Developmental Psychology

Priming Feminine Typical Traits Does Not Change Autobiographical Memory Narrative Content

Azriel Grysman 1, Qi Wang 2

1 Psychology Department, Dickinson College, Carlisle, PA, USA, 2 Department of Human Development, Cornell University, Ithaca, NY, USA

Keywords: socio-cultural development, narrative coding, gender differences, autobiographical memory

Gender differences in autobiographical memory have been reported in many studies using narrative coding of features including emotion word use, connectedness to others, and event specific details, with women using more of these narrative features than men. The current pair of studies explored if these narrative tendencies are linked to a sense of self being feminine, by priming female participants to rate themselves on stereotypically feminine trait terms related to emotionality and relationships. In study 1, an online study, participants provided one high point and one low point memory narrative, either in a priming or control group. Narrative coding only differed between the prime and control groups on one of five narrative coded variables, and order effects suggested that the prime may not have been strong enough to influence the second narrative or low point narratives. Study 2 remedied methodological concerns by conducting in-person data collection, strengthening the prime, and eliciting only high point narratives. In study 2, no effects of the prime were found on narrative coded variables. Findings suggest these narrative variables are not influenced by priming femininity and contribute to a growing literature that is deepening an understanding of the source of gender differences in autobiographical memory.

Our sense of self influences how we remember our personal experiences (Conway & Pleydell-Pearce, 2000; Wang, 2016). Gender, as an integral part of individual identity, has been regarded as an important factor in shaping autobiographical memory (for a review, see Grysman & Hudson, 2013). Studies have identified gender differences in the detail (e.g., Bohanek & Fivush, 2010; Bohn & Berntsen, 2008; Herlitz & Rehnman, 2008; Wang, 2013), retrieval speed (e.g., Davis, 1999; Robinson, 1976), self-reported phenomenology (e.g., Compère et al., 2018; Grysman et al., 2017), and content (e.g., Bauer et al., 2003; Grysman et al., 2016; Wang, 2013) of autobiographical memories. These differences are believed to originate from early socialization practices where parents, especially mothers, share memories more in a detailed, emotionally rich, and relationship-oriented manner when conversing with daughters than with sons (Nelson & Fivush, 2004). It is further proposed that broad cultural norms and expectations encourage developing girls to report autobiographical memories with greater elaboration, especially about thoughts, feelings, and the more interpretive aspects of experience as well as the roles of others (Brody & Hall, 1993; Reese et al., 1993); in turn, girls, and later women, report more of these elements than boys and men when recalling memories (Grysman & Hudson, 2013).

However, the relevance of the sociocultural approach to gender has recently been challenged by a meta-analysis of mother-child studies that reported that mothers do not, in fact, elaborate more with their daughters than with their sons (Waters et al., 2019), and a second meta-analysis (Aznar & Tenenbaum, 2020) suggesting mothers do not speak more about emotion with their daughters outside of the autobiographical memory context. These meta-analyses raise questions about the source of gender differences in autobiographical memory. Furthermore, cultural research has shown that gender egalitarianism is increasing across many societies (Varnum & Grossmann, 2017), which may reduce and eventually diminish gender-identity related differences in memory and cognition.

There is little experimental research to test the effect of gender on autobiographical memory. Experimental work would help to clarify if gender differences in memory narratives are linked to expectations about gender, and especially feminine gender norms, or reflect patterns in remembering. The present research aims to explore this by priming feminine gender identity and examining the subsequent memory recall in two samples of young women.
Culture, Gender, and Autobiographical Memory

Support for the socio-cultural approach to autobiographical memory has come from numerous sources. For example, Reese et al. (2010) reported longitudinal correlations between maternal reminiscing style in early childhood and various aspects of adolescents’ memories, suggesting long-term influences of these early conversations (work with fathers is more limited, see Fivush et al., 2009; Zaman & Fivush, 2015). Cross-cultural comparisons have contrasted the reminiscing styles of Euro-American and Asian-American parents (e.g., Wang, 2006, 2007; see Ross & Wang, 2010), showing concurrent and long-term connections between greater elaboration and focus on autonomy among European American parents and a corresponding greater report of memory details in their children, when compared with Chinese parents and children. Pertaining to gender, Wang (2013) and Grysman (2017) suggest that gender differences in recalling event details may take place at the encoding stage. They observed that women recalled more details about daily events than did men shortly after the events occurred, and the differences persisted over time. The content of memories, on the other hand, appears to be reconstructed later at recall in alignment with gendered expectations (Wang, 2015). In addition, women are found to not only recall more detailed memories but also imagine more detailed future personal events (Wang et al., 2011).

Importantly, because self-reporting gender in a study almost always confounds gender identity with biological sex, researchers are left with little evidence as to the source of gender differences in autobiographical memory. Four studies specifically addressed this question by examining gender differences in autobiographical memory alongside measures of gender typicality using the Bem Sex Role Inventory (BSRI, Bem, 1974) and the Personality Attributes Questionnaire (PAQ, Spence & Helmreich, 1978). Using the BSRI, Compère et al. (2018) found that higher feminine-typical scores predicted higher episodic specificity (i.e., more event details) and emotionality (i.e., more emotional expressions) in memory recall, and Compère et al. (2021) extended these findings to brain imaging data. Grysman (2017) reported the same findings regarding emotionality using the PAQ. Furthermore, Grysman (2018) found that gender differences in many aspects of memory recall, including affect, connectedness, factual elaboration, and interpretive elaboration, were stronger when comparing women with high PAQ feminine-typicality scores and men with low PAQ feminine-typicality scores than when comparing women and men with similar scores. These studies extend our understanding of gender differences in autobiographical memory as correlated with self-perceived feminine typicality, although the causal connection remains unclear and further experimental research is required.

Situational Influences on Memory Retrieval

Remembering entails an active constructive process that is influenced not only by individuals’ self-goals and cultural knowledge (Conway & Pleydell-Pearce, 2000; Wang, 2016), but also by the situational demands of the retrieval context (Wang et al., 2017). As such, memory recall can change from one telling to another, depending on the dynamic situational influences. For example, Wang (2008) asked Asian American college students to describe themselves as either being American (i.e., American-identity prime) or being Asian (i.e., Asian-identity prime). Participants then recalled autobiographical memories of personal importance. Those whose American identity was activated prior to the recall reported more self-focused memories, whereas those whose Asian identity was made salient recalled more socially oriented memories. Memory content thus reflected the particular identity being primed. Similarly, Grysman & Hudson (2011) administered an identity questionnaire to prime thoughts about the self. They found that participants who recalled autobiographical memories after this prime described their memories with a more self-referent stance, including greater use of first-person pronouns. Zator & Katz (2017) compared musical cues to lifetime period cues, showing that the latter raised the use of personal pronouns and emotion words, whereas the former raised motor-perceptual words in their recall. Grysman & Denney (2017) asked participants to recall difficult life events in the presence of either a male or female experimenter. Men reported longer narratives and women included more internal state propositions in the presence of a female than a male experimenter.

In all, these empirical examples show scenarios in which influences at the time of recall can change the memory narrative produced, depending both on naturalistic changes in the environment of recall or changes imposed by the researcher. In other words, retrieval contexts and situational cues condition people’s “mindsets” or cognitive frames, making certain memories or certain aspects of memories more accessible for recall (Wang et al., 2017). If gender identity, and especially feminine gender identity, is responsible for the previously observed gender differences in autobiographical memory (Brody & Hall, 1993; Reese et al., 1993), and in particular, if a sense of self-as-female contributes to more elaborate, emotional, and socially-oriented recall among women (Compère et al., 2018; Grysman, 2018), then priming the female identity or femininity is likely to increase the retrieval of such memories among female participants. Furthermore, according to the current literature (Grysman & Hudson, 2015), certain dimensions of memories may be particularly sensitive to such priming.

Dimensions of Memory

Research has revealed relatively consistent findings of gender differences in four dimensions of memory, in each of which women have demonstrated higher production than men (e.g., Grysman, 2018; Grysman et al., 2017). All these dimensions appear to be associated with gender identity (Brody & Hall, 1993; Reese et al., 1993).

Emotion

When sharing memories with their children, mothers speak about emotions differently with sons than with daughters; whereas they use emotions to explain behaviors with boys, they include emotions as matter of fact with girls (Cervantes & Callanan, 1998). Mothers also discuss emotions more frequently with daughters than with sons.
Primed Feminine Typical Traits Does Not Change autobiographical memory narrative content

(e.g., Fivush et al., 2000). Grysman & Hudson (2013) suggested that this pattern of speech may differentially scaffold the role of emotions in autobiographical remembering, modeling to girls that emotions are integral to any event, whereas to boys that emotions are only relevant in certain circumstances. Accordingly, girls and women tend to include more emotional terms in their memory recall than do boys and men (e.g., Boals, 2010; Bohanek & Fivush, 2010; Niedzwieksa, 2003; Rice & Pasupathi, 2010). If sharing emotions is part of the gender norms for women and is encouraged through early socialization, making female identity salient should guide women to include more emotional information in their recall.

**Connectedness**

Theories by Bakan (1966), Gilligan (1982), and McAdams (1985) have long argued that women define themselves more centrally than men in terms of their relations to other people. This gender difference in social connectedness is often expressed in autobiographical memories (e.g., McAdams et al., 2006; Peterson, 2011; Thorne & McLean, 2002; Wang, 2015). Furthermore, women’s and girls’ expressions of relatedness to others in memory narrative has been shown to predict feelings of closeness to a romantic partner (Alea & Bluck, 2007), self-esteem (McLean & Breen, 2009), and psychological well-being (Thorne & Michaelieu, 1996). Priming female identity in women should thus highlight a sense of self as relationally connected, thus raising the amount of social connectedness expressed in memory recall.

**Subjective perspective**

Extensive research has revealed that parents often highlight subjective perspectives more to girls than to boys through reminiscing, where they share their interpretations of past events, focus conversations on what the self and others were thinking and feeling, and highlight what information listeners find interesting (Fivush et al., 2006; Fivush & Wang, 2005; Reese et al., 2010; Wang, 2007). That parents more frequently discuss social interactions with daughters than sons may facilitate among girls perspective taking and the subjective component of remembering. Indeed, women more centrally than men in terms of their relations to other people. This gender difference in social connectedness is often expressed in autobiographical memories (e.g., McAdams et al., 2006; Peterson, 2011; Thorne & McLean, 2002; Wang, 2015). Furthermore, women’s and girls’ expressions of relatedness to others in memory narrative has been shown to predict feelings of closeness to a romantic partner (Alea & Bluck, 2007), self-esteem (McLean & Breen, 2009), and psychological well-being (Thorne & Michaelieu, 1996). Priming female identity in women should thus highlight a sense of self as relationally connected, thus raising the amount of social connectedness expressed in memory recall.

**Event-specific details**

Women are better able to recall event-specific details such as where, when, and what from past episodes than men (Davis, 1999; Grysman, 2018; Pasupathi & Wainryb, 2010; Pillemer et al., 2003; Ross & Holmberg, 1990; Wang et al., 2011). Such gender difference in episodic specificity occurs at the encoding stage as the event unfolds (Wang, 2013) and emerges early in development (Nelson & Fivush, 2004). That parents more frequently and more elaborately share memories with their daughters than with their sons may help girls better organize, represent, and retrieve event details than boys (Davis, 1999; Fivush & Buckner, 2003). Making female identity salient for women may activate the elaborate way of remembering.

The Present Study

The current study tested the effect of female gender identity on autobiographical memory by recruiting women, priming some to think about feminine norms, and then examining the content of the participants’ subsequent autobiographical recall. Specifically, prior to the recall task, participants were asked to complete the Personality Attributes Questionnaire feminine subscale (Spence & Helmreich, 1978), with those in the priming condition being told that the trait words were often used to describe women. We hypothesized that the memory dimensions discussed above would show increases after the prime, when compared with a control group of participants.

Importantly, work with this kind of gender-related prime has not been conducted in the past, and careful steps were taken to analyze how participants responded to it. Alerting female participants to traits that are traditionally conceptualized as female—typical might not elicit homogenous responses such that some traits (e.g., understanding) might be valued whereas others (e.g., emotional) might be perceived as inaccurate gender stereotypes. Nevertheless, participants would be made aware of a common societal perception of femininity that is still prevalent today (Löckenhoff et al., 2014), and we assessed the effect of this awareness on their narratives.

**Study 1**

**Methods**

**Participants**

Data were collected from 126 participants, all of whom self-reported as female as part of their registration with the psychology department’s participant pool at Hamilton College. Self-reported age ranged from 18 to 26 years, (\( M = 19.19, SD = 1.30 \)). Self-reported ethnicity included 92 White, 1 Black, 9 Hispanic/Latinx, 1 Indian, and 17 Asian participants. An additional 6 participants identified as Biracial, including 4 White/Asian, 1 Black/Latinx, and 1 participant who did not specify specific ethnicities.

**Materials**

**Priming procedure.** Participants completed the Personality Attributes Questionnaire, feminine subscale (Spence & Helmreich, 1978), including self-ratings of eight traits terms identified as typically female (emotional, devotes self, gentle, kind, aware of feelings, understanding, warm, helpful) on a 1–9 scale, ranging from ‘not at all like me’ to ‘exactly like me’. Participants in the priming condition completed this scale with the following instructions: “Below are some words that are often used to describe women. Please indicate how well each of the following traits describes you.” Participants in the control condition saw only the second of those two sentences. This scale was chosen because the scores have been found in two emerging adult samples to not differ by gender (Grysman, 2018; Grysman et al., 2017),
and so was considered ideal to measure perceptions about gender rather than actual differences. Internal consistency of this measure is discussed in the results section.

Memory prompts. Two memory prompts were adapted from McAdams’ (1995) Life Story Interview, including the high point and low point memories. These prompts were chosen because they provide an opportunity for participants to report salient, self-relevant events. Gender differences have been found in these memories in previous research (Grysman, 2018; Grysman et al., 2016).

Low Point. A “nadir” is a low point in your life. Please think about your entire life. Try to remember a specific experience in which you felt extremely negative emotions, such as despair, disillusionment, terror, profound guilt, shame, etc. You should consider this experience to represent one of the “low points” in your life story. Even though this memory is unpleasant, we would still appreciate an attempt on your part to be honest and straightforward and to provide us with as much detail as possible.

High Point. Many people report occasional “peak experiences.” These are generally moments or episodes in a person’s life in which he or she feels a sense of great uplifting, joy, excitement, contentment, or some other highly positive emotional experience. Indeed, these experiences vary widely. Some people report them to be associated with religious or mystical experience. Others find great joy or excitement in vigorous athletics, reading a good novel, artistic expression, or in love or friendship. A peak experience may be seen as a “high point” in your life story – a particular experience that stands out in your memory as something that is extremely positive.

For both prompts. Please describe below in some detail a peak experience (low point) that you have experienced sometime in your life. Make sure that this is a particular and specific incident (e.g., happened at a particular time and in a particular place) rather than a general “time” or “period” in your life. Please describe this event as if you were telling it to a friend in a conversation. Your description should be at least two paragraphs in length.

Memory Ratings. To assess various phenomenological properties of memories, participants were asked to indicate on 5-point scales how often they had thought or talked about each event, its personal importance, level of detail, emotional intensity, valence, and the event’s source (internal or hearing from others). Participants also reported their age at the time of each event.

Procedure

Participants signed up for the study via the departmental participant pool at Hamilton College. Only women were eligible to complete this study, but that information was based on pre-screen demographics in the participant pool and so was not publicly available to participants. They received a link and instructions to complete the survey in one sitting, and that it would take approximately 30 minutes. As outlined in the pre-registration of this study on Open Science Framework (https://osf.io/ab26y/), 20 participants who completed it in more than 90 minutes were excluded. After providing consent, participants were directed to complete the feminine subscale of the PAQ (Spence & Helmreich, 1978), with either the control or prime instructions. After completing this rating scale, participants read instructions for the two memory reports, the high and low points, which were presented in randomized order. After reporting these two memories, participants completed the ratings for each memory and provided demographics.

Memory Content Coding

Memories were content coded using the following coding systems. For all coding systems, two research assistants completed coding at least 20% of narratives. After achieving acceptable reliability scores and resolving disagreements through conversation and discussion with the first author, one assistant completed the remainder of the coding. Two coders completed 57 narratives and coded these three elements simultaneously. Intraclass correlation coefficient was used to index inter-coder reliability as it is sensitive to differences in rater thresholds (e.g., Wang et al., 2018).

Emotion. Using the coding system of Bauer et al. (2003) and Fivush et al. (2012), memory narratives were coded for every occurrence of an emotion word (e.g., sad, happy, Intraclas correlation coefficient (ICC) = .98) or affective term (e.g., ‘it was great,’ ICC = .86), although these two categories were combined after coding. All terms were subsequently coded as either positive or negative.

Connectedness. This coding system is designed to evaluate the role that relationships with close others play in a narrative. As in previous research (e.g., Andrews et al., 2015), a scale of 0 to 3 assessed whether important relationships were presented or elaborated upon in a narrative. A score of 0 was assigned when no meaningful interactions with others were included in the narrative. A score of 1 was assigned when interactions with others are presented but the meaning of the relationship or interaction is not explained. A score of 2 was assigned when people who are identified as important to the narrator are included, but the narrative is event-focused rather than relationship-focused. Finally, a score of 3 is assigned if the narrator explains the meaning of the relationship in one of several ways. A full manual is obtainable from the first author. ICC for coding of 54 narratives was .91.

Event-specific details. This coding system was based on Wang & Song (2018) and on Hudson et al. (1992) to tackle the “factual” components of memories. Every instance of a new detail was coded, but repeated details were not. As a rule of thumb, event-specific details refer to such information as what happened, where, and when. It was divided into two categories. First, action statements or quoted speech (ICC = .92) that reflect event happenings (e.g., ‘when I traveled with my best friend...’). The other category included additional details about the event (ICC = .96), including character (e.g., ‘my mom said...’) or object (e.g., ‘I had my new watch’) introduction, temporal (e.g., ‘on Tuesday...’) or location (e.g., ‘we got together at the mall’) information, and other factual descriptors (e.g., ‘my eyes were red’).

Subjective perspective. Subjective perspective includes references to emotions, thoughts, desires, and subjective evaluations, whether as background or as they were occurring in the event (ICC = .97). Following Hudson et al. (1992),
subjective evaluations included intensifiers, negations, exclamation points, metaphors, and qualifiers as indicators of subjective perspective. Different from the emotion code, this code captures all mental states that reflect subjective perspectives, appraisals, and evaluations of the memory event. They are conceptually distinct narrative codes based on prior research, both relevant to gender (see Grysman et al., 2016; Pasupathi & Wainryb, 2010; Wang & Song, 2018).

**Power Analysis**

A sample size of 128 was planned based on a power analysis using G*Power 3.1.9.2 for a between groups comparison of two groups (prime and control), power of .80, and a medium effect size of Cohens’s $d = .5$. This effect size was chosen for the following reasons: There is little precedent for a study like this because we tested narrative coded variables that are commonly used in gender comparisons. In those studies, effect sizes have ranged vastly, including Grysman’s (2018) report of $d$ values ranging from .67 to 1.27. Wang (2013) reported small to medium effect sizes, as have various researchers on the topic (e.g., Pillemer et al., 2003; Wang et al., 2011). Grysman and Hudson reported medium effects sizes of their priming procedure (partial $\eta^2 = .10$, and .12), and Grysman & Denney (2017) found that the effects of experimenter gender on narrative content yielded Cohen’s $d$ scores in the medium to large range. Because high and low point narratives represent repeated measures, this sample is also sufficient to test between by within interactions with this variable. After data collection and coding were completed, researchers noticed that there were two duplicate values in the data file, hence explaining the reduction from 128 participants to 126, and reducing power from .80 to .79.

**Results**

**Manipulation Check**

Preliminary analyses found no difference between prime ($M = 7.13$, $SD = .93$) and control ($M = 7.33$, $SD = 1.01$) participants on mean PAQ-F scores, $t(124) = 1.13, p = .259$. Further analyses were conducted to examine the variability in each participant’s responses on this measure for the prime and control groups, respectively. The reasoning for the analyses is that it is possible, especially on a liberal campus, that female participants might bristle at the suggestion that certain traits are being considered feminine and thus respond either by reporting lower scores or by providing less consistent responses across items. As a first test of this possibility, a standard deviation score of the eight trait items were computed for each participant. Higher scores would represent more intra-individual variability. Priming participants ($M = 1.29$, $SD = 0.57$) had higher standard deviations of their responses than control participants ($M = 1.03$, $SD = 0.64$), $t(124) = 2.37, p = .019$, Cohen’s $d = 0.43$.

As a further effort to understand possible differences between the responses of the control and priming groups, separate internal consistency reliability analyses were conducted for the eight items on the PAQ-F questionnaire. Amongst the control group, Cronbach’s $\alpha = .85$; in contrast, the priming group demonstrated Cronbach’s $\alpha = .74$. Feldt & Kim (2006) constructed a test comparing Cronbach’s alpha scores to each other, using the statistic $W = (1-\alpha_1)/(1-\alpha_2)$, with an F distribution with $N_1$-1, $N_2$-1 as df values. Thus critical $F(65, 61) = 1.525$ for this test in the current sample, whereas $W = 1.68$, suggesting a significant difference between reliability the two conditions.

Finally, in addition to increased intrapersonal variability distinguishing the priming and control groups, exploratory factor analyses were conducted in an attempt to understand any underlying patterns in the way the questionnaire was completed. For each condition, a factorial analysis was conducted with a Varimax rotation. In the control condition, a two-factor solution emerged, KMO = .77, with the two factors accounting for 45.8% and 20.4% of the variance in the rotated solution, respectively. As can be seen in Table 1, six of the eight items loaded onto the first factor, with responses to devotes self and helpful loading onto the second factor. In the priming condition, a three-factor solution emerged, KMO = .62, with the three factors accounting for 32%, 19.8%, and 17.1% of the variance, respectively. As can be seen in Table 1, the second factor now included the items emotional and aware of feelings, whereas the third includes the same items as in the control condition. Predictions were not made about which items would load with each other, and so inferences will not be attempted. However, it is notable that in the priming group, the factor solution includes lower KMO, lower variance explained by the largest factor, and more factors than in the control condition. We tentatively conclude that the prime was effective, but that its effect was not a straightforward, mean-level influence. Instead, it led to different intra-individual variability in patterns of responses to the PAQ-F.

**Core Analyses**

For the purposes of descriptive statistics, the age of memories reported was examined in a mixed-model ANOVA with memory type (high point, low point) as a within-subject predictor and condition (control, priming) as a between-subject predictor. It was found that low point events ($M = 41.39$ months, $SD = 38.73$) were older than high point events ($M = 24.67$ months, $SD = 30.52$), $F(1, 125) = 20.01, p < .001$. Age of memories did not significantly differ by condition ($p = .08$), nor did it interact with event type ($p = .11$).

Descriptive statistics for narrative-coded variables can be seen in Table 2. Analyses began with a 2 (condition: priming, control) X 2 (memory type: low point, high point) ANOVA for each of the five narrative-coded variables: emotion, connectedness, subjective perspective, and event-spe-

---

1 Note: the analyses reported in this subsection were not predicted in advance in the OSF registration. They were conducted based on comments from reviewers, and were added only after both Study 1 and 2 had been fully conducted and analyzed.
Table 1. Components of Varimax-rotated factor analyses of responses to the PAQ-F across the three distinct groups completing this scale.

<table>
<thead>
<tr>
<th>PAQ-F Item</th>
<th>Study 1: Control</th>
<th>Study 1: Priming</th>
<th>Study 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Emotional</td>
<td>.690</td>
<td>.032</td>
<td>.077</td>
</tr>
<tr>
<td>Devotes self</td>
<td>.026</td>
<td>.908</td>
<td>.038</td>
</tr>
<tr>
<td>Gentle</td>
<td>.870</td>
<td>.142</td>
<td>.802</td>
</tr>
<tr>
<td>Helpful</td>
<td>.306</td>
<td>.766</td>
<td>.543</td>
</tr>
<tr>
<td>Kind</td>
<td>.749</td>
<td>.273</td>
<td>.805</td>
</tr>
<tr>
<td>Understanding</td>
<td>.810</td>
<td>.052</td>
<td>.655</td>
</tr>
<tr>
<td>Aware of feelings</td>
<td>.657</td>
<td>.264</td>
<td>.151</td>
</tr>
<tr>
<td>Warm</td>
<td>.824</td>
<td>.223</td>
<td>.719</td>
</tr>
</tbody>
</table>

Table 2. Descriptive data for the narrative-coded variables appearing in both study 1 and study 2. For each study, data represent a score averaged across both conditions. In study 1, scores are averaged across high points and low points; in study 2, they are averaged across control and priming conditions.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1, N = 126</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjective Perspective</td>
<td>32.80</td>
<td>15.73</td>
<td>3.50</td>
<td>89.00</td>
</tr>
<tr>
<td>Action</td>
<td>9.90</td>
<td>6.16</td>
<td>2.00</td>
<td>40.00</td>
</tr>
<tr>
<td>Connectedness</td>
<td>1.94</td>
<td>0.76</td>
<td>0.00</td>
<td>3.00</td>
</tr>
<tr>
<td>Emotion</td>
<td>6.96</td>
<td>3.35</td>
<td>0.50</td>
<td>19.00</td>
</tr>
<tr>
<td>Event detail</td>
<td>26.80</td>
<td>14.91</td>
<td>2.50</td>
<td>86.50</td>
</tr>
</tbody>
</table>

Study 2, N = 56

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjective Perspective</td>
<td>23.62</td>
<td>9.50</td>
<td>6.50</td>
<td>55.50</td>
</tr>
<tr>
<td>Action</td>
<td>12.39</td>
<td>4.21</td>
<td>4.00</td>
<td>22.50</td>
</tr>
<tr>
<td>Connectedness</td>
<td>1.60</td>
<td>0.74</td>
<td>0.00</td>
<td>3.00</td>
</tr>
<tr>
<td>Emotion</td>
<td>5.44</td>
<td>2.71</td>
<td>1.00</td>
<td>12.00</td>
</tr>
<tr>
<td>Event detail</td>
<td>28.56</td>
<td>8.62</td>
<td>12.00</td>
<td>47.00</td>
</tr>
</tbody>
</table>

Specific details. Condition was a between-subject factor, and memory type was a within-subject factor. Results from this initial analysis showed a main effect of condition for event details, F(1,122) = 4.45, p = .037, partial $\eta^2 = .035$, with participants including more details in the prime condition ($M = 28.93$, $SD = 14.59$) than in the control condition ($M = 24.60$, $SD = 15.03$). No statistically significant main effects of condition were found on the other four narrative dependent variables. Analyses also indicated statistically significant main effects of memory type for emotions, F(1,122) = 4.28, p = .041, partial $\eta^2 = .054$, subjective perspective, F(1,122) = 14.48, p < .001, partial $\eta^2 = .106$, and connectedness, F(1,122) = 10.94, p = .001, partial $\eta^2 = .082$, although it should be noted that the subjective perspective coding includes emotion content, and thus these two effects contain some overlap. For all of these, participants’ low point narratives received higher scores than their high point narratives. No interactions of condition by memory type were present (all p’s > .21).

These results show an effect of the prime in only one of the five coding systems, namely event details. They also show a substantial difference between high point and low point narratives, with differences found on three of the five coding systems. Next, an exploratory analysis added order as an independent variable to the current analysis. The reasoning for this exploration was a concern that either the prime was not sufficiently strong to influence the second narrative or that participants were losing interest and putting less effort into the second narrative, especially as the study was being completed remotely. A 2 (condition: priming, control) X 2 (memory type: low point, high point) X 2 (order of high point and low point) ANOVA was conducted for each of the five narrative variables. Statistically significant interactions of memory type by order emerged for subjective perspective, F(1,124) = 7.25, p = .008, partial $\eta^2 = .055$, and event details, F(1,124) = 5.45, p = .021, partial $\eta^2 = .042$. As can be seen in Table 3, both facts and sub-
subjective perspective decreased in high point narratives when this narrative was second compared to when it was first. No substantial differences were found in low point narratives based on order.

Exploratory analyses were conducted on all the memory ratings questionnaires using 2 X 2 ANOVA's with condition and narrative as independent variables. No main effects of condition or interaction with this variable emerged on any of the tests. As a further exploratory follow-up analysis, we considered whether participants’ ratings on the PAQ-F would predict their scores on the various narrative measures. When PAQ-F scores were added to the analyses as a covariate, no statistically significant findings emerged, indicating that priming-group participants’ scores on the PAQ-F did not influence their memories in comparison with the control group.

Further exploratory analyses were conducted based on reported ethnicity because of prior work suggesting that various non-White ethnic groups may value interdependence more highly than White groups (see Fitzgerald, 2010). Our sample included 92 White and 34 non-White participants (see demographics). First, PAQ-F score was used as a DV in 2 (ethnicity: non-White, White) X 2 (Condition: control, priming) ANOVA. Main effects and the interaction did achieve statistical significance, but the main effect of ethnicity, \( F(1,122) = 3.01, p = .085 \), showed mean-level difference between non-White participants (\( M = 6.99, SD = .95 \)) and White participants (\( M = 7.32, SD = .97 \), Cohen’s \( d = .54 \)). Another 2 X 2 ANOVA was conducted with the same independent variables and intraindividual variability as the DV; no main effect or interaction with ethnicity emerged, \( p’s > .75 \). Finally, a 2 (condition: priming, control) X 2 (memory type: low point, high point) X 2 (ethnicity: White, non-White) ANOVA for each of the five narrative-coded variables was conducted. All main effects and interactions involving ethnicity were nonsignificant except for one interaction between condition and ethnicity for event details, \( F(1,120) = 6.14, p = .015 \). As can be seen in Table 4, White participants provided more event details in the priming condition than in the control condition, \( t(90) = 2.76, p = .007 \), and in the priming condition, White participants provided more event details than non-White participants, \( t(62) = 2.70, p = .009 \).

Finally, in an attempt to assess if scores in both conditions may have been primed from the scale presented, the data were compared to high point and low point narratives reported by Grysman (2018) in which the scale was not used. These memory data were chosen because they came from a very similar sample and were coded using the same measures of connectedness and affect language (although not the other three coding measures). No significant differences emerged on these coding measures when compared to Grysman’s (2018) data: for connectedness in low points, \( t(220) = - .91, p = .36, Cohen’s d = .12 \), and in high points, \( t(225) = .26, p = .79, Cohen’s d = .02 \); for affect in low points, \( t(219) = .58, p = .56, Cohen’s d = .08 \), and in high points, \( t(225) = .42, p = .68, Cohen’s d = .05 \) suggesting that a lack of significant findings cannot be attributed to accidental priming of both groups.

**Table 3.** Means (95% confidence intervals) for facts and subjective perspective statements, organized by event type and the order of completing the narratives.

<table>
<thead>
<tr>
<th>Event Details</th>
<th>Order</th>
<th>High Point</th>
<th>Low Point</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HP First</td>
<td>29.96 (17.51)</td>
<td>25.50 (15.85)</td>
</tr>
<tr>
<td></td>
<td>LP First</td>
<td>24.84 (15.24)</td>
<td>27.26 (17.56)</td>
</tr>
<tr>
<td>Subjective Perspective</td>
<td>HP First</td>
<td>33.39 (18.45)</td>
<td>34.61 (20.95)</td>
</tr>
<tr>
<td></td>
<td>LP First</td>
<td>28.33 (12.67)</td>
<td>35.36 (16.84)</td>
</tr>
</tbody>
</table>

**Table 4.** Means (standard deviations) for event-specific details, organized by condition and by participant ethnicity in Study 1.

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Control</th>
<th>Priming</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-White</td>
<td>78.82 (61.86)</td>
<td>57.24 (27.23)</td>
<td>34</td>
</tr>
<tr>
<td>White</td>
<td>64.89 (31.10)</td>
<td>85.43 (39.67)</td>
<td>92</td>
</tr>
</tbody>
</table>

**Discussion**

Some preliminary conclusions can be drawn from these findings. First, only event details demonstrated the predicted main effect of the prime when compared the control group. Although this is consistent with the general finding that females tend to outperform males in episodic specificity (Herlitz & Rehnman, 2008; Wang, 2013), it is only one predicted finding in the context of multiple null results. The remaining narrative variables showed no priming effect. Exploratory analyses of differences in order and of high versus low points were conducted to determine if there were any methodological factors that may have suppressed potential influence of the prime or if the prime simply didn't influence narrative content. These exploratory analyses showed differences between high and low point events, with narrative coding of high points more variable than low points based on order. The order effects showed that participants were elaborating less on high points when they were second in the order. These effects were small, and \( p \) values were between .01 and .05, but are reported to emphasize that because high points are more susceptible to methodological...
factors, they represent the greatest potential in detecting a priming effect, were it present. In sum, study 1 showed limited effects of priming feminine gender identity on narrative content. Study 2 was designed to examine if these results would change with in-person data collection, a focus on high point events, and a priming manipulation that was more distinct from the control group.

Study 2

To alleviate the design concerns of study 1, the second study used a within-subject design, with every participant reporting two high-point (HP) events, because these events showed more priming effects in the first study. Participants completed the survey in the laboratory, instead of online. Memories were elicited on separate days to counterbalance order of the prime and control conditions without concern that prime would influence the control condition when it is second. Finally, the manipulations were adjusted to make a stronger distinction between the priming and control conditions.

Methods

Participants

Data were collected from 56 participants, all of whom self-reported as female as part of their registration with the psychology department's participant pool at Hamilton College. Self-reported age ranged from 18-22, \( M = 18.32, \ SD = 1.98 \). Self-reported ethnicity included 33 White, 5 Black, 5 Hispanic/Latinx, 1 American Indian, and 10 Asian participants. One additional participant identified as 'White, Latino, and Asian,' and 1 participant declined to answer.

Materials

Priming procedure. In the priming condition, participants again read the PAQ-F (Spence & Helmreich, 1978), but with the following introductory prompt:

How do you define yourself AS A WOMAN? Please indicate how well each of the following FEMININE traits describes you: (emotional, devotes self; gentle, kind, aware of feelings, understanding, warm, helpful)

Each trait was then rated on a 1-9 scale, as in study 1. In the control condition, participants read the following prompt:

What do you think about YOUR NATURAL ENVIRONMENT in which you live on campus? Please indicate how well each of the following words describes your current environment:

beautiful, polluted, peaceful, limited resources, green, natural, pristine, balanced

Again, each term was rated on a 1-9 scale.

The narrative prompt was the same as the high point prompt from study 1, however participants completed two separate high point narratives.

Procedure. Narratives were elicited on separate days to counterbalance order of the prime and control conditions without concern that prime will influence the control condition when the prime was delivered first. Participants signed up for the study through the department subject pool. They were told that the study included two parts and that part 2 must be completed between 2 and 7 days after part 1. The same questionnaires as study 1 were completed after report of the memories. Coding of memory content was identical as Study 1. Participants only completed the PAQ-F once as part of the priming procedure.

Power analysis. To simulate the power of the current study, for each variable, 10,000 pairs of correlated data were generated using R, with means, standard deviations, and within-subject correlation values from study 1 (see Table 5). These correlations were considered an acceptable minimum because they were for one HP and one LP, whereas study 1 will include two HP's. From these 10,000 pairs, 1000 random samples (\( N = 55 \)) were drawn for pairs of the five variables, and paired-samples t-tests were conducted. The target, to control for multiple tests, was that the number of tests achieving significance would remain above 80% after multiplying all five percentages by each other, using an effect size of .2 SD, the smallest difference of interest. At \( N = 55 \), total power = .85, and thus was chosen for the current study.

Results

As a preliminary analysis, PAQ-F scores in study 2 (\( M = 7.15, SD = 0.90 \)) did not differ from those in study 1 (\( M = 7.25, SD = 0.97 \)). The prime in study 2 was examined according to the same metrics as in study 1, though there was no control group in study 2, due to the within-subjects design. Notably, the intrapersonal variability on the PAQ-F in study 2 demonstrates similar trends as in the priming condition in study 1 (\( M = 1.26, SD = .66 \)), as was the internal consistency reliability, Cronbach’s \( \alpha = .73 \). The factor analysis for study 2 yielded only a 2-factor solution, KMO = .66, with the two factors accounting for 32.3% and 23% of the variance in the rotated solution, respectively. The second factor in study 2 included emotional, devotes self, and aware of feelings, three of the four items appearing in the latter two factors from the priming study. Again, we will not make inferences from the specific items in the different components, but we will emphasize that the KMO value and the lower variance explained by the largest factor follow the pattern from the priming group in study 1 and suggest less consistency in responses this questionnaire in the priming conditions.

Additionally, correlations for each of the five narrative variables between the two time points in study 2 were conducted to test for within-subject consistency. Facts (\( r = .46 \)), affect (\( r = .45 \)), and subjective perspective (\( r = .53 \)) were significantly correlated (\( p < .05 \)) across sessions. Connectedness (\( r = .15 \)) and actions (\( r = .05 \)) were not. Data analysis began with five paired-samples t-tests (one for each narrative variable). Descriptive statistics are presented in Table 6. No t-test achieved statistical significance (all \( p's > .21 \)). PAQ-F scores did not correlate to any of the narrative variables. Word count also did not differ by session, \( t(55) = 1.16, p = .25 \).

As with study 1, analyses were conducted by ethnicity, as study 2 included 36 White and 20 non-White partici-
Table 5. Descriptive statistics used in the power simulation to prepare for study 2, including actual means (standard deviations) from study 1 and those projected for 1000 simulations of study 2 with an effect size of $d = .2$. Correlation values represent the correlation between high and low point narratives, and were thus used as a minimum for simulations of study 2.

<table>
<thead>
<tr>
<th>Coded Variable</th>
<th>Study 1</th>
<th>Study 2 - Projected Priming</th>
<th>Number of statistically significant simulations, $N = 55$ ($p &lt; .05$)</th>
<th>$r(124)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affect</td>
<td>6.56 (3.72)</td>
<td>7.30 (3.72)</td>
<td>926</td>
<td>.33</td>
</tr>
<tr>
<td>Connectedness</td>
<td>1.77 (0.99)</td>
<td>1.97 (0.99)</td>
<td>964</td>
<td>.24</td>
</tr>
<tr>
<td>Facts</td>
<td>27.12 (15.65)</td>
<td>30.40 (15.65)</td>
<td>1000</td>
<td>.61</td>
</tr>
<tr>
<td>Actions</td>
<td>9.86 (7.12)</td>
<td>11.28 (7.12)</td>
<td>960</td>
<td>.45</td>
</tr>
<tr>
<td>Subjective Perspective</td>
<td>30.58 (16.42)</td>
<td>33.71 (16.42)</td>
<td>996</td>
<td>.67</td>
</tr>
</tbody>
</table>

Table 6. Means (standard deviations) of narrative coding of high points in study 2, grouped by whether they were reported after a prime or not. No comparisons were statistically significant.

<table>
<thead>
<tr>
<th>Coded Variable</th>
<th>Priming</th>
<th>Control</th>
<th>Cohen’s $d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affect</td>
<td>5.65 (3.50)</td>
<td>5.25 (2.93)</td>
<td>.12</td>
</tr>
<tr>
<td>Connectedness</td>
<td>1.69 (1.00)</td>
<td>1.51 (0.96)</td>
<td>.18</td>
</tr>
<tr>
<td>Facts</td>
<td>29.25 (15.65)</td>
<td>27.44 (15.65)</td>
<td>.18</td>
</tr>
<tr>
<td>Actions</td>
<td>12.13 (6.06)</td>
<td>12.65 (5.68)</td>
<td>.09</td>
</tr>
<tr>
<td>Subjective Perspective</td>
<td>24.16 (10.38)</td>
<td>22.56 (11.21)</td>
<td>.15</td>
</tr>
</tbody>
</table>

Despite all these efforts, no effect of the prime emerged in study 2.

**Effectiveness of Priming**

Although the priming task did not produce mean-level differences between the priming and control groups, analyses of intraindividual variability and internal consistency reliability suggest the PAQ-F was answered differently in the two groups. This finding is informative both for the purposes of our study and for considerations in use of this scale in the future. First, the PAQ-F was designed as a singular subscale of the broader PAQ that measures feminine typical traits (Spence & Helmreich, 1978), our studies found that, regardless of group, participants answered this questionnaire in a manner that implied at least two separate factors. This finding might inform future research exploring this measure and similar ones (e.g., BSRI, Bem, 1974) as societal attitudes towards femininity change (Varnum & Grossmann, 2017). Second, the analyses of the scale found that after being primed to think of themselves as women, participants responded on the PAQ with more intraindividual variability, yielding different factor structures and less consistency in how they responded to the eight trait terms. Such a pattern, replicated across two studies, suggests that women draw more substantive distinctions between traits, especially emotional and aware of feelings. Given the liberal orientation of the students at the institution, we suggest that the priming procedure led to a heightened awareness...
of feminine stereotypes, leading women to more carefully consider each feminine typical trait, and possibly influencing the degree to which they endorsed them. Although the priming did not generate a mean-level difference, this pattern of responses raises confidence in the prime achieving its goal, namely to make female participants aware of their feminine gender identity. Despite our interpretation of these patterns, we caution that the analyses were post hoc, and encourage research on gender typicality and perception of gender stereotypes to consider this measurement issue in future work.

No Effect of Priming on Narratives

The simplest explanation of the general lack of effect of priming on narratives is that gender differences in autobiographical memory are not necessarily related to a sense that these narrative tendencies are feminine-typical. In other words, even on an implicit level, gender-linked narrative differences may not be connected to a female self-concept, and thus are not influenced by our priming. Another possibility is that memory narratives are relatively stable, reflect how information is encoded in memory, and cannot be influenced by simple priming procedures. Although these are both possible interpretations, there is not sufficient evidence from these two studies to support such broad claims. We consider each one in turn.

Regarding the possibility that gender differences in AM are not connected to a feminine self-concept, there is insufficient evidence to support such a broad conclusion. First, only two event types were tested (high point and low points), and it is possible that gender-based connections to memory are more subtle or only apparent in certain event types. Researchers might explore AM narratives that are more explicitly connected to a sense of self, such as self-defining memories (Singer & Salovey, 1993), or events connected to gender-typed emotions, such as anger and sadness (see Adams et al., 1995; Bird & Reese, 2006). Another possibility is that different forms of narrative coding would be more open to influence from priming. Although we used narrative coding that has been shown to vary by gender in multiple studies, other narrative coding approaches might be more open to the influence of a prime. For example, Grysman et al. (2017) found that PAQ-F scores predicted scale ratings of emotional intensity but not the amount of affect used in AM recall. They raised the possibility that scale measures might be more connected to identity, because answering a survey item is a more explicit choice than narrative style. However, the current studies found no impact of the prime on either narrative or scale variables. Finally, the unique response patterns to the PAQ following the prime suggest that feminine gender identity is a complex construct in the contemporary society and that its relation to memory narratives could be complex as well. Thus, although the current findings show that AM narrative variables were not influenced by a gender-related prime, we caution against generalizing too broadly from these findings and call for additional research.

The second possible explanation was that AM narratives are broadly stable and are not open to situational influence by something like a priming procedure. We find this possibility doubtful because studies have shown that various situational factors can influence the way an event is narrated, whether they are listener input (Jennings et al., 2014), gender of the listener (Grysman & Denney, 2017), or cues presented prior to recall (Wang, 2008; Zator & Katz, 2017). Extensive effects of audience on the ways in which narratives are told have been reported (McLean & Jennings, 2012; McLean & Mansfield, 2011; Pasupathi & Hoyt, 2009), and several researchers have emphasized how the context of telling one’s stories shapes the way events are told and conceptualized for the self (e.g., McLean et al., 2007; Ross & Wang, 2010; Wang et al., 2017). Thus, we think it unlikely that AM narratives are not open to priming, and more likely that this particular prime was not effective in influencing narrative content.

Limitations

Despite the clarity of the null findings from these studies, some caution is warranted. First, the power of study 1 was borderline acceptable (.79), and although study 2 improved on that power, it did so with a small sample using repeated measures. Additionally, the exploratory analyses of ethnicity data suggest that the construct of gender may vary by ethnic identity, but there were not enough participants, and many values above .05 but below .10, thus preventing a clear interpretation. We further caution against including all non-White participants in a single analysis, but, given the data, we have reported what was available to us for the sake of full disclosure.

Conclusion

In conclusion, results from two studies point to a lack of women’s narratives being influenced by a priming manipulation that highlighted feminine self-concept. We began this paper by situating gender-based predictions in models of sociocultural development such that parents speak differently to boys than to girls, but noted current meta-analyses contradicting these models and changes in societal attitudes towards gender as potential countervailing forces. We thus view the current study as another step towards establishing boundary conditions on when attitudes towards femininity are linked to autobiographical memory.

Acknowledgements

The authors declare no conflict of interest and no fund-

2 Although political affiliation was not collected in the current data, a recent study using the same participant pool (Meisels & Grysman, 2020) found that, on a 7-point political ideology scale ranging from very conservative to very liberal, only 5% of students self-identified on the conservative end.
ing for this work.

The authors would like to thank Aspacelia Geranios, Robert Chapman, Alexandra Stetter, Sarah Young, and Amanda Leonhard for their work coding narratives.

Both studies were preregistered on the Open Science Framework, https://osf.io/st86x/?view_only=7f56a609c8cb42b5be2f1ab98d43239

The first author would like to thank the Dickinson College Research and Development Committee for its co-support of open access publishing fees.

Contributions

The studies and coding plan were conceptualized jointly through conversation. Azriel Grysman oversaw data collection, interrater reliability, and conducted statistical analyses. Qi Wang consulted on decisions regarding power analysis and data analysis. Both authors jointly wrote the manuscript.

Data Accessibility Statement

Data from both studies have been uploaded to the OSF page referenced in describing the preregistration.

Submitted: June 05, 2020 PDT, Accepted: April 12, 2021 PDT

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license’s legal deed at http://creativecommons.org/licenses/by/4.0 and legal code at http://creativecommons.org/licenses/by/4.0/legalcode for more information.
REFERENCES


SUPPLEMENTARY MATERIALS

Peer review history