gender × Emotional valence ANOVA revealed main effects of gender and emotional valence (Gender: $F(1,59) = 12.86$, $p = .003$, $\eta^2 = .12$; Valence: $F(2, 59) = 4.64$, $p = .04$, $\eta^2 = .09$). Across valence conditions, females used more internal states terms than males ($p = .004$). Across gender groups, pairwise comparisons revealed that more internal states terms were used in negative narratives than neutral narratives ($p = .04$). The use of internal states terms in positive narratives did not differ from negative or neutral narratives (ps > 0.10). Significant main effects of gender and emotional valence were also found in the ANCOVA (Gender: $F(1,58) = 20.27$, $p < .001$, $\eta^2 = .004$; Valence: $F(2, 58) = 7.69$, $p = .001$, $\eta^2 = .004$). Thus the gender effect remained. With the length of narrative controlled, pairwise comparison revealed that across gender groups, participants included more total internal state terms in their narratives about both positive and negative events than neutral events (*p*s<.05); the number of internal state terms used in positive and negative events did not differ. The interaction was not significant.

Mental terms. There was a main effect of gender, F(1,59) = 5.7, $p = .02$, $\eta^2 = .09$. Across emotional narratives, females used more emotion terms than males. The effect remained significant in the ANCOVA, F(1,58) = 8.52, $p = .005$, $\eta^2 = .003$. Whether or not narrative length was controlled, the main effect of emotional valence was not reliable. The effects were not qualified by interactions.

Emotion terms. There was a main effect of gender, F(1,59) = 7.32, $p = .01$, $\eta^2 = .10$. Across emotional narratives, females used more emotion terms than males. The effect remained in the ANCOVA, F(1,58) = 9.32, $p = .003$, $\eta^2 = .01$. Whether or not narrative length was controlled, the main effect of emotional valence was not reliable.

Evaluation terms. In contrast to the other internal state analyses, there was not a main effect of gender for use of evaluation term, either without or with the control for narrative length. There was a main effect of emotional valence, F(2,59) = 3.21, $p = .05$, $\eta^2 = .04$. Pairwise comparisons revealed that participants used more evaluation terms in narratives about negative events than in narratives about neutral events ($p = .05$). The use of evaluation terms in narratives about positive events did not differ from narratives about either negative or neutral events. The ANCOVA also revealed a main effect of emotional valence, F(2,58) = 4.78, $p = .01$, $\eta^2 = .006$; the pattern of relations among valence conditions was the same as in the ANOVA. The effects were not qualified by an interaction.

Multilevel Model of Pupillometry

Because the length of time participants spent typing varied (see above), we standardized the writing interval. To do so, for each of the 6 narratives, we divided the overall duration of typing by 100. To examine changes in pupillometry over time, for each 1% of overall duration, we averaged together the observations that fell into that bin. The average number of observations in each bin was 147.95 (SD = 73.61; range = 13-434). The trajectory of change in pupil size over time is depicted in Figure 1.

A longitudinal analysis was conducted in which gender and emotional valence were used to predict the change of pupil size over time. The females' pupil size during production of neutral narratives was arbitrarily set as the baseline. To find the model that best reflected the relation between the factors and pupil size, two two-level conditional growth models were built and compared: (1) the linear model reflected the linear change of pupil size over time, and (2) the quadratic model reflected the quadratic change of pupil size over time. A dummy variable, Gender, was created for the two-level categorical variable gender. Two dummy variables, Diff-Positive and Diff-Negative, were created for the three-level categorical variable of emotion. Diff-Positive refers to the difference between positive and neutral conditions, and the Diff-Negative refers to the difference between negative and neutral conditions. Improvements in model fit were evaluated by comparing fit statistics (-2 log likelihood statistic) across the two nested models. Results revealed that the quadratic model significantly improved model fit, $\chi^2(8)$

= 1202.5, p <.001, implying that affected by the factors of gender and emotional valence, the change of pupillary size showed a curvilinear trend over time.

For the quadratic model, the fixed effects from factors are provided in Table 3. As shown in the table, fixed effects predicting the change in pupil size showed significant effects for both the linear slope (the "Time" variables) and quadratic slope (the "$Time^2$" variables). Significant effects of gender difference were observed in the linear and quadratic slopes (shown as Gender*Time and Gender*$Time^2$ in the table), but not in the initial state (shown as Gender in the table). Thus, as suggested by inspection in Figure 1, the pupillary diameters of females and males did not differ at the beginning of the narrative production, but differed in a curvilinear trend over time.

For females, the effect of emotion was tested by comparing positive and negative narratives to neutral narratives labelled by Diff-Positive and Diff-Negative. Reflected by the significant contributions from DN, DN*Time, and DN* $time^2$, the pupillary diameter during production of narratives about negative events is apparent in the initial state, the linear slope, and the quadratic slopes. That is, retrieving and elaborate negative experiences elicited more dilated pupillary responses from the beginning of narrative production. The difference increased over the course of narrative production. For production of narratives about positive events, no statistically significant contributions were observed from DP, DP*time, and DP*$time^2$, indicating that the pupillary response during the production of narratives about positive events did not differ from the neutral narrative production.

Because the analysis reported in Table 3 used females' neutral condition as the baseline, the emotion's effect described above is for females only. To examine whether the effect of emotion on males is identical to the effect on females, we conducted a Wald test. The fixed effects of males' coefficients were calculated by adding the baseline (females) and Gender (gender difference). For example, the males' pupillary response in the linear time slope was calculated as Time + Gender*Time. Results revealed that males' pupillary responses during production of narratives about both positive and negative events differed from their pupillary responses during production of narratives about neutral events in the initial state, the linear slope, and the quadratic slope.

Analysis of Relations between Pupillary Response and Internal States Terms

The correlations between average pupil size and the length of the narratives and duration of typing were not statistical reliable (*p*s > .05). Nevertheless, because participants who produced longer narratives also tended to include more internal states terms in their narratives, in analyses of relations between pupil size and internal states measures, we controlled for the total number of words in the narrative.

With the number of words controlled, correlation analysis was conducted between the average pupil size during narrative production and the number of internal states terms across narrative types. In males only, pupil size was positively correlated with use of evaluation terms, $r(8) = .60$, $p = .06$, and negatively correlated with use of emotion, terms to describe events, $r(8) = -0.58$, $p = .07$. No significant correlations were found in females between pupil size and any categories or the sum of internal states terms.

Possible Effects of Fatigue on Performance

In this study, fatigue may have affected participants' performance in two ways. First, participants could experience greater fatigue at the end of each narrative relative to the beginning. Second, participants could experience greater fatigue while typing the last narrative relative to the first narrative. To explore these potential effects, the behavioral performance and pupillary diameter were analyzed within the narrative production, and across different narratives.

We examined the possibility of fatigue effects over the course of typing of a narrative by examining the speed of typing over the course of a narrative. As shown in Figure 2, typing speed was calculated after the standardization of duration. Typing speed was stable over most of the narrative, decreasing only at the very end of the narrative production. The sharp decrease of typing speed at the end corresponds with participants' decision to finish the current narrative, and end current session. Because a fatigue effect should have been reflected in lower speed of typing over the course of the narrative, the observed result suggests that fatigue was not observed within a narrative.

We also examined possible fatigue effects over the course of the study by comparing the curves of pupil dilation during participants' first and last narrative productions. The results are shown in Figure 3. An independent samples *t* test of the average pupil size was conducted between the first and last narrative production task. The pupil size was significantly constricted in the last session relative to the first session ($t(35.95) = 2.16$, $p = 0.04$). However, the effect was not consistent with a fatigue explanation, per se. In previous studies, fatigue effects have been observed as fluctuations of pupil size (see Sirevaag and Stem, 2000 for a review), which was not the pattern found in our data. As well, although the pupil was more constricted in the last task, the changes of pupil size in the first and last tasks were identical. Taken as a whole, the findings suggest that fatigue effects did not affect the pattern of the pupil dilation.

Discussion

In the present study, we developed the method of pupillometry to examine the process of AM retrieval. A longitudinal analysis of the pupillary response over time was conducted to depict the temporal process of AM retrieval. We also evaluated correlations between pupil size and use of internal states terms in the narratives. Based on the results, several major findings are found, which in turn validated the availability of the pupillometry in AM study.

First, reflected by the change of pupil size, the pattern of cognitive effort changed over the course of AM retrieval. Based on current models of memory processes (Conway & Pleydell-Pearce, 2000; Tulving, 1983), retrieval of autobiographical memories requires a series of cognitive activities. The to-be-retrieved memory must be sought out, accessed, and selected. Simultaneously, other competing memories must be inhibited. Then the recovery of the selected memory occurs (Tulving, 1983). The activated information, which forms the retrieved AM, will be maintained and elaborated based on the ongoing task. This view is supported by the observed pattern of pupillary response. The peak of cognitive effort was at the beginning of the retrieval process, corresponding to the cognitive activities of search, access, activation, and inhibition. Gradually, less cognitive effort was expended, presumably as participants reported aspects of the memory, thus eliminating the need for them to be maintained. The longitudinal analysis also revealed that a quadratic model can better predict the change of pupil size during the AM retrieval. This means the change of cognitive effort over time is best described by a curvilinear trend.

The second major finding from the present research concerned gender differences in AM retrieval. Consistent with previous studies, as they produced their narratives, females spent more time, and wrote more words than males. However, multi-level analysis indicated that the gender difference was not present at the initial state of AM retrieval. Instead, it emerged during the narrative production. Significant gender difference also changed in the linear and quadratic slope, reflecting a difference in the acceleration speed of pupillary dilation. Given that retrieval of AM is considered to entail re-experiencing of events (Tulving, 1983), the onset of the gender difference might be located in the elaboration process, which suggested females devoted more cognitive effort as they re-experienced the events than males. Similar result can be found in Ellermeier and Westphal's study (1995). While experiencing painful pressure, females' pupils dilated more than males as the pressure level of stimuli elevated.

Females and males also differed in the sources of their cognitive effort. Analysis of the frequency of using internal states terms revealed that females used more internal states terms than males, overall. More emotion terms and mental terms were used by females than males as they narrated their personal experiences. This finding is consistent with previous studies (e.g. Bauer, Stennes, & Height, 2003). Females and males both used more evaluation terms in negative narratives; no gender difference was observed. However, correlational analysis between pupil size and the use of internal states terms indicate that only males' pupillary dilation was negatively correlated to the emotion terms. Given that cognitive effort must be spent to inhibit or to execute a response, the effect may reveal that males might make extra effort to talk in an objective view, and avoid emotional expression.

The present study proved the availability of monitoring the cognitive process of autobiographical memory over time, with the combination of pupillometry and longitudinal analysis. As the statistic method for series of observations obtained over time (Fitzmaurice, Laird, & Ware, 2012), longitudinal analysis has been widely used in longitudinal studies, like the children development of memory (e.g. see Bauer & Wewerka, 1997; Jorm, Christensen, Korten, Jacomb, & Henderson, 2001). Coupled with the pupillometry, which has been confirmed as the index of memory process, this method can be broadly utilized in memory studies focusing on the dynamic cognitive process. It shows promise for informing the dynamic processes underlying effects of gender and emotional valence on memory retrieval. Moreover, the source and trace of changing cognitive effort during AM retrieval might be better understood with collaborative use of pupillometry and neuro-imaging methods.

Finally, because the duration of narrative production was hundreds of seconds, we evaluated the possible effects of fatigue on the pupillary dilation. Several factors led us to reject fatigue effects as the controlling source of variance in the pupillary response. Firstly, the paradigm allowed participants to decide when to terminate narrative production, which avoided the fatigue effect brought by forced typing. Secondly, if the fatigue effects emerged in the narrative production task, it would not only affect the change of pupil size, but also slow the speed of typing. However, we found stable typing speed across the whole narrative production task, as shown in the Figure 2. Thirdly, the fatigue effect, if any, should emerge after some time spent in narrative production. Other than the rising trend at the beginning of narrative production, we did not observe such an effect, such as the fluctuations of pupil dilation, on the change of pupil dilation. And fourth, the gender difference reflected in the pupil dilation cannot be interpreted by the effect of fatigue.

The pupillometry method used in narrative production task in this study successfully revealed the change of cognitive effort over time. Yet one question this study raised is how much of the cognitive effort was spent in language processing per se, as opposed to narrative production. Organizing language is indispensable for narrative production task, but whether the corresponding cognitive effort is stable or not during the task is unknown. To ensure the cognitive effort reflected by the pupillary response is accurately representing the AM retrieval, the manipulation of the cognitive load of language processing in the task should be considered in the future direction.

In conclusion, the present study makes clear that a curvilinear trend of cognitive effort was made during the retrieval of autobiographical memory. The gender difference reflected in the pupil size emerged overtime. Moreover, the correlation between pupil size and internal state terms coding suggests males made extra effort in avoiding emotional expression and talking in an objective view.

References

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance.*Annu. Rev. Neurosci.*, *28*, 403-450.

Bauer, P. J., Stennes, L., & Haight, J. (2003). Representation of the inner self in autobiography: Women's and men's use of internal states language in personal narratives. *Memory*, *11*(1), 27-42.

Bauer, P. J., & Wewerka, S. S. (1997). Saying is revealing: Verbal expression of event memory in the transition from infancy to early childhood.

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology, 45*(4), 602-607.

Bauer, P., Stennes, L., & Haight, J. (2003). Representation of the inner self in autobiography: Women's and men's use of internal states language in personal narratives. *Memory*, 11(1), 27-42.

Cabeza, R., & St Jacques, P. (2007). Functional neuroimaging of autobiographical memory. *Trends in cognitive sciences, 11*(5), 219-227.

Comblain, C., D'Argembeau, A., & Van der Linden, M. (2005). Phenomenal characteristics of autobiographical memories for emotional and neutral events in older and younger adults. *Experimental aging research*, *31*(2), 173-189.

Conway, A. R., Skitka, L. J., Hemmerich, J. A., & Kershaw, T. C. (2009). Flashbulb memory for 11 September 2001. *Applied cognitive psychology*,*23*(5), 605-623.

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological review*, *107*(2), 261.

Daselaar, S. M., Rice, H. J., Greenberg, D. L., Cabeza, R., LaBar, K. S., & Rubin, D. C. (2008). The spatiotemporal dynamics of autobiographical memory: neural correlates of recall, emotional intensity, and reliving. *Cerebral Cortex, 18*(1), 217-229.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). *Applied longitudinal analysis* (Vol. 998): John Wiley & Sons.

Goldinger, S. D., & Papesh, M. H. (2012). Pupillary dilation reflects the creation and retrieval of memories. *Current Directions in Psychological Science, 21*(2), 90-95.

Goldwater, B. C. (1972). Psychological significance of pupillary movements.*Psychological bulletin*, *77*(5), 340.

Grysman, A., & Hudson, J. A. (2013). Gender differences in autobiographical memory: Developmental and methodological considerations. *Developmental Review, 33*(3), 239-272.

Heitz, R. P., Schrock, J. C., Payne, T. W., & Engle, R. W. (2008). Effects of incentive on working memory capacity: Behavioral and pupillometric data. *Psychophysiology, 45*(1), 119-129.

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science, 132*(3423), 349-350.

Jorm, A., Christensen, H., Korten, A., Jacomb, P., & Henderson, A. (2001). Memory complaints as a precursor of memory impairment in older people: a longitudinal analysis over 7–8 years. *Psychological medicine, 31*(03), 441-449.

Kahneman, D., & Beatty, J. (1966). Pupillary diameter and load on memory. *Science, 154*(3756), 1583-1585.

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry A Window to the Preconscious? *Perspectives on Psychological Science*, *7*(1), 18-27.

Marshall, S. P. (2002). The index of cognitive activity: Measuring cognitive workload. In *Human factors and power plants, 2002. proceedings of the 2002 IEEE 7th conference on* (pp. 7-5). IEEE.

Morad, Y., Lemberg, H., Yofe, N., & Dagan, Y. (2000). Pupillography as an objective indicator of fatigue. *Current eye research*, *21*(1), 535-542.

Papesh, M. H., & Goldinger, S. D. (2011). Your effort is showing! Pupillary dilation reveals memory heuristics. *Constructions of remembering and metacognition*, 215-224.

Papesh, M. H., Goldinger, S. D., & Hout, M. C. (2012). Memory strength and specificity revealed by pupillometry. *International Journal of Psychophysiology*, *83*(1), 56-64.

Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology*, *47*(3), 560-569.

Rajkowski, J., Kubiak, P., & Aston-Jones, G. (1993). Correlations between locus coeruleus (LC) neural activity, pupillary diameter and behavior in monkey support a role of LC in attention. *Society for neuroscience abstracts* , 19, 974.

Rajkowski, J., Majczynski, H., Clayton, E., & Aston-Jones, G. (2004). Activation of monkey locus coeruleus neurons varies with difficulty and performance in a target detection task. *Journal of Neurophysiology*, *92*(1), 361-371.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin, 124*(3), 372.

Rayner, K., & Pollatsek, A. (1992). Eye movements and scene perception. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *46*(3), 342.

Reese, E., Haden, C. A., Baker-Ward, L., Bauer, P., Fivush, R., & Ornstein, P. A. (2011). Coherence of personal narratives across the lifespan: A multidimensional model and coding method. *Journal of Cognition and Development*, *12*(4), 424-462.

Rubin, D. C. (2005). A basic-systems approach to autobiographical memory. *Current Directions in Psychological Science, 14*(2), 79-83.

Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews Neuroscience, 10*(3), 211-223.

Sirois, S., & Brisson, J. (2014). Pupillometry. *Wiley Interdisciplinary Reviews: Cognitive Science, 5*(6), 679-692.

St. Jacques, P. L., Conway, M. A., & Cabeza, R. (2011). Gender differences in autobiographical memory for everyday events: retrieval elicited by SenseCam images versus verbal cues. *Memory, 19*(7), 723-732.

Sterpenich, V., D'Argembeau, A., Desseilles, M., Balteau, E., Albouy, G., Vandewalle, G., . . . Maquet, P. (2006). The locus ceruleus is involved in the successful retrieval of emotional memories in humans. *The Journal of neuroscience, 26*(28), 7416-7423.

Tulving, E. (1983). Elements of episodic memory. Clarendon: Oxford.

Appendix A

Examples of terms coded into each internal states category. Emotion terms only were coded and as to whether the emotional state was explicitly noted or was implied.

I. *Emotion*

  a. "This made me feel *angry* and *sad*."

  b. "I was so *glad* I decided to come here."

II. *Evaluation*

  a. "His cancer became more *serious*."

  b. "She did not act which was *odd*."

III. *Mental*

  a. "I *knew* that I wasn't totally at fault."

  b. "I still don't *understand* why."

Tables

Table 1. Means (Standard Deviations in Parenthesis) of the Word Count and Duration of Narrative Production

Table 2. The Use of Internal States Terms by Females and Males in Positive, Negative, and Neutral Narratives

Table3. Fixed Effects from the Two-level Quadratic Model with Pupil Size as Outcome

Table 1

Means (Standard Deviations in Parenthesis) of the Word Count and Duration of Narrative Production

Gender	Valence	Word Count (SD)	Duration by Typing (second) (SD)
Female	Neutral	148.57(58.77)	194.83(80.84)
	Positive	227.17(59.48)	288.03(101.57)
	Negative	241.28(51.94)	303.42(110.51)
	Total	210.24(68.01)	267.48(108.11)
Male	Neutral	143.55(81.33)	169.75(111.04)
	Positive	162.94(66.10)	171.08(70.78)
	Negative	167.80(51.76)	181.73(74.84)
	Total	159.60(65.00)	174.54(82.36)
Total	Neutral	146.36(68.07)	183.80(94.02)
	Positive	196.94(69.77)	233.00(105.37)
	Negative	207.88(63.14)	248.11(112.81)
	Total	187.12(70.98)	225.05(107.31)

Table 2

The Use of Internal States Terms by Females and Males in Positive, Negative, and Neutral Narratives

Valence Of Narratives	Categorical Phase	Participant Group					
		Female		Male		Across gender groups	
		Mean	SD	Mean	SD	Mean	SD
Total	Total	18.18	1.12	13.12	0.92	15.71	0.76
	Emotion	3.97	0.43	2.42	0.25	3.21	0.26
	Evaluation	9.29	0.63	8.36	0.82	8.83	0.51
	Mental	5.48	0.47	3.97	0.32	4.74	0.29
Neutral	Total	13.67	2.29	11.13	1.9	12.47	1.5
	Emotion	2.39	0.53	2.25	0.57	2.32	0.38
	Evaluation	6.72	1.7	6.63	1.68	6.68	0.96
	Mental	4.56	0.97	3.38	0.47	4	0.56
Positive	Total	20.32	1.71	12.32	1.31	16.32	1.23
	Emotion	5.09	0.82	2.05	0.32	3.57	0.49
	Evaluation	9.95	1.03	9.5	1.44	9.73	0.88
	Mental	5.72	0.86	3.45	0.5	4.59	0.52
Negative	Total	19.73	1.64	15.48	1.58	17.65	1.17
	Emotion	4.14	0.7	2.95	0.42	3.56	0.42
	Evaluation	10.73	1.01	8.48	1.18	9.63	0.78
	Mental	6	0.61	4.95	0.58	5.49	0.43

Table 3

Fixed Effects from the Two-level Quadratic Model with Pupil Size as Outcome. Female's neutral narratives were selected as the baseline. The Diff-Positive and Diff-Negative dummy variables were shown as DP and DN for short. The Gender variable refers to the difference between males and females.

Variable	Estimate	SE	t	p-value
Intercept	3.604*	0.117	30.84	<.001
Gender	0.039	0.158	0.25	0.806
Time	-0.01*	0.001	-13.98	<.001
$Time^2$	<.001*	<.001	8.18	<.001
Gender*Time	-0.011*	0.001	-10.12	<.001
$Gender*Time^2$	<.001*	<.001	8.9	<.001
DP	-0.036	0.022	-1.6	0.109
Gender*DP	-0.131*	0.031	-4.21	<.001
DP*Time	<.001	0.001	-0.06	0.949
$DP*Time^2$	<.001	<.001	0.33	0.741
Gender* DP*Time	0.007*	0.001	5.21	<.001
$Gender*DP*Time^2$	<.001*	<.001	-4.85	<.001
DN	0.073*	0.022	3.26	0.0011
Gender*DN	-0.145*	0.031	-4.62	<.001
DN*Time	-0.004*	0.001	-4.18	<.001
$DN*Time^2$	<.001*	<.001	3.95	<.001
Gender*DN*Time	0.007*	0.001	4.92	<.001
$Gender*DN*Time^2$	<.001*	<.001	-4.6	<.001

Figures

Figure 1.

The curve of pupillary dilation in standardized typing duration

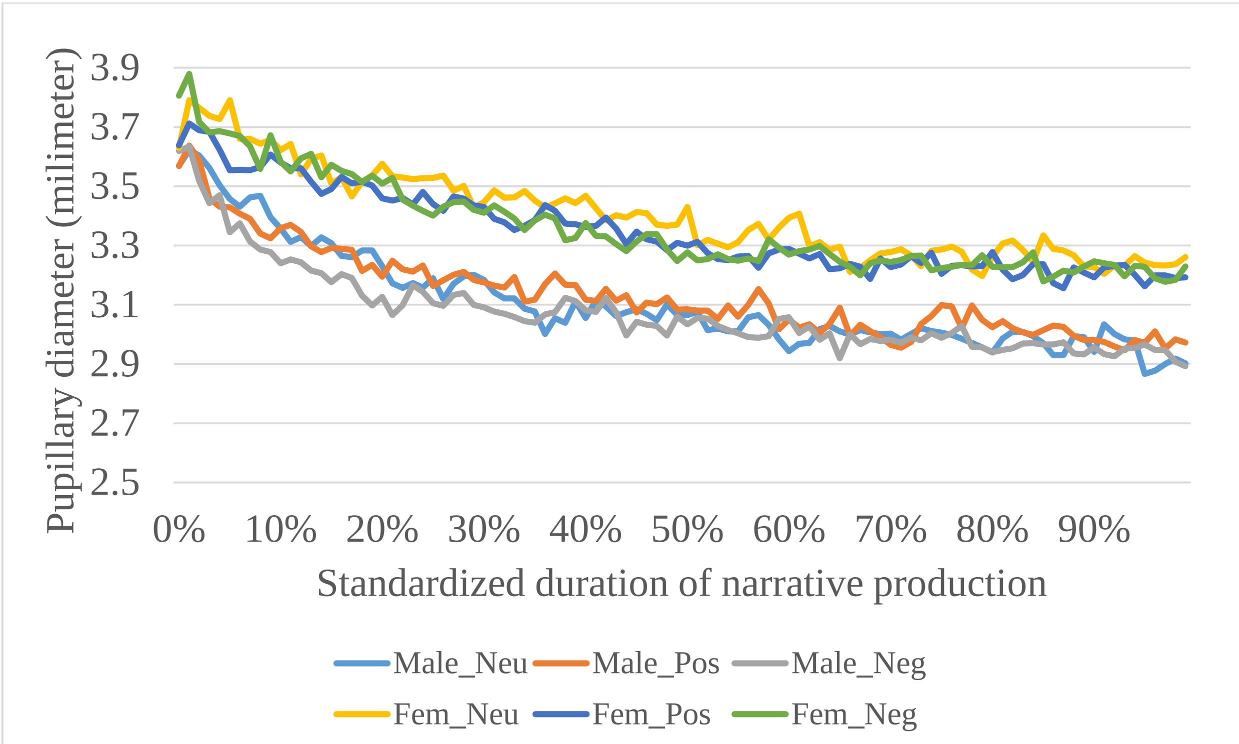

Figure 2.

The distribution of number of letters typed during narrative production

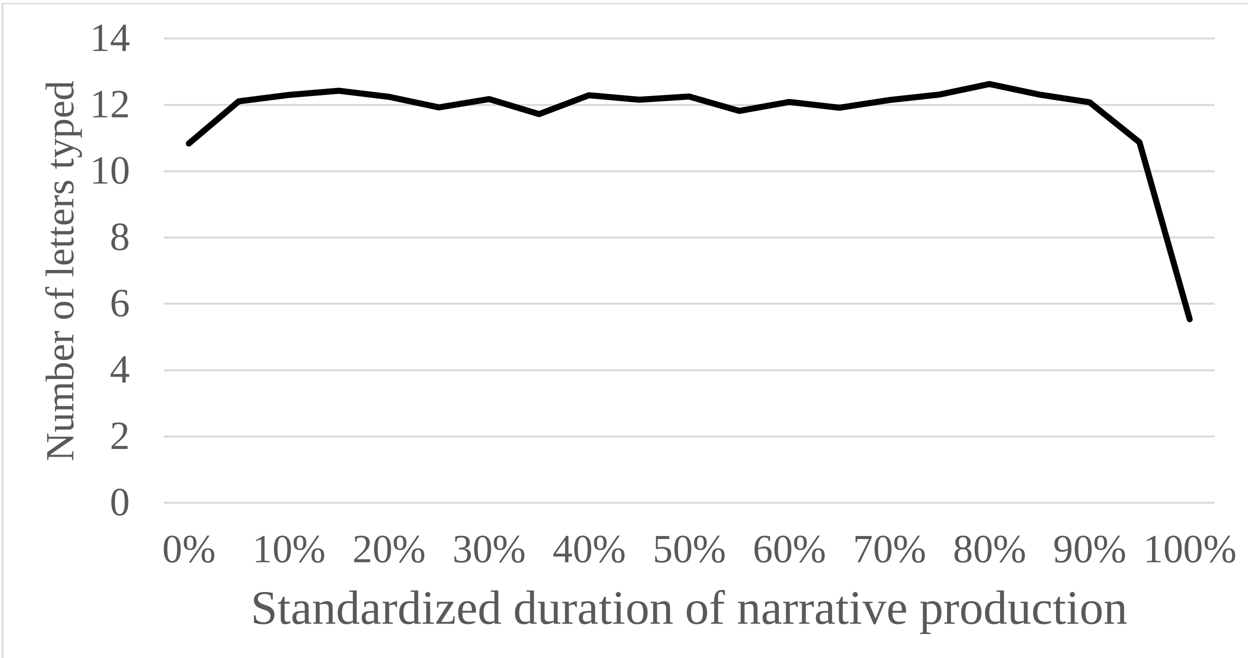

Figure 3.

The curve of pupillary dilation in the first and last narrative production