The Context Repetition Effect: Role of prediction in new memory formation.

Honors Research Thesis

Presented in partial fulfillment of the requirements for graduation *with honors research distinction* in Psychology in the undergraduate offices of The Ohio State University

by

Chad Gossett

The Ohio State University

Spring 2013

Project Advisor: Per Sederberg, Department of Psychology

Abstract

Many theories posit that the associative process at the core of episodic memory binds the content of an experience to the context in which we experience it. Here, context can be broadly defined as the mental representation capturing our recent experience. We recently discovered the context repetition effect (CRE), which shows that repeating a context once leads to greater memory performance for an item learned within that context even if the item does not occur again. Currently, we have conducted three studies to test the CRE. Experiment 1 was a complete replication of the original experiment that first discovered the CRE, save that there were multiple repetitions of a context instead of just one. We found that the presentation of a context and item, followed by two repetitions of the context with a new item each time, resulted in a near significant boost in memory and confidence in memory of subjects for the original item. Experiment 2 replaced words with scenes and faces. Subjects associated male and female faces with indoor and outdoor scenes. Subjects showed trends reduced performance and no demonstration of the CRE. Lack of power for performance results possibly due to difficulty in encoding faces relative to words. Experiment 3 replicated Experiment 2, save that there was an additional repetition. Results trended toward those found in Experiment 1.

Background

During the many moments we experience in our daily lives, our minds are continuously forming associations between different external constructs: a black key on a keyboard and the tactile sensation it produces, a bark and the image of the family dog, a family gathering and the face of someone you've never met before, the words one speaks and how they hint at what will be said next. This ability to form associations in memory between our experiences has been a crucial component in the development of humanity, and questions about its mechanisms and the reach of its effects have been the subject of numerous scientific studies over the years.

One concept at the core of our understanding of how we form and retrieve these associations is *context*, which can be broadly defined as the mental representation capturing our recent experience. In its attempt to explain the mental processes that underlie episodic memory, the temporal context model (TCM; Howard & Kahana 2002; Sederberg et al. 2008) claims that the primary associative mechanism guiding memory formation is the binding of the items we experience to the mental context in which we experience them. Due primarily to this simple associative mechanism, the binding of items and context, TCM has been shown to account for a large number of phenomena in episodic memory, including the pattern of responses observed for effects of recency and contiguity across multiple time scales (Sederberg et al. 2008).

Despite its success, there is one major aspect of cognition that is absent from TCM and, in fact, most other memory models--the role of prediction shaping what we learn. To fill this void, Sederberg and Smith (2011) recently introduced a new model inspired by normative models of reinforcement learning (Dayan, 1993) and recent neuroimaging work in visual perception (Turk-Browne, Sederberg, & Simon, 2011) that builds on the concepts from TCM. Specifically, the Temporal Context Update and Prediction (T-CUP) model links the successor representation from reinforcement learning to TCM and incorporates prediction and prediction error as key components of the learning mechanism (Gershman et al., 2012). The T-CUP model proposes that the episodic memory system uses the current mental representation of context to make a prediction, compares the prediction to the current state of events (i.e., the context and stimuli in the world), and then uses the resulting prediction error to determine what to encode from our current experience and how strongly to encode it.

Because the T-CUP model takes into account how well the memory system has learned the current environment, it makes a unique prediction about what should happen when items are observed and learned within a repeated or familiar context. Intuitively, one would expect that, on average, similar events that are only experienced once should be equally memorable, and this is what all major memory models predict. However, what if those two events occurred in a similar context? Does the fact that the context is repeated change the memorability of either one of the events? According to the T-CUP model, when a context is repeated the memory system predicts that the event(s) that have previously occurred within that context will occur again, and this prediction becomes part of the what is encoded in memory during the second occurrence of the context. Thus, the model makes the counterintuitive prediction that repeating a context enhances memory for a once-presented stimulus that was previously presented in that context relative to other once-presented stimuli that are not experienced in a repeated or familiar context. Corroborating this is a study conducted by Turk-Browne et al., in which it was demonstrated that humans form associative predictions based upon repetitions within a stream of images (Turk-Browne et al. 2010). Hockley also provided evidence for our hypothesis through his studies on the mediating effect of familiarity between environmental context and recognition using words and real-world photographs (Hockley 2008).

This prediction was confirmed in a series of four behavioral experiments (Smith, Hasinski, & Sederberg, submitted) where participants were presented with lists of words or images to study one at a time. Unknown to the participants, the lists were structured such that they were constructed from triplets of stimuli (e.g., ABC DEF ... ABJ GHI ...), with the first two stimuli in each triplet acting as the context within which the third stimulus (the "item") was experienced. Within the stream of stimuli, some items were presented once, either in a context that was repeated (repeated context, novel item: RCNI) or in a context that was never repeated (novel context, novel item: NCNI), and some were presented twice, again either in a repeated context (RCRI) or a novel context (NCRI).

According to the T-CUP model, items that were learned in a repeated, familiar context (items C and J in the above example) should be remembered better than items learned in a novel context (items F and I in the above example). The results of the study followed this prediction to the letter. Items that were learned in repeated contexts were recognized more frequently by participants as a previously studied item than items that were learned in a random, novel context. Shown below in **Fig.1** are the results from these experiments:

Fig.1

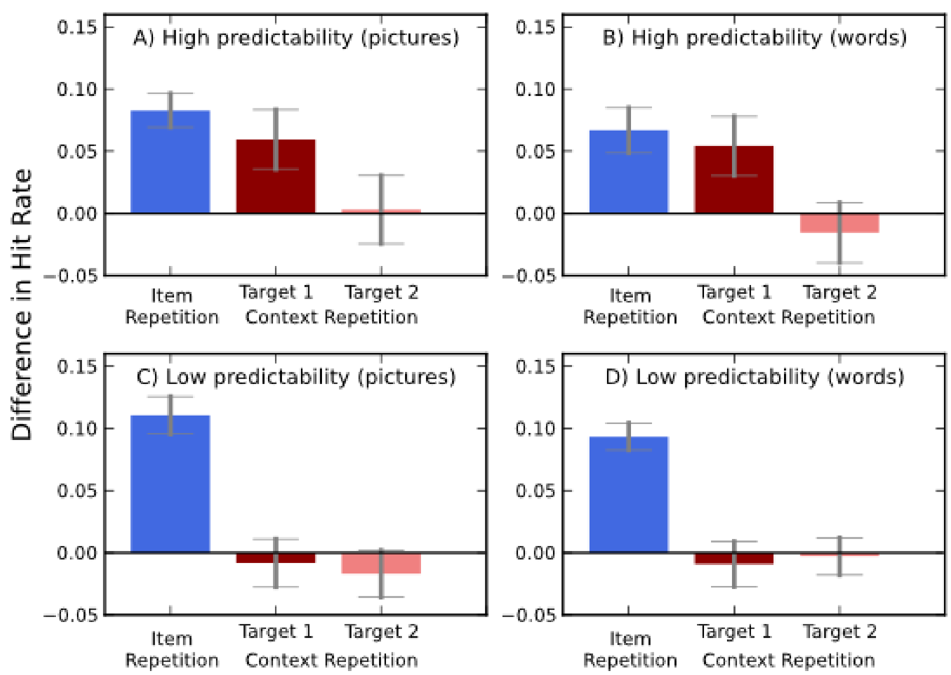


Figure 1. Item and context repetition effects for all experiments. (A) Experiment 1: high degree of predictability when encoding pictures; (B) Experiment 2: high degree of predictability when encoding word lists; (C) Experiment 3: low degree of predictability when encoding pictures; (D) Experiment 4: low degree of predictability when encoding word lists. Item repetition effect is the contrast between hit rates for all twice presented targets (NCRT and RCRT) vs. once presented targets (NCNT and RCNT). Context repetition effects are the contrasts between hit rates for once-presented targets preceded by a repeated context (RCNT) vs. those preceded by a non-repeated context (NCNT), calculated separately for the first and second targets associated to the repeated context and matched novel context. Error bars are standard errors of the mean calculated across subjects.

Introduction

There are many questions left to answer about the T-CUP and what it says about the CRE. How does the T-CUP predict the CRE will react to multiple repetitions of a context? What does it say of the CRE and faces, given that faces are much more difficult to encode (Diamond & Carey 1986)? These questions formed the basis of the two overarching goals that we wished to accomplish with these three experiments. First, we wished to better elucidate how the CRE held up across multiple repetitions of a context. To accomplish this, we designed Experiment 1 to replicate the study that originally discovered the CRE, save that it contained an additional repetition of context. Second, we desired to make the experiment more applicable to real life. As a result, instead of words, Experiment 2 used scenes as contexts and faces as items in order to simulate the everyday experience of viewing individuals in different, yet repeated, settings. In order to see how the CRE reacted to multiple repetitions under more realistic settings (a combination of goals 1 and 2), Experiment 3 both repeated contexts multiple times *and* used scenes as contexts and faces as items.

Multiple Predictions

As mentioned earlier, an essential component of the T-CUP is context-based prediction and how errors in this prediction lead to improved memory performance. In other words, when we observe an item in a particular context and then return to that context at a later date, we unconsciously and automatically predict that the original item will be found within the context. According to the T-CUP, this prediction causes us to activate the original item so strongly that it becomes enhanced. While this was confirmed by our lab (Smith, Hasinski, & Sederberg, submitted), the degree to which this effect holds up across multiple repetitions was not examined. The T-CUP model predicts that multiple repetitions of a context, an experimental scenario that is more applicable to the constant entering and reentering of contexts that make up our daily lives, will lead to multiple predictions about what will appear in a repeated context. To put it another way, when we see Item X in Context A, a second visit to Context A will cause us to predict Item X's presence and thus enhance our memories of Item X even if it is not present. However, if in three visits to Context A we saw Item X on the first visit, Item Y on the second visit, and Item Z on the third visit, our second visit would cause us to be predict Item X's presence and our third visit would cause us to predict both Item X's presence *and* Item Y's presence. The effect of this second prediction upon our memories is crucial to our understanding of the CRE. By creating a scenario in which novel items are observed within repeated contexts, we will be able to better understand how differential placement of items within a repeated context, which leads to multiple predictions about those items, affects memory for them.

Prediction Violation

Additionally, we have not examined how the severity of prediction violation affects memory. In other words, how does a severe prediction violation affect the memory of an item relative to a moderate prediction violation? To answer this, we added into Experiments 1 & 3 the conditions RCRI-112, which raises a severe violation of prediction, and RCRI-121, which raises a weak violation of prediction. In RCRI-112, we present context and item, repeat the presentation, and then present the repeated context with a new item. Despite seeing the same item bound with the same context twice and violating the produced prediction, the T-CUP predicts that the observation of a completely new item within the repeated context will not greatly improve memory performance for the new item. In RCRI-121, we present context and item once and then repeat the context with a new item. According to the T-CUP, this very moderate violation of the prediction elicited by the context will also not lead to an increase in memory performance.

Methods

Experiment 1

Subjects. 54 volunteers were recruited from a pool of undergraduate students enrolled in Ohio State University Department of Psychology courses who participate in the Research Experience Program (REP) for partial credit. Participants were required to be at least 18 years of age. All participants provided consent in accordance with the requirements of the OSU Institutional Review Board.

Exclusions. Fourteen participants were excluded due to incompletion of experiment and/or poor performance, which was identified by negative d’ values.

Design. This experiment had six conditions. The first condition, the RCNI condition (Repeated Context, Novel Item), was shown as ABC…ABD…ABE, in which two words acting as a context (AB) are repeated throughout the block in conjunction with a unique and novel item (C,D,E). This condition appeared four times throughout a block, with the context words A and B being repeated each time.

The next four conditions were variations of the RCRI condition (Repeated Context, Repeated Item). The first variation, RCRI-112, was modeled ABC…ABC…ABD, in which triplet ABC was repeated once before a RCNI trial; the second variation, RCRI-122, was ABC…ABD…ABD, which is a reverse of the first variation; the third variation, RCRI-121, was ABC…ABD…ABC, in which the triplet ABC is presented once before an RCNI trial and once after an RCNI trial; finally, the fourth variation, RCRI-111, was ABC…ABC…ABC, in which the triplet ABC is presented eight times in a block.

The sixth condition was the NCNI condition (Novel Context, Novel Item) and is modeled as ABC…DEF…GHI, in which both the context and item in a trial are completely novel (i.e. never seen previously by the subject).

Materials. Stimuli consisted of 736 words from personal collection. All words were random and the valence of the words was neutral.

Procedure. The experiment we have designed consisted of eight blocks, with each phase, study and test, having four blocks.

Study Phase: Within each block were four stimulus sets that consisted of 21 trials each and contained all experimental conditions. Stimulus sets were repeated four times in order to generate the appropriate effect size. This resulted in a total of 84 study trials per block and 336 study trials total. A single trial consisted of a word appearing on the computer screen. The subject then had 1500 ms to make a Living/Nonliving judgment about the word. The word then disappeared and the subject was given 700 ms before the next pairing plus an additional jitter that lasted between 200 - 400 ms. In summary, an individual trial lasted around 2.5 seconds. With 84 trials in one study phase block and 336 study trials total, a single block lasted 3.5 minutes and the study phase as a whole lasted 14 minutes. Below are examples of both how a condition (in this case, the RCNI condition) was structured within the Study Phase and what a single trial of the Study Phase looked like to a participant.

Fig. 2 - Structure of RCRI-112 Condition within Study Phase

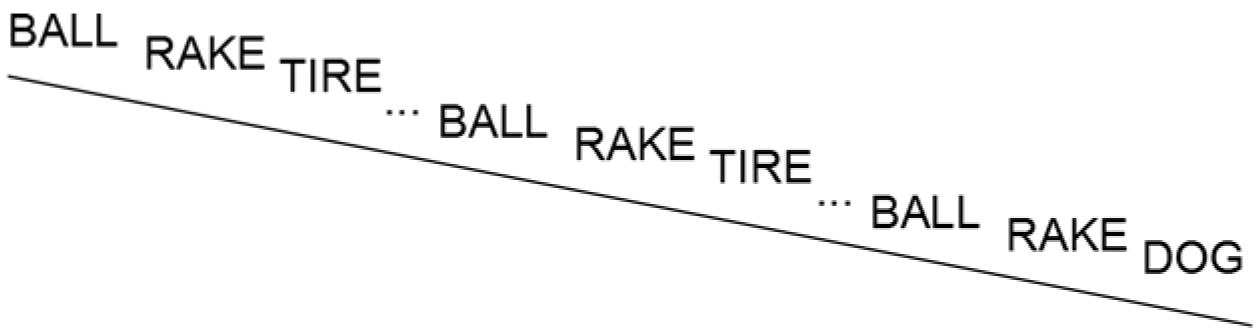


Figure 2: Above are triplets of words as they appeared in Study Phase, all three of which belonged to the RCRI-112 condition.. Words were presented one at a time in sequence with judgment options underneath (see **Fig. 2**). Number of words between triplets varied across Study Phase, hence the ellipsis marks separating triplets.

Fig. 3 - Single Trial of Study Phase

DOG

Living - J   Nonliving - K

Figure 3: Above is an example of an individual trial as it appeared to subjects. Subjects were asked to make Living/Nonliving judgments about words in order to ensure they were paying attention.

Test Phase: In the test phase, trials consisted of a single target word appearing on the participant's computer screen to respond to with one of four possible judgments: Definitely OLD (studied in Study Phase), Possibly OLD, Possibly NEW (not studied in Study Phase), or Definitely NEW. Participants had 2000 ms to observe and respond to faces, with a 500 ms ITI between faces. As a result, a single trial in the Test Phase lasted 2500 ms. With 112 trials per block (448 trials total), a single block took 4.7 minutes to complete and the test phase itself, consisting of 4 blocks, lasted 18.7 minutes.

Fig. 4 - Single Trial of Test Phase

DOG

Definitely OLD - J   Possibly OLD - K   Possibly NEW - L   Definitely New - ;

Figure 4: Above is an example of a single trial as it appeared in the Test Phase. Subjects were asked to identify items as either OLD (previously seen in the Study Phase) or NEW (not seen in the study phase). The degree to which they correctly identified items as OLD or NEW was used to determine their overall memory performance. Additionally, they were also allowed to rate their confidence in their answer.

Measures. To assess the ability of participants to identify different items as either OLD or NEW (i.e. memory performance), we used d'values. This measure was chosen in order to accommodate the discriminability between rates of subjects correctly identifying items as OLD and rates of subjects incorrectly identifying lures (novel items similar in appearance to OLD items) as OLD. We also utilized a linear mixed effects model to account for variance generated by random effects, such as between-subject differences.

Results. The independent variables in this experiment were the repetition of contexts and items and the presentation of items at different points before or after a repeated context. Dependent variables were the subjects' memory performance and confidence scores. The data was analyzed using R (R Development Core Team, 2009), the R packages lme4 (Bates & Maechler, 2009), and *languageR* (Baayen, 2009; cf. Baayen, 2008). Data was analyzed using linear mixed effect regression models to estimate p-values, which were used to determine significance. Linear mixed effect regression models are used for continuous variables and produce t-statistics. We also used generalized linear mixed models, which estimated p-values via the Laplace approximation. Generalized linear mixed models are used for categorical data and to produce t-statistics.

The use of mixed models is important because, by treating these differences as random effects, it does not ignore random differences that exist in stimuli or in participants. When variables are included as random effects, they are assumed to have a normal distribution (Bates, 2012). Therefore, their variance does not contribute to whether or not a result is found to be significant. An example of a random effect in stimuli is that one stimulus may be inherently more recognizable than another. Likewise, one participant may be predisposed to naturally feel more strongly about a word than another participant. Isolating these differences as random effects reduces the likelihood of a Type 1 error, which refers to erroneously concluding a result is significant when it is not (Judd, Westfall, & Kenny, 2012).

We used the following model to determine if memory performance for items was affected by the condition in which the items belonged to:

correct ~ cond * trial_id + (1|subject) + (1|stim_name)

In this model, "correct", which refers to memory performance, is the dependent variable. The independent variable “cond” refers to the condition of interest and the independent variable “trial_id” refers to the placement of the item with respect to the number of times a context has been repeated. Again, where mixed effects model become more powerful is in their ability to account for random effects. Here, we account for inter-subject variability with “(1|subject)” and inter-stimuli variability with “(1|stim_name)”.

We used the following model to determine if confidence for items was affected by the condition in which the items belonged to:

confidence ~ cond * trial_id + (1|subject) + (1|stim_name)

In this model, “confidence”, which refers to the confidence scores provided by participants ranked on a linear scale (Definitely OLD: 3, Possibly OLD: 1, Possibly NEW: -1, Definitely NEW: -3), is the dependent variable. As before, the independent variables “cond” and “trial_id” refer to the condition and item placement. Again,we account for inter-subject variability with “(1|subject)” and inter-stimuli variability with “(1|stim_name)”.

For RCRI-112 and RCRI-121, we excluded “trial_id” as a fixed effect because each condition had only one trial placement of interest. Therefore, we used the following models for memory performance and confidence, respectively:

correct ~ cond + (1|subject) + (1|stim_name)

confidence ~ cond + (1|subject) + (1|stim_name)

The memory performance and confidence results for the conditions of interest - RCNI, RCRI-112, and RCRI-121- are shown below in **Figs. 5**, **6**, and **7**.

Fig. 5 - Memory Performance RCNI vs. NCNI

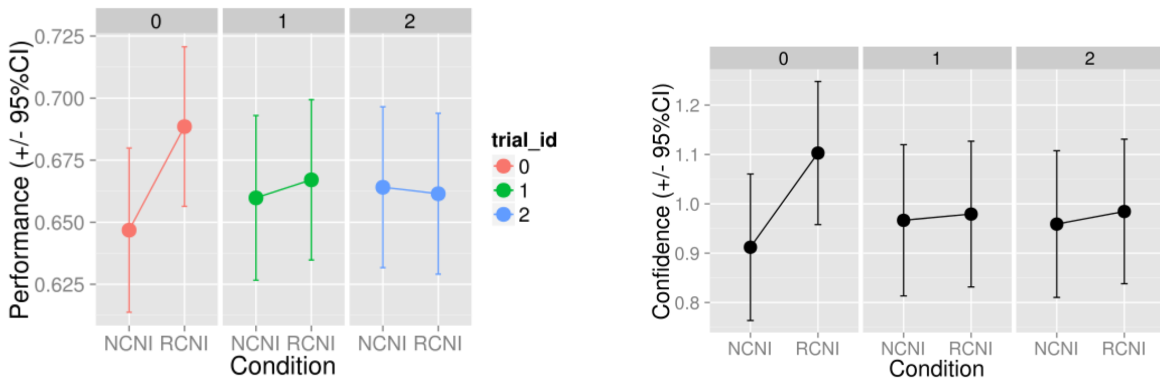


Figure 5: Memory performance (p<0.06) and confidence (p<0.06) for novel items bound to a repeated context in the RCNI condition trended upwards relative to NCNI items.

Fig. 6 - Memory Performance RCRI-112 vs. NCNI

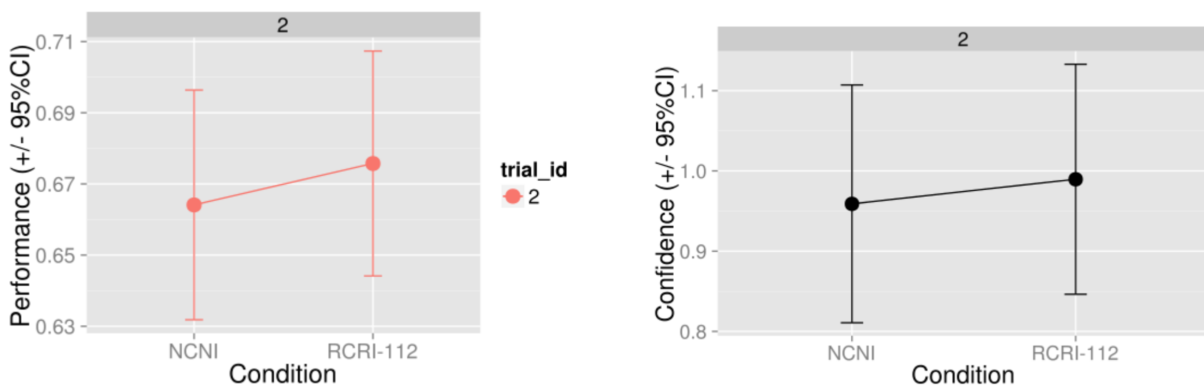


Figure 6: Memory performance (p>0.6) and confidence (p>0.7) for third items bound to the repeated context in the RCRI-112 condition were not significantly higher than for third items bound to a novel context in the NCNI condition.

Fig. 7 - Memory Performance RCRI-121 vs. NCNI

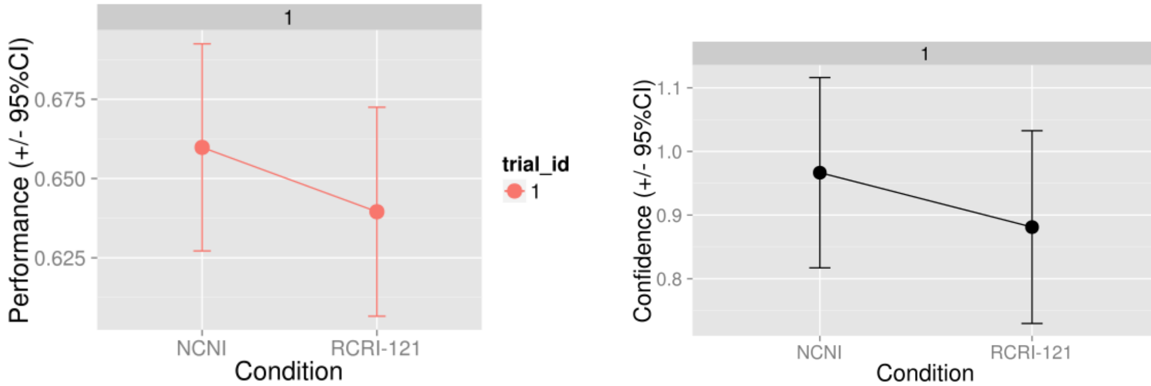


Figure 7: Memory performance (p>0.35) and confidence (p>0.35) for second items bound to the repeated context in the RCRI-121 condition was not significantly different from second items bound to a novel context in the NCNI condition.

Discussion. Just as the T-CUP predicted, memory performance for RCNI items trended upward in comparison to NCNI items (t=1.996, p<0.06). Subject confidence in memory also trended upwards (t=2.022, p<0.06). This is a complete replication of the CRE shown by Smith, Hasinski, & Sederberg and further evidence supporting our claim that prediction is a crucial component of context-based memory formation.

However, both memory performance (t = 0.444, p>0.62) and confidence in memory performance (t = 0.286, p>0.72) for the third item of the RCRI-112 condition did not improve at all compared to that of NCNI items. Furthermore, subjects demonstrated no significant increase in memory performance (t = -0.85, p>0.35) or confidence (t = -0.874, p>0.37) for RCRI-121. Both results suggest that, in the case of multiple repetitions of a context, neither a moderate violation nor a severe violation of one's prediction about what will be observed in a context is enough to produce a significant boost in either performance or confidence in performance.

Experiment 2

Subjects. 65 volunteers were recruited from a pool of undergraduate students enrolled in Ohio State University Department of Psychology courses who participate in the Research Experience Program (REP) for partial credit. Participants were required to be at least 18 years of age. All participants provided consent in accordance with the requirements of the OSU Institutional Review Board.

Exclusions. Five participants were excluded from final analysis due to technical issues and incompletion of experiment.

Design. This experiment had four conditions. The first condition, the RCNI condition (Repeated Context, Novel Item), was shown as AB…AC, in which a scene acting as a context (A) is repeated throughout the block in conjunction with a unique and novel item (B, C, D). This condition appeared eight times throughout a block, with the context A being repeated each time.

The second condition was the NCNI condition (Novel Context, Novel Item) and was modeled as AB…CD, in which both the scenes and faces in a trial are completely novel (i.e. never seen previously by the subject). The condition appeared eight times throughout a block.

The third condition was the RCRI condition (Repeated Context, Repeated Item) and was modeled as AB…AB, in which the pairing AB is repeated sixteen times times in a block. This condition appeared sixteen times within a block, with the context A and target B repeated each time.

The fourth condition was the NCRI condition (Novel Context, Repeated Item) and was modeled as AB...CB, in which the scenes were completely novel and the faces repeated. The condition appeared eight times throughout a block, with the target face B being repeated each time.

Materials. Stimuli consisted of 512 images from personal collection consisting of both faces and scenes. All pictures of faces were random and the valence of the images of faces was neutral.

Procedure. The experiment we have designed consisted of eight blocks, with each phase, study and test, having four blocks.

Study Phase: Within each block were four stimulus sets that consisted of 20 trials each and contained all experimental conditions. Stimulus sets were repeated four times in order to generate the appropriate effect size. This resulted in a total of 80 study trials per block and 320 study trials total. A single trial consisted of a scene appearing on one half of a computer screen for 1500 ms. A face then appeared alongside the scene. The subject then had 1000 ms to make a Male/Female judgment about the face. The pairing then disappeared and the subject was given 700 ms before the next pairing plus an additional jitter that lasted between 200 - 400 ms. In summary, an individual trial lasted 3.35 seconds. With 80 trials in one study phase block and 320 study trials total, a single block lasted 4.5 minutes and the study phase as a whole lasted nearly 18 minutes.

Fig. 8 - Sample Sequence from Study Phase

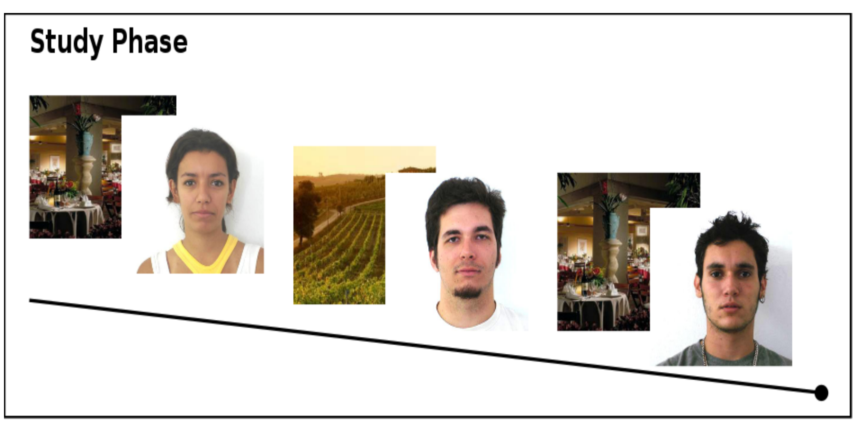


Figure 8: Above are pairings of scenes and faces as they appeared in the Study Phase. Scenes and faces were presented together in sequence with judgment options underneath (see **Fig. 9**).

Fig. 9 - Sample Trial from Study Phase

Indoor - J Outdoor - K Male - J Female - K

Figure 9: Above is a sample trial from the Study Phase. At the beginning of the trial, a scene appeared on the left-hand side of the screen with an Indoor/Outdoor judgment beneath it. After judgment was made, or 500 ms passed, the judgment prompt disappeared and a face appeared on the right-hand side of the screen with a Male/Female judgment underneath. Both images and prompt disappeared after judgment was made or 1000 ms passed.

Test Phase: In the test phase, trials consisted of a single face appearing on the participant's computer screen to respond to with one of four possible judgments: Definitely OLD (studied earlier in the experiment), Possibly OLD, Possibly NEW (not studied in the experiment), or Definitely NEW. Participants had 2000 ms to observe and respond to faces, with a 500 ms ITI between faces. As a result, a single trial in the Test Phase lasted 2500 ms. With 128 trials per block (512 trials total), a single block took 5.3 minutes to complete and the test phase itself, consisting of 4 blocks, lasted slightly over 21 minutes.

Fig. 10

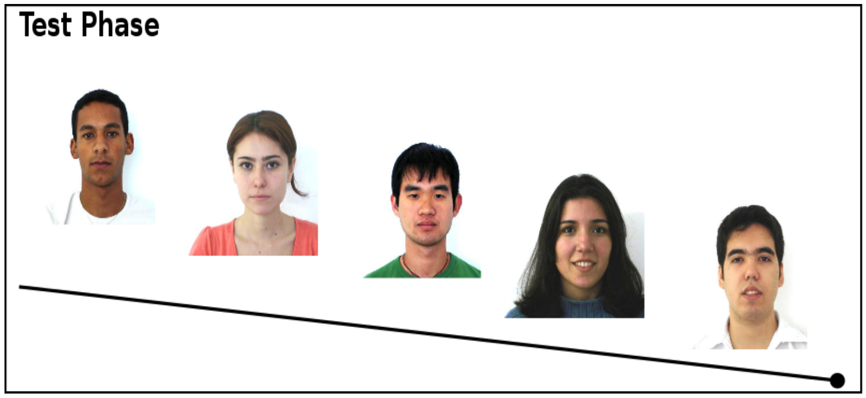


Figure 8: During Test Phase, faces appeared one at a time on screen. Subjects asked to make OLD/NEW judgments, as well as rate their confidence in their answer (see **Fig. 9**).

Fig. 11

Definitely OLD - J   Possibly OLD - K   Possibly NEW - L   Definitely New - ;

Figure 11: Above is an example of a single trial as it appeared in the Test Phase. Subjects were asked to identify items as either OLD (previously seen in the Study Phase) or NEW (not seen in the study phase). The degree to which they correctly identified items as OLD or NEW was used to determine their overall memory performance. Additionally, they were also allowed to rate their confidence in their answer.

Measures. We used d' values to assess the ability of participants to identify different items as either OLD or NEW. This measure was chosen in order to accommodate the discriminability between rates of subjects correctly identifying items as OLD and rates of subjects incorrectly identifying lures (novel items similar in appearance to OLD items) as OLD. We also utilized a linear mixed effects model to account for variance generated by random effects, such as between-subject differences.

Results. The independent variables in this experiment were the repetition of contexts and items and the presentation of items at different points before or after a repeated context. Dependent variables were the subjects' memory performance and confidence scores. The data was analyzed using R (R Development Core Team, 2009), the R packages lme4 (Bates & Maechler, 2009), and *languageR* (Baayen, 2009; cf. Baayen, 2008). Data was analyzed using linear mixed effect regression models to estimate p-values, which were used to determine significance. Linear mixed effect regression models are used for continuous variables and produce t-statistics. We also used generalized linear mixed models, which estimated p-values via the Laplace approximation.

We used the following model to determine if memory performance for items was affected by the condition in which the items belonged to:

correct ~ cond + (1|subject) + (1|stim_name)

In this model, "correct", which refers to memory performance, is the dependent variable. The independent variable, "cond", refers to the condition. Again, where mixed effects model become more powerful is in their ability to account for random effects. Here, we account for inter-subject variability with "(1|subject)" and inter-stimuli variability with "(1|stim_name)". We excluded "trial_id", which refers to the placement of items, as a fixed effect because an insignificant likelihood ratio test revealed that it's presence would not significantly affect the results.

We used the following model to determine if confidence for items was affected by the condition in which the items belonged to:

confidence ~ cond + (1|subject) + (1|stim_name)

In this model, "confidence", which refers to the confidence scores provided by participants ranked on a linear scale (Definitely OLD: 3, Possibly OLD: 1, Possibly NEW: -1, Definitely NEW: -3), is the dependent variable. The independent variable, "cond", refers to the condition. Again,we account for inter-subject variability with "(1|subject)" and inter-stimuli variability with "(1|stim_name)". Unlike Experiment 1, we excluded "trial_id" as a fixed effect because an insignificant likelihood ratio test revealed that it's presence would not significantly affect the results.

Fig. 12

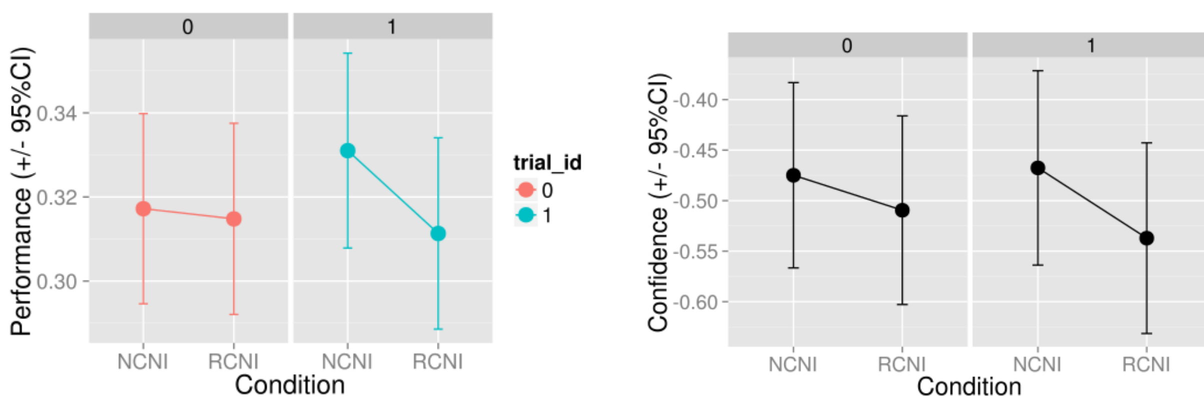


Figure 12: Memory performance (p>0.4) and confidence (p>0.2) for novel items bound to a repeated context in the RCNI condition was not significantly different from NCNI items.

Discussion. Unlike Experiment 1, memory performance (t = -0.83, p>0.4) and confidence scores (t = -1.025, p>0.25) for RCNI items were neither significantly enhanced or suppressed relative to NCNI items as a result of multiple repetitions. For first items associated with repeated contexts, there was almost no change in performance (t = -0.12, p>0.85) or confidence (t = -0.668, p>0.5). This could be due to the greater difficulty in encoding faces relative to words. Furthermore, subjects were asked to constantly switch judgment tasks between scenes and faces, which would further reduce performance.

Experiment 3

Subjects. 59 volunteers were recruited from a pool of undergraduate students enrolled in Ohio State University Department of Psychology courses who participate in the Research Experience Program (REP) for partial credit. Participants were required to be at least 18 years of age. All participants provided consent in accordance with the requirements of the OSU Institutional Review Board.

Exclusions. Seven participants were excluded from final analysis due to incompletion of experiment.

Design. This experiment had six conditions. The first condition, the RCNI condition (Repeated Context, Novel Item), was shown as AB…AC…AD, in which a scene acting as a context (A) is repeated throughout the block in conjunction with a unique and novel item (B,C,D). This condition appeared four times throughout a block, with the context A being repeated each time.

The next four conditions were variations of the RCRI condition (Repeated Context, Repeated Item). The first variation was modeled AB…AB…AC, in which the pairing between scene A and a face B (AB) was repeated before a RCNI trial; the second variation was AB…AC…AB, in which the pairing AB is repeated after an RCNI trial; the third variation was AB…AC…AC, in which the pairing AC is repeated; finally, the fourth variation was AB…AB…AB, in which the pairing AB is repeated eight times in a block.

The sixth condition was the NCNI condition (Novel Context, Novel Item) and is modeled as AB…CD…DE, in which both the scenes and faces in a trial are completely novel (i.e. never seen previously by the subject).

Materials. Stimuli consisted of 512 images from personal collection consisting of both faces and scenes. All pictures of faces were random and the valence of the images of faces was neutral.

Procedure. The experiment we have designed consisted of eight blocks, with each phase, study and test, having four blocks.

Study Phase: Within each block were four stimulus sets that consisted of 21 trials each and contained all experimental conditions. Stimulus sets were repeated four times in order to generate the appropriate effect size. This resulted in a total of 84 study trials per block and 336 study trials total. A single trial consisted of a scene appearing on one half of a computer screen with prompt asking whether the scene was Indoor or Outdoor. A face then appeared alongside the scene following the judgment or 500 ms. The subject then had 1000 ms to make a Male/Female judgment about the face. The pairing then disappeared and the subject was given 700 ms before the next pairing plus an additional jitter that lasted between 200 - 400 ms. In summary, an individual trial lasted around 2.5 seconds. With 84 trials in one study phase block and 336 study trials total, a single block lasted 3.5 minutes and the study phase as a whole lasted 14 minutes. See **Figs. 8** and **9** for structure of study phase and example of a single study phase trial.

Test Phase: In the test phase, trials consisted of a single face appearing on the participant's computer screen to respond to with one of four possible judgments: Definitely OLD (studied earlier in the experiment), Possibly OLD, Possibly NEW (not studied in the experiment), or Definitely NEW.

Participants had 2000 ms to observe and respond to faces, was given 200 ms before the next pairing plus an additional 200-400 ms between faces. As a result, a single trial in the Test Phase lasted around 2300 ms. With 112 trials per block (448 trials total), a single block took 4.3 minutes to complete and the test phase itself, consisting of 4 blocks, lasted close to 17.2 minutes. See **Figs. 10** and **11** for structure of test phase and example of a single trial from test phase.

Measures. We used d' values to assess the ability of participants to identify different items as either OLD or NEW. This measure was chosen in order to accommodate the discriminability between rates of subjects correctly identifying items as OLD and rates of subjects incorrectly identifying lures (novel items similar in appearance to OLD items) as OLD. We also utilized a linear mixed effects model to account for variance generated by random effects, such as between-subject differences.

Results. The independent variables in this experiment were the repetition of contexts and items and the presentation of items at different points before or after a repeated context. Dependent variables were the subjects' memory performance and confidence scores. The data was analyzed using R (R Development Core Team, 2009), the R packages lme4 (Bates & Maechler, 2009), and *languageR* (Baayen, 2009; cf. Baayen, 2008). Data was analyzed using linear mixed effect regression models to estimate p-values, which were used to determine significance. Linear mixed effect regression models are used for continuous variables and produce t-statistics. We also used generalized linear mixed models, which estimated p-values via the Laplace approximation.

We used the following model to determine if memory performance for items was affected by the condition in which the items belonged to:

correct ~ cond * scale(lure_conf_mean) + (1|subject) + (1|stim_name)

In this model, “correct”, which refers to memory performance, is the dependent variable. The independent variable, “cond”, refers to the condition. Again, where mixed effects model become more powerful is in their ability to account for random effects. Here, we account for inter-subject variability with “(1|subject)” and inter-stimuli variability with “(1|stim_name)”. We excluded “trial_id”, which refers to the placement of items, as a fixed effect because an insignificant likelihood ratio test revealed that it’s presence would not significantly affect the results.

We used the following model to determine if confidence for items was affected by the condition in which the items belonged to:

confidence ~ cond * scale(lure_conf_mean) + (1|subject) + (1|stim_name)

In this model, “confidence”, which refers to the confidence scores provided by participants ranked on a linear scale (Definitely OLD: 3, Possibly OLD: 1, Possibly NEW: -1, Definitely NEW: -3), is the dependent variable. The independent variable, “cond”, refers to the condition. Again,we account for inter-subject variability with “(1|subject)” and inter-stimuli variability with “(1|stim_name)”. We excluded “trial_id”, which refers to the placement of items, as a fixed effect because an insignificant likelihood ratio test revealed that it’s presence would not significantly affect the results.

The memory performance and confidence results for the conditions of interest - RCNI, RCRI-112, and RCRI-121- are shown below in **Figs. 13**, **14**, and **15**:

Fig. 13 - Memory Performance RCNI vs. NCNI

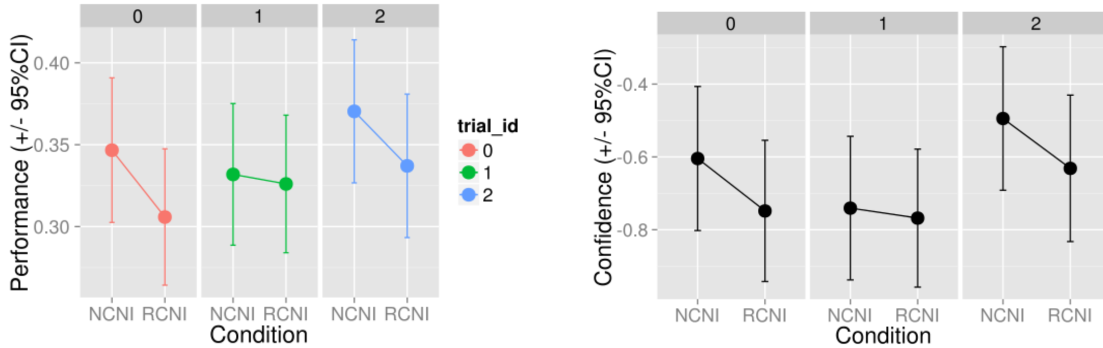


Figure 13: Memory performance and confidence for novel items bound to a repeated context in the RCNI condition was not significantly different from that of NCNI items(p>0.1). Performance for the first RCNI item trended downward relative to NCNI items, but not significantly so (p>0.1).

Fig. 14 - Memory Performance RCRI-112 vs. NCNI

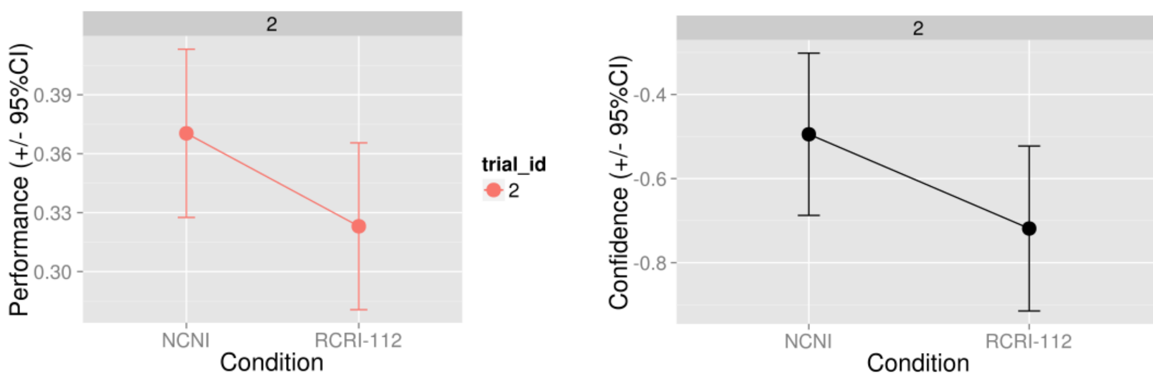


Figure 14: Memory performance (p>0.1) and confidence (p>0.1) for third items bound to the repeated context in the RCRI-112 condition were not significantly different from third items bound to a novel context in the NCNI condition.

Fig. 15 - Memory Performance RCRI-121 vs. NCNI

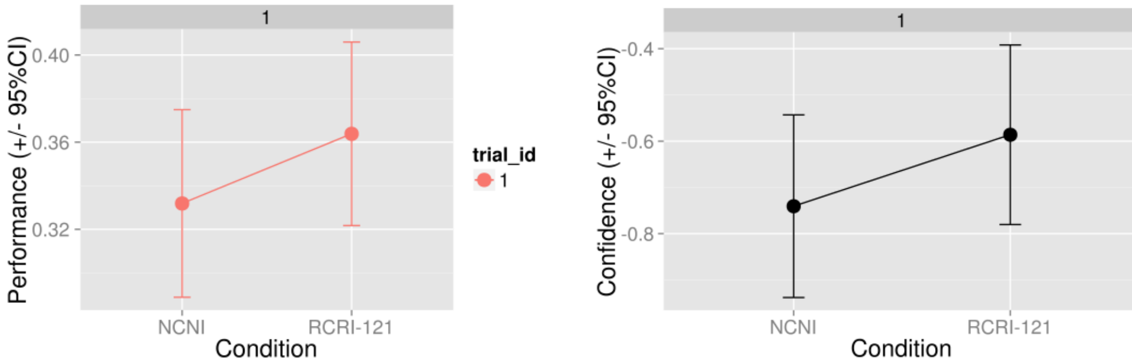


Figure 15: Memory performance (p>0.2)and confidence (p>0.4) for second items bound to the repeated context in the RCRI-121 condition were not significantly different from second items bound to a novel context in the NCNI condition.

Discussion. Unlike Experiment 1, memory performance and confidence scores for RCNI items were neither significantly enhanced or suppressed relative to NCNI items as a result of multiple repetitions (t = -1.521, p>0.1). For first items associated with repeated contexts, there were trends towards suppression in memory performance (t = -1.404, p>0.1) and confidence (t = -1.130, p>0.25), but they were not significant. Given the low performance demonstrated by subjects in Experiment 2, which also used scenes as contexts and faces as items, it is understandable that an additional repetition would further weaken any traces of the CRE and lower overall performance.

RCRI-112 items also demonstrated no significant difference from NCNI items in either memory performance (t = -1.425, p>0.1) or confidence (t = -1.473, p>0.1). RCRI-121 items showed no significant difference in either performance (t = 1.056, p>0.2) or confidence (t = 1.004, p>0.4).

General Discussion

We live cyclic existences. Contexts, and the myriad items we observe within them, are repeated and revisited constantly. By studying the Context Repetition Effect, we might better understand the mechanisms underlying how context affects our memories, as well as our ability to suppress or enhance that which we have experienced. More specifically, we can use the T-CUP model to better understand and predict how context-based predictions shape our own memories.

After viewing an item within a context and then seeing that context repeated twice with two new items, the T-CUP model claims that participants predicted the original item would be present. These predictions would then theoretically enhance memory for the original item, even if it was never seen again. As we demonstrated in Experiment 1, memory performance and confidence trended towards being significantly higher than that of NCNI items. This was a complete replication of the CRE as it was discovered in the original experiment by Smith, Hasinski, and Sederberg (Smith, Hasinski, & Sederberg, submitted). When either severe or moderate prediction violations were induced, however, we saw no substantial increase in memory performance or confidence. This suggests that, as the T-CUP model claims, prediction violation does not lead to memory enhancement for items that violate context-based predictions.

In Experiments 2 and 3, we found no evidence of the CRE when using scenes as contexts and faces as target items. Indeed, the prediction elicited by repeated contexts did not appear to produce any enhancement for the original items associated with those contexts. This could simply mean that, because faces are in general much harder to encode than words (Diamond & Carey 1986), the low memory performance displayed by participants in Experiments 2 and 3 suppressed the CRE. It could also indicate that participants did not properly bind faces to scenes, which is a crucial component of the

CRE. Furthermore, subjects in Experiments 2 and 3 had to shift encoding tasks on every trial from scene judgments (“Indoor/Outdoor”) to face judgments (“Male/Female”), which expended mental energy and reduced performance.

References

Baayen, R. H. (2008). languageR: Data sets and functions with “Analyzing Linguistic Data: A practical introduction to statistics”. *R package version 0.953*.

Bates, D. (2012). Linear mixed model implementation in lme4. *Ms., University of Wisconsin*.

Bates, D., Maechler, M., & Bolker, B. (2009). lme4: Linear mixed-effects models using S4 classes R package version 0999375-31 Available at h ttp. *CRANR-projectorg/package= lme4. Accessed on April*, *17*, 2009.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, *115*(2), 107-117.

Gershman, S. J., Moore, C. D., Todd, M. T., Norman, K. A., & Sederberg, P. B. (2012). The successor representation and temporal context. *Neural computation*, *24*(6), 1553-1568.

Hockley, William E. (2008). The effects of environmental context on recognition memory and claims of remembering. *Journal of Experimental Psychology*, 34(6):1412-1429.

Howard, M. and Kahana, M. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46(3):269–299.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103,* 54-69.

Norman, K. A., Newman, E. L., & Detre, G. (2007). A neural network model of retrieval-induced forgetting. *Psychological review*, *114*(4), 887-953.

Sederberg, P., Howard, M., and Kahana, M. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review*, 115(4):893–912.

Sederberg, P. and Smith, T. (in press). Modeling the role of context and prediction error in encoding variability.

Turk-Browne, Nicholas B., Scholl, Brian J., Johnson, Marcia K., and Chun, Marvin M. (2010) Implicit perceptual anticipation triggered by statistical learning. *Journal of Neuroscience*, 30(33):11178-11184.

Turk-Browne, Nicholas B., Sederberg, P., Simon, M.G. (submitted). Scene representations in parahippocampal context depend on temporal context.

Appendix

Fig.1 - Memory Performance RCNI vs. NCNI - Original Experiment

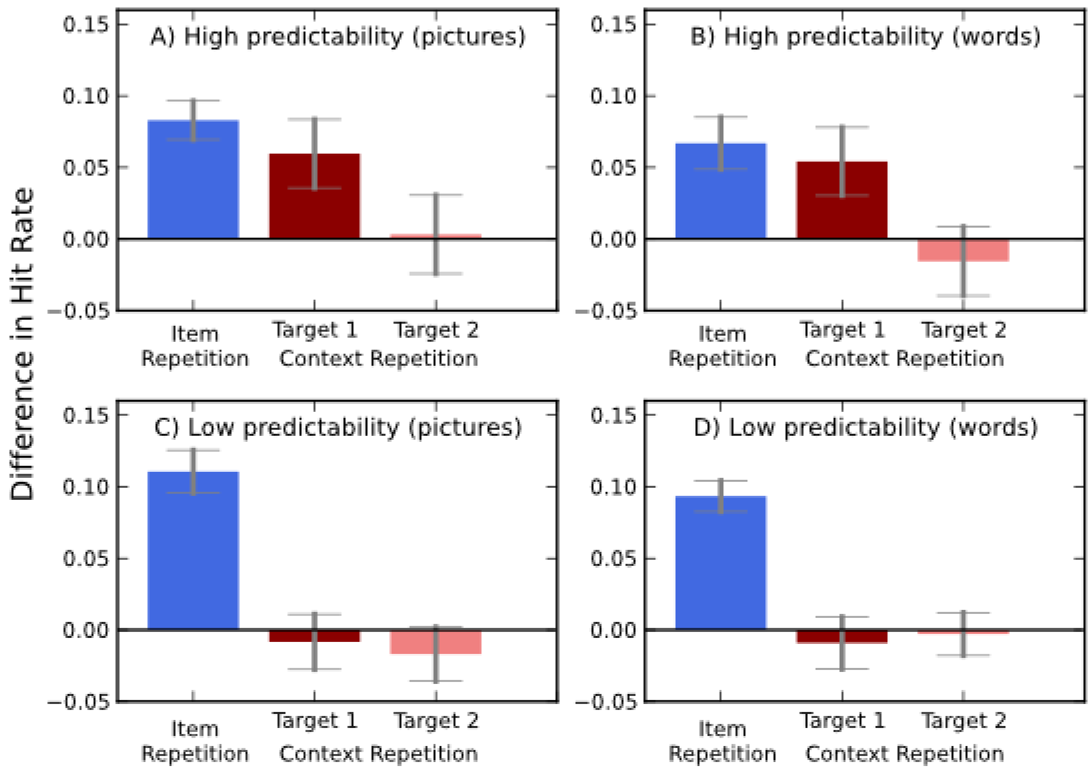


Fig. 2 - Structure of RCRI-112 Condition within Study Phase - Experiments 1

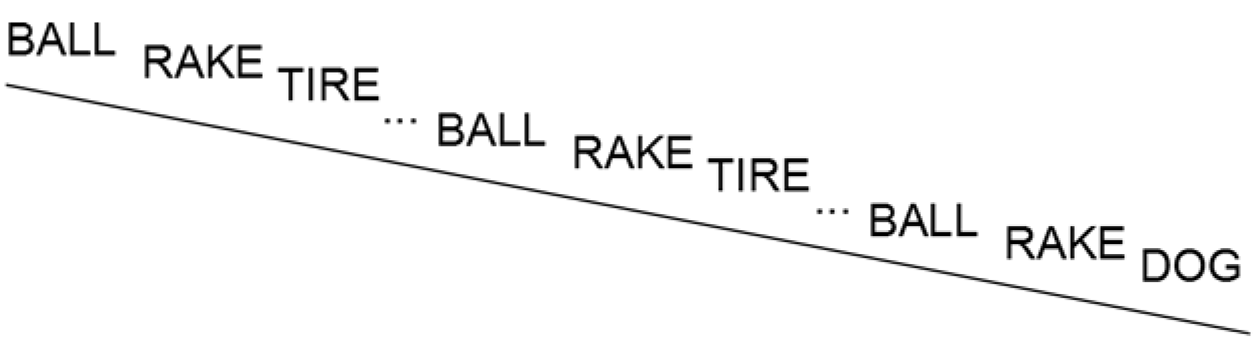


Fig. 3 - Single Trial of Study Phase - Experiments 1

DOG

Living - J   Nonliving - K

Fig. 4 - Single Trial of Test Phase - Experiments 1

DOG

Definitely OLD - J   Possibly OLD - K   Possibly NEW - L   Definitely New - ;

Fig. 5 - Memory Performance RCNI vs. NCNI - Experiment 1

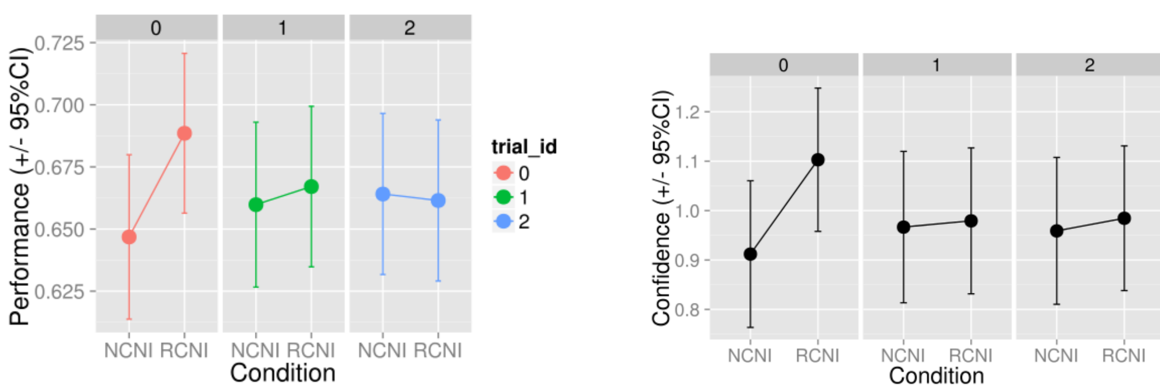


Fig. 6 - Memory Performance RCRI-112 vs. NCNI - Experiment 1

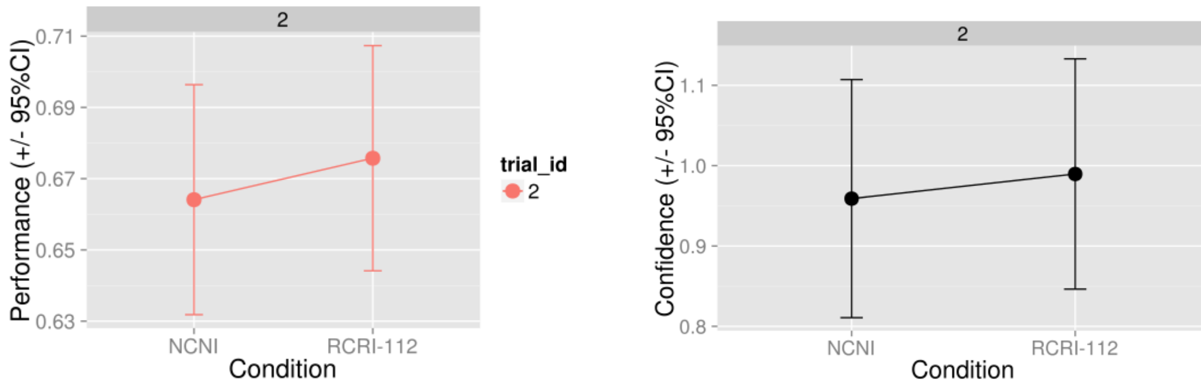


Fig. 7 - Memory Performance RCRI-121 vs. NCNI - Experiment 1

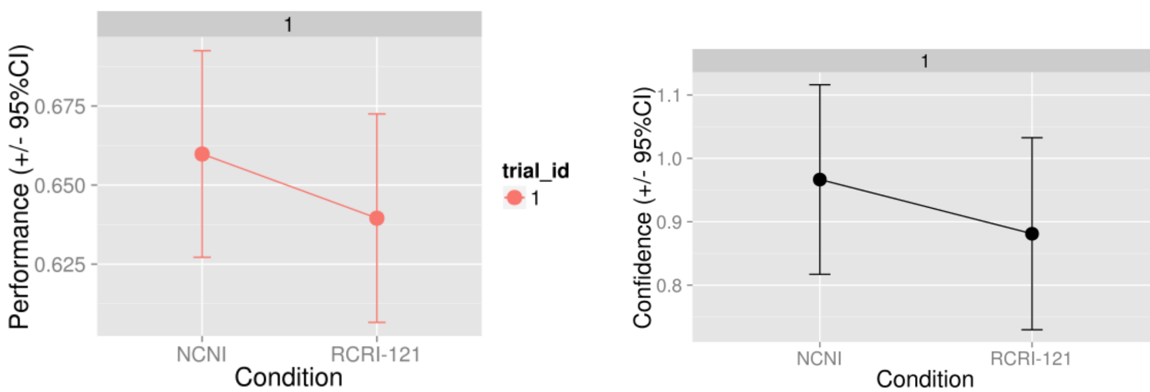


Fig. 8 - Sample Sequence from Study Phase - Experiments 2 & 3

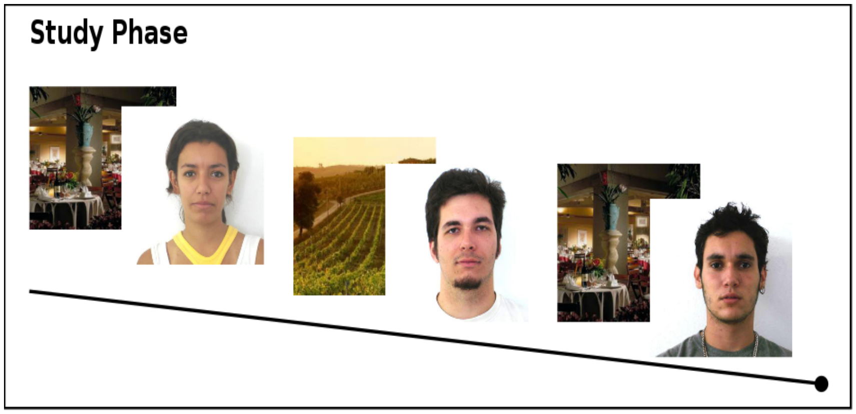


Fig. 9 - Sample Trial from Study Phase - Experiments 2 & 3

Indoor - J Outdoor - K Male - J Female - K

Fig. 10 - Sample Sequence from Test Phase - Experiments 2 & 3

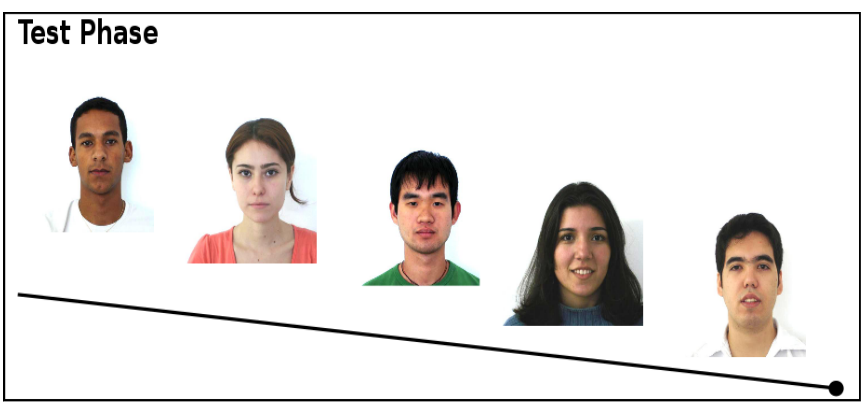
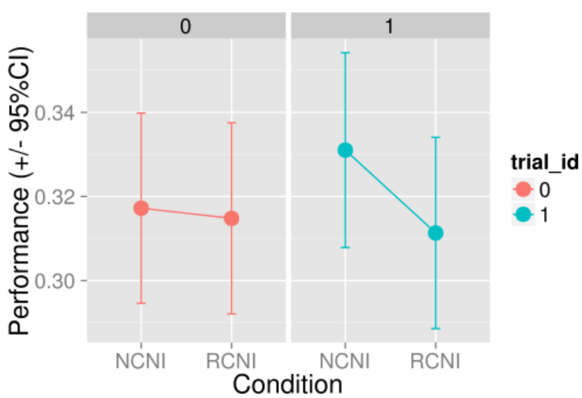


Fig. 11 - Sample Trial from Test Phase - Experiments 2 & 3

Definitely OLD - J Possibly OLD - K Possibly NEW - L Definitely New - ;

Fig. 12 - Memory Performance RCNI vs. NCNI - Experiment 2



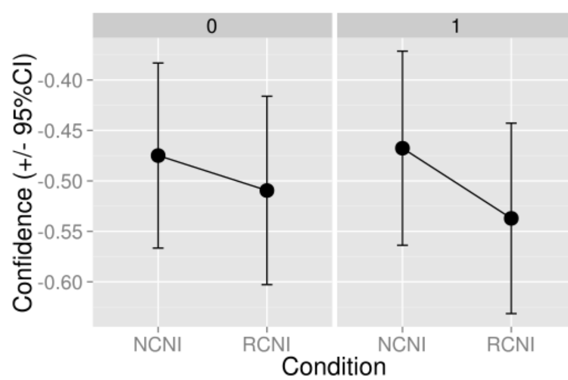


Fig. 13 - Memory Performance RCNI vs. NCNI - Experiment 3

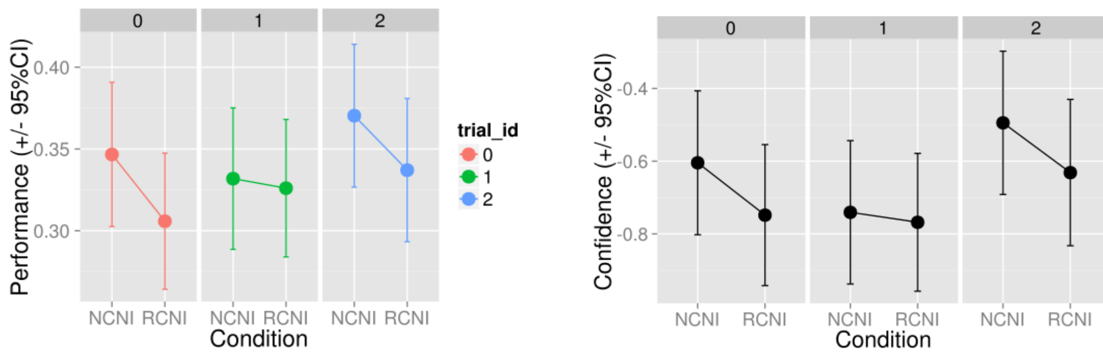


Fig. 14 - Memory Performance RCRI-112 vs. NCNI - Experiment 3

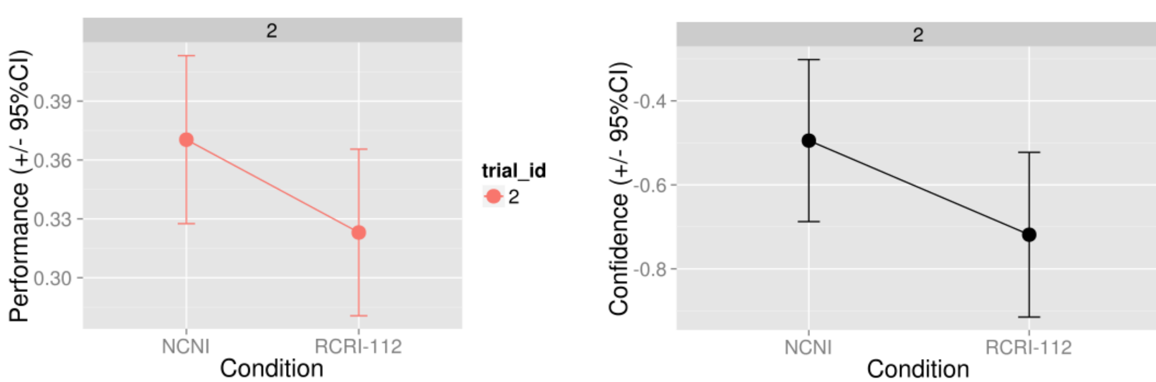
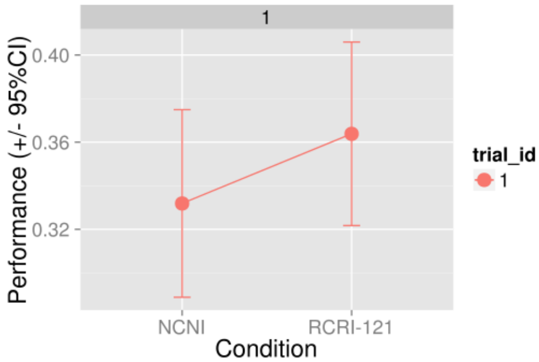
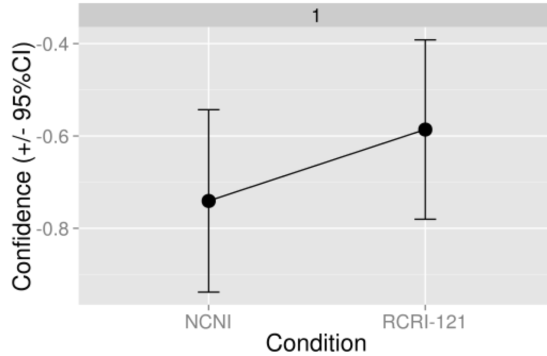


Fig. 15 - Memory Performance RCRI-121 vs. NCNI - Experiment 3


1
Performance (+/- 95%CI)
0.40
0.36
0.32
NCNI
RCRI-121
Condition
trial_id
1


1
Confidence (+/- 95%CI)
-0.4
-0.6
-0.8
NCNI
RCRI-121
Condition