

Is The Self Reference Effect a Retrieval Practice Phenomenon?

Shifa Maqsood ¹

¹Department of Psychology, Brooklyn College of CUNY

Author Note

The materials data and analysis scripts are available from the project repository. This research paper is a two-semester undergraduate honor thesis advised by Dr. Matthew J. C. Crump, and is submitted to the Department of Psychology in partial fulfillment of the requirements for graduating with Departmental Honors in Psychology.

Correspondence concerning this article should be addressed to Shifa Maqsood, Department of Psychology, Brooklyn College of CUNY, 2900 Bedford Avenue, Brooklyn, NY 11210

Abstract

In levels of processing framework, Craik and Tulving (1975)'s experiments results demonstrated that subjects recalled more of the words which were encoded under the semantic tasks than other encoding conditions as semantic tasks require elaborate analysis of stimulus than judging the size of it. In 1977, Rogers et al. conducted an experiment in which self-reference effect was found to have led to superior recall. Subsequent research on underlying mechanisms of SRE have led to inconclusion. The present study takes a different perspective at why SRE promotes learning and proposes that the SRE effect emerges as a retrieval practice phenomenon as self is frequently engaged to recall information. We hypothesized that if participants practice recall of the words during study, there will be no difference between the self and other conditions. Our second hypothesis stated that retrieval practice will only enhance the existing differences in each condition. The results revealed that practicing recall does not help produce more words at the end and there was no self reference effect observed in no-recall condition. We infer that if there had been more delay in time, retrieval practice would have led to superior retention of the stimuli. It is also possible that our sample size was too small and that participants did not follow instructions. Future research will focus on conducting the experiment on Brooklyn College students.

Keywords: Self Reference, Retrieval Practice

Is The Self Reference Effect a Retrieval Practice Phenomenon?

Memory performance can be enhanced by mnemonic devices that help learners retain information for later recall. A few effective devices include the method of loci, acronyms, and alphabetic cuing. The method of loci involves associating an arbitrary list of to-be-remembered items with the spatial layout of a familiar place. For example, in the case of memorizing a grocery list, one could imagine walking through rooms in their house and mentally place each item on the grocery list in different rooms. When later at the grocery store, the list of items can be recalled by mentally simulating walking through the rooms which provide a retrieval cue for each grocery item. Acronyms involve arranging the first letters of a list of to-be-remembered words to form a new memorable word. For example, each letter in “ROY G. BIV” refers to the colors on the visible spectrum. Similarly, alphabetic cueing is a device that may help people recover from momentary memory lapses, like tip-of-the tongue, where a person knows a word or name, but is unable to recall it. Alphabetic cuing involves systematically considering whether the word begins with “a”, or “b”, and so on. The process of considering each letter one at a time strategically varies the retrieval cue, and can lead to successful recall (Nelson & Archer, 1972). This thesis examines theoretical accounts of a mnemonic device called the self-reference effect that has been forwarded as a particularly powerful memory aid (Rogers et al., 1977).

The self reference effect

All of the above mnemonic devices provide support for a very general memory principle: making information meaningful makes it more memorable (Bransford & Johnson, 1972). The self-reference effect provides an example of this principle that may directly tap a deep source of meaning for individuals. Simply put, prior research on the self-reference effect suggests that the act of relating an item to oneself makes the item more memorable. This finding is consistent with theories suggesting that the self is a complex and dynamic process that gives meaning to new experiences by relating them to past experiences (Markus, 1977). So, the act of considering something in relation to the self

may therefore imbue the object with multiple layers of meaning leading to an enhancement in memorability.

Prior research has found robust evidence for self-reference effects in laboratory procedures (for a meta-analysis see, Symons & Johnson, 1997). Most prior research assumes that self-referential processing strengthens a memory trace while it is encoded, which is hypothesized to increase memorability. The purpose of this thesis is to propose and examine an alternative explanation for the self-reference effect that explains memorability benefits in terms of memory retrieval processes.

Evidence for the self-reference effect

Rogers et al. (1977) laboratory procedure originally showed evidence of a self-reference effect in recall for adjectives. Their experiments were intended to further extend Craik and Lockhart's (1972) levels of processing principle to determine whether relating information to oneself could promote deeper processing and enhance memorability. Previous results supporting the levels of processing principle had shown that memory for words depended on "depth" of processing during encoding. For example, Craik & Tulving (1975) showed participants words for a later memory test, but systematically varied how the words were encoded using a word judgment task. Some judgments were about physical (letter size) or acoustic (rhyming) properties of the words, and others were about category membership or whether the word would fit meaningfully in a sentence. Memory performance improved for words that were given deeper levels of encoding (e.g., category and sentence judgments) compared to more shallow levels of encoding (e.g., physical and acoustic judgments). Broadly speaking, Craik and Tulving demonstrated that encoding words in a more meaningful way improved memory compared to words encoded in a more superficial way.

Rogers et al. (1977) extended Craik and Tulving's design to test whether deeper levels of processing could be achieved using a self-reference judgment task. The laboratory experiment included 40 personality trait adjectives presented on a computer screen one at

a time along with a cue, in the form of a question, asking participants to make a judgment about the adjective. There were four judgment tasks: structural, phonemic, semantic, and self-reference. In the structural task, participants judged whether the physical size of the letters was the same size as the cue question. In the phonemic task, participants judged whether the adjective rhymed with another word in the cue question. In the semantic task, participants judged whether the adjective had the same meaning as another word in the cue question. Last, in the self-reference task, participants judged whether the adjective described themselves or not. Each cue question was followed by a 500-ms interval and then the target adjective was presented. Participants were instructed to respond “yes” or “no” by pressing a button. After the encoding phase, participants were asked to recall as many words as possible.

Rogers et al. (1977) found evidence consistent with the levels of processing principle. The number of words recalled was the least for the structural condition, slightly more in the rhyming condition, more yet in the semantic condition, and the most in the self-reference condition. Furthermore, there was a significant difference between the self-reference and semantic conditions, suggesting that the self-reference manipulation caused a deeper level of more meaningful processing than the already effective semantic condition. Throughout this paper, the self-reference effect will refer specifically to evidence showing memorability enhancements for self-referential conditions compared to other semantic conditions.

Participants gave the same ‘yes’ or ‘no’ response in all judgment tasks, and another finding was that recall memory was better for ‘yes’ than ‘no’ response words. This finding suggests that the self-reference effect is enhanced when subjects agreed that the adjective did describe themselves. The basic pattern of results was also replicated in a second experiment using a modifying design involving a rating scale rather than a yes/no judgment.

Replicating the self-reference effect

According to a meta-analysis (Symons & Johnson, 1997), the self-reference effect is a robust phenomena that has been replicated across numerous labs. The analysis included 129 studies that used similar designs to Rogers et al. (1977). Some of the characteristics of the included studies were that they used 39 participants on average, undergraduates being the majority of the participant population. The studies used within-subject designs 75% of the time and between-subjects designs 25% of the times. These studies also primarily used free recall in their designs and trait adjectives. Collapsing across studies, the meta-analysis found that the self-reference effect had an average effect size of .45.

More recently, Bentley et al. (2017) replicated the self-reference effect as an online experiment through Amazon's Mechanical Turk. They expected the online version would allow the researchers to study larger sample sizes and not just be limited to undergrad populations. There were four versions of the standard self-reference experiment. Study 1 was a direct replication of Rogers et al. (1977) with the exception that participants could proceed to the next question on their own and not wait for five seconds for each question. There was also a filler task at the end of the encoding phase before the recall. Conservative analysis of the results showed that there was not a significant difference in mean recall of words between SR and semantic tasks. It was inferred that participants did not spend more time during encoding. Study 2 tested this inference and instead each question was set to be present for five seconds. The longer duration of stimulus resulted in participants engaging deeply with stimulus as the conservative analyses showed were more self-referent words recalled than the words in semantic tasks. Study 3 tested the effect of recognition and free recall on memory recall. The results showed a diminished self reference effect for recognition and no significant difference between semantic and self reference. Study 4 tested if the recall of the trait adjectives across the encoding task is affected by prior knowledge of the recall. There was no significant difference between incidental and informed recall conditions.

Extensions and applications of the self-reference effect

Self referential processing may be a useful mnemonic device in educational settings to promote meaningful learning. For example, children in primary school were able to retain a list of words better when they used themselves as subjects in relation to the to-be-remembered words in sentences. Spelling accuracy was also judged to be higher in self-reference conditions than other-reference conditions (Turk et al., 2015).

People may engage in self-referential processing even without an explicit instruction to link a current stimulus with the self. Turk et al. (2008) included an incidental condition in which the participants were only instructed to respond if the stimuli was placed above or below the cue. The referent cue was either the image or name of the participant or an image or name of a well known celebrity. The findings showed participants remembered adjectives associated with themselves more than the other in the incidental condition.

In another replication of the self-reference paradigm, Bower and Gilligan (1979) hypothesized that presenting adjective noun phrases such as “broken bone” or “a boring lecture”, would enhance recall equally for trait adjectives as the adjective phrases would arouse a personal experience in the participant’s memory. The results showed that there was no difference in recall and recognition in the self-reference condition between adjective trait-adjectives and adjective phrase tasks. However, other experiments have deviated from the usage of adjectives only and they have instead tried to demonstrate the use of self-reference in everyday life. In one study, for example, participants were instructed to retain as much as information about profiles of unknown people. They were then instructed to recall as much as information they could remember about the profiles. Results showed that the subjects correctly remembered birthdays of the unknown people that were in the same month as their own birthday (Kesebir & Oishi, 2010).

Neuro-imaging techniques such as fMRI have also been used to investigate and locate the neural correlates of the self. The medial prefrontal cortex along with other brain regions including anterior cingulate were found to be activated during self-reference trials

(Macrae et al., 2004). The activity of medial prefrontal cortex was also more active during encoding and for the remembered words which implicates a role in memory formation. A meta-analysis of self-referential processing in the brain reveals that cortical midline structures which also include medial prefrontal cortex were found to be active during self related tasks during a number of domains such as verbal, spatial and emotional domains (Northoff et al., 2006).

The literature review of the self-reference effect has shown that relating information to oneself can effectively improve memory recall in a variety of settings. In fact, the meta-analysis across the self-reference effect literature has also shown that despite the increase in memory load (the amount of information to be remembered), self reference proved to be comparatively more efficient than semantic encoding in some study designs (Symons & Johnson, 1997). As yet, the self-reference effect has been found in a variety of manipulations of the standard self-reference paradigm; however, the underlying mechanisms which facilitate this cognitive process have not been well understood.

Theoretical explanations of the self-reference effect

What makes relating information to oneself more memorable? This section describes major theoretical explanations of the self-reference effect. First, theories that focus on encoding processes are reviewed, followed by theories that focus on retrieval processes. The section concludes with a novel retrieval focused theory that sets the stage for the experiments presented in the thesis.

Encoding focused theories of the SRE

As previously mentioned, the Rogers et al. (1977) study was an attempt to replicate and extend empirical work (Craik & Tulving, 1975) providing evidence for the levels of processing principle (Craik & Lockhart, 1972). The levels of processing principle is the very general idea that memorability is influenced by depth of processing during encoding. Specifically, deeper processing during encoding yields better memory performance. For example, in the Craik and Tulving (1975) experiments, participants were presumed to

encode words in the semantic conditions more deeply than words in the physical attribute or rhyming conditions. Similarly, in the self-reference effect, relating information to oneself may cause an even deeper level of processing than the semantics condition. On this view, the self-reference effect occurs because relating information to oneself strengthens memory traces through a deep level of meaningful processing. The strengthened trace is assumed to be less vulnerable to decay and thus more available for later recall.

Organizational encoding processes have also been forwarded as an explanation of the self-reference effect. Klein and Kihlstrom (1986) posited that self reference promotes categorical organization as the participants organize the words into two categories during the encoding phase: words that describe them and words that do not describe them. The act of categorizing words during encoding was assumed to provide additional cues that would facilitate later recall. They provided support for this idea by showing no difference in recall when semantics and self-reference were both equated to promote active categorization of list words.

Nevertheless, it can not be argued that organization is solely responsible for self-referential processing since non-depressed subjects remembered more non-depressed content than mildly depressed subjects who remembered both non-depressed and depressed content (Kuiper & Derry, 1982). These findings support a role for elaboration processes in the self-reference effect. For example, mildly depressed subjects remembered more depressed adjectives than non-depressed subjects. This finding is consistent with the idea that depressed adjectives were recalled more because they were compatible with the self-schema of the depressed individuals leaving an elaborate trace for recall. Therefore, it is possible that both, elaborative and organizational processes contribute to great recall of self-referential information.

Klein and Loftus (1988) found that self-reference tasks performed similarly to both organizational and elaborative processes, however they were unable to infer to which degree either process contributes to the SRE. In elaborative processing, the word is encoded in

reference to “item-specific” information present in the memory which leads to more than one retrieval route established for each stimulus word during encoding for later recall.

Whereas in organization processing, the stimulus list words are organized in category labels in a relation to each other which are later recalled in a cluster.

There have been other attempts to understand the basis for SRE, one of which was proposed by Ganellen and Carver (1985) who argues for several other mediators being responsible for the incidental encoding induced by the self-reference effect. Ganellen and Carver (1985) hypothesized that self-referent encoding is dependent on emotions the stimulus produces during encoding, the importance of the stimulus to the subject, and its distinctiveness. However their results yielded no relation of these predictor variables to the self-reference effect in their experiment which led to the possible implication that cognitive processes underlying self-reference are distinct from judgmental or emotional processes.

Also diverging from the depth of processing paradigm, Bellezza and Hoyt (1992) proposed self-schema as an organization of internal cues which then associate personal experiences with new information and these cues are then retrieved during recall. However, Bellezza (1984) showed that participants only recalled about half of the trait words they had associated with their personal experiences. To conclude, the research on the processes behind the self-reference seems to be inconclusive despite it being a robust phenomenon which has been replicated in many studies.

Retrieval focused theories of the SRE

There has been more emphasis on encoding than retrieval processes in the self-reference effect. Retrieval refers to processes that bring information from the past to mind in the current context. In general, retrieval of memory traces is facilitated when cues in the current context match well with cues from the encoding context, and retrieval is impaired when cues between the current context and a previous one mismatch. This matching rule is supported by context-dependent memory phenomena (Godden & Baddeley, 1975), the encoding specificity principle (Tulving & Thomson, 1973), and the

transfer-appropriate processing principle (Morris et al., 1977).

This thesis proposes that retrieval processes may play an important role in the self-reference effect. For example, according to the matching rule, participants will have better recall when the retrieval context matches well with the encoding context. In general, studies of the self-reference effect do not manipulate retrieval conditions. Typical studies involve a free-recall phase where participants generate as many words from the study list as possible. During this phase participants are free to rely on any cues, including self-generated ones, to prompt recall. If participants were naturally biased toward thinking about themselves during the recall phase, this bias would match well with the processing associated with words presented in the self-reference condition, and could provide additional support for retrieving words in the self-reference condition compared to other conditions.

Retrieval processes may also be involved during the encoding phase. For example, retrieval practice is known to improve memorability more than multiple study attempts. For example, Roediger III and Karpicke (2006) had participants study paragraphs about undergraduate level concepts. Some paragraphs were allowed to be studied twice, and other paragraphs were only studied once followed by a quiz about the concepts. They found that concepts that were quizzed once were better remembered over the long term than concepts that were studied twice. A general interpretation of the retrieval practice effect is that remembering a piece of information is like a skill that requires specific cognitive actions. Quizzing prompts people to engage in the specific actions necessary to retrieve information, and this form of practice develops the skill necessary to retrieve the information at a later date.

A related explanation is that as the learner tries to bring the information to their mind during each test, the diagnosticity of the retrieval cues improves by the strengthening of the match between the cue and its target (Karpicke, 2012). Proponents of elaborative encoding have argued that elaboration of the stimuli during retrieval produces mediators

and additional retrieval cues which facilitate the recall during testing. However, Karpicke and Smith (2012) observed that retrieval practice is a separate mechanism which can not be attributed to elaborative encoding. In their experiment 4, there were identical twin pairs presented during learning to inhibit mediators and the repeated retrieval condition score was reported to be 99.6%. Furthermore, retrieval practice also produced higher recall than elaborative studying even when subjects were instructed to draw concept maps, which are largely associated with elaboration, at the final test. (Karpicke & Blunt, 2011).

This thesis proposes and tests the possibility that the retrieval practice may be contributing to the self-reference effect. Specifically, it is possible that the self-reference encoding condition is confounded with retrieval practice. For example, consider the subjective experience involved in a typical self-reference judgment asking participants whether an adjective, such as “nice”, describes themselves. Participants may begin an internal narrative asking themselves, “Am I nice?”, “Maybe, I’m nice, but I’m not that nice”, etc. This narrative involves repeatedly stating the to-be-remembered word, which provides multiple retrieval practice attempts. Participants may not engage in such a complex inner monologue for the other conditions, especially when the judgment can be made more directly. On this view, the self-reference effect could reflect an advantage for words that were given retrieval practice in the self-reference condition compared to words that were not practiced in the other conditions. The purpose of the experiments reported in this thesis is to investigate this hypothesis directly by explicitly manipulating retrieval practice across all conditions.

Experiment 1a: Pilot

The first experiment was designed to test the hypothesis that the self-reference effect may reflect a confound due to retrieval practice. The design closely resembled the first experiment from Rogers et al. (1977). Participants were presented with trait adjectives during an encoding phase followed by free recall. During the encoding phase each word was paired with one of three judgment tasks: letter case, semantic, and

self-reference. The novel manipulation was to explicitly manipulate retrieval practice during the encoding phase for half of the words. For example, half of the encoding trials were followed by a secondary $n-1$ recall task, where participants were instructed to recall the previous word they had judged on the immediately preceding trial.

Predictions for performance based on the major hypothesis are shown in Figure @ref(fig:predictions).

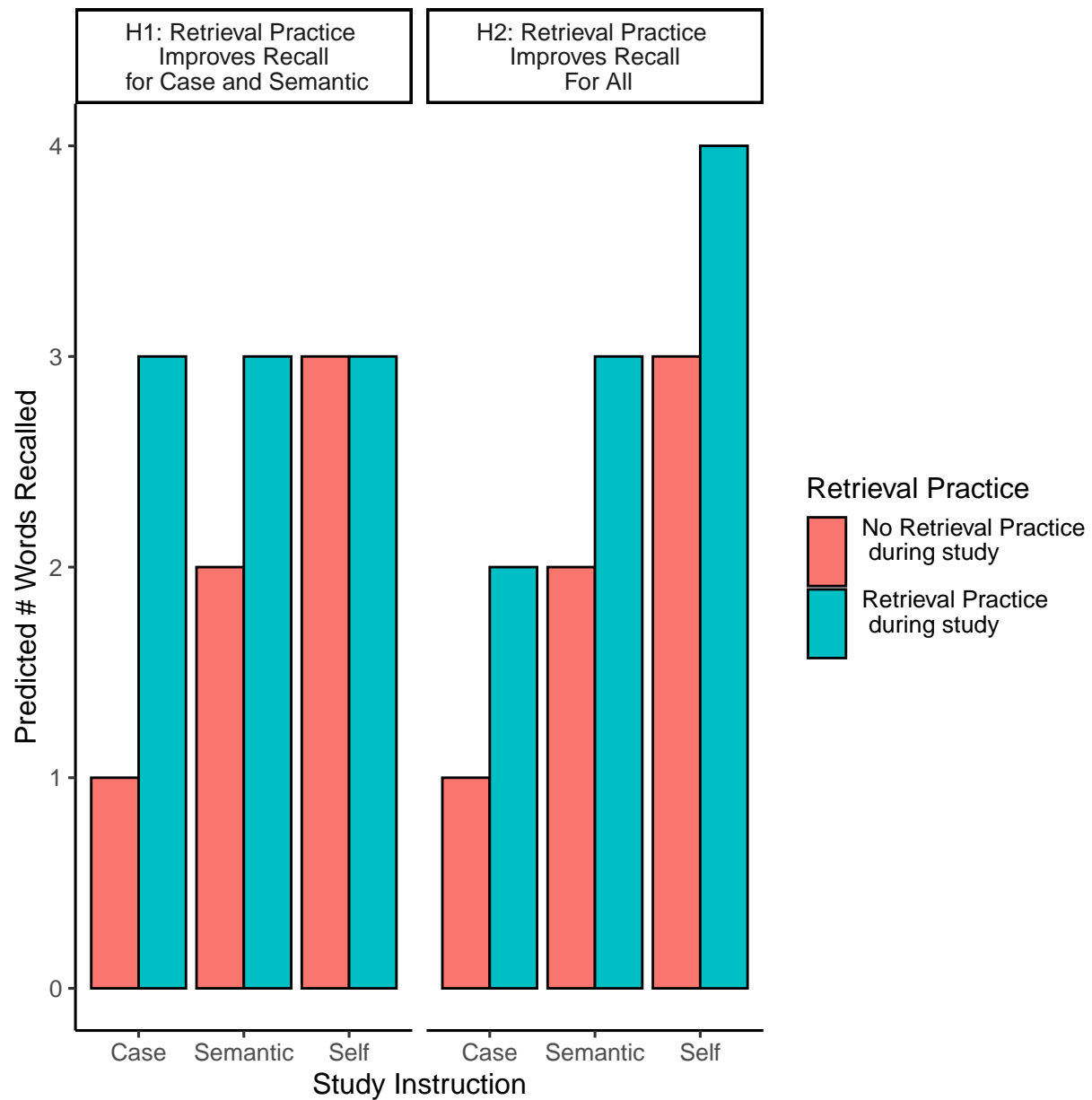
Predictions from the primary hypothesis are shown in the H1 panel. The expectation for the no retrieval practice condition was to replicate the standard self-reference effect. Specifically, recall rates should increase with depth of encoding, such that recall would be highest in the self-reference condition, and lowest in the case-judgment condition ($\text{case} < \text{semantic} < \text{self}$).

Two possible outcomes were considered for the retrieval practice condition. First, it is possible that participants already engage in retrieval practice (by mentally repeating the word) for words in the self-reference condition; and, that they do not engage in retrieval practice for the case or semantic judgments (which can be answered immediately without mentally repeating the word). In this case, retrieval practice may improve recall for words in the case and semantic conditions compared to the self-reference condition which may already be benefiting from retrieval practice. This could result in equivalent recall rates across all conditions ($\text{case} = \text{semantic} = \text{self}$). In other words, if the self-reference condition has superior recall rates because of retrieval practice, then we expected that delivering retrieval practice for items in the case and semantic conditions would improve recall to the level of the self-reference condition in those conditions.

Another possibility is that retrieval practice provides benefits to recall for all words in all conditions. This possibility is shown in the H2 panel of Figure @ref(fig:predictions). Here, we there was a main effect of retrieval practice, and no interaction with the encoding question factor.

Experiment 1a is considered a pilot experiment. To foreshadow the results, the

Figure 1



participants recruited from Amazon's Mechanical Turk did not provide analyzable data. The design and participant recruitment is improved upon in the remaining experiments.

Methods

Participants.

Participants were recruited from Amazon's Mechanical Turk, but due to several issues it was unclear how many legitimate human participants were recruited. The experiment was approved by the Brooklyn College Institutional Review Board.

Materials.

A list of person-descriptive words was obtained from (Chandler, 2018). The list of words was filtered to exclude any words containing punctuation, and words considered derogatory and otherwise objectionable. The total stimulus set had 375 words. These words were categorized by Chandler (2018) in terms of likeableness. There were 173 high likeable words, and 202 low likeable words.

Design.

The experimental design was a 3x2x2 mixed factorial. The encoding judgment task was a within-subject factor with three levels: structural, semantic, and self-reference. Retrieval practice was also a within-subject factor with two levels: retrieval practice vs. no retrieval practice. Last the encoding judgment tasks were presented to participants in either a blocked or mixed design. The dependent variable was the number of words recalled.

12 unique words (6 high and 6 low likeable) were randomly chosen from the stimulus set for each of the three conditions (36 words in total). Four additional filler items were inserted at the beginning and end of the encoding phase. The intention was to randomly assign half of the words for each judgment task to the retrieval practice condition, and the other half to the no-retrieval practice condition. However, a mistake in the experiment script caused .33 of the words to be assigned to the retrieval practice condition and .67 words to the no-retrieval practice condition. This error was corrected in the following experiments.

Apparatus.

The experiment was programmed in JavaScript and HTML using JsPsych (de Leeuw, 2015). Just Another Tool for Online Studies (JATOS) was used to host this experiment online (Lange et al., 2015). The source code for the experiments is available from the project website.

Procedure.

Participants were recruited from Amazon's Mechanical Turk. They completed a consent form followed by a demographics questionnaire and task instructions. The task instructions stated, "In this phase you will make judgments about words that describe personality characteristics or traits. Sometimes you will be asked to remember words that were just shown to you. There are 44 words to judge. After this phase you will be given a memory test for all of the words you saw. Press any key to begin."

Following the instructions, participants completed the encoding phase. On each trial a cue question indicating the word judgment task was presented on screen for 3000 ms. There were three possible cue questions: "Is the following word written in upper case?", "Does the following word have a positive meaning?", and "Would you use the following word to describe yourself?". Next, the target word appeared on screen underneath the cue question, along with two buttons to respond "yes" or "no", all of which remained on screen until a response was made. There was an intertrial interval of 1000 ms. For words assigned to the retrieval practice condition a recall prompt appeared prior to the next trial. The recall prompt stated, "Recall the last word, type the last word that you judged." Participants were shown a box with a cursor and proceeded to type their response. They pressed a button to continue to the next trial. There were a total of 44 encoding trials.

In phase 2 of the experiment, participants were instructed to recall as many of the words as they could. Participants were shown an instruction screen stating, "Phase II instructions. In this phase you will be given a recall memory test. You made judgments about 44 words. Your task is now to recall as many words as you can by typing out each

word. Press any key to begin.” Next participants were shown a screen with instructions and a text entry box. The instructions read, “Recall each word. You were shown 44 words, try to recall as many words as possible. Type each word separated by a space. Your responses will be shown in the box below.”

Participants pressed a button to continue when they could not recall additional words. At this time, a second free recall phase began and participants were asked to try one more time to recall any additional words. The experiment ended after the final recall phase.

Data-Exclusion.

Our experiment was conducted online and it was completed by participants on their web browsers. Although there were specific instructions at the beginning of the experiment, there was a possibility that participants completed and submitted the experiment without following the instructions. Thus, we set criteria for excluding those participants whose responses indicated that they did not follow the instructions to complete the tasks. During the study phase, part one of the experiment, participants were instructed to click on either, “Yes” or “No” button. There were three conditions: structural, semantics and self. In the structural condition, the questions asked if the word displayed on screen was in uppercase or lowercase. We judged the response to the questions in this condition to be the baseline for our criteria as this condition served as the participant’s ability to attend to instructions. We excluded participants who scored lower than 70% on structural judgement questions. A total of 16 participants were excluded. We also checked for response bias to make sure that the participants were not randomly choosing the “Yes” or “No” option as their responses to the question.

Results and Discussion

The data analysis suggested that many of the participants may not have been completing the study as intended. First, a large proportion of participants were scored at or near chance on the letter size judgment task. Second, recall rates were very low overall, and several participants typed nonsense responses. Third, the response patterns suggested that

some number of participants were completing the experiment multiple times under different user names. The data collected for the pilot study also revealed an error in the program which did not assign an equal number of trials to the retrieval practice and no-retrieval practice conditions. For these reasons, the data for the pilot study are not reported.

There are several known issues with recruiting participants from Amazon's Mechanical Turk (Rodd, 2023). The pilot experiment recruited participants with minimal filtering. The next experiment addressed these issues with recruitment.

Experiment 1b and 1c

Experiment 1b and 1c were the same as the pilot experiment with the following exceptions. Experiment 1b recruited participants from Amazon's Mechanical Turk using different filters than the pilot experiment. Experiment 1c recruited participants from the Brooklyn College undergraduate community. Additionally, for both experiments the experiment script was corrected to ensure an equal number of retrieval practice and no-retrieval practice trials within each encoding condition.

In addition, experiment 1b and 1c included further instructions that participants should not use memory aids such as writing down words during the experiment. And, the experiment concluded with a series of debriefing questions.

Methods

Participants.

For experiment 1b, A total of 53 participants were recruited from Amazon's Mechanical Turk. Mean age was 46.1 (range = 29 to 69). There were 30 females, and 23 males. There were 49 right-handed participants, and no left or both handed participants. 28 participants reported normal vision, and 25 participants reported corrected-to-normal vision. 52 participants reported English as a first language, and 1 participant reported English as a second language.

For experiment 1c, A total of 65 participants were recruited from Amazon's Mechanical Turk. Mean age was 21.6 (range = 18 to 49). There were 48 females, and 16

males. There were 56 right-handed participants, and 8 left or both handed participants. 30 participants reported normal vision, and 28 participants reported corrected-to-normal vision. 42 participants reported English as a first language, and 22 participants reported English as a second language.

Materials.

The same materials as described in the pilot experiment were used in experiment 1b and 1c.

Design.

The same 3x2x2 mixed factorial design as described in the pilot experiment was used in experiments 1b and 1c.

Apparatus.

The apparatus was the same as the pilot experiment.

Procedure.

The procedure was the same as before with the following differences.

The instructions given during the encoding phase further specified that, “Important. There is a memory test after this phase. Please do not use external aids to help your performance. For example, please do not write down the words or take screenshots. The experiment is aimed at measuring unaided memory processes. Press any key to begin.”

At the end of the experiment participants were presented with the following debriefing statement and questions: “Thanks for participating, we are planning on running similar experiments in the future, and we are interested in your feedback. Did you use any external memory aids like writing down the words or taking screenshots? Thanks for letting us know. Were you able to give this task your full attention, or did you get interrupted? Thanks for letting us know. Did you run into issues with the task, or other issues that might have impacted your performance?”

Data-Exclusion.

We used the same exclusion procedure as described previously. No participants were removed from experiment 1b and 1 participant was removed from experiment 1c.

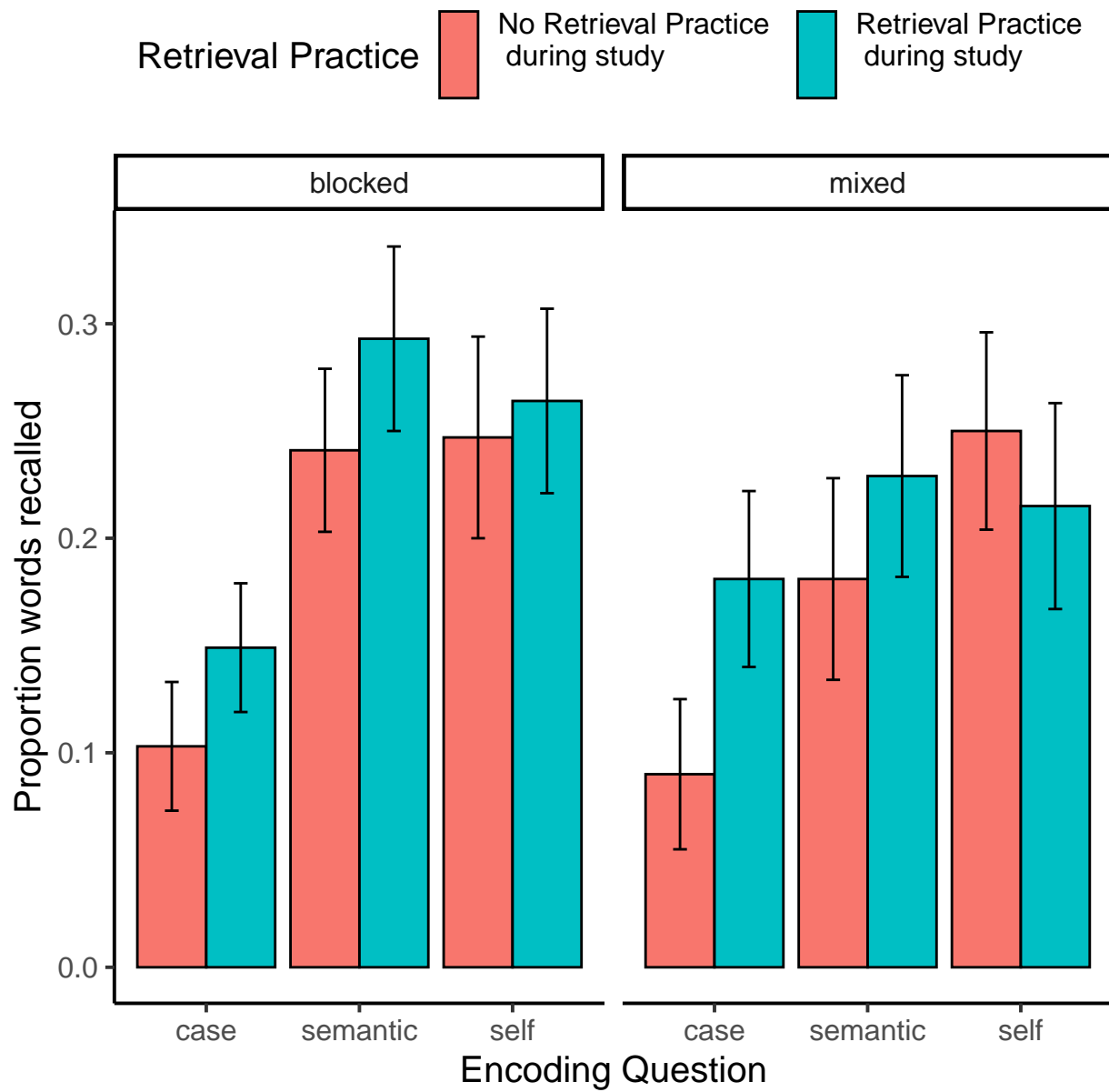
We used R (Version 4.3.1; R Core Team, 2022) and the R-packages *conflicted* (Version 1.2.0; **R-conflicted?**), *data.table* (Version 1.14.8; Dowle & Srinivasan, 2022), *dplyr* (Version 1.1.3; Wickham, François, et al., 2023), *flextable* (Version 0.9.3; **R-flextable?**), *forcats* (Version 1.0.0; Wickham, 2023), *ftExtra* (Version 0.6.1; **R-ftExtra?**), *ggplot2* (Version 3.4.2; Wickham, 2016), *jsonlite* (Version 1.8.7; Ooms, 2014), *knitr* (Version 1.44; **R-knitr?**), *lubridate* (Version 1.9.2; Grolemund & Wickham, 2011), *officer* (Version 0.6.2; **R-officer?**), *papaja* (Version 0.1.1; Aust & Barth, 2022), *purrr* (Version 1.0.2; Wickham & Henry, 2023), *readr* (Version 2.1.4; Wickham, Hester, et al., 2023), *stringr* (Version 1.5.0; Wickham, 2022), *tibble* (Version 3.2.1; Müller & Wickham, 2022), *tidyr* (Version 1.3.0; Wickham, Vaughan, et al., 2023), *tidyverse* (Version 2.0.0; Wickham et al., 2019), *tinylabls* (Version 0.2.3; Barth, 2022), and *xtable* (Dahl et al., 2019; Version 1.8.4; **R-flextable?**) for all our analyses. An alpha criterion of .05 was adopted for all statistical tests.

Results***Experiment 1b: Amazon Mechanical Turk Master's***

We computed proportion of words correctly recalled for each participant in each condition of the design. The proportions were submitted to a 3 (Encoding Question: Case, Semantic, Self) x 2 (Retrieval Practice: Yes, No) x 2 (Question Order: Blocked, Mixed) mixed factorial design with Encoding Question, and Retrieval Practice as within-subject factors, and Question Order as the sole between-subject factor. Mean proportions of correctly recalled words in each condition are shown in Figure @ref(fig:E1Bresults).

There was a main effect of encoding question, $F(2, 102) = 12.99$, $MSE = 0.03$, $p < .001$, $\eta_G^2 = .059$. Mean proportion recall was lowest in the case condition ($M = 0.131$, $SEM = 0.017$), and at similar higher levels for the semantic ($M = 0.239$, $SEM = 0.022$), and self condition ($M = 0.245$, $SEM = 0.023$).

Figure 2



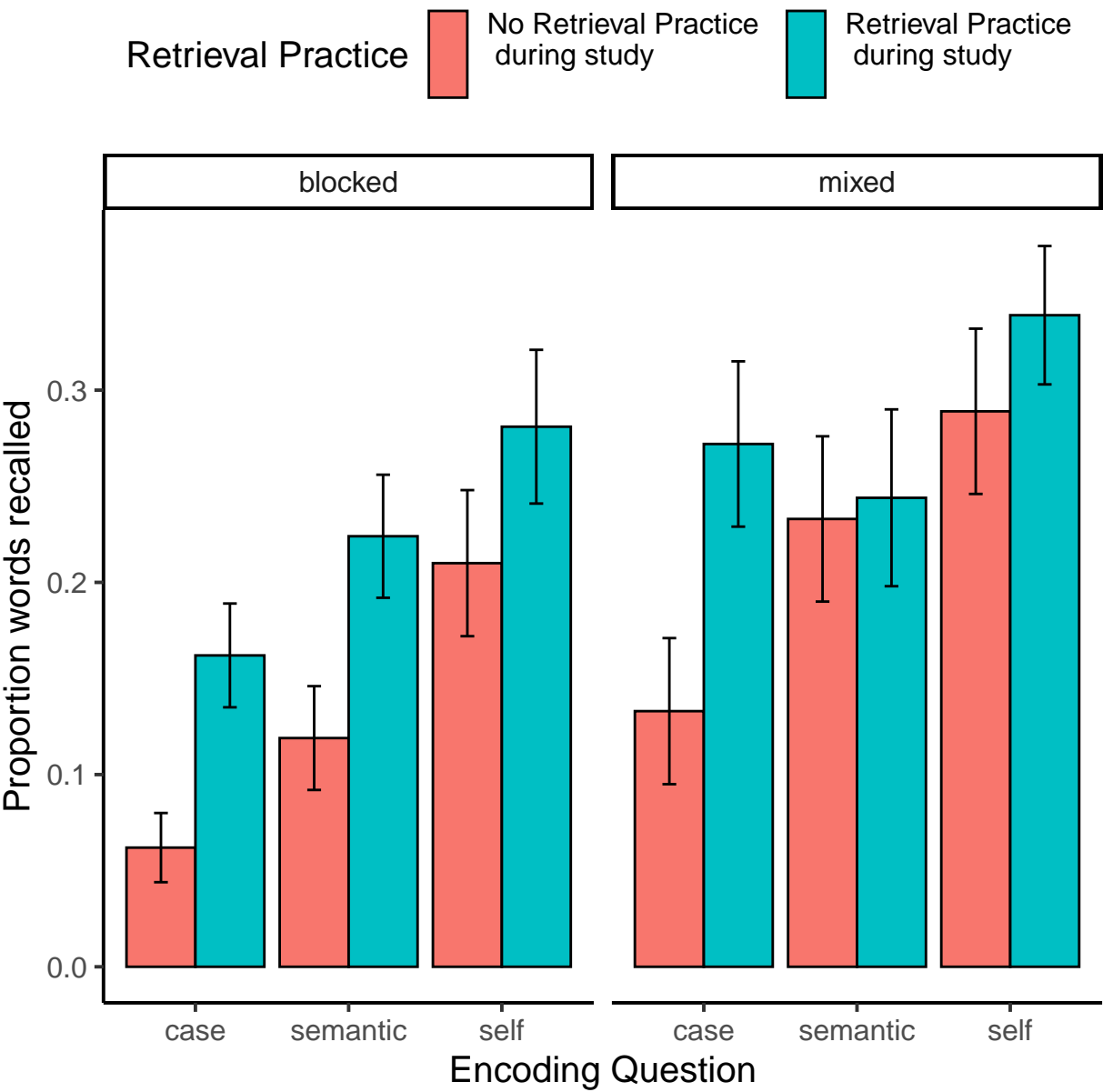
The retrieval practice effect was also significant, $F(1, 51) = 4.18$, $MSE = 0.03$, $p = .046$, $\hat{\eta}_G^2 = .008$. Mean proportion recall was lower for items that did not receive retrieval practice ($M = 0.187$, $SEM = 0.017$), compared to items that did receive retrieval practice ($M = 0.223$, $SEM = 0.017$).

No other main effects or interactions reached significance.

Experiment 1c: Brooklyn College SONA

The results from experiment 1c were submitted to the same analysis as above. Mean proportions of correctly recalled words in each condition are shown in Figure @ref(fig:E1Cresults).

Figure 3



There was a main effect of encoding question, $F(2, 126) = 12.71$, $MSE = 0.04$,

$p < .001$, $\hat{\eta}_G^2 = .059$. Mean proportion recall was lowest in the case condition ($M = 0.154$, $SEM = 0.017$), higher for the semantic condition ($M = 0.203$, $SEM = 0.019$), and highest for the self condition ($M = 0.277$, $SEM = 0.02$).

The retrieval practice effect was also significant, $F(1, 63) = 26.54$, $MSE = 0.02$, $p < .001$, $\hat{\eta}_G^2 = .038$. Mean proportion recall was lower for items that did not receive retrieval practice ($M = 0.171$, $SEM = 0.015$), compared to items that did receive retrieval practice ($M = 0.251$, $SEM = 0.016$).

Finally, there was a main effect of block type, $F(1, 63) = 5.41$, $MSE = 0.10$, $p = .023$, $\hat{\eta}_G^2 = .033$. Mean proportion recall was lower for the blocked ($M = 0.176$, $SEM = 0.014$), than mixed conditions ($M = 0.252$, $SEM = 0.017$).

No other main effects or interactions reached significance.

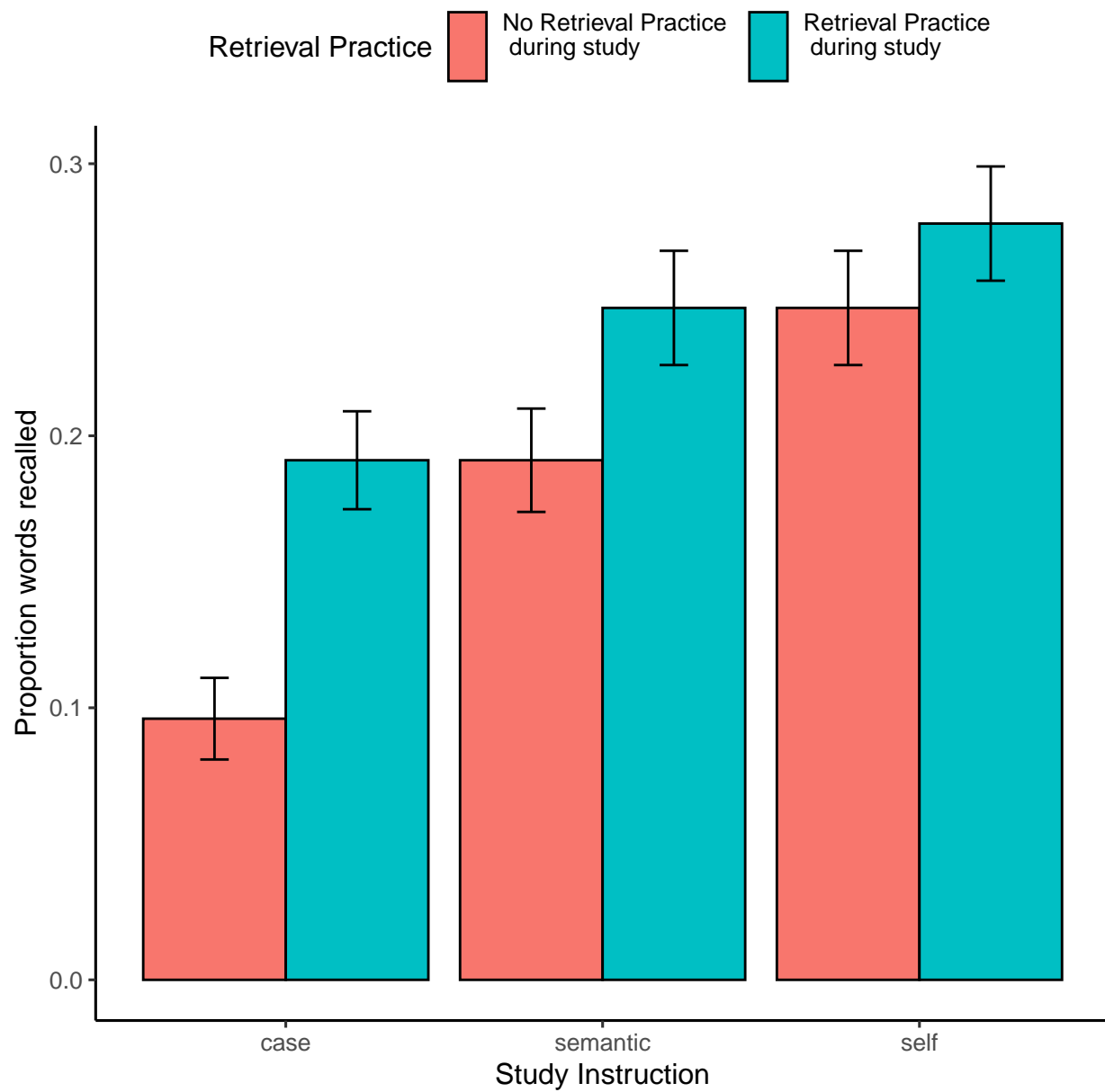
Post-hoc power analysis

The results from Experiment 1b were used to conduct a simulation-based post-hoc power analysis. The purpose of the power analysis was to estimate the smallest effect size of interest that the design was capable of detecting. Experiment 1b did not show evidence of the expected self-reference effect, specifically measured as a difference between the self and semantic encoding conditions. One possibility is that our design was not sensitive enough to detect a difference between those conditions.

The results of the power analysis indicated that the design had greater than power .8 to detect differences in recall proportions greater than .1 with $n = 100$. The design had power greater than .8 to detect differences in recall proportions greater than .05 with $n = 200$.

The power analysis suggests that individual experiment may have been under-powered to detect the self-reference effect. To increase power by increasing the number of participants, the following analysis collapsed over the results from experiment 1b and 1c.

Figure 4



Combined analysis

There was a main effect of encoding question, $F(2, 234) = 23.25$, $MSE = 0.04$, $p < .001$, $\hat{\eta}_G^2 = .052$. Mean proportion recall was lowest in the case condition ($M = 0.143$, $SEM = 0.012$), higher for the semantic condition ($M = 0.219$, $SEM = 0.014$), and highest for the self condition ($M = 0.263$, $SEM = 0.015$).

The retrieval practice effect was also significant, $F(1, 117) = 26.36$, $MSE = 0.02$, $p < .001$, $\hat{\eta}_G^2 = .021$. Mean proportion recall was lower for items that did not receive retrieval practice ($M = 0.178$, $SEM = 0.011$), compared to items that did receive retrieval practice ($M = 0.239$, $SEM = 0.012$).

The interaction between retrieval practice and encoding question approached significance, $F(2, 234) = 2.58$, $MSE = 0.02$, $p = .078$, $\hat{\eta}_G^2 = .004$.

The main hypotheses from figure @ref(fig:predictions) were investigated with simple effects comparisons. There are two ways to evaluate whether the self-reference effect was influenced by the retrieval practice manipulation.

First, looking solely at the self-reference condition, recall proportions in the retrieval practice condition were not higher than the no-retrieval practice, $M_D = -0.03$, 95% CI $[-0.08, 0.02]$, $t(117) = -1.33$, $p = .185$. This absence of a retrieval practice benefit for the self-reference condition is consistent with the hypothesis that participants were already engaging in retrieval practice as a part of making the self-reference judgment.

Second, a self-reference benefit over the semantic judgment was observed in the no-retrieval practice condition, $M_D = 0.06$, 95% CI $[0.01, 0.10]$, $t(117) = 2.50$, $p = .014$. However, a self-reference benefit over the semantic judgment was not observed in the retrieval practice condition, $M_D = 0.03$, 95% CI $[-0.02, 0.08]$, $t(117) = 1.23$, $p = .220$.

General Discussion

The present experiments tested competing hypotheses about the potential role of retrieval practice in the self-reference effect. The first hypothesis was that retrieval practice will improve recall in the case and semantic judgment conditions, but not improve recall in the self judgment condition. The reasoning was that in self reference condition, participants were already repeating the words to themselves in the no-retrieval condition which would make retrieval practice redundant. The second hypothesis stated that retrieval practice will improve recall in all conditions. To be noted, according to both hypotheses we expected to observe the self-reference effect in the no-retrieval practice

condition. The results for our pilot experiment 1a pointed out technical issues which were discussed earlier, and due to participants not following instructions results were not reported for this experiment. For experiment 1b and 1c, a combined analysis collapsing across the blocked and mixed conditions showed that the self-reference effect was only observed in the no-retrieval practice condition, and not observed in the retrieval practice condition. This pattern of results supported the first hypothesis. The remainder of the general discussion considers the main hypothesis, limitations, and more general memory-based explanations of the self-reference effect.

To reiterate the main hypothesis, although the self-reference effect has often been attributed to a deep level of meaningful processing during encoding, the reported experiments examined a somewhat more superficial basis for superior recall in the self-reference condition: repeating or not repeating the target word using inner speech during a judgment. In the self judgment condition, participants may pose questions to themselves using their inner voice as a part of judging whether the target adjective relates to themselves. These questions could involve repeating the target adjective which would provide retrieval practice and improve recall for words in the self judgment condition compared to the case and semantic judgment conditions. Another assumption was that participants do not need to repeat target words as part of an inner monologue while completing case or semantic judgments. As a result, retrieval practice during encoding may be confounded across conditions. The experiments controlled whether or not retrieval practice occurred across all judgments, and found that the self-reference effect disappeared when overt retrieval practice was equivalent across conditions.

The results are consistent with the primary hypothesis, but there are also limitations on the kinds of inferences that can be drawn from this study alone. The main assumption was that participants are naturally repeating words to themselves in the self judgment condition. However, the design did not measure whether participants actually repeat words to themselves during the encoding phase of the experiment. One suggestion

for future work is to ask participants about their subjective experience while performing the task. According to the hypothesis, the expectation is that participants will self-report repeating the target words in the self condition, but not the other conditions.

The experiment did not attempt to directly control inner voice processes during judgments. Instead, the n-back task following some encoding trials was used to force retrieval practice and act as a proxy processing experience equivalent to repeating a word to oneself during the actual judgment. A converging manipulation for future work would be to control inner voice processes more directly. For example, verbal suppression could be manipulated instead of retrieval practice. Verbal suppression involves having participants repeat a word—like “the” over and over—out loud while completing a task. The expectation is that verbal suppression would interrupt inner narrative processing during self judgments, reduce or eliminate inner voice repetitions of the target word, and thus eliminate the self-reference effect.

A last concern is the possibility of ceiling effects. On the one hand, there may be a ceiling on recall rates due to the nature of the stimuli and general capacity limitations for this design. As a result, it is possible the absence of a self-reference effect in the retrieval practice condition reflects some kind of ceiling on recall in this task. On the other hand, ceiling effects may not be a concern. For example, recall rates were overall very low, and it was entirely possible for participants to recall many more words than they did on average.

Despite these limitations, the results indicate there are some theoretical and practical implications to be considered. The results showed that mean proportion recall was lower for words that did not receive retrieval practice. This findings can be interpreted to suggest that self-reference may not be about depth of processing. As discussed earlier, the paradigm has heavily focused on “durability” of memory traces which are determined by how deeply the stimuli is processed. Accordingly, less durable traces are prone to decaying and not available for retrieval.

Although the current study offers support for retrieval practice as a mechanism

contributing to the self-reference condition, it remains possible that the self reference effect has multiple causes. For example, the role of retrieval practice could conceivably exist in tandem with the other encoding based explanations mentioned in the introduction. In addition, it is possible that retrieval processes may play a larger role than described so far. For example, the transfer appropriate processing principle (Morris et al., 1977) suggests that memory performance facilitated when processing demands during retrieval match well with processing demands encountered during encoding. It may be the case that participants naturally think about themselves during the free recall task, and such a default perspective could match well with the processing demands encountered during the self-reference task versus other tasks. On this view, the self-reference effect may be enhanced or diminished by manipulating how recall is framed during retrieval. For example, if participants were instructed to separately recall words from the case, semantic, and self-reference conditions, then the self-reference effect may be reduced because recall in the other conditions would benefit from explicit cues that matched words in those encoding conditions.

Conclusion

Relating information to yourself can make it more memorable and help you retain important information. The self-reference effect has been reliably observed in literature in various manipulations; however its underlying mechanisms are still not understood. The current study proposed an alternative explanation: self-reference could be a retrieval practice phenomenon as it also involves repeating information to yourself. Our pattern of results supported the first hypothesis as the recall was improved for tasks other than self reference. However there are a number of limitations due to which we can not presume if our results could be attributed to retrieval practice only. Future research can improve upon this research and test several improvements mentioned in this paper.

References

Aust, F., & Barth, M. (2022). *papaja: Prepare reproducible APA journal articles with R Markdown*. <https://github.com/crsh/papaja>

Barth, M. (2022). *tinylabels: Lightweight variable labels*.

<https://cran.r-project.org/package=tinylabels>

Bellezza, F. S. (1984). The Self as a Mnemonic Device: The Role of Internal Cues. *Journal of Personality and Social Psychology*, 47, 506–516.

<https://doi.org/10.1037/0022-3514.47.3.506>

Bellezza, F. S., & Hoyt, S. K. (1992). The Self-Reference Effect and Mental Cueing. *Social Cognition*, 10(1), 51–78. <https://doi.org/10.1521/soco.1992.10.1.51>

Bentley, S. V., Greenaway, K. H., & Haslam, S. A. (2017). An online paradigm for exploring the self-reference effect. *PLOS ONE*, 12(5), e0176611.

<https://doi.org/10.1371/journal.pone.0176611>

Bower, G. H., & Gilligan, S. G. (1979). Remembering information related to one's self. *Journal of Research in Personality*, 13(4), 420–432.

[https://doi.org/10.1016/0092-6566\(79\)90005-9](https://doi.org/10.1016/0092-6566(79)90005-9)

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 717–726. [https://doi.org/10.1016/S0022-5371\(72\)80006-9](https://doi.org/10.1016/S0022-5371(72)80006-9)

Chandler, J. (2018). Likeableness and meaningfulness ratings of 555 (+ 487) person-descriptive words. *Journal of Research in Personality*, 72, 50–57.

<https://doi.org/10.1016/j.jrp.2016.07.005>

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684.

[https://doi.org/10.1016/S0022-5371\(72\)80001-X](https://doi.org/10.1016/S0022-5371(72)80001-X)

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104(3), 268–294.

<https://doi.org/10.1037/0096-3445.104.3.268>

Dahl, D. B., Scott, D., Roosen, C., Magnusson, A., & Swinton, J. (2019). *Xtable: Export tables to LaTeX or HTML*. <https://CRAN.R-project.org/package=xtable>

- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47(1), 1–12.
<https://doi.org/10.3758/s13428-014-0458-y>
- Dowle, M., & Srinivasan, A. (2022). *Data.table: Extension of 'data.frame'*.
<https://CRAN.R-project.org/package=data.table>
- Ganellen, R. J., & Carver, C. S. (1985). Why does self-reference promote incidental encoding? *Journal of Experimental Social Psychology*, 21(3), 284–300.
[https://doi.org/10.1016/0022-1031\(85\)90021-6](https://doi.org/10.1016/0022-1031(85)90021-6)
- Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, 66(3), 325–331.
<https://doi.org/10.1111/j.2044-8295.1975.tb01468.x>
- Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, 40(3), 1–25. <https://www.jstatsoft.org/v40/i03/>
- Karpicke, J. D. (2012). Retrieval-Based Learning: Active Retrieval Promotes Meaningful Learning. *Current Directions in Psychological Science*, 21(3), 157–163.
<https://doi.org/10.1177/0963721412443552>
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval Practice Produces More Learning than Elaborative Studying with Concept Mapping. *Science*, 331(6018), 772–775.
<https://doi.org/10.1126/science.1199327>
- Karpicke, J. D., & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory and Language*, 67(1), 17–29.
<https://doi.org/10.1016/j.jml.2012.02.004>
- Kesebir, S., & Oishi, S. (2010). A Spontaneous Self-Reference Effect in Memory: Why Some Birthdays Are Harder to Remember Than Others. *Psychological Science*, 21(10), 1525–1531. <https://doi.org/10.1177/0956797610383436>
- Klein, S. B., & Kihlstrom, J. E. (1986). Elaboration, Organization, and the Self-Reference Effect in Memory. *Journal of Experimental Psychology: General*, 115, 26–38.

<https://doi.org/10.1037/0096-3445.115.1.26>

Klein, S. B., & Loftus, J. (1988). The nature of self-referent encoding: The contributions of elaborative and organizational processes. *Journal of Personality and Social Psychology*, 55(1), 5–11. <https://doi.org/10.1037/0022-3514.55.1.5>

Kuiper, N. A., & Derry, P. A. (1982). Depressed and nondepressed content self-reference in mild depressives. *Journal of Personality*, 50(1), 67–80.

<https://doi.org/10.1111/j.1467-6494.1982.tb00746.x>

Lange, K., Kühn, S., & Filevich, E. (2015). "Just Another Tool for Online Studies" (JATOS): An Easy Solution for Setup and Management of Web Servers Supporting Online Studies. *PLOS ONE*, 10(6), e0130834.

<https://doi.org/10.1371/journal.pone.0130834>

Macrae, C. N., Moran, J. M., Heatherton, T. F., Banfield, J. F., & Kelley, W. M. (2004). Medial Prefrontal Activity Predicts Memory for Self. *Cerebral Cortex*, 14(6), 647–654.

<https://doi.org/10.1093/cercor/bhh025>

Markus, H. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology*, 35, 63–78.

<https://doi.org/10.1037/0022-3514.35.2.63>

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, 16(5), 519–533. [https://doi.org/10.1016/S0022-5371\(77\)80016-9](https://doi.org/10.1016/S0022-5371(77)80016-9)

Müller, K., & Wickham, H. (2022). *Tibble: Simple data frames*.

<https://CRAN.R-project.org/package=tibble>

Nelson, D. L., & Archer, C. S. (1972). The first letter mnemonic. *Journal of Educational Psychology*, 63(5), 482–486. <https://doi.org/10.1037/h0033131>

Northoff, G., Heinzl, A., Greck, M. de, Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain—A meta-analysis of imaging studies on the self. *NeuroImage*, 31(1), 440–457.

<https://doi.org/10.1016/j.neuroimage.2005.12.002>

Ooms, J. (2014). The jsonlite package: A practical and consistent mapping between JSON data and r objects. *arXiv:1403.2805 [Stat.CO]*. <https://arxiv.org/abs/1403.2805>

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>

Rodd, J. M. (2023, April 13). *Moving Experimental Psychology Online: How to Maintain Data Quality When We Can't See Our Participants*.

<https://doi.org/10.31234/osf.io/2fhcb>

Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255.

<https://doi.org/10.1111/j.1467-9280.2006.01693.x>

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, 35(9), 677.

<https://doi.org/10.1037/0022-3514.35.9.677>

Symons, C. S., & Johnson, B. T. (1997). *The Self-Reference Effect in Memory: A Meta-Analysis*. <https://doi.org/10.1037/0033-2909.121.3.371>

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80(5), 352–373.

<https://doi.org/10.1037/h0020071>

Turk, D. J., Cunningham, S. J., & Macrae, C. N. (2008). Self-memory biases in explicit and incidental encoding of trait adjectives. *Consciousness and Cognition*, 17(3), 1040–1045. <https://doi.org/10.1016/j.concog.2008.02.004>

Turk, D. J., Gillespie-Smith, K., Krigolson, O. E., Havard, C., Conway, M. A., & Cunningham, S. J. (2015). Selfish learning: The impact of self-referential encoding on children's literacy attainment. *Learning and Instruction*, 40, 54–60.

<https://doi.org/10.1016/j.learninstruc.2015.08.001>

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New

York. <https://ggplot2.tidyverse.org>

Wickham, H. (2022). *Stringr: Simple, consistent wrappers for common string operations*.

<https://CRAN.R-project.org/package=stringr>

Wickham, H. (2023). *Forcats: Tools for working with categorical variables (factors)*.

<https://CRAN.R-project.org/package=forcats>

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R.,
Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E.,
Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani,
H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686.

<https://doi.org/10.21105/joss.01686>

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A grammar of data manipulation*. <https://CRAN.R-project.org/package=dplyr>

Wickham, H., & Henry, L. (2023). *Purrr: Functional programming tools*.

<https://CRAN.R-project.org/package=purrr>

Wickham, H., Hester, J., & Bryan, J. (2023). *Readr: Read rectangular text data*.

<https://CRAN.R-project.org/package=readr>

Wickham, H., Vaughan, D., & Girlich, M. (2023). *Tidyr: Tidy messy data*.

<https://CRAN.R-project.org/package=tidyr>