Emotional Tone Embedded in Leadership Items and Memory Processing: Implications for Leadership Measurement

Mengying Li, Bryan Acton

Leadership & Organizational Science

Bernard M. and Ruth R. Bass Center for Leadership Studies

School of Management | State University of New York at Binghamton

Abstract

Leadership research has predominantly relied on recollective self-report questionnaires to collect leadership ratings. Understanding respondents' memory processing plays an important role in enhancing response process validity and facilitating better alignment between theory and measurement in leadership research. However, we have limited understanding in terms of leadership item-level factors that may influence respondents' memory processing. Moreover, extant research has largely focused on the impact of the cognitive (vs. affective) aspect of memory processing on leadership measurement. This study fills the void by elucidating the effect of emotional tone embedded in leadership items on memory processing as well as how the permutation of leadership items and their response scale format (i.e., Likert versus binary) may moderate this relationship. The episodic versus semantic memory paradigm and the parallel remember versus know judgment are employed to capture memory processing. The counterintuitive findings from the mixed-effects model shed light on alternative mechanisms that may be at play in memory processing and bear important implications for future research.

Keywords: emotion, memory processing, leadership measurement

Emotional Tone Embedded in Leadership Items and Memory Processing: Implications for Leadership Measurement

Introduction

Leadership research has predominantly relied on recollective self-report questionnaires to collect leadership ratings (Fischer, Hambrick, Sajons, & Van Quaquebeke, 2020; Hunter, Bedell-Avers, & Mumford, 2007). Because followers typically answer questionnaires based on their recall of past events or draw on retrospective judgments, the emphasis on memory recognition in leadership measurement is critical (Hansbrough et al., 2021). However, research shows that leadership ratings collected through questionnaires tend to capture information other than actual leader behaviors, such as followers' leadership schemas (Lord, Epitropaki, Foti, & Hansbrough, 2020) and generalized affective reactions (Yammarino, Cheong, Kim, & Tsai, 2020). It is not surprising that there is widespread concern in the area of leadership and the broader field of organizational behavior regarding the effectiveness of questionnaires in capturing actual leader behaviors (Banks, Fischer, Gooty, & Stock, 2021; Banks, Woznyj, & Mansfield, 2021).

One way to mitigate such concern and enhance leadership measurement is to promote response process validity (Borsboom, Mellenbergh, & van Heerden, 2004; Cronbach & Meehl, 1955; Hughes, 2018). Response process validity implies that if a leadership item is intended to measure specific leadership behaviors vs. generalized leadership impressions, then ideally respondents should tap into their corresponding memory sources that store such information. Specifically, items used to measure leadership behaviors should elicit episodic memory, while items intended to measure generalized leadership impressions should elicit semantic memory (Tulving, 1972, 2002; Hansbrough et al., 2015, 2021; Shondrick et al., 2010). As such, the enhancement of response process validity in leadership measurement requires a thorough

understanding of the factors that influence memory processing. Extant research in this area mainly focuses on rater factors (e.g., Schacter et al., 2008, 2012), item linguistic characteristics (e.g., Semin & Fiedler, 1991), and interventions (e.g., Martell & Evans, 2005). Scholars have paid disproportionate attention to the cognitive path vs. the affective path of memory processing (schemas, e.g., Lord et al., 2020; Shondrick et al., 2010). It is surprising to know that the influence of emotional tone embedded in leadership items on respondents' memory processing remains elusive while most leadership items are emotion-laden. Considering the fact that emotion plays a salient role in memory processing (Blaney, 1986) and the acknowledgement that affective processing is closely intertwined with cognitive processing (Allen, Kaut, & Lord, 2008), this study aims at unpacking the relationship between emotional tone and memory processing as well as its implications for leadership measurement. Figure 1 depicts the conceptual model.

Insert Figure 1 about here

Positive and negative emotional tone tend to have disparate implications for memory processing in leadership measurement. Upon memory encoding, positive (negative) emotions engage relational (item-specific) processing which enables information to be stored in a generalized (detailed) manner (Bless, 2001; Bless, Clore, et al., 1996; Bless, Schwarz, & Wieland, 1996; Clore, Gasper, & Garvin, 2001). Consequently, negative emotion tends to have a recall advantage in terms of specific details of the events during memory retrieval (e.g., Baumeister et al., 2001). Because episodic (semantic) memory is context-specific (context-general) in nature and concerns specificity (generalized impressions), negative (positive) emotions are more likely to evoke episodic (semantic) memory. Moreover, since episodic retrieval follows a pattern-

completion process where specific details of the events are recalled, respondents tend to incur a longer reaction time when they activate episodic memory (Horner et al., 2015; Tulving, 1983).

The design of leadership questionnaire, particularly permutation of leadership items (i.e., positive leadership items followed by negative leadership items and vice versa) and response scale format (i.e., Likert vs. binary scale) may further moderate the relationship between emotional tone and memory processing. Research suggests that episodic and semantic memory processing are not dichotomous but rather highly dynamic (Winocur & Moscovitch, 2011; Winocur, Moscovitch, & Bontempi, 2010) and episodic memory processing has a carryover effect on semantic memory processing (Evans et al., 2015; Nadel & Moscovitch, 1997). Considering the association between negative (vs. positive) emotional tone and episodic (vs. semantic) memory processing, the current study investigates if a parallel spillover effect in terms of memory processing exists from negative to positive leadership items. Another design element of interest is the response scale format of the leadership items, which follows either a 5-point Likert scale (from 1 – Strongly Disagree to 5 – Strongly Agree) or a "Yes" or "No" binary scale. Although Likert scale is the most popular response scale format in leadership research, the use of binary scale is not without its merits (e.g., increase response rates and reduce respondent fatigue, Dolnicar et al., 2011). From the perspective of memory processing, Likert (vs. binary) scale may engender semantic (vs. episodic) memory because Likert (vs. binary) scale asks about generalized impressions (vs. a concrete and definite answer which requires support from the recollection of specific details) (Hansbrough et al., 2021).

The current study makes several contributions. First, it enriches the extant literature studying factors that affect memory processing in leadership measurement by considering the influence of emotional tone embedded in leadership items. This is not only about adding another leadership item-level factor to the checklist that impacts the memory source of leadership ratings,

but also concerns balancing the disproportionate attention paid to affective (vs. cognitive) processing. Moreover, as will be discussed later, the influence of other rater-level (e.g., dyadic duration) and leadership item-level (e.g., base rate of specific leadership items) factors on memory processing should be considered to develop a better understanding of the role played by emotional tone.

Second, disentangling the leadership item-level factors that influence memory processing could facilitate the attainment of response process validity and generate insights for leadership scale construction and questionnaire design, which ultimately facilitate leadership measurement. Specifically, researchers should consider the extent to which emotional tone, permutation, and response scale format of the leadership items used in the questionnaire may influence respondents' memory processing to avoid introducing biases and incurring "false alarms" (e.g., ascribing specific behaviors to a leader when the leader did not engage in such behaviors) in leadership measurement (e.g., Hansbrough et al., 2015, 2021; Shondrick et al., 2010).

Third, the current study enhances the understanding of the carryover effect from episodic to semantic memory processing in leadership measurement. The findings indicate a more nuanced association between negative (positive) leadership items and episodic (semantic) memory processing than hypothesized, which calls for further investigation into the relative recall (dis)advantage between positive and negative emotions. Particularly, the issue concerns whether negative leadership events are recalled better than positive leadership events (e.g., Baumeister et al., 2001) or vice versa (e.g., Sedikides & Green, 2009).

Theoretical Development and Hypotheses

Memory Systems and Remember vs. Know Judgments

According to Tulving (1972, 2002), declarative memory can be classified as episodic versus semantic memory. Semantic memory encompasses general, context-independent representations of a target. Episodic memory is context-specific and concerns details of the events and emotions experienced during encoding (Allen et al., 2008; Tulving, 2002). Tulving (1985) argues that individuals' metacognitions (i.e., memory monitoring, Hampton et al., 2020) enable them to distinguish between "remember" and "know" judgments, whereby remembering is based on a vivid recollection of a specific event (episodic memory), and knowing is based on a general feeling or impression (semantic memory). We follow previous research and use the remember/know paradigm to capture respondents' memory processing.

Emotion and Memory Processing

Emotion impacts both memory encoding and retrieval (Clore, Schwarz, & Conway, 1994; Kensinger & Corkin, 2003, 2004; Morris, 1989). Positive emotions induce information to be coded using general knowledge structures (i.e., relational processing), whereas negative emotions prompt information to be coded in a detailed manner (i.e., item-specific processing) (Bless, 2001; Bless, Clore, et al., 1996; Bless, Schwarz, et al., 1996; Clore et al., 2001). As such, more details about the negative events are available upon retrieval, which gives rise to an overall recall advantage for negative events and details (Rimmele, Davachi, Petrov, Dougal, & Phelps, 2011; Sharot, Delgado, & Phelps, 2004; Sharot & Yonelinas, 2008). This is in line with the conclusion in Baumeister and colleagues' (2001) review that negative memories are recalled better than positive memories. In this vein, I propose that positive (negative) leadership items stimulate respondents to recall positive (negative) leadership events, thus are more likely to trigger semantic (episodic) memory processing. Research demonstrates that episodic memory processing tends to take more time because it follows a pattern completion process (Horner & Burgess, 2013, 2014) such that all

elements related to the specific event or behavior are recalled (Horner, Bisby, Bush, Lin, & Burgess, 2015; Tulving, 1983). As such, we posit that:

H1a: Positive leadership items will have fewer remember judgments than negative leadership items.

H1b: Positive leadership items will incur a shorter reaction time than negative leadership items.

Permutation and Memory Processing

Episodic and semantic memory are not dichotomous or mutually exclusive entities (Renoult, Irish, Moscovitch, & Rugg, 2019). Research suggests that there is marked neural overlap between the brain regions that store episodic memory (i.e., the hippocampal structures and adjacent regions of the medial temporal lobe) and semantic memory (i.e., other neocortical regions) (Dudai, 2012; Moscovitch, Cabeza, Winocur, & Nadel, 2016; Sekeres, Winocur, & Moscovitch, 2018; Squire, Genzel, Wixted, & Morris, 2015; Winocur & Moscovitch, 2011). Using experiments, Evans et al. (2015) have shown the progressive neural insult of key structures engaged in episodic and semantic memory processing, such that when respondents switch from an episodic memory task to a perceptual task, the cognitive processes relevant to the episodic task were carried over and active during the perceptual task. Put differently, episodic memory processing will spill over to semantic memory processing that follows. Along a similar vein, we hypothesize that there will be a spillover effect in terms of episodic memory processing associated with negative leadership items to positive leadership items when respondents answer negative leadership items followed by positive leadership items. In other words, respondents are more likely to use episodic memory processing for the positive items that come after the negative leadership items. As a result, the

ensuing positive leadership items will have more remember judgments and incur a longer reaction time. We thus propose:

H2a: The effect of positive leadership items on remember judgments will be weaker when answered second.

H2b: The effect of positive leadership items on reaction time will be weaker when answered second.

Response Scale Format and Memory Processing

A binary scale is likely to trigger episodic memory processing because it primes respondents to recall concretely (a definite "Yes" or "No") whether they have experienced specific leadership events or behaviors, which demands the recollection of specific leadership events or behaviors (Hansbrough et al., 2021). In contrast, a Likert scale that asks respondents the extent to which they agree with a particular statement tends to be associated with summary evaluations and relies on schemas and heuristics in the semantic memory system (Hansbrough et al., 2021). We posit that response scale format will moderate the relationship between emotional tone and memory processing. Specifically, when rating positive leadership items, binary scale prompts respondents to use episodic memory processing to a larger extent as compared to Likert scale, therefore incur a longer reaction time. However, the influence of binary scale on triggering episodic memory processing is different between positive and negative leadership items. Particularly, the power of binary scale to incur episodic memory processing on negative leadership item is expected to be weaker than that on positive leadership items, because the influence of negative emotion on episodic memory processing of negative leadership items is expected to be strong enough to weaken the role played by binary scale. This leads to our third hypothesis:

H3a: The effect of positive leadership items on remember judgments will be weaker when using binary scale.

H3b: The effect of positive leadership items on reaction time will be weaker when using binary scale.

Methods

Sample and Procedure

The current study concerns a three-factor mixed design (see Table 1), with emotional tone embedded in leadership items (negative vs. positive) being a within-subjects factor and permutation (positive first vs. positive second) and response scale format (binary vs. Likert) being between-subjects factors.

Insert Table 1 about here

Because the focal predictions of the current study contain main effects and two-way interactions, we rely on a power analysis of a 2 x 2 design with a power of .8 using "pwr2" package (Lu, Liu, Koestler, & Lu, 2017) in R. The sample size per cell is estimated to be 50, which is consistent with previous recommendations concerning randomization in experiments (Lonati, Quiroga, Zehnder, & Antonakis, 2018). We have pre-registered with AsPredicted (https://aspredicted.org/47B_1KB) to enhance the transparency of the current study and the credibility of the results before any data has been collected for the study.

In line with the power analysis (see Appendix 1), we recruited 200 participants from a crowdsourcing site, Prolific. The criteria for participation were as follows: Participants were required to be native English speakers, working full time, and be able to answer the questionnaire using a computer or laptop with a normal keyboard and not any other type of device such as tablets

and phones. Having all the participants as native English speakers is to control the potential impact of cultural differences on affective and cognitive processing (e.g., Oyserman & Lee, 2008; Schimmack et al., 2002; Schwarz et al., 2010). Having all the participants as full-time employees is to make sure that they have experiences with current leaders. The requirement for a normal keyboard is because reaction time will be captured using JavaScripts through Qualtrics, which necessitates the use of keyboards. We pre-tested the survey to provide an estimated completion time (5 – 8 minutes). The obtained median completion time of 7.68 minutes falls within the pre-test range. Four responses were removed from the sample because the corresponding respondents failed the data integrity check by either stating that they are not sure if their data should be included or that their data should be excluded from subsequent analysis. Another 15 responses were dropped from the sample because the respective respondents demonstrated that they did not understand the differences between remember and know judgments when answering the questionnaire. For instance, one respondent wrote that remember judgment is chosen "Because I can remember that quality in my manager but don't have a specific example", which depicts a know judgment. In the end, I have 181 respondents (90 females, 88 males, and 3 non-binary).

The average respondent was 37 years old, has been working for his/her current supervisor for more than two years, and interacts with his/her current supervisor on a daily basis. 162 respondents (90%) reported a positive sentiment associated with the last interaction with their immediate supervisors, while 19 respondents (10%) experienced negative sentiment. Regarding educational attainment, 48 (27%) respondents had some college education, 81 (45%) had a bachelor's degree, 30 (17%) had a master's or professional degree, 3 (2%) had a doctorate degree, whereas 19 (10%) had a high school diploma. Respondents reported a wide array of occupations including manager, sales specialist, consultant, auditor, educator, accountant, teacher, web

developer, engineer, financial analyst. In terms of position within the organization, 95 participants are non-managerial employees, 73 are middle managers, and 13 are upper-level managers. 50 respondents (28%) rated their immediate supervisors of the same gender and 131 (72%) evaluated immediate supervisors of a different gender.

All respondents were asked to rate their current leader on 10 positive and 10 negative leadership items. Respondents were randomly assigned based on two dimensions: the permutation and the response scale format of leadership items. Regarding the permutation, respondents either rated the 10 positive leadership items first followed by 10 negative leadership items (i.e., positive first) or vice versa (i.e., positive second). Respondents were also randomly assigned to a binary scale ("Yes" or "No") or a 5-point Likert scale (ranging from 1 - Strongly Disagree to 5 - Strongly Agree).

Right after a respondent rated his/her immediate supervisor on a specific leadership item, he/she was asked to reflect on the type of memory processing engaged when performing the rating. As detailed below, following Tulving's (1985) methodology, respondents were provided with definitions of remember and know judgments each time they answered the memory processing questions. To mitigate the concern that respondents may regard one type of judgment as more desirable than the other, it has been emphasized in the instructions that both judgments are useful and the judgments do not differ in terms of their confidence or uncertainty. The instructions read as follows:

> "We have two different ways that we make judgments about other people, remembering and knowing. **Remembering is based on a vivid recollection of a specific event.** For example, we might describe someone as outgoing because we can recall specific examples of their behavior. Alternatively, **knowing is based on a general feeling or impression**

about a person. It is important to note that both types of memory are useful and that one is not inherently better than the other. Moreover, remember and know judgments do not differ in terms of their confidence or certainty. For example, we can be equally confident about a judgment even though we might not associate it with a specific event. For each of the following items please rate your supervisor and then using the definitions above indicate whether your rating reflects a remember or a know judgment."

Selection of Leadership Items

We started with a dictionary that consists of seven established leadership scales with 126 leadership items that are coded as either positive or negative (see Table 2). We used k-means clustering (i.e., partitioning leadership items into k clusters so that each leadership item belongs to the cluster with the nearest mean, Steinley, 2006) to extract 20 positive and 20 negative items (see Table 3) that are similar in terms of emotional valence (Blaney, 1986), self-reference (Jiga-Boy, Clark, & Semin, 2013), and concreteness (Marschark & Cornoldi, 1991; Paivio, 1991), which pertain to linguistic factors identified by previous research to influence memory processing. To check whether the actual interpretations of the emotional tone embedded in the selected leadership items align with the perceptions that we intend to elicit, we engaged subject matter experts[1] in a manipulation check survey by asking them whether they think each leadership item reflects a positive or negative attribute of a leader. From there, we removed the leadership items with ambiguous emotional tones and the final survey ends up with 10 positive and 10 negative leadership items that are similar along the dimensions of emotional valence, self-reference, and concreteness (see Table 4). Table 5 demonstrates the measures of the three linguistic characteristics and Table 6 shows the corresponding descriptive statistics.

[1] PhD students from the Leadership and Organizational Science Department and Bernard M. and Ruth R. Bass Center for Leadership Studies at Binghamton University, State University of New York.

Insert Tables 2 - 6 about here

Measures

Dependent Variables. The dependent variable concerning semantic vs. episodic memory processing was measured as the percentage of remember judgments (i.e., remember judgments out of overall remember and know judgments) for 10 negative and 10 positive leadership items respectively. Another dependent variable is reaction time, which was captured using Qualtrics in units of seconds.

***Independent Variables*.** The emotional tone of leadership items was measured using a binary scale with negative encoded as 0 and positive as 1. Permutation of leadership items was captured using a binary scale with positive first (10 positive leadership items followed by 10 negative leadership items) represented by 0 and positive second (10 negative leadership items followed by 10 positive leadership items) by 1. Response scale format was denoted using a binary scale with Likert = 0 and binary = 1.

Statistical Analysis

Following Shiverdecker and LeBreton (2019), we employed the model building/comparison approach to conducting a two-level analysis (i.e., within-individual and between-individual levels of analysis) using a mixed-effects model (MEM) framework. The two-level model reflects the hierarchical structuring of repeated measures of memory processing associated with different emotional tones nested within individual respondents who are exposed to different permutations and response scale formats. Although the ANOVA model remains to be one of the most commonly used analysis tools for experimental design, scholars have pointed out the merits of using MEM (e.g., Hoffman & Rovine, 2007; Quené & van den Bergh, 2004). MEM

is appropriate for the current study because this analytic approach (a) takes account of the nonindependence and covariance of observations because memory processing is performed and reported by the same subject, (b) allows initial values (i.e., intercepts) and rates of change (i.e., slopes) to vary among individuals to capture potential individual differences, and (c) enables the investigation of predictors at both within- and between-individual level (Hoffman & Rovine, 2007; Quené & van den Bergh, 2004). We used the multilevel package in R (Bliese, 2022) to conduct the MEM analysis. For brevity, we will use the percentage of remember judgments as the dependent variable for the illustration of MEM. The same modeling procedures apply when reaction time serves as the dependent variable.

Hypothesis 1a posits that positive leadership items have a smaller percentage of remember judgments than negative leadership items. The Level-1 equation reflects the overall change in the percentage of remember judgments at the within-individual level between negative and positive leadership items. Following the notation provided by Raudenbush and Bryk (2002), the Level-1 equation of the current study is as follows:

$$\text{Level 1:}\quad \text{Remember Percentage}_{ij} = \beta_{0j} + \beta_{1j}\left(\text{emotional tone}_{ij}\right) + r_{ij}$$

where $Remember\ Percentage_{ij}$ is the percentage of remember judgments for leadership items with emotional tone i as reported by individual respondent j; β_{0j}, β_{1j}, and r_{ij} are the intercept, slope, and residual score for leadership items with emotional tone i on the percentage of remember judgments of individual respondent j, respectively. The intercept refers to the initial value of the percentage of remember judgments engaged by individual respondent j, and the slope represents the changes in the percentage of remember judgments for negative vs. positive emotional tone. As such, the estimation of the coefficient on emotional tone (β_{1j}) will be observed to test the main effect postulated in Hypothesis 1a.

Hypotheses 2a and 3a propose cross-level moderating effects of permutation and response scale format on the above relationship, respectively. To test Hypotheses 2a and 3a, we add between-individual level variables, namely permutation and response scale format, to the Level-2 equations. The Level-2 equations are used to explore the conjecture that permutation and response scale format are related to the intercept and slope coefficient variation of the Level-1 equation. Specifically, these random components reflect the cross-level moderating effects of permutation and response scale format on changes of the percentage of remember judgments, resulting in the Level-2 equations for the model in this study demonstrated below:

$$\text{Level 2:}\quad \beta_{oj} = \gamma_{00} + \gamma_{01}(permutation_j) + \gamma_{02}(response\ scale\ format_j) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(permutation_j) + \gamma_{12}(response\ scale\ format_j) + u_{1j}$$

The first Level-2 equation tests whether permutation and response scale format can explain some of the variance in the intercept of the emotional tone and the percentage of remember judgments relationship. The first Level 2 equation states that the mean level of respondent *j*'s percentage of remember judgments on emotional tone i can be modeled as a function of two components. The first component of the first equation is the mean level of the percentage of remember judgments (γ_{00}) for all respondents. The second and third components of the first equation are coefficients associated with the main effects of permutation (γ_{01}) and response scale format (γ_{02}) on the percentage of remember judgments.

The second Level-2 equation determines whether permutation and response scale format can explain some of the variance in the slope of the emotional tone and the percentage of remember judgments relationship. The slope of leadership items with emotional tone i on the percentage of remember judgments of individual respondent j (β_{1j}) is a function of an overall slope estimate (γ_{10}), coefficients associated with permutation (γ_{11}) and response scale format (γ_{12}), and an error

term (u_{0j}). The estimations on coefficients γ_{11} and γ_{12} allow the assessment of the cross-level moderating effects of permutation and response scale formats on the relationship between emotional tone and the percentage of remember judgments. The full notation of the MEM equations are as follows:

$$
\begin{aligned}
&Remember\ Percentage_{ij} \\
&\quad = \gamma_{00} + \gamma_{01}(permutation_j) + \gamma_{02}(response\ scale\ format_j) \\
&\quad + \gamma_{10}(\text{emotional tone}_{ij}) + \gamma_{11}(\text{emotional tone}_{ij})(permutation_j) \\
&\quad + \gamma_{12}(\text{emotional tone}_{ij})(response\ scale\ format_j) \\
&\quad + u_{1j}(\text{emotional tone}_{ij}) + u_{0j} + u_{1j}
\end{aligned}
$$

Results

Descriptive Statistics. Table 7 reports the means, standard deviations, and correlations of variables. The percentage of remember judgments is significantly correlated with permutation.

Insert Table 7 about here

Hypothesis Testing. Table 8 shows the results of the mixed-effects model. The model building/comparison approach (Shiverdecker & LeBreton, 2019) reveals that a MEM model with both random intercepts and random slopes fits better than a random intercept and fixed slope model (p = 0.0174) based on the -2log likelihood test, which is widely known to be conservative (Pinheiro & Bates, 2000) and is generally acceptable to halve the p-values as if the tests were one-tailed (LaHuis & Ferguson, 2009).

Insert Table 8 about here

The slope associated with emotional tone on the percentage of remember judgments is not significant ($\gamma_{10} = -0.05, p = 0.26$), as such Hypothesis 1a 2qw not supported. The slope of emotional tone on reaction time is statistically significant ($\gamma_{10} = 1.87, p < 0.001$). However, the positive coefficient suggests that positive leadership items incur a longer reaction time than negative leadership items, which goes against Hypothesis 1b.

The cross-level interaction between permutation and emotional tone on the percentage of remember judgments is negative and statistically significant ($\gamma_{01} = -0.15, p < 0.001$). The interaction plot is available in Figure 2a. As can be observed, permutation only negatively moderates the relationship between negative emotional tone and the percentage of remember judgments. In other words, negative leadership items answered after positive leadership items have a higher percentage of remember judgments than negative leadership items answered before positive leadership items. In contrast, permutation is not found to have a moderating effect on positive leadership items. Therefore, Hypothesis 2a was not supported. The interaction plot demonstrating the cross-level interaction between permutation and emotional tone on reaction time is available in Figure 2b. As can be observed, permutation positively (negatively) moderates the relationship between negative (positive) emotional tone and reaction time ($\gamma_{01} = 2.03, p < 0.001$). Leadership items answered second, both positive and negative, had shorter reaction time than leadership items answered first. As such, Hypothesis 2b was not supported.

Insert Figures 2a and 2b about here

The cross-level interaction between response scale format and emotional tone on the percentage of remember judgments ($\gamma_{12} = 0.04, p = 0.36$) as well as reaction time ($\gamma_{12} = -0.27, p = 0.31$) are not statistically significant. Hypotheses 3a and 3b were not supported.

Discussion

Overall, the hypotheses were not supported through a two-level (i.e., within-individual and between-individual levels) mixed effects model. However, this does not necessarily mean that the spillover effect from episodic to semantic memory processing does not hold. Rather, the findings that go against the hypotheses may be contributed by the broken link between episodic (semantic) memory processing and negative (positive) leadership events. There are several possible reasons to explain why the hypotheses were not supported.

First, negative leadership events tend to have low base rates and episodic memory processing associated with negative leadership items may fail to reach a critical mass to be carried forward to influence the ensuing positive leadership items. The questionnaire contains ten negative leadership items from toxic, abusive, and exploitative leadership scales. However, negative items, such as "My manager publicly belittles subordinates" and "My manager has explosive outbursts", tend to represent low base-rate events. Theoretically, these leadership behaviors may seldom be witnessed in the workplace because a leader who engages in negative behaviors often most likely would not remain to be a leader for long (Banks, Woznyj, et al., 2021). The low base rate of negative vs. positive leadership events is also implied by empirical data. In the current study, the average response for positive and negative leadership items using Likert scale is 3.58 ($sd = 0.31$) and 1.77 ($sd = 0.25$) respectively, and t.test for the mean responses of positive and negative leadership items are statistically significant ($p < 0.01$). For leadership items using binary scale, on average, 73.16% (vs. 16.63%) of the respondents selected “Yes” for positive (vs. negative) leadership items, and the chi-square test suggests that the distribution of responses for positive and negative leadership items are statistically different ($p < 0.01$). Moreover, when respondents are asked to provide an example of a “remember” item, 46% of the respondents described a positive

leadership event (e.g., “My supervisor helped someone in the community.”), whereas 15% of the respondents described a negative leadership event (e.g., “I picked that my boss manipulates to get results because I remember a time where data was changed.”)[2]. In fact, Walker and colleagues (2003) have already pointed out the asymmetrical distribution of positive versus negative events in people's lives: positive events (50%) are about twice as frequent as negative events (25%). The same distribution pattern may hold when it comes to the leadership domain. After all, the linkage between negative leadership events and episodic memory processing may be weakened due to the low base rate. 89.5% of the respondents reported a positive experience on the last interaction with their respective leaders.

Second, the relationship between negative leadership items and semantic memory processing may be strengthened if respondents use semantic processing (i.e., “know judgment”) to rate their leaders when they have not encountered specific leader events described in the leadership item. For instance, one respondent mentioned that "I know my boss would not belittle someone because it doesn't seem like something he would do. I have not seen him belittle anyone or recall him stating explicitly that it was something he wouldn't do." Apparently, due to the inaccessibility of specific instances of the leader belittling subordinates, the respondent falls back on semantic memory processing to make inferences based on generalized impressions of the leader. Along this vein, because respondents are less likely to encounter negative (vs. positive) leadership events, they may engage in semantic memory processing instead when answering negative leadership items in the questionnaire.

Third, the relatively short dyadic duration between the leader and the follower and the high base rate may strengthen (undermine) the association between positive leadership items and

[2] The remaining 38% of the respondents provided more objective description of what “remember” means, such as " I chose remember when I had a specific memory of a situation pertaining to the statement.”

episodic (semantic) memory processing. As explained earlier, there exists a dynamic interplay between episodic and semantic memory processing. However, it takes time and experience for episodic memory to be transformed into a more semantic form ("systems consolidation", Winocur et al., 2010; Winocur & Moscovitch, 2011). On average, the respondents spent 3.44 years working with their current leaders and around 50% of the respondents have a dyadic duration of fewer than two years. Yet it takes a long period of time for (autobiographical) memories to become more semantic with repetitive recollection (Cermak, 1984). Moreover, it is less likely for memory to lose its contextual (i.e., episodic) aspects (Winocur & Moscovitch, 2011) within this dyadic duration. In other words, with a relatively short dyadic duration, both positive and negative leadership items are more (less) likely to be associated with episodic (semantic) memory. As explained earlier, the linkage between negative leadership items and episodic memory processing may be hampered due to low base rates. Consequently, the impact of short dyadic duration on memory processing associated with negative leadership items is likely compromised. In contrast, the correlation between positive leadership items and episodic (semantic) memory processing may be reinforced (hindered) because of high base rates and short dyadic duration.

Last but not least, social desirability bias and savouring may result in the recall advantage of positive vs. negative leadership events and contribute to the null finding of the main effect between emotional tone and memory processing. Social desirability can be defined as a distortion of responses in a socially desirable direction as a result of self-deception and other-deception (Nederhof, 1985). Although the survey is purely anonymous, individuals' innate desire to construct and maintain a positive self-image may prevent respondents from honestly reporting negative leadership behaviors. The extrinsic motivation to recall negative leadership episodes hardly exists as well, as an average respondent received US$1.5 in return for participating in the questionnaire.

The combination of reporting negative leadership behaviors with the fact that one is still working with such a leader probably signals that it is beyond the individual's power to change the status quo, which contradicts the intention to be the master of one's life. For the same reason, respondents might even be motivated to recall specific positive behaviors of the leader to demonstrate that they are doing great in terms of their career choice. Research shows that memory can serve as a self-protective mechanism such that positive memories are recalled better than negative memories (Sedikides & Green, 2009). Savouring denotes a process through which individuals recall positive experiences, cherish positive events in the present, and anticipate future positive events (Sthapit, Björk, Jiménez-Barreto, & Stone, 2021). Research shows that savoring helps individuals maintain positive affect (Quoidbach, Berry, Hansenne, & Mikolajczak, 2010; Speer, Bhanji, & Delgado, 2014), which may prohibit respondents from recalling negative leadership events. After all, episodic memory associated with positive (negative) leadership items has been boosted (hindered).

To sum up, the arguments so far demonstrate that (1) the association between negative (positive) leadership items and episodic memory has been weakened (strengthened) due to the low (high) base rate, social desirability, and savouring; (2) the linkage between positive leadership items and semantic memory has been hampered because of the short dyadic duration and the high base rate. The spfmiillover effect of episodic on semantic memory processing may still hold, just that the progressive order is now from positive to negative (instead of the hypothesized negative to positive) leadership items due to the reasons above. Recall that episodic memory processing tend to incur a longer reaction time than semantic memory processing. These may explain the null finding of H1a (no main effect between emotional tone and memory processing), the counterintuitive finding of H1b (positive leadership items incur a longer reaction time than negative leadership items), and the contradictory finding of H2a (negative leadership items

answered second incur more episodic memory processing than negative leadership items answered first).

Along the same vein, it should be expected that negative leadership items answered after positive leadership items will incur a longer reaction time than negative leadership items answered before positive leadership items due to the spillover effect from episodic to semantic memory processing. However, results from H2b suggest that leadership items answered second, whether positive or negative, incur a shorter reaction time. Moreover, positive leadership items answered first incur a long reaction time than negative leadership items answered second ($p < 0.001, mean_{pos\,rt} = 3.33, mean_{neg\,rt} = 1.61$), whereas negative leadership items answered first also incur a longer reaction time than positive leadership items answered second ($p < 0.001, mean_{pos\,rt} = 1.71, mean_{neg\,rt} = 3.64$). Rather than providing implications for the relationship between emotional tone and memory processing, the H2b results may suggest a general response pattern where people tend to spend more time on the front compared to the back end of the questionnaire, either due to limited cognitive resources and attention span or increased familiarity with the questionnaire design.

It is also interesting to observe that response scale format does not have a significant impact on memory processing associated with the leadership items nor significantly moderate the impact of emotional tone on memory processing. Although the findings accord much more importance to permutation than response scale format as a moderator, the extent to which factors such as (1) base rate, (2) dyadic duration, (3) social desirability and savouring, influence the role played by response scale format warrants further investigation.

Limitations and Future Research Directions

There are several limitations associated with this study. First, individual differences and leadership item discrepancies may prevent the current study from reaching its full potential. The questionnaire is designed such that respondents refer to their respective leaders when answering the questions, and the lack of a common leader runs the risk of leaving individual differences in terms of leadership experiences unaccounted for. For instance, some respondents may have a shorter dyadic duration with their leaders and have fewer leadership episodes to recall; some leaders may incur fewer negative behaviors than others. At the leadership item level, factors other than emotional tone, such as the base rate of leadership items, may influence memory processing. Indeed, the MEM results indicate that both the base rate of leadership items and dyadic duration are important elements to consider before delving into the role played by emotional tone embedded in leadership items. One way to address this concern would be to employ an experimental vignette approach, whereby respondents watch a video of leadership behaviors conducted by either human actors or avatars.

Second, the original theorizing focus on the notion that negative memories are recalled better than positive memories following the "bad is stronger than good" principle (Baumeister et al., 2001). However, there is evidence from research on autobiographical memory that unpleasant life events are recalled more poorly than pleasant life events (Sedikides & Green, 2009). The MEM results seem to be more in line with the second stream of assertions. As such, more effort should be expended to delineate the recall discrepancy between negative and positive memories and its implications on leadership measurement.

Third, the use of remember/know paradigm (Tulving, 1985) to capture episodic/semantic memory processing follows the dual-proecss theory (i.e., individual recognition decisions are based on either remember or know judgment, Yonelinas, 1994). However, supporters of a single-

process account of recognition (i.e., individual recognition decisions are a function of both remember and know judgment, Slotnick & Dodson, 2005) suggest a more nuanced memory processing system which can be better revealed using functional magnetic resonance imaging (fMRI). Future research may incorporate tools such as fMRI to ensure that memory processing is accurately captured.

Conclusion

The current study intends to contribute to the literature on the intersection of leadership measurement and memory processing by examining leadership item-level factors that may influence respondents' memory processing when answering leadership questionnaires, specifically the emotional tone embedded in leadership items as well as the permutation and response scale format of leadership items. The counterintuitive findings from the current study highlight the need to develop a deeper understanding of the roles played by individual differences in terms of leadership experiences and other potential factors at the leadership item level that influence memory processing. The current study also brings to attention the disparate arguments concerning the recall disparity of positive vs. negative leadership events which merit further investigation to advance leadership measurement.

References

Allen, P. A., Kaut, K. P., & Lord, R. R. 2008. Emotion and episodic memory. ***Handbook of Behavioral Neuroscience***, 18: 115–132.

Banks, G. C., Fischer, T., Gooty, J., & Stock, G. 2021. Ethical leadership: Mapping the terrain for concept cleanup and a future research agenda. ***The Leadership Quarterly***, 32(2): 101471.

Banks, G. C., Woznyj, H. M., & Mansfield, C. A. 2021. Where is "behavior" in organizational behavior? A call for a revolution in leadership research and beyond. ***The Leadership Quarterly***, 101581.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. 2001. Bad is Stronger than Good. ***Review of General Psychology***, 5(4): 323–370.

Blaney, P. H. 1986. Affect and memory: A review. ***Psychological Bulletin***, 99(2): 229–246.

Bless, H. 2001. The consequences of mood on the processing of social information. ***Blackwell Handbook of Social Psychology: Intraindividual Processes***, 391–412.

Bless, H., Clore, G. L., Schwarz, N., Golisano, V., Rabe, C., et al. 1996. Mood and the use of scripts: Does a happy mood really lead to mindlessness? ***Journal of Personality and Social Psychology***, 71(4): 665–679.

Bless, H., Schwarz, N., & Wieland, R. 1996. Mood and the impact of category membership and individuating information. ***European Journal of Social Psychology***, 26(6): 935–959.

Bliese, P. 2022. ***Multilevel Modeling in R (2.7)***, 1–83.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. 2004. The Concept of Validity. ***Psychological Review***, 111(4): 1061–1071.

Brown, M. E., Treviño, L. K., & Harrison, D. A. 2005. Ethical leadership: A social learning perspective for construct development and testing. ***Organizational Behavior and Human Decision Processes***, 97(2): 117–134.

Cermak, L. S. 1984. The episodic-semantic distinction in amnesia. ***Neuropsychology of Memory***, 55–62.

Clore, G. L., Gasper, K., & Garvin, E. 2001. Affect as information. ***Handbook of Affect and Social Cognition***, 121–144.

Clore, G. L., Schwarz, N., & Conway, M. 1994. Affective causes and consequences of social information processing. ***Handbook of social cognition: Basic processes; Applications, Vols. 1-2, 2nd ed***: 323–417. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Cronbach, L. J., & Meehl, P. E. 1955. Construct validity in psychological tests. ***Psychological Bulletin***, 52(4): 281–302.

Dolnicar, S., Grun, B., Leisch, F., & Rossiter, J. 2011. Three good reasons NOT to use five and seven point Likert items. ***Faculty of Commerce - Papers (Archive)***. https://ro.uow.edu.au/commpapers/775.

Dudai, Y. 2012. The Restless Engram: Consolidations Never End. ***Annual Review of Neuroscience***, 35(1): 227–247.

Evans, L. H., Herron, J. E., & Wilding, E. L. 2015. Direct Real-Time Neural Evidence for Task-Set Inertia. ***Psychological Science***, 26(3): 284–290.

Fischer, T., Hambrick, D. C., Sajons, G. B., & Van Quaquebeke, N. 2020. Beyond the ritualized use of questionnaires: Toward a science of actual behaviors and psychological states. ***The Leadership Quarterly***, 31(4): 101449.

Gaddy, M. A., & Ingram, R. E. 2014. A meta-analytic review of mood-congruent implicit memory in depressed mood. ***Clinical Psychology Review***, 34(5): 402–416.

Graen, G. B., & Uhl-Bien, M. 1995. Relationship-based approach to leadership: Development of leader-member exchange (LMX) theory of leadership over 25 years: Applying a multi-level multi-domain perspective. ***The Leadership Quarterly***, 6(2): 219–247.

Hampton, R. R., Engelberg, J. W. M., & Brady, R. J. 2020. Explicit memory and cognition in monkeys. ***Neuropsychologia***, 138: 107326.

Hansbrough, T. K., Lord, R. G., Schyns, B., Foti, R. J., Liden, R. C., et al. 2021. Do you remember? Rater memory systems and leadership measurement. ***The Leadership Quarterly***, 32(2): 101455.

Hoffman, L., & Rovine, M. J. 2007. Multilevel models for the experimental psychologist: Foundations and illustrative examples. ***Behavior Research Methods***, 39(1): 101–117.

Horner, A. J., Bisby, J. A., Bush, D., Lin, W.-J., & Burgess, N. 2015. Evidence for holistic episodic recollection via hippocampal pattern completion. ***Nature Communications***, 6(1): 7462.

Horner, A. J., & Burgess, N. 2013. The associative structure of memory for multi-element events. ***Journal of Experimental Psychology: General***, 142(4): 1370–1383.

Horner, A. J., & Burgess, N. 2014. Pattern Completion in Multielement Event Engrams. ***Current Biology***, 24(9): 988–992.

Hughes, D. 2018. Psychometric validity: Establishing the accuracy and appropriateness of psychometric measures. ***The Wiley handbook of psychometric testing: A multidisciplinary approach to survey, scale and test development***. John Wiley & Sons Ltd.

Hunter, S. T., Bedell-Avers, K. E., & Mumford, M. D. 2007. The typical leadership study: Assumptions, implications, and potential remedies. ***The Leadership Quarterly***, 18(5): 435–446.

Jiga-Boy, G. M., Clark, A. E., & Semin, G. R. 2013. Situating Construal Level: The Function of Abstractness and Concreteness in Social Contexts. ***Social Cognition***, 31(2): 201–221.

Kensinger, E. A., & Corkin, S. 2003. Effect of Negative Emotional Content on Working Memory and Long-Term Memory. ***Emotion***, 3(4): 378–393.

Kensinger, E. A., & Corkin, S. 2004. Two routes to emotional memory: Distinct neural processes for valence and arousal. ***Proceedings of the National Academy of Sciences***, 101(9): 3310–3315.

LaHuis, D. M., & Ferguson, M. W. 2009. The Accuracy of Significance Tests for Slope Variance Components in Multilevel Random Coefficient Models. ***Organizational Research Methods***, 12(3): 418–435.

Liden, R. C., Wayne, S. J., Zhao, H., & Henderson, D. 2008. Servant leadership: Development of a multidimensional measure and multi-level assessment. ***The Leadership Quarterly***, 19(2): 161–177.

Lonati, S., Quiroga, B. F., Zehnder, C., & Antonakis, J. 2018. On doing relevant and rigorous experiments: Review and recommendations. ***Journal of Operations Management***, 64: 19–40.

Lord, R. G., Epitropaki, O., Foti, R. J., & Hansbrough, T. K. 2020. Implicit Leadership Theories, Implicit Followership Theories, and Dynamic Processing of Leadership Information. ***Annual Review of Organizational Psychology and Organizational Behavior***, 7(1): 49–74.

Lu, P., Liu, J., Koestler, D., & Lu, M. P. 2017. ***Package "pwr2."***

Marschark, M., & Cornoldi, C. 1991. Imagery and verbal memory. ***Imagery and cognition***: 133–182. Springer.

Martell, R. F., & Evans, D. P. 2005. Source-Monitoring Training: Toward Reducing Rater Expectancy Effects in Behavioral Measurement. ***Journal of Applied Psychology***, 90: 956–963.

Morris, W. N. 1989. ***Mood: The frame of mind***. Springer Science & Business Media.

Moscovitch, M., Cabeza, R., Winocur, G., & Nadel, L. 2016. Episodic Memory and Beyond: The Hippocampus and Neocortex in Transformation. ***Annual Review of Psychology***, 67(1): 105–134.

Nederhof, A. J. 1985. Methods of coping with social desirability bias: A review. ***European Journal of Social Psychology***, 15(3): 263–280.

Neider, L. L., & Schriesheim, C. A. 2011. The Authentic Leadership Inventory (ALI): Development and empirical tests. ***The Leadership Quarterly***, 22(6): 1146–1164.

Oyserman, D., & Lee, S. W. 2008. Does culture influence what and how we think? Effects of priming individualism and collectivism. ***Psychological Bulletin***, 134(2): 311.

Paivio, A. 1991. ***Images in mind: The evolution of a theory.*** Harvester Wheatsheaf.

Pinheiro, J., & Bates, D. 2000. ***Mixed-Effects Models in S and S-PLUS***. New York, NY: Springer-Verlag.

Quené, H., & van den Bergh, H. 2004. On multi-level modeling of data from repeated measures designs: A tutorial. ***Speech Communication***, 43(1): 103–121.

Quoidbach, J., Berry, E. V., Hansenne, M., & Mikolajczak, M. 2010. Positive emotion regulation and well-being: Comparing the impact of eight savoring and dampening strategies. ***Personality and Individual Differences***, 49(5): 368–373.

Raudenbush, S. W., & Bryk, A. S. 2002. ***Hierarchical Linear Models: Applications and Data Analysis Methods***. SAGE.

Renoult, L., Irish, M., Moscovitch, M., & Rugg, M. D. 2019. From Knowing to Remembering: The Semantic–Episodic Distinction. ***Trends in Cognitive Sciences***, 23(12): 1041–1057.

Rimmele, U., Davachi, L., Petrov, R., Dougal, S., & Phelps, E. A. 2011. Emotion enhances the subjective feeling of remembering, despite lower accuracy for contextual details. ***Emotion***, 11(3): 553–562.

Schacter, D. L., Addis, D. R., & Buckner, R. L. 2008. Episodic Simulation of Future Events. ***Annals of the New York Academy of Sciences***, 1124(1): 39–60.

Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., et al. 2012. The Future of Memory: Remembering, Imagining, and the Brain. ***Neuron***, 76(4): 677–694.

Schimmack, U., Oishi, S., & Diener, E. 2002. Cultural influences on the relation between pleasant emotions and unpleasant emotions: Asian dialectic philosophies or individualism-collectivism? ***Cognition & Emotion***, 16(6): 705–719.

Schmid, E. A., Pircher Verdorfer, A., & Peus, C. 2019. Shedding Light on Leaders' Self-Interest: Theory and Measurement of Exploitative Leadership. ***Journal of Management***, 45(4): 1401–1433.

Schmidt, A. A. 2008. ***Development and validation of the toxic leadership scale***. University of Maryland, College Park.

Schwarz, N., Oyserman, D., & Peytcheva, E. 2010. Cognition, communication, and culture: Implications for the survey response process. ***Survey Methods in Multinational, Multiregional, and Multicultural Contexts***, 175–190.

Sedikides, C., & Green, J. D. 2009. Memory as a Self-Protective Mechanism. ***Social and Personality Psychology Compass***, 3(6): 1055–1068.

Sekeres, M. J., Winocur, G., & Moscovitch, M. 2018. The hippocampus and related neocortical structures in memory transformation. ***Neuroscience Letters***, 680: 39–53.

Semin, G. R., & Fiedler, K. 1991. The Linguistic Category Model, its Bases, Applications and Range. ***European Review of Social Psychology***, 2(1): 1–30.

Sharot, T., Delgado, M. R., & Phelps, E. A. 2004. How emotion enhances the feeling of remembering. ***Nature Neuroscience***, 7(12): 1376–1380.

Sharot, T., & Yonelinas, A. P. 2008. Differential time-dependent effects of emotion on recollective experience and memory for contextual information. ***Cognition***, 106(1): 538–547.

Shiverdecker, L. K., & LeBreton, J. M. 2019. A primer on multilevel (random coefficient) regression modeling. In S. E. Humphrey & J. M. LeBreton (Eds.), ***The handbook of multilevel theory, measurement, and analysis.***: 389–422. Washington: American Psychological Association.

Shondrick, S. J., Dinh, J. E., & Lord, R. G. 2010. Developments in implicit leadership theory and cognitive science: Applications to improving measurement and understanding alternatives to hierarchical leadership. ***The Leadership Quarterly***, 21(6): 959–978.

Slotnick, S. D., & Dodson, C. S. 2005. Support for a continuous (single-process) model of recognition memory and source memory. ***Memory & Cognition***, 33(1): 151–170.

Speer, M. E., Bhanji, J. P., & Delgado, M. R. 2014. Savoring the Past: Positive Memories Evoke Value Representations in the Striatum. ***Neuron***, 84(4): 847–856.

Squire, L. R., Genzel, L., Wixted, J. T., & Morris, R. G. 2015. Memory Consolidation. ***Cold Spring Harbor Perspectives in Biology***, 7(8): a021766.

Steinley, Douglas. 2006. K-means clustering: A half-century synthesis. ***British Journal of Mathematical and Statistical Psychology***, 59(1): 1–34.

Sthapit, E., Björk, P., Jiménez-Barreto, J., & Stone, M. J. 2021. Spillover effect, positive emotions and savouring processes: Airbnb guests' perspective. ***Anatolia***, 32(1): 33–45.

Tepper, B. J. 2000. Consequences of abusive supervision. ***Academy of Management Journal***, 43(2): 178–190.

Tulving, E. 1972. Episodic and semantic memory. ***Organization of memory*** (E. Tulving&W. Donaldson (Eds.)): 381–403. Oxford, England: Academic Press.

Tulving, E. 1983. ***Elements of Episodic Memory***. Oxford University Press.

Tulving, E. 1985. How many memory systems are there? ***American Psychologist***, 40(4): 385–398.

Tulving, E. 2002. Episodic Memory: From Mind to Brain. ***Annual Review of Psychology***, 53(1): 1–25.

Walker, W. R., Skowronski, J., Gibbons, J., Vogl, R., & Thompson, C. 2003. On the emotions that accompany autobiographical memories: Dysphoria disrupts the fading affect bias. ***Cognition and Emotion***, 17(5): 703–723.

Winocur, G., & Moscovitch, M. 2011. Memory Transformation and Systems Consolidation. ***Journal of the International Neuropsychological Society***, 17(5): 766–780.

Winocur, G., Moscovitch, M., & Bontempi, B. 2010. Memory formation and long-term retention in humans and animals: Convergence towards a transformation account of hippocampal–neocortical interactions. ***Neuropsychologia***, 48(8): 2339–2356.

Yammarino, F. J., Cheong, M., Kim, J., & Tsai, C.-Y. 2020. Is Leadership More Than "I Like My Boss"? In M. Ronald Buckley, A. R. Wheeler, J. E. Baur, & J. R. B. Halbesleben (Eds.), ***Research in Personnel and Human Resources Management***, vol. 38: 1–55. Emerald Publishing Limited.

Yeomans, M. 2021. A concrete example of construct construction in natural language. ***Organizational Behavior and Human Decision Processes***, 162: 81–94.

Yonelinas, A. P. 1994. Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. ***Journal of Experimental Psychology: Learning, Memory, and Cognition***, 20: 1341–1354.

Figure 1

Conceptual Model

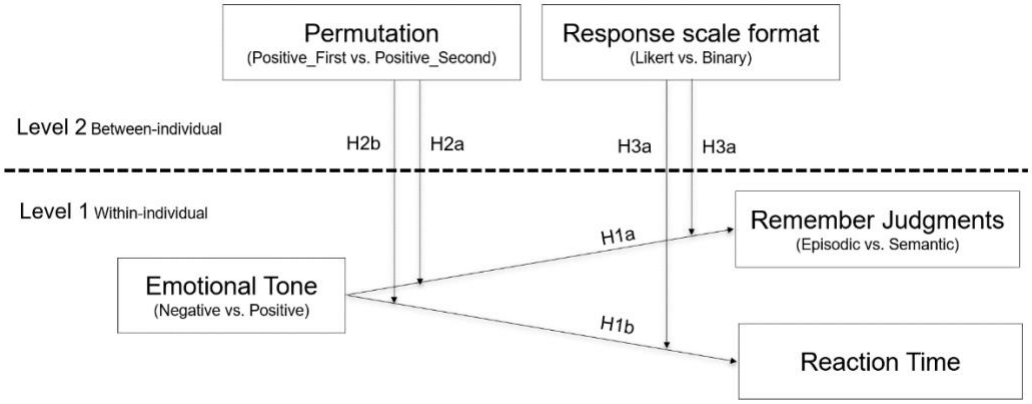


Table 1

Research Design

Response Scale Format (between) / Emotional Tone (within) / Permutation (between)	Binary		Likert	
	Positive	Negative	Positive	Negative
Positive First	Group 1		Group 3	
Positive Second	Group 2		Group 4	

Table 2

Leadership Scales

Leadership Scale	Number of Items	Source
Ethical Leadership	10	Brown et al., 2005
LMX	12	Graen & Uhl-Bien, 1995
Servant Leadership	28	Liden et al., 2008
Abusive Leadership	15	Tepper, 2000
Exploitative Leadership	15	Schmid et al., 2019
Toxic Leadership	30	Schmidt, 2008
Authentic Leadership	16	Neider & Schriesheim, 2011
Total	126	

Table 3

List of 20 Positive and 20 Negative Leadership Items

20 Positive Leadership Items	
Ethical Leadership (Brown, Treviño, & Harrison, 2005)	
Ethical 4	My manager has the best interests of employees in mind.
Ethical 5	My manager makes fair and balanced decisions.
Ethical 6	My manager can be trusted.
LMX (Graen & Uhl-Bien, 1995)	
LMX 3	My manager is the kind of person one would like to have as a friend.
LMX 10	My manager is a lot of fun to work with.
LMX 11	I am willing to apply extra efforts beyond those normally required to meet my manager's work goals.
Servant Leadership (Liden & Maslyn, 1998)	
Servant 1	My manager can tell if something is going wrong.
Servant 4	My manager seems to care more about my success than their own.
Servant 8	My manager is able to effectively think through complex problems.
Servant 12	My manager is always honest.
Servant 14	My manager is always interested in helping people in our community.
Servant 15	My manager has a thorough understanding of our organization and its goals.
Servant 19	My manager would not compromise ethical principles in order to achieve success.
Servant 22	My manager can solve work problems with new or creative ideas.
Servant 26	My manager values honesty more than profits.
Authentic Leadership (Neider & Schriesheim, 2011)	
Authentic 1	My manager solicits feedback for improving his/her dealings with others.
Authentic 4	My manager asks for ideas that challenge his/her core beliefs.
Authentic 6	My manager admits mistakes when they occur.
Authentic 9	My manager shows that he/she understands his/her strengths and weaknesses.
Authentic 16	My manager encourages others to voice opposing points of view.

Table 3 (Cont'd)

List of 20 Positive and 20 Negative Leadership Items

20 Negative Leadership Items	
Toxic Leadership (Schmidt, 2008)	
Toxic 2	My manager denies responsibility got mistakes made in their unit.
Toxic 4	My manager accepts credit for successes that do not belong to him/her.
Toxic 6	My manager ridicules subordinates.
Toxic 8	My manager is not considerate about subordinates' commitments outside of work.
Toxic 10	My manager publicly belittles subordinates.
Toxic 11	My manager reminds subordinates of their past mistakes and failures.
Toxic 12	My manager tells subordinates they are incompetent.
Toxic 13	My manager has explosive outbursts.
Toxic 14	My manager allows their current mood to define the climate of the workplace.
Toxic 15	My manager expresses anger at subordinates for unknown reasons.
Toxic 16	My manager allows their mood to affect their vocal tone and volume.
Toxic 18	My manager causes subordinates to try to read their mood.
Toxic 19	My manager affects the emotions of subordinates when impassioned.
Toxic 23	My manager believes that he/she is an extraordinary person.
Toxic 24	My manager thrives on compliments and personal accolades.
Toxic 27	My manager does not permit subordinates to approach goals in new ways.
Abusive Leadership (Tepper, 2000)	
Abusive 10	My manager expresses anger at me when he/she is mad for another reason.
Exploitative Leadership (Schmid et al., 2017)	
Exploitative 3	My manager values the achievement of his or her own goals over the needs of the employees.
Exploitative 9	My manager gives me boring routine tasks when he or she can benefit from it.
Exploitative 14	My manager manipulates others to reach his or her goals.

Table 4

List of 10 Positive and 10 Negative Leadership Items

10 Positive Leadership Items	
Ethical Leadership (Brown et al., 2005)	
Ethical 4	My manager has the best interests of employees in mind.
Ethical 5	My manager makes fair and balanced decisions.
Ethical 6	My manager can be trusted.
LMX (Graen & Uhl-Bien, 1995)	
LMX 3	My manager is the kind of person one would like to have as a friend.
Servant Leadership (Liden et al., 2008)	
Servant 1	My manager can tell if something is going wrong.
Servant 14	My manager is always interested in helping people in our community.
Authentic Leadership (Neider & Schriesheim, 2011)	
Authentic 1	My manager solicits feedback for improving his/her dealings with others.
Authentic 4	My manager asks for ideas that challenge his/her core beliefs.
Authentic 6	My manager admits mistakes when they occur.
Authentic 9	My manager shows that he/she understands his/her strengths and weaknesses.

10 Negative Leadership Items	
Toxic Leadership (Schmidt, 2008)	
Toxic 2	My manager denies responsibility got mistakes made in their unit.
Toxic 8	My manager is not considerate about subordinates' commitments outside of work.
Toxic 10	My manager publicly belittles subordinates.
Toxic 12	My manager tells subordinates they are incompetent.
Toxic 13	My manager has explosive outbursts.
Toxic 15	My manager expresses anger at subordinates for unknown reasons.
Toxic 27	My manager does not permit subordinates to approach goals in new ways.
Abusive Leadership (Tepper, 2000)	
Abusive 10	My manager expresses anger at me when he/she is mad for another reason.
Exploitative Leadership (Schmid et al., 2019)	
Exploitative 9	My manager gives me boring routine tasks when he or she can benefit from it.
Exploitative 14	My manager manipulates others to reach his or her goals.

Table 5

Measures of Affect, Self-reference, and Concreteness

Measure	Source
Emotional valence	Affect from the LIWC dictionary
Self-reference	i measure (i.e., first-person pronoun) from the LIWC dictionary after trimming[3]
Concreteness	doc2concrete library in R (Yeomans, 2021)

Table 6

Affect, Self-reference, and Concreteness of the Final 20 Leadership Items

Emotional Tone	N	Affect		Self-reference		Concreteness	
		Mean	SD	Mean	SD	Mean	SD
Negative	10	13.9	4	1.6	3.38	2.7	0.19
Positive	10	14.6	4.08	0	0	2.6	0.23
t.test (p-value)		0.692	-	0.168	-	0.305	-

[3] Some leadership items start with "My manager…" whereby "My" is not considered an i measure (e.g., My manager can be trusted). As such, the leadership items are trimmed first to remove such first-person pronouns that are not appropriate to be considered as an i measure.

Table 7

Means, standard deviations, and correlations with confidence intervals

Variable	*M*	*SD*	1	2	3	4
Emotional Tone	0.50	0.50				
Permutation	0.51	0.50	.00 [-.10, .10]			
Response Scale Format	0.52	0.50	.00 [-.10, .10]	-.01 [-.11, .10]		
Percentage of Remember Judgments	0.44	0.28	.06 [-.05, .16]	-.18** [-.28, -.08]	-.01 [-.11, .10]	
Reaction Time	2.58	2.00	-.03 [-.14, .07]	.05 [-.05, .15]	-.00 [-.10, .10]	.04 [-.06, .14]

Note. M and *SD* are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). * indicates $p < .05$. ** indicates $p < .01$.

Table 8

Results for the Mixed-Effects Model

	Percentage of Remember Judgments		Reaction Time	
Predictors	*Estimates*	*std. Error*	*Estimates*	*std. Error*
Fixed Effects				
Intercept (γ_{00})	0.51***	0.04	1.54***	0.25
Emotional Tone (γ_{10})	-0.05	0.04	1.87***	0.24
Permutation (γ_{01})	-0.15***	0.04	2.03***	0.28
Response Scale Format (γ_{02})	-0.03	0.04	0.13	0.28
Emotional Tone x Permutation (γ_{11})	0.11*	0.05	-3.66***	0.27
Emotional Tone x Response Scale Format (γ_{12})	0.04	0.05	-0.27	0.27
Random Effects				
Level-1 residual variance (σ^2)	0.01		0.51	
Level-2 intercept variance (τ_{00})	0.07		2.95	
Variance in slopes (τ_{11})	0.07		2.19	
Observations	362		362	

Note. Level-1 N = 2; Level-2 N = 181. All entries are unstandardized regression coefficients.
*p < .05, ** p < .01, *** p < .001

Figure 2a

Moderating Effect of Permutation on Emotional Tone – Percentage of Remember Judgments

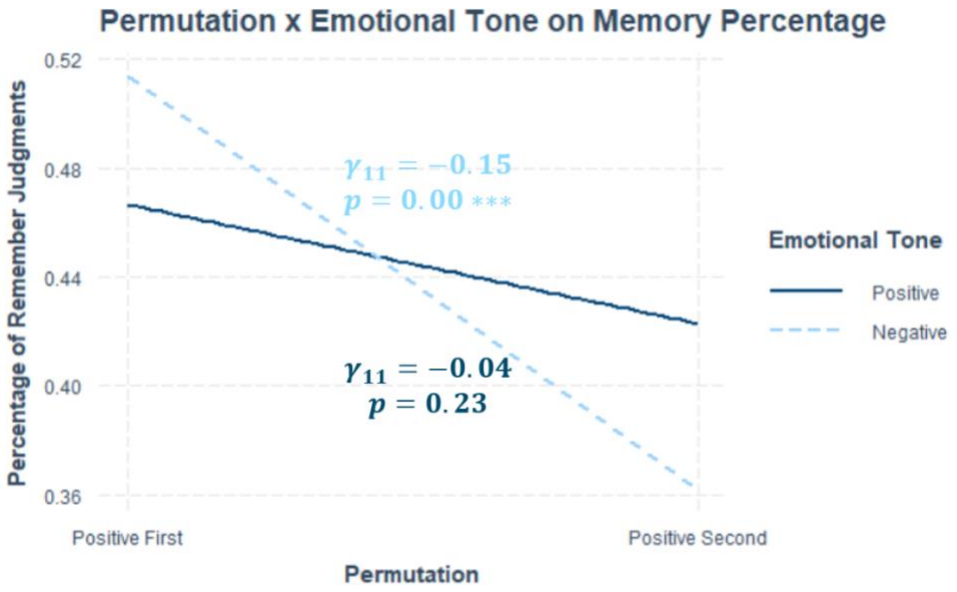
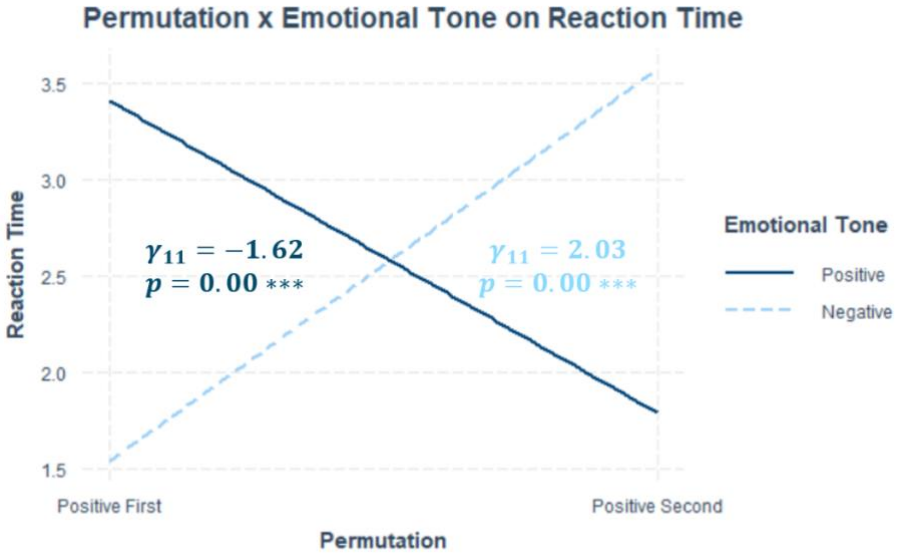


Figure 2b

Moderating Effect of Permutation on Emotional Tone – Reaction Time



Appendix 1

Power Analysis

Power analysis is conducted using the pwr2 package (Lu et al., 2017) in R, specifically calculating the sample size using ss.2way for balanced two-way ANOVA. Alpha is set at 0.05, beta at 0.2. Regarding the effect size for each dimension of order and response scale format, no approximate effect size has been observed yet due to a lack of relevant research with a similar factorial design. However, effect sizes for research investigating the impact of emotions on memory revolve around 0.3 (e.g., Gaddy & Ingram, 2014). Taking a conservative approach, I use 0.2 as the effect size for both of the two-way interactions. Below is the R code, which shows that 50 samples are required for each cell, with 200 samples in total.

```
library(pwr2)
ss.2way(a = 2, b = 2, alpha = 0.05, beta = 0.2, f.A = 0.2, f.B = 0.2, B = 100)
Output:
Balanced two-way analysis of variance sample size adjustment
a = 2
b = 2
sig.level = 0.05
power = 0.8
n = 50
```

NOTE: n is the sample size for each group, total participants required amounts to 200.