UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

EXAMINING WHETHER COLLABORATION CAN IMPROVE MEMORY FOR BINDING OF REAL-WORLD OBJECT FEATURES

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

In partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

JESSICA BECERRA

Norman, Oklahoma

2023

EXAMINING WHETHER COLLABORATION CAN IMPROVE MEMORY FOR BINDING OF REAL-WORLD OBJECT FEATURES

A DISSERTATION APPROVED FOR THE

DEPARTMENT OF PSYCHOLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Scott Gronlund, Chair

Dr. Lauren Ethridge

Dr. Hairong Song

Dr. Ziho Kang

Acknowledgements

I would first like to extend my sincere appreciation to my advisor, Dr. Scott Gronlund, for his support, guidance, and mentorship throughout this journey. I am deeply grateful for his contribution to my academic and personal growth. I would also like to thank Mackenzie Riggenbach, Samantha England, Miguel Castillo, Tessa Giammona, Claudia and Perla Avalos, and Dayana Olvera, who have supported me endlessly throughout this endeavor. Finally, I would like to thank my beloved parents, Maria and Fidencio, and my sister, Rosa. Their love, support, and sacrifices were my driving force throughout this journey. Thank you.

Table of Contents

Abstract

Individuals can accurately store thousands of objects in their visual long-term memory. However, when objects vary on numerous features, previous research found that individuals struggle to bind the objects to their correct states (e.g., state of the studied coffee mug: full or empty). We tested whether collaboration could serve to overcome chance-level exemplar-state binding by conducting three recognition memory experiments. In Experiment 1A, participants completed 2-AFC tests, in which they had to identify either which exemplars or which exemplar-state conjunctions they had studied. Similar to previous research, we found that when participants needed to identify the exemplar-state information together, they struggled to bind this information and performed near-chance performance. In Experiment 1B, we used a within-subject design and tested whether collaboration could enhance memory for exemplar-state binding at retrieval. To accommodate our design, we divided each task into two blocks, cutting each task in half. We found that participants who remembered individually, and those who worked collaboratively, demonstrated the ability to remember exemplars and the states of exemplars they studied. Surprisingly, they were able to successfully remember this information as a bound unit. In Experiment 2, we tested whether we could replicate this ability to successfully bind when the task becomes more challenging. Using an old/new recognition test, we found that participants who collaborated were able to discriminate above chance performance for both tasks. Thus, we found evidence that exemplar-state binding is possible by individuals who remember individually and that binding performance can be improved when individuals collaborate to remember. However, it seems apparent that the amount of information participants are required to bind impacts this ability.

Keywords: collaborative memory, binding, error-pruning, recognition memory

Examining Whether Collaboration Can Improve Memory for Binding of Real-World Object Features

People are able to store thousands of objects in visual long-term memory with high accuracy (Brady, Konkle, Gill, Oliva, & Alvarez, 2013; Utochkin & Brady, 2019), despite objects being multidimensional, varying on numerous features. How objects and their features are stored in and retrieved from memory has been a longstanding question (Gauthier & Tarr, 2016; Utochkin & Brady, 2019). Are the multiple features making up an object stored as a bounded entity (perceived as a single unit) or are the component features each stored independently? For example, do I have a single memory of an object (a coffee mug that is full) or do I have a collection of memories of the features that make up an object (a memory of the coffee mug plus a memory of the state of that mug—full)?

One way to examine our memory of objects is by testing recognition memory. Recognition memory refers to our ability to judge whether we have been exposed to an object, person, or event, previously. Many dual-process models have been proposed to explain recognition memory (e.g., Atkinson & Juola, 1973; Jacoby, 1991; Jacoby & Dallas, 1981; Mandler, 1980; Wixted & Mickes, 2010; Yonelinas, 1994). These models assume that recognition memory judgments reflect two retrieval processes: familiarity and recollection (Yonelinas, Aly, Wang, & Koen, 2010). Familiarity refers to the feeling that something has been previously encountered whereas recollection refers to retrieving the contextual details about the past experience (Yonelinas, 2002). Research has found that recollection typically requires more time and cognitive capacity compared to familiarity (Diana, Yonelinas, & Ranganath, 2007) and that recollection is critical to the retrieval of bound information compared to familiarity (Daselaar & Cabeza, 2014).

Memory binding refers to an individual's ability to bind different features of an object, an event, or a person, into one coherent representation (Daselaar & Cabeza, 2014). When we perceive the world, we are constantly exposed to complex objects with varying features (e.g., shapes, size, colors), and we are faced with the challenge of keeping track of which features go with which object. That is, we must keep track of how an object's features are combined (e.g., a jacket that is pink), which requires identifying whether features belong to object A vs. object B when we have multiple objects in view (Treisman, 1996). This problem of sorting and keeping track of features does not only relate to perception but is a general problem of cognitive processing (Zimmer, Mecklinger, & Lindenberger, 2012). For example, when individuals are thinking and remembering information regarding different events they experience, they must consider which features of an event are bound to others and which are separated from features that belong to a different event (Zimmer et al., 2012). Binding is critical during both encoding and retrieval (Zimmer et al., 2012). A stimulus can be encoded differently depending on how it is processed (Diana et al., 2007). If memory components are not bound at encoding, the features must be re-bound during the retrieval process (Nader, 2003).

Research on binding in long-term memory has examined simple features such as the color of an item, and demonstrated that, for example, the color of an item maintained some independence in long-term memory (Chalfonte & Johnson, 1996; Hicks & Starns, 2015). However, the featural relationships amongst the stimuli used in these studies were arbitrary, with no semantic relationship to one another. Brady et al. (2013) and Utochkin and Brady (2019) made the important point that real-world objects possess semantically meaningful featural relationships (e.g., whether a cookie is whole or a bite is missing) rather than arbitrary featural relationships (e.g., a word and the color the word is printed in). Brady, Konkle, Alvarez, et al.

(2013) studied meaningful semantic features of objects using exemplars and the state of the exemplar (e.g., which cup of coffee did you see, was the cup full or empty) and found that different features were forgotten separately over time and that forgetting one feature did not always mean you forgot the other feature. For example, you may forget which cup you saw but not forget that whatever cup you saw was full. Their findings suggest that the features of real-world objects are stored independently rather than as a single bound unit, or at least that the features of objects are forgotten independently, which means that an object is not stored as a single unit but rather as a collection of features.

Recent research on real-world objects also has found that memory of object exemplars and their states are not bound to one another. Utochkin and Brady (2019) tested whether different features of objects were stored independently by having participants study both exemplars and exemplar-states of real-world objects. They had participants complete two tasks in a randomized order. During the exemplar task, exemplars were studied and participants had to complete a recognition test where they needed to identify which exemplar they studied (e.g., I saw this chair). During the exemplar-state task, exemplars (e.g., two different cooking pots, two different doors) were studied either in the same state or in different states (e.g., lid on/lid off, door open/door closed), and at test participants needed to identify which exemplar-state combination they had studied (e.g., I saw this cooking pot and the lid was off), which required both the exemplar and the state of the object to be retrieved. During the test phase, each test trial consisted of four objects, with two objects on the top row and two objects on the bottom row; each row required a 2-alternative forced-choice (2AFC) decision. In the exemplar task, participants had to select the two exemplars (one from each row) from each category that they had studied. In the exemplar-state task, participants had to select the correct state for each exemplar from each category they had studied. In both their Experiments 1A and 1B, they found that participants had a strong memory of the object exemplars they had studied and a strong memory of whether the state of the two object exemplars were displayed in the same or different states. However, when participants needed to remember which state went with which object exemplar, they struggled to bind the state to its exemplar information and their performance was near chance (Utochkin & Brady, 2019). This suggests that real-world objects (the exemplar and its state) are independently stored rather than bound holistically.

Utochkin and Brady's (2019) study design is the foundation for our current study. The current study will extend their work by examining whether collaboration can serve to overcome chance-level exemplar-state binding. We review next why we think that it will.

Recognition Memory and Collaboration

Previous research has found that collaborating to remember improves recognition memory performance. This advantage stems from individuals being able to discuss what they remember and reviewing whether a stimulus was previously studied. Consequently, the goal of the present research is to explore if collaboration can enhance memory performance for exemplar-state binding at retrieval. Collaborating with another individual should allow for the dynamics during retrieval to facilitate what is remembered. For example, person A may remember the exemplar that was studied, but person B may remember the state of that exemplar. Thereby, the memories of the two collaborators can ascertain the correct exemplar and its state during the test phase.

Previous research found that recognition accuracy is better for groups compared to those working alone (Hinsz, 1990; Vollrath, Sheppard, Hinsz, & Davis, 1989). The study designs used to examine the influence that collaboration has on recognition memory vary depending on the researcher's question. Common questions regarding collaboration and recognition memory are: How does collaborating-to-remember impact individual recognition decisions? How do individual recognition decisions differ compared to group decisions (where the groups consist of individuals who first make a decision alone and then work with others to make a final group decision)? Lastly, how do groups make their group decisions (e.g., majority vote, reliance on groups' best member)? The main difference among the study designs is whether there is an individual recognition decision prior to a collaborative decision. For instance, some studies have the individuals assigned to the collaborative condition make an individual recognition decision prior to making a collaborative recognition decision. Conversely, other studies have the collaborative groups discuss the tested stimuli with one another, but after this discussion, the researchers collect individual recognition decisions rather than a group decision. Next, we review the literature regarding collaboration and recognition memory, discussing reasons why collaborating can be beneficial.

The question of whether a recognition advantage is due to collaborative or non-collaborative group processes was first studied by Clark, Hori, Putnam, and Martin (2000). This question arose from the free-recall literature, which found that the recall of nominal groups exceed collaborative groups (Basden, Basden, Bryner, & Thomas, 1997). Nominal groups are created by pooling the reports of same-size sets of individuals but counting the items reported by multiple individuals only once. Because the recall of nominal groups exceeds collaborative groups, it is clear that the nominal group *recall* advantage arises from pooling performance (i.e., noncollaborative) rather than collaborative processes.

Clark et al. (2000) examined whether the advantage found in recognition is due to the benefits of collaborating (e.g., discussing amongst group members) or if groups are resource pooling (i.e., majority vote, reliance on groups' best members), which would imply that non-collaborative aggregation was responsible for the advantage. To study this, Clark et al. (2000, Experiment 1) had participants in the collaborative condition first make an individual recognition decision followed by a collaborative recognition decision. The collaborative condition consisted of groups of either two or three participants; participants in the noncollaborative condition worked alone. In the collaborative condition, each individual first made his or her own old-new recognition decision about each item, and then the group made a single collaborative decision about each item. In the noncollaborative condition, participants worked alone to complete the same recognition task. The collaborative recognition decision was compared with 1) the average group member's original responses, 2) the majority rule (only for groups of 3), and 3) the group's best member's decision. Collaborative groups performed better than individual participants in the noncollaborative condition; the larger the group the better they did. Because this study examined the majority votes for groups and for the group's best member's decision, Clark et al. concluded that the advantage in recognition memory for groups was not just because the groups went with the majority answer nor because the groups went with the best group member's response. Specifically, the collaborative advantage was rooted in an increase in hit rates (previously studied items that the group correctly reported 'old'). That is, the presence of the test item in recognition facilitates performance when one group member can make a compelling argument about a test item that was previously studied, perhaps cuing the other individuals' memory. Clark et al. (2000) pointed out that composing an argument regarding an item that was *not* studied is much more difficult.

Other researchers have examined how prior collaboration influences individual recognition memory when a consensus decision is not required (Rajaram & Pereira-Pasarin, 2007). To study this, researchers group participants in the collaborative condition into groups of three individuals for the test phase and ask them to discuss the studied items. Following the discussion of each item, each group member made their own individual decision. These participants were instructed that the researchers were most interested in their individual responses even if their responses were contrary to the group. They found that the collaborative condition resulted in better individual recognition performance than those in the noncollaborative condition. A second experiment included test delays of 1h, 48 hrs., and 1 week and found that collaboration again led to better recognition performance compared to those in the noncollaborative condition. These findings demonstrated that group consensus is not required to generate the benefit of collaboration on individual recognition (Rajaram & Pereira-Pasarin, 2007).

A major concern of individuals collaborating to remember is that inaccurate information may be reported, which could contribute to memory errors being passed along and contaminating other groups member's memory. This phenomenon is referred to as social contagion (Meade & Roediger, 2002; Roediger, Meade, & Bergman, 2001). When participants are forced to take turns during recall, without being able to discuss with their partner, participants make more errors because there is no way to correct each other's errors (Rajaram, Maswood, & Pereira-Pasarin, 2020; Rajaram & Pereira-Pasarin, 2010). However, when discussions are allowed, this can reduce the number of errors (Rajaram & Pereira-Pasarin, 2010). The benefit of reducing memory errors is believed to occur because of the opportunity to error prune: listening and discussing with other group members allows an individual to receive feedback and prune out errors that they might have otherwise reported (Barber, Rajaram, & Aron, 2010).

Over the years, researchers have manipulated the type of stimuli presented and have found that certain types of stimuli facilitate error pruning while others hamper it. Studies that use unrelated lures or categorically-related lures have found support for error pruning (Rajaram, 2011; Rajaram et al., 2020; Rajaram & Pereira-Pasarin, 2007). These types of lures do not tend to activate the same false memory from person to person. For example, if participants study a list of fruits (e.g., strawberry, lemon, pear), it is not likely that both participants will share the same false memory lure item (e.g., mango), which would make it more likely for a lure to be rejected if only one person from the group remembers it (Rajaram et al., 2020). In contrast, error pruning is less likely when studies use stimuli that increase false memories, such as those involving semantically-related words (e.g., seat, stool, sit) that align with a single critical lure (chair) (Deese, 1959; Roediger & McDermott, 1995). These word lists tend to induce increased false alarms for individual recognition tasks despite collaboration (Basden, Reysen, & Basden, 2002); if all group members share the same false memory of a critical lure (Rajaram & Pereira-Pasarin, 2010), it is likely that the group will fail to reject the critical lure.

Given what we know about recognition memory and collaboration, we will next consider how collaboration might enhance binding at retrieval. For starters, it is known that binding is a cognitively demanding task in which recollection is vital. When two individuals collaborate, the cognitive load can be shared between the pair while simultaneously doubling the cognitive resources available to the group. During recognition tasks, group members can use recollection to justify their decision to their partner, which can facilitate the binding process. Discussing the context for a memory of an item can help collaborators identify the correct item and reject the wrong item, especially when the task is harder (e.g., exemplar-state). For example, if both Partner A and B provide context to support their memory for the same item, it makes a correct decision more likely. Conversely, individuals working alone may not always have the cognitive resources to invest when attempting to retrieve the memory of exemplars and their correct states.

Collaboration has been found to aid recognition memory through error-pruning. During the exemplar-state task, participants must identify both the correct exemplar and the state in which it was studied. This requires participants to bind those features together; mistakenly remembering either the exemplar or the state results in unsuccessful binding. When individuals collaborate, they are given the opportunity to have discussions with their partners and provide feedback on each other's decisions and this should improve binding performance. To illustrate, consider an individual working alone on the exemplar-state task. If the individual working alone erroneously selects an exemplar or an erroneous state, that trial is incorrect, even if one feature was correct. Now consider a collaborative pair in which both individuals agree on an exemplar, but one collaborator mistakenly remembers the wrong state. If the collaborator who erroneously remembered the state was alone, they would have missed this binding trial. However, because they have a partner who recollects the correct state, the pair is able to achieve the correct binding decision.

Current Studies

The current studies are motivated by Utochkin and Brady (2019, Experiments 1 and 2). In Experiment 1A, we sought to replicate Utochkin and Brady's (2019) Experiment 1A. We made the decision to run a replication after it became apparent that our modified version of the experiment (our Experiment 1B) resulted in above chance binding performance. We did this to ensure we were following the original Utochkin and Brady experiment correctly and did not make any mistake in implementing their procedure. We present our replication study first for ease of interpretation. In Experiment 1B, we will test whether collaborating can improve binding performance. Lastly, we should mention that one additional decision was collected from participants. In our Experiments 1A and 1B, participants provide a confidence judgment after each individual recognition decision; confidence judgments were not collected in the original studies (Utochkin & Brady, 2019, Experiments 1 and 2). Therefore, we acknowledge this modification in our methodology and note that our studies are not exact replications.

Experiment 1A

Method

Participants

To determine the sample size, a statistical power analysis using GPower 3.0.10 (Faul, Erdfelder, Lang, & Buchner, 2007) was performed. We estimated sample size using the same effect size estimates used by Utochkin and Brady's (2019), which was .67 for a medium effect. Their study consisted of 20 participants. Participants in our study consisted of 19 undergraduate students at the University of Oklahoma participating in exchange for SONA credit or a $10 Amazon gift card. There were 16 females and 3 males (M = 28.21, SD = 7.01). All participants gave written consent and were debriefed following the completion of the study.

Materials

Qualtrics was used to present the stimuli to participants and for the completion of all tasks. There will be two image sets used in this experiment. The first set of 480 images to be used for the exemplar-state task come from Brady et al. (2013) and contains 100 different object categories. Twenty additional categories with images was found via a Google search. Each of the categories contains two exemplars (e.g., two different cabinets) that are each in two different states (each cabinet with the doors open and doors closed). The states varied for each category as there was a wide variety of combinations (e.g, open/closed, whole donut vs. bite taken from donut, full coffee cup vs. empty coffee cup). The second set of 480 images to be used for the exemplar task came from Konkle, Brady, Alvarez, and Oliva (2010). The 120 categories had no overlap with the categories from the exemplar-state task. Each of the categories contained four variations but had no state differences (e.g., four different coffee cups, all full).

Procedure

Participants arrived at the lab and were seated at a computer desk where they completed all of the tasks. Participants were first shown the study stimuli and asked to memorize them. Then during the test phase, they completed a series of two 2-AFC tests. Participants were asked to identify which items they had previously studied during each trial accompanied by a confidence judgment (using a scale with the options remember, know, guess). Remember, guess, know judgments can be used as a subjective confidence index, representing the participant's level of belief in their own memory (Dunn, 2004). After completing the first task, participants were shown the other set of study stimuli and completed the second set of 2-AFC tests. Each of the tasks will next be explained in detail.

Exemplar Task

The exemplar task is included to verify that participants are able to successfully remember information about the exemplars they studied. Otherwise, it would be impossible to determine whether poor binding performance results from an inability to bind or is due to an inability to remember exemplars. During the study phase, participants were shown 120 categories with two images per category (e.g., two different wheelbarrows). The study phase of the exemplar task is depicted in the top left panel of Figure 1. Each image was displayed for 2 seconds followed by a 1-second blank screen. During the test phase, participants were presented with four exemplars (e.g., two previously studied pies and two never before seen pies) from each category and were instructed to choose the exemplar from each row that they previously studied. The top right panel of Figure 1 shows the test phase with the four options for each exemplar from which participants need to identify which two pies, and which two wheelbarrows, they studied previously. One object in each row always had been previously studied.

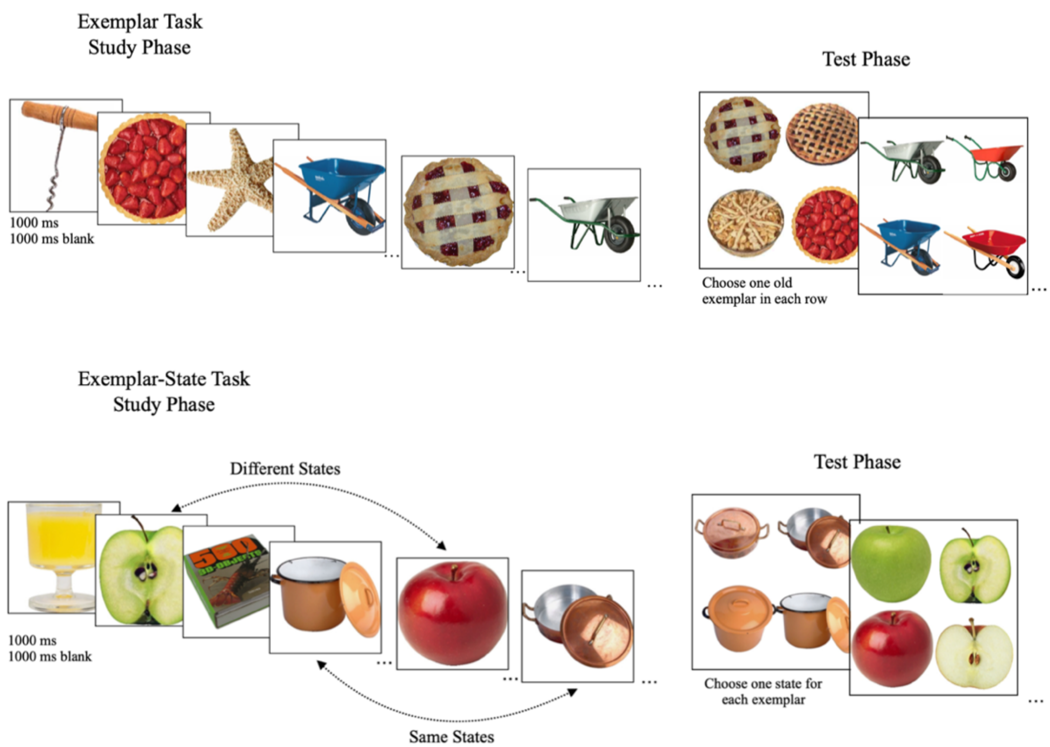

Figure 1. During the *exemplar task* (top panels), participants memorize a set of objects that include two object exemplars from each category (e.g., two pies). During the test phase, participants complete two 2-AFC tasks, one for each row, and must choose the exemplar they had previously studied, one exemplar from each row. During the *exemplar state task* (bottom panels), participants memorize a set of objects that include two object exemplars from each category. The object exemplars from the same category will either be presented in the same or different states (two cooking pots with lids off, half of an apple and a whole apple, respectively). During the test phase, participants will complete two 2-AFC tasks by selecting the correct state, selecting one object from each row. This procedure outline and images taken from Figure 1 in Utochkin and Brady (2019, p.5).

Exemplar-state task

The exemplar-state task also had a study phase followed by a test phase. During the study phase (bottom left panel of Figure 2), participants were shown 240 images of exemplars in various states, with each image shown for 2 seconds followed by a 1-second blank screen. The two object exemplars from a given category were displayed in either the same state (e.g., cooking pot A and B both have the lid off) or in two different states (e.g., apple A is cut in half and apple B is whole). Whether the states were the same or different was randomized for each condition. We also randomized across stimulus categories ensuring the studied exemplar-state varied across participants. During the test phase, participants completed two 2-AFC tasks per trial. Each trial contained both studied exemplars and the two different states (i.e., 4 exemplars were displayed, two on the top row and two on the bottom row). The participants had to select the correct state for each exemplar that they studied. Utochkin and Brady (2019) argued that presenting both studied exemplars in the same trial limits the possibility of swap errors where participants retrieve one exemplar from a given category but then mistake its state with the state of the other exemplar from that category.

Analysis Plan

The dependent measures are accuracy of remembering exemplars, the accuracy of remembering the state, and the accuracy of exemplar-state binding. The memory for same states and different states is estimated by a participant's ability to correctly indicate whether the two exemplars from a category were in the same or different states (e.g., two different cabinets both with open or closed doors, or two different sneakers, one with tied shoelaces and one with untied shoelaces). We compared these accuracies to chance performance (50%).

To measure the boundedness between state and exemplar memories, we computed how often the participant's state responses were correct for each exemplar in the exemplar-state task (e.g., they studied a green apple that was cut in half and correctly identified the green apple and that it was cut in half, see Figure 2). If exemplars and states are bound, participants can correctly pick the correct state for each exemplar as often as they remember the state information overall. Evidence in support of the state being unbound from the object would occur if participants were unable to correctly pick the correct state for each exemplar when the states they originally studied were presented in different states, because studying the exemplars in different states would require successful binding at retrieval. Moreover, if the participants only had a memory of the object and not for the state of the object (or vice versa), this would indicate the memory of the object and the state are independent of the other.

We conducted *t*-tests to evaluate these effects and used JASP software to conduct the analysis. We also report Cohen's *d*'s and d_z's as estimates of effect size with their 95% confidence intervals.

Results

The current data violated the normality assumption, therefore, a logit transformation was applied to the data. The logit transformation fixed the nonnormality in all of the data. Here we report the results from the *t*-tests but include the means and standard deviations from the raw data for ease of interpretation.

Accuracy in remembering exemplars. During the Exemplar condition, when asked to remember exemplars without requiring state memory, accuracy was high and above chance, M = .80, SD = .13, $t(18)$ = 4.71, p < .001, d = 1.08, 95% CI [.50, 1.64], indicating participants remembered exemplar information when it was not required to be bound to state information. Table 1 shows the proportion correct for each of the tasks.

Accuracy in remembering state. In order to determine how well participants remembered state information on its own, we examined performance in picking the correct state when both objects were shown in the same state. When both objects in a category were shown in the same state, participants were above chance at choosing this state, $M = .67$, $SD = .20$, $t(18) = 3.61$, $p = .002$, $d = .83$, 95% CI [.30, 1.34], which indicates that participants were good at remembering states of objects when binding was not required. Additionally, participants also were good at discriminating between the objects in the same state versus in different states. Participants selected the same state for two exemplars more often when the two exemplars were indeed studied in the same states (same states: $M = .76$, $SD = .18$; different states: $M = .36$, $SD = .19$)(see Figure 2B); paired samples *t*-test: $t(18) = 6.42$, $p < .001$, $d = 1.47$, 95% CI [.80, 2.12].

Accuracy in exemplar-state binding. Given that participants had a good memory for exemplar information and a good memory for state information independently, we can now examine whether exemplar information and state information were bound. Participants were significantly worse at remembering the state of each exemplar when the exemplars were shown in different states ($M = .58$, $SD = .18$) than in the same state ($M = .67$, $SD = .20$); paired samples *t*-test comparison: $t(18) = 1.94$, $p = .069$, $d = .44$, 95% CI [.03, .91] (see Figure 2A). Participants were no better than chance ($M = .58$, $SD = .18$, vs. .50) when selecting states of exemplars that had been studied in different states; (0.50; one-sample $t(18) = -.87$, $p = .396$, $d = -.20$, 95% CI [-.65, .26], as compared with the same state condition, reported above, which was significantly better than chance. Finally, on .30 of the trials, participants reported choosing one of two correct states, which did not differ from the proportion of selecting one correct answer in the same state condition, *M* = .25, *SD* = .17, *t*(18) = 1.16, *p* = .261, *d* = .23, 95% CI [.20, .72] (see Figure 2C). In other words, the difference in overall accuracy between the same state and different states was the result of the capacity to accurately report both or neither of the states correctly.

Table 1

Experiment 1A Proportion correct in the two tasks

	M
Exemplar task	
Exemplar correct	.80(.13)
Exemplar-state task	
Correct same	.67(.20)
Correct different	.58(.18)
Reported two same states; studied two same states	.76(.18)
Reported two same states; studied two different states	.36(.19)

Note. The standard deviation is provided in parentheses.

Experiment 1A

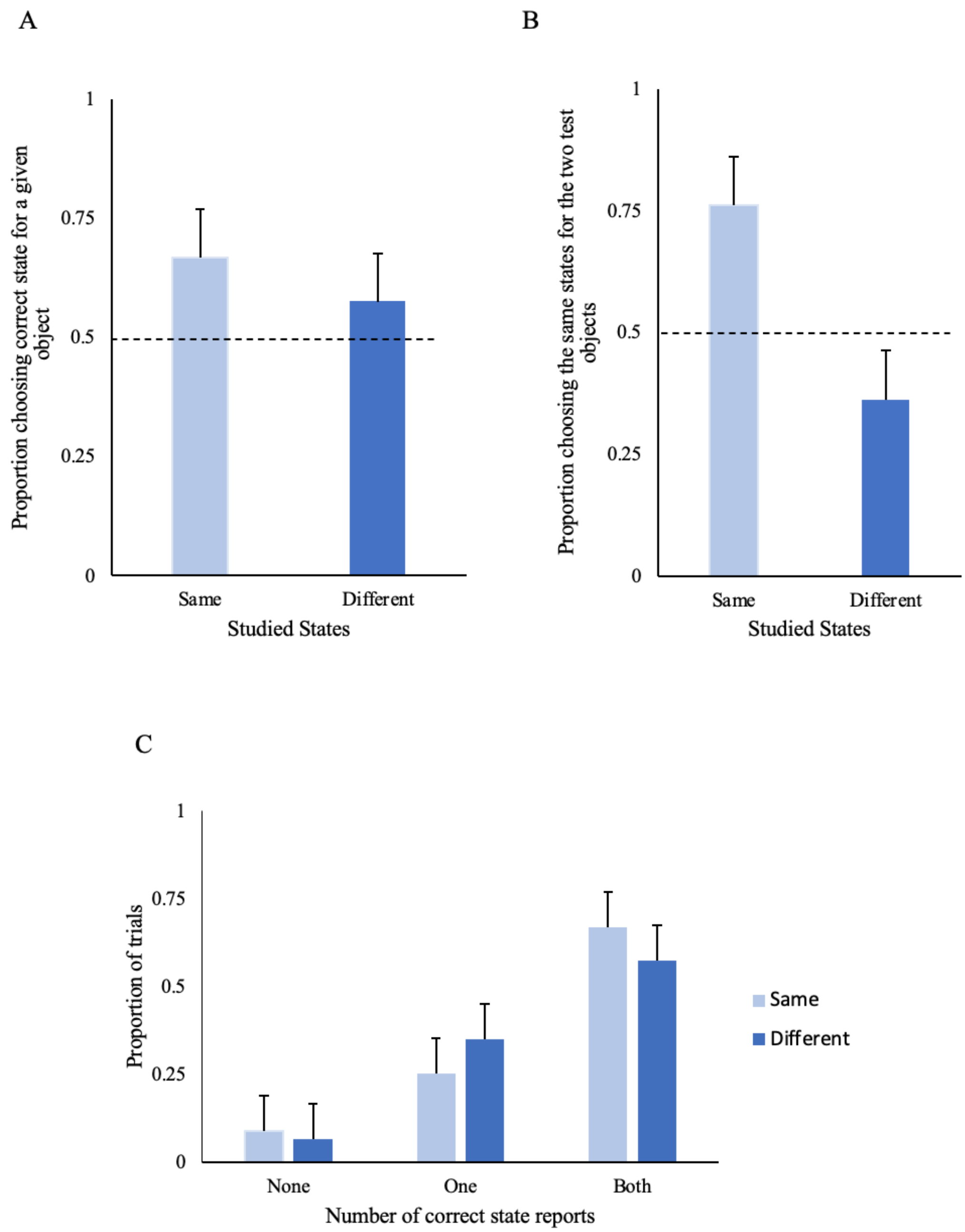

Figure 2. (A) The proportions selecting the correct state for a given exemplar when the two studied objects were shown in the same state (left-hand bar in panel A ; no binding required) or different states (right-hand bar in panel A; binding required); (B) The proportions selecting the same states for the two test objects (regardless of whether these states are correct or incorrect). The dashed lines show chance levels. (C) The proportion of participants in a test trial who reported both correct states, one correct state, or no correct states, as a function of the study condition.

Confidence judgments. The proportion correct for remember, know, and guess judgments were computed for each task. For the exemplar task, participants were correct 86% when they reported remembering the exemplar, 86% when they reported knowing they saw the exemplar, and 70% when they guessed the exemplar. For the exemplar-state task, participants were correct: 87% when they reported remembering the exemplar-state, 79% when they reported knowing they saw the exemplar-state, and 59% when they guessed the exemplar-state, see Figure 3.

Experiment 1A

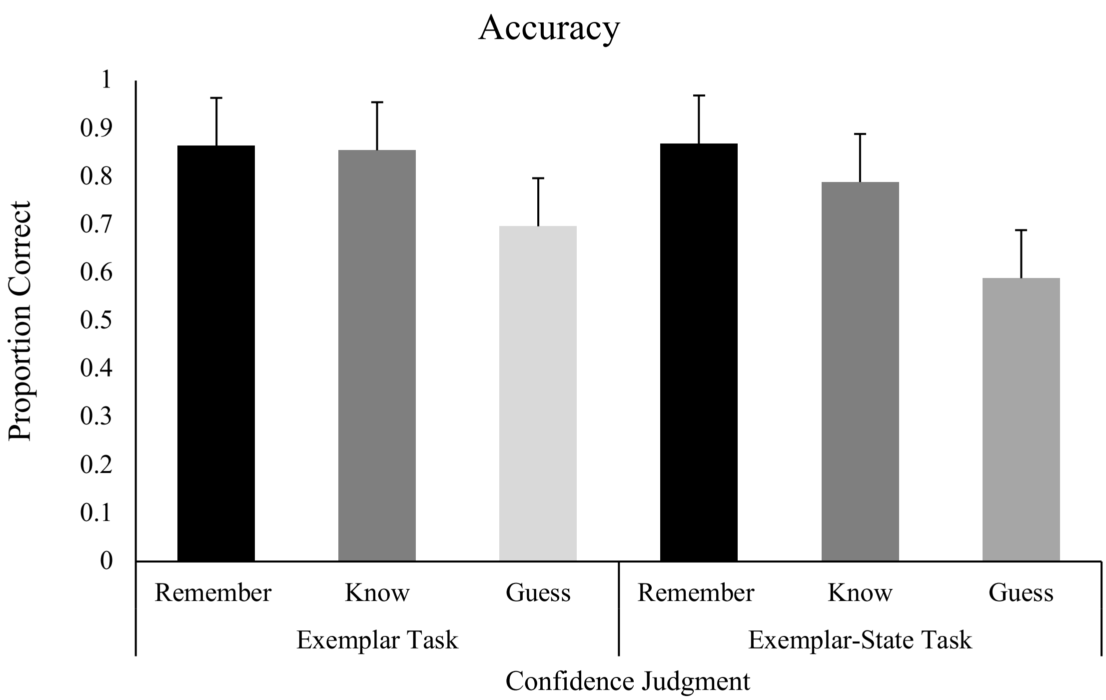

Figure 3. 2-AFC accuracy for remember, know, and guess judgment (left panel, exemplar task; right panel, exemplar-state task).

Thus, our study replicates the finding of Utochkin and Brady (2019), indicating that individuals struggle to bind exemplar information with state information. Now that we have replicated the original study's findings, we will proceed to address our main question of interest, whether collaboration can help to overcome chance-level, exemplar-state binding.

Experiment 1B

Next, we tested whether collaborating could improve binding performance. We expected that individuals remembering alone would perform at chance levels when required to correctly bind which state goes with which exemplar, replicating our Experiment 1A (and Utochkin & Brady, 2019, Experiments 1A and 1B). Of primary interest, we examined whether working with a partner to make recognition memory judgments could enhance performance in binding states with exemplars. Given what is known about collaboration and recognition memory, we hypothesized that working with a partner during the exemplar-state recognition test would improve performance in terms of tying which state corresponds to which exemplar. Collaborating with a partner allows for discussion as to whether either individual recognized an exemplar and its particular state, while also being able to provide the context of their memory to strengthen their argument (e.g., "I remember we saw the chocolate that had a bite missing because my sister used to always leave half-eaten chocolates in the box."). Collaborators could also correct and challenge each other's incorrect decisions, which could help collaborating pairs error-prune exemplars that were not studied in the state-exemplar task and/or to identify more correct studied objects (Rajaram et al., 2020). Specifically, we expected to find that recognition memory accuracy for bound information would be greater for those who collaborate compared to those who perform the task individually. As for the exemplar task, we also expected to find that the collaborative condition would outperform the individual condition in identifying previously studied exemplars.

If collaboration improves the ability to bind exemplars to their correct states, we will attempt to understand how the group made its decision. Previous research has examined how often collaborative groups go with the decision of their best group member (Clark et al., 2000). If we find a collaborative advantage, we will undertake similar analyses.

Method

Participants

Following the power analysis described in Experiment 1A, a different group of 42 undergraduate students at the University of Oklahoma participated in exchange for SONA credit or a $10 Amazon gift card. There were 30 females, 11 males, and 1 non-binary individual (M = 22.5, SD = 9.24). The alone and noncollaborative conditions contained all 42 individuals; however, when they made collaborative conditions, they were with a partner, resulting in 21 collaborative pairs in total (21*2 to make a pair).

Design

The independent variables in this study are condition: individual and collaborative, and task: exemplar and exemplar-state. Both independent variables follow a within-subject design.

Materials

The stimulus materials were the same as Experiment 1A.

Procedure

Study sign-ups requested two participants per session. When participants arrived at the lab, they were taken to a computer desk where they completed all of the tasks. The two participants sat next to each other with a cubicle divider between the desks. Participants were shown the study stimuli and were asked to memorize them.

During the test phase, they completed a series of two 2-AFC tests, just as in Experiment 1A. However, this time, each task was divided into two blocks to allow for the within-subject design (individual and collaborative), making a total of four blocks total (individual exemplar task, collaborative exemplar task, individual exemplar-state task, collaborative exemplar-state task; the order was randomized) see Figure 4. However, we should note that the collaborative condition test trials had to be presented in the same order so that the pair could work on the same exemplar as their partner. When participants were completing an individual condition, they completed all tasks on their own, and when participants were assigned to the collaboration condition, each person first made an individual decision (accompanied by a remember, know, guess confidence judgment) and then the two participants made a collaborative decision for each test.

Following the completion of the first task, all participants were shown the other set of study stimuli and completed the second set of 2-AFC tests. The stimuli used during each block did not repeat exemplars between conditions. Each of the tasks will next be explained in detail.

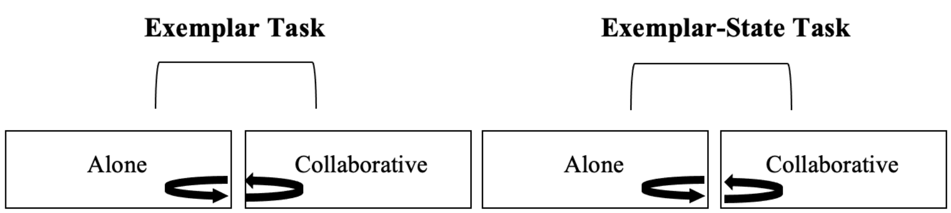

Figure 4. The procedure is shown above for the two tasks. The exemplar task and the exemplar-state tasks each contained two blocks, one block was done alone while the second block was done collaboratively. The alone and collaborative conditions as well as the tasks were randomized throughout the study.

Exemplar Task

The exemplar task was the same as Experiment 1A except that during the study phase, participants were shown 60 categories with two images per category (e.g., two different wheelbarrows) in contrast to the 120 categories studied in Experiment 1A. In other words, Experiment 1B used 60 categories for the alone condition and the other 60 were used for the collaborative condition. The 60 categories were counterbalanced across both conditions.

Exemplar-state task

The exemplar-state task also followed the same procedure as Experiment 1A, however, again, 120 images were used for the individual condition study stimuli and 120 images were used for the collaborative condition study stimuli instead of 240 images.

Analysis Plan

The same analyses were conducted as in Experiment 1A. However, this set of analyses also compares noncollaborative decisions to collaborative decisions. A combination of *t*-tests and mixed ANOVAs were conducted to evaluate the effects, and the conclusions will be interpreted using Bayes factors (e.g., Rouder, Speckman, Sun, Morey, & Iverson, 2009). We expected to find that the collaborative condition would have greater accuracy when remembering exemplars and states, independently, compared to individuals remembering alone. Furthermore, we expect to find that collaborative pairs will perform better than chance when it comes to exemplar-state binding, whereas individuals working alone will still perform at a chance level, as they did in Experiment 1A.

Results

The results are presented in two sections. First, we present the results from the task done alone. Second, the results comparing the collaborative condition decisions, which consist of the noncollaborative decisions and collaborative pair decisions, will be presented.

The current data violated the normality assumption; therefore, a logit transformation was applied to the data. The logit transformation fixed the normality of all of the data. Here we report the original results from the *t*-tests and mixed ANOVAs but report the means and standard deviations from the raw data for ease of interpretation.

Accuracy in remembering exemplars—Alone Condition. During the Exemplar condition, when asked to remember exemplars without requiring state memory, participants in the alone condition performed well above chance, $M = .89$, $SD = .08$, $t(41) = 12.05$, $p < .001$, $d = 1.86$, 95% CI [1.35, 2.36], suggesting they remembered exemplar information when it was not required to be bound to state information. Table 2 shows the proportion correct for each of the tasks.

Accuracy in remembering state—Alone Condition. Similarly to Experiment 1A, we conducted two tests to gauge how well participants were able to remember state information. When both exemplars from a category were displayed in the same state, participants were well above chance at choosing this state, $M = .72$, $SD = .17$, $t(41) = 4.26$, $p < .001$, $d = .66$, 95% CI [.32, .99], which suggests that participants were able to remember states when binding was not necessary. Additionally, participants were also good at discriminating between when the objects were shown in the same state versus in a different state. Notably, the proportion of the time participants chose the same states for the two category exemplars was much higher for the objects that actually were presented in the same states compared to objects presented in different states; paired samples *t*-test: $t(41) = 11.20$, $p < .001$, $d_z = 1.73$, 95% CI [1.24, 2.20]. In both the same and different state conditions, the proportions differed from the chance level of .50, indicating that participants were good at choosing the same correct states when they studied the same states (same states: $M = .79$, $SD = .14$, $t(41) = 7.55$, $p < .001$, $d = 1.17$, 95% CI [.77, 1.55]) and participants were able to correctly discriminate when they were presented in different states: $M = .29$, $SD = .16$, $t(41) = 10.70$, $p < .001$, $d = 1.65$, 95% CI [1.18, 2.11]. Given that participants studied same-state exemplars, it would be expected that reporting different states for these exemplars would be below chance because this the wrong decision, and this is exactly what these results show (see Figure 5B). In sum, we replicated Experiment 1A; participants remembered exemplars and participants remembered states. But is binding still at chance?

Accuracy in exemplar-state binding—Alone Condition. How often did participants correctly remember the state of each exemplar when the exemplars were shown in different states? Our results showed that participants were significantly worse at remembering the state of each exemplar when the exemplars were in different states ($M = .66$, $SD = .19$) than in the same state ($M = .72$, $SD = .17$); paired samples t-test comparison: $t(41) = 2.12$, $p = .040$, $d_z = .33$, 95% CI [.01, .64]; Figure 5A. However, memory for the different state condition was above chance (0.50; one-sample $t(41) = 2.19$, $p = .034$, $d = .34$, 95% CI [.02, .65], as was the same state condition we reported above, $M = .72$, $SD = .17$, $t(41) = 4.26$, $p < .001$, $d = .66$, 95% CI [.32, .99]. The proportion of trials that reported one of the two correct states was .23, which did not differ from the proportion of choosing one correct answer in the same state condition, $M = .25$, $SD = .17$, $t(18) = 1.28$, $p = .206$, $d = .20$, 95% CI [.11, .50] (see Figure 5C). This indicates that the difference in overall accuracy between the same state and different state stems almost entirely from the proportion of trials in which participants correctly identify both exemplars; this is evident in Figure 5C.

In sum, we found that individuals are able to remember information *both* about the object itself and the state of the object. This finding supports the idea that object features can be bound to one another, contrary to the findings from Experiment 1A.

Table 2

Experiment 1B Proportion correct in the two tasks for the alone condition

	M
Exemplar task	
Exemplar correct	.89(.08)
Exemplar-state task	
Correct same	.72(.17)
Correct different	.66(.19)
Reported two same states; studied two same states	.79(.14)
Reported two same states; studied two different states	.29(.16)

Note. The standard deviation is provided in parentheses.

Experiment 1B

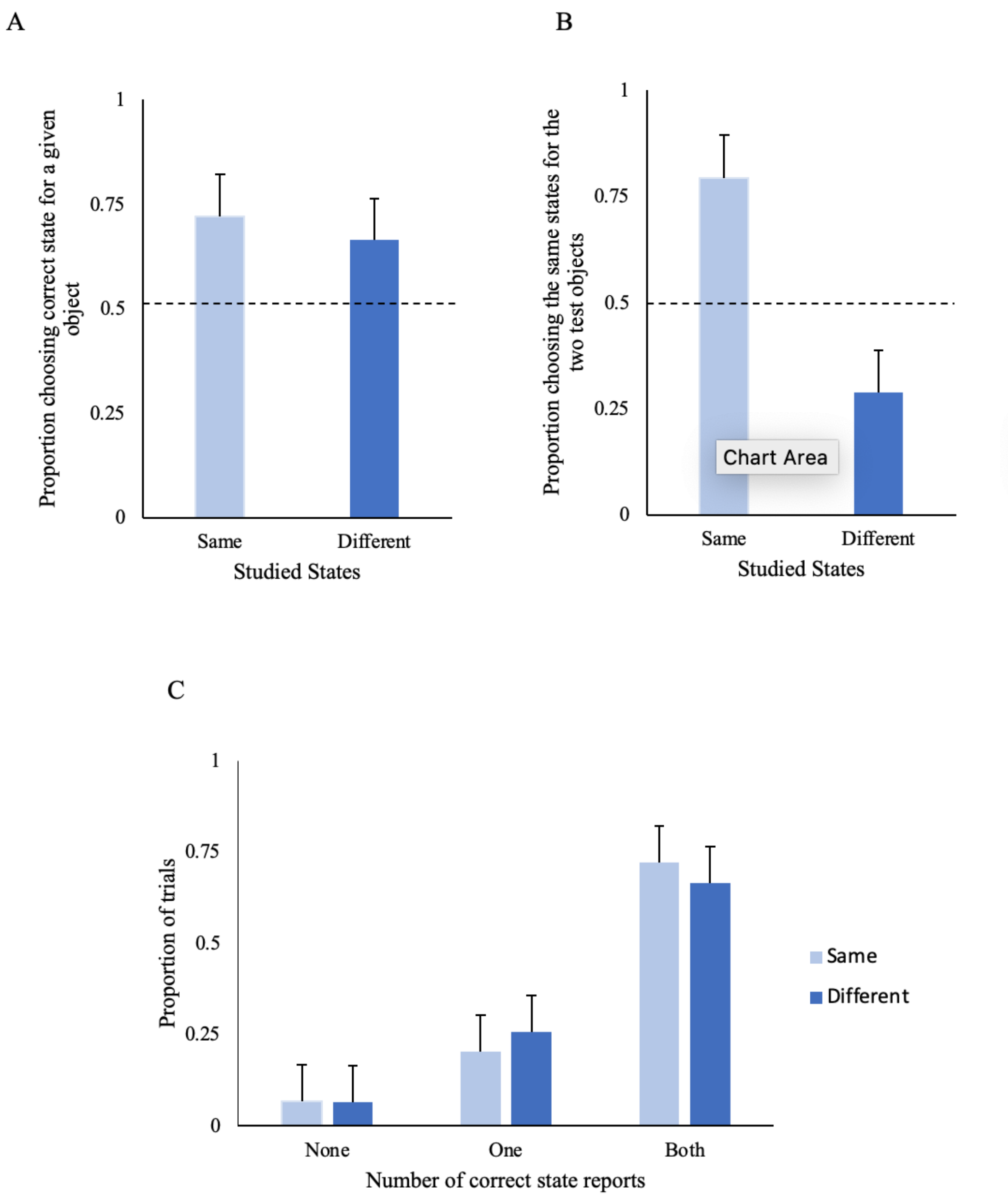

Figure 5.(A) The proportions selecting the correct state for a given exemplar when the two studied objects were shown in the same state (left-hand bar in panel A; no binding required) or different states (right-hand bar in panel A; binding required). (B) The proportions selecting the same states for the two test objects (regardless of whether these states are correct or incorrect). The dashed lines show chance levels. (C) The proportion of participants in a test trial who reported both correct states, one correct state, or no correct states, as a function of the study condition.

Confidence judgement —Alone Condition. The proportion correct for remember, know, and guess judgments were computed for each task. For the exemplar task, participants in the alone condition were correct 95% when they reported remembering the exemplar, 88% when they reported knowing they saw the exemplar, but only 59% when they guessed the exemplar. For the exemplar-state task, participants were correct 89% when they reported remembering the exemplar-state, 75% when they reported knowing they saw the exemplar-state, and 56% when they guessed the exemplar-state, see Figure 6.

Experiment 1B

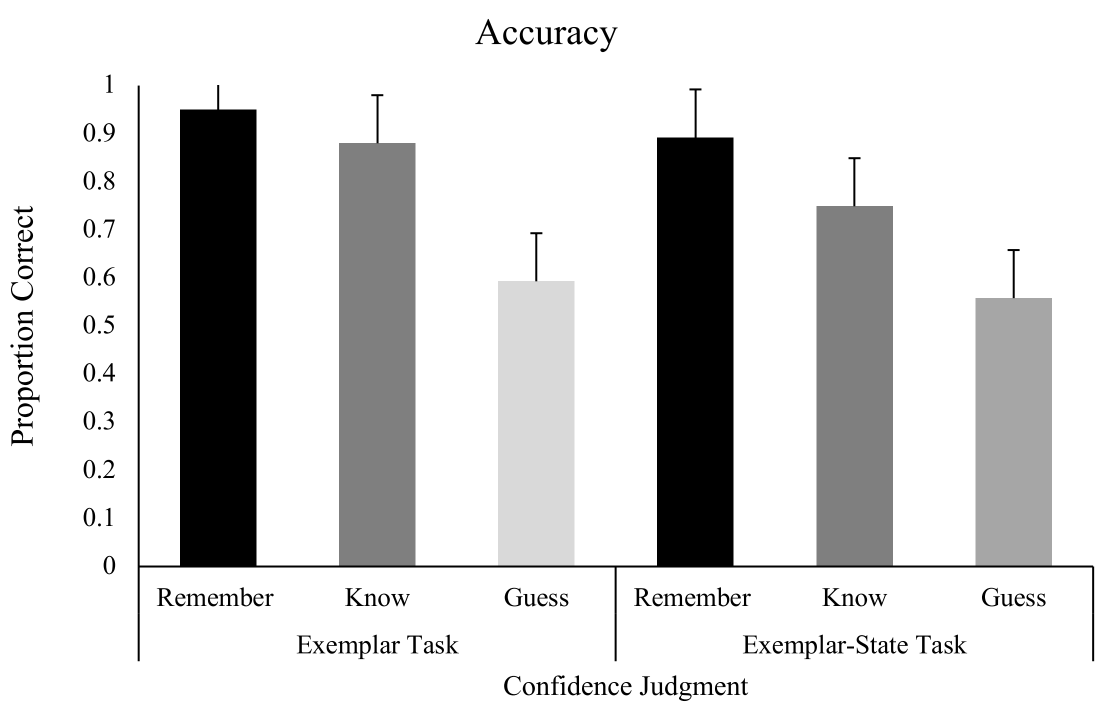

Figure 6. 2-AFC accuracy for remember, know, and guess judgment for the alone condition (left panel, exemplar task; right panel, exemplar-state task).

Accuracy in remembering exemplars—Collaborative Condition. In the Exemplar condition, participants performed above chance during their noncollaborative individual decision, $M = .90$, $SD = .09$, $t(41) = 13.15$, $p < .001$, $d = 2.03$, 95% CI [1.49, 2.56], suggesting that exemplar information was well remembered when it was not required to be bound to state information. Collaborative pairs also performed above chance, $M = .94$, $SD = .05$, $t(20) = 12.62$, $p < .001$, $d = 2.75$, 95% CI [1.80, 3.70]. A comparison between the two noncollaborative decisions and the collaborative decision of the exemplar task was made using a paired sample *t*-test. In order to make this comparison, the two noncollaborative decisions were averaged for each pair and compared to their collaborative decision. The results showed that the averaged noncollaborative decision ($M = .90$, $SD = .07$) performed worse compared to the collaborative decision ($M = .94$, $SD = .05$), $t(20) = 11.43$, $p < .001$, $d = 2.49$, 95% CI [1.61, 3.37]. Table 3 shows the proportion correct for each of the tasks.

Accuracy in remembering state—Collaborative Condition. When both exemplars in a category were shown in the same state, noncollaborative decisions were above chance at choosing this state, $M = .79$, $SD = .16$, $t(41) = 6.32$, $p < .001$, $d = .98$, 95% CI [.60, 1.34]. Similarly, collaborative pairs were also above chance at choosing this state, $M = .86$, $SD = .09$, $t(20) = 8.56$, $p < .001$, $d = 1.87$, 95% CI [1.14, 2.58], suggesting that participants were good at remembering states when binding was not required. A comparison between the averaged noncollaborative decision and the collaborative decision for choosing two exemplars in the same state, using a paired samples *t*-test, showed that the averaged noncollaborative decision ($M = .79$, $SD = .16$) was worse compared to the collaborative decision ($M = .86$, $SD = .09$), $t(60.12) = 3.16$, $p = .005$, $d = .69$, 95% CI [.21, 1.16].

Additionally, participants performed better when the exemplars were shown in the same state versus in different states. A mixed ANOVA was conducted to compare the proportion of times participants in the noncollaborative and collaborative condition selected the same states versus different states for exemplars that actually were shown in the same states. The 2 (condition: noncollaborative decision vs. collaborative decision) X (exemplar-state type: two same states vs. two different state) revealed a significant main effect for exemplar-state type, F(l, 61) = 346, , $p < .001$, $\eta_p^2 = .85$. The main effect for condition was nonsignificant, F(l, 61) = .082 $p = .775$, $\eta_p^2 = .001$, as well as the interaction, $F(1,61) = 1.22$, $p = .274$, $\eta_p^2 = .020$. These findings indicate that the proportion of times that participants selected the same two states (M = .84, SD = .11) for the two exemplars was higher if the exemplars actually were shown in the same states compared to those shown in different states (M = .19, SD = .12). This was true for the noncollaborative decision and the collaborative decision. In both conditions, the proportions were above the chance level, according to an one-sample t-test; for the noncollaborative decision (same states: M = .83, SD = .13, $t(41) = 8.67$, $p < .001$, d = 1.34, 95% CI [.92, 1.75]; different states: M = .20, SD = .13, $t(41) = 15.61$, $p < .001$, d = 2.41, 95% CI [1.80, 3.01]), and for the collaborative decision (same states: M = .87, SD = .08, $t(20) = 9.25$, $p < .001$, d = 2.02, 95% CI [1.26, 2.77]; different states: M = .15, SD = .10, $t(20) = 12.31$, $p < .001$, d = 2.69, 95% CI [1.75, 3.61]), see Figure 7B. Similar to the alone condition above, this finding suggests that participants remembered whether objects from a particular category were presented in the same state or in different states. Additionally, individuals who collaborated were better at correctly choosing two exemplars in the same state than for the noncollaborative decision. In sum, individuals, working by themselves or with a collaborator, have a good memory of both the exemplars themselves and the state of the exemplar.

Accuracy in exemplar-state binding—Collaborative Condition. We found evidence that individuals have good memory for both exemplars and their state information, independently, and when these individuals collaborate, they have even better memory for exemplars and their state information. So, now we will examine whether their features are remembered as a bound unit. As in the alone condition above, participants' memory in the different state condition was above chance for both the noncollaborative decision *M* = .71, *SD* =.15, *t*(41) = 4.40, *p* < .001, *d* = .68, 95% CI [.34, 1.01] and the collaborative decision *M* = .78, *SD* =.10, when choosing the states of exemplars that had been presented in different states, *t*(20) = 6.63, *p* < .001, *d* = 1.45, 95% CI [.82, 2.06].

To test whether collaboration helped with binding, we compared the proportion correct for the same state and different state trials by running a 2 (condition: noncollaborative decision vs. collaborative decision) X (exemplar-state correct: same state correct vs. different state correct) mixed ANOVA. The results revealed a marginally significant main effect for condition, *F*(l, 61) = 3.78, *p* = .057, η_p^2 = .06, and a significant main effect for exemplar-state correct, *F*(l, 61) = 30.95, *p* < .001, η_p^2 = .34. The interaction was nonsignificant, *F*(1,61) = .02, *p* = .891, η_p^2 = .3.079× 10^{-4}. When participants collaborated (same states: *M* = .86, *SD* = .09, different states: *M* = .78, *SD* = .10), they correctly selected the state of exemplars better in comparison to their original noncollaborative decision (same states: *M* = .79, *SD* = .16, different states: *M* = .71 *,SD* = .15), see Figure 7A. However, it should be noted that performance was fairly high for both decisions (noncollaborative individual and collaborative). Furthermore, post hoc analyses were conducted using the Bonferroni correction, which indicated participants were significantly better at remembering the state of each exemplar when the exemplars were shown in the same state (*M* = .81, *SD* = .14) than in different states (*M* = .74, *SD* = .14), *p* < .001.

Finally, we calculated the proportion of trials that reported one of the two correct states in the noncollaborative condition, .19, which did not differ from the proportion of choosing one correct answer in the same state condition, *M* = .17, *SD* = .13, *t*(41) = 1.08, *p* = .285, *d* = .17, 95% CI [.14, .47]; see Figure 7C. Additionally, the collaborative condition reported .16 trails in which one of the two correct state was reported, which did not differ from the proportion of choosing one correct answer in the same state condition, *M* = .13, *SD* = .09, *t*(20) = 1.59, *p* = .127, *d* = .35, 95% CI [.10, .79]. Therefore, the difference in overall accuracy between same state and different state arises almost entirely from the proportion of trials in which participants get both exemplars correct. This can be seen clearly in Figure 7C.

Table 3

Experiment 1B Proportion correct in the two tasks for the collaborative condition

	Noncollaborative	Collaborative
Exemplar task		
Exemplar correct	.90(.07)	.94(.05)
Exemplar-state task		
Correct same	.79(.11)	.86(.09)
Correct different	.71(.10)	.78(.10)
Reported two same states; studied two same states	.83(.09)	.87(.08)
Reported two same states; studied two different states	.20(.13)	.15(.10)

Note. The standard deviation is provided in parentheses.

Experiment 1B

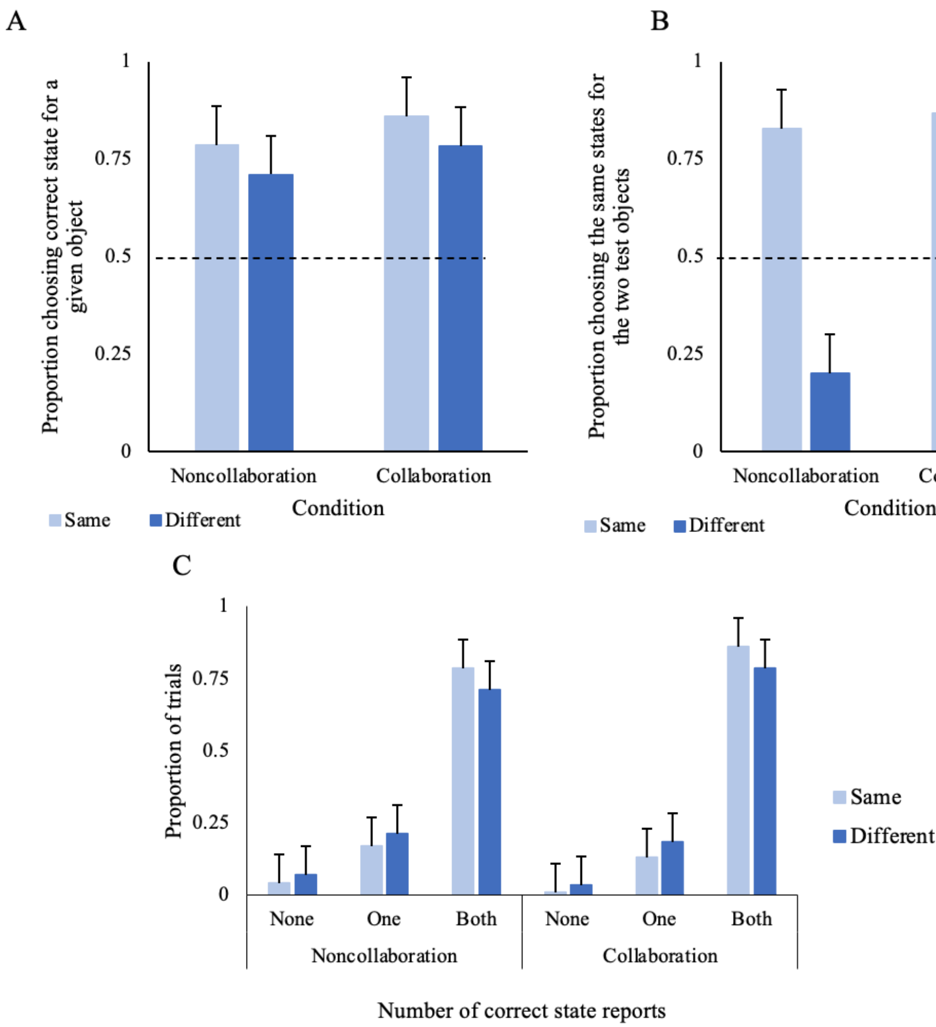

Figure 7. (A) The proportions selecting the correct state for a given exemplar when the two studied exemplars were shown in the same state (left-hand bar in panel A; no binding required) or different states (right-hand bar in panel A; binding required) for the noncollaborative decision and the collaborative decision. (B) The proportions selecting the same states for the two test exemplars (regardless of whether these states are correct or incorrect) for the noncollaborative decision and the collaborative decision. The dashed lines show chance levels. (C) The proportion of participants in a test trial who reported both correct states, one correct state, or no correct states, as a function of the study condition.

Confidence judgement—Collaborative Condition. The proportion correct for remember, know, and guess judgments were computed for each task. For the exemplar task, participants in the noncollaborative condition were correct 97% when they reported remembering the exemplar, 86% when they reported knowing they saw the exemplar, and 71% when they guessed the exemplar. For the exemplar-state task, participants were correct 91% when they reported remembering the exemplar-state, 80% when they reported knowing they saw the exemplar-state, and 55% when they guessed the exemplar-state, see Figure 8.

Experiment 1B

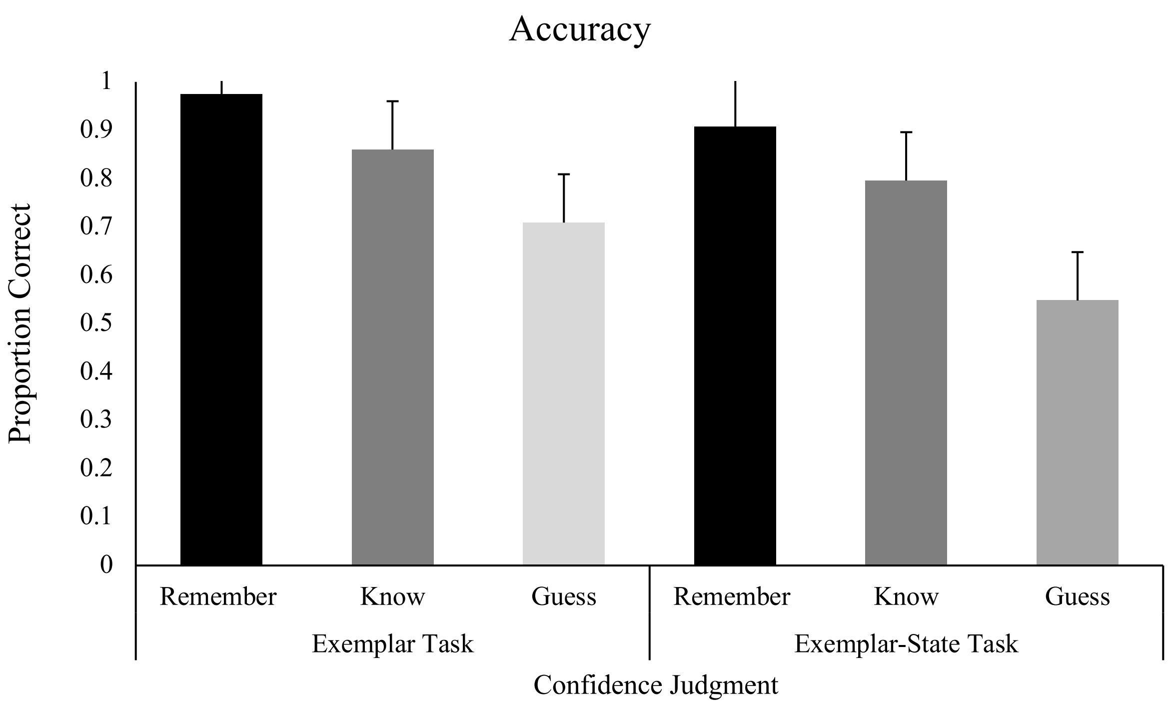

Figure 8. Noncollaborative 2-AFC accuracy for remember, know and guess judgment (left panel, exemplar task; right panel, exemplar-state task).

Examining how collaboration helped binding at retrieval. Finally, we were curious to understand how collaborating aided in binding exemplars and state information at retrieval. We focused on when the collaborative condition was better (the advantage was marginal) than the noncollaborative at correctly identifying exemplar-state when the objects were in different states (which is what they studied). We examined how collaborative groups made their 2-AFC decisions by calculating how often the groups went with the decision of their best group member. Reviewing this allows us to rule out if the collaboration benefit arises from one partner's ability to enhance the group's performance. We determined which group member had the highest proportion correct for identifying exemplar-states that were presented in different states during their noncollaborative decisions. We then used the best group members for each pair to evaluate the proportion of disagreements that were resolved that were consistent with the best group member's decisions. Because there were two exemplar-state stimuli for each category, we examined the top-row and bottom-row responses independently. We discovered that the collaborative pairs went with the best group members' decision 50% of the time, which indicates that the collaborative advantage was not exclusively attributable to the group's best member.

Next, we determined whether the group's top performer from the noncollaborative condition outperformed those in the collaborative condition's performance when it came to correctly identifying exemplar-states that were studied in different states. We ran paired samples *t*-tests on the ability to correctly identify exemplar-states that were shown in different states and found that the collaborative condition ($M = .78$, $SD = .10$) performance did not significantly differ from the group's best member's performance ($M = 79$, $SD = .08$): $t(21) = -.39$, $p = .702$, $d = -.09$, 95% CI [-.51, .35]. These findings indicate that the group's best member was actually able to bind the exemplar and state information of exemplar-states displayed in different states. Consequently, the noncollaborative condition performed worse than the collaborative decision, which suggests that the best performer did not always stick with their correct decisions when working with a partner. To further understand, we calculated the group's best performers willingness to change their correct different exemplar-state decisions to incorrect decisions when they collaborated. We found that the group's best performers changed the correct noncollaborative decision to incorrect when collaborating with a partner 17% of the time. Furthermore, 33% of the time, the group's best performer carried their correct answer along with them during the collaborative test.

In sum, we find support for individuals remembering both exemplar information as well as state information. We also showed that individuals were able to successfully remember the state information when they are presented the exemplars in different states, which provides evidence that exemplar features are bound together. Furthermore, collaboration benefited the accuracy of remembering exemplar information, state information, and the binding of exemplar information to the correct state.

In Experiment 1A, we replicated Utochkin and Brady (2019) and found evidence for chance-level exemplar-state binding, despite individuals' ability to successfully remember exemplars when they were not required to bind and state information when exemplar information was not required. However, in Experiment 1B, we split the tasks into smaller blocks of studied stimuli to accommodate a within-subjects design (individual and collaborative) and found that this design modification resulted in participants successfully binding exemplars and their states at above chance levels. The results from Experiment 1A, coupled with the results of this experiment, suggest that the binding of object features is negatively impacted by having to study a large number of stimuli. As a reminder, Experiment 1A used 120 categories, which replicates Utochkin and Brady (2019), whereas Experiment 2 used 60 categories per block. When participants studied half the number of stimuli, they could bind the object and state information at better than chance levels.

We also compared the noncollaborative decision to the collaborative decision and found that collaborating significantly improved recognition performance. First, memory for exemplars was significantly better when individuals collaborated in comparison to averaged noncollaborative decision. Second, collaborating significantly improved the ability to correctly choose two exemplars in the same state when that was how they were presented during the study phase. Third, when participants collaborated, they correctly selected the state of exemplars better overall (both, same state, and different state) in comparison to their original noncollaborative decision.

To better understand whether the evidence for exemplar-state information being bound exists under different testing procedures, we conducted another experiment, this time using old-new recognition testing. This switch in testing was undertaken to further explore the benefit that collaboration had on exemplar-state binding. To address this, we will examine individuals' and collaborative pairs’ ability to differentiate between bound stimuli that they studied and new stimuli that they have not studied.

Experiment 2

In Experiments 1A and 1B, the 2-AFC test provided participants with four possible options, two of which were always correct. With forced-choice designs, individuals make decisions based on how familiar the two options are, simply choosing the alternative that yields the greatest familiarity (Hicks & Marsh, 1998). This design allowed us to examine recognition memory accuracy while reducing confusion because it supports comparisons among the possible choices. In Experiment 1B, we found that both individuals remembering alone and individuals collaborating to remember were able to successfully bind exemplars to their states. We also found that collaborating led to better binding performance in comparison to the noncollaborative decision. However, if collaboration helps with binding at retrieval, would this remain true if the test phase required participants to make a decision when less information is provided at test? For example, instead of providing participants two options, where one option is always correct, what if participants only have one option, and it could be correct or incorrect? In Experiment 2, we conducted a traditional old-new recognition test, a more challenging task that requires a response criterion. The requirement of setting and maintaining a criterion poses a strain on memory resources (Benjamin, Diaz, & Wee, 2009). In fact, some researchers state that 2AFC tests are criterion-free (Hicks & Marsh, 1998; Macmillan & Creelman, 2004). According to Signal Detection Theory, (SDT; Green & Swets, 1966), which is commonly used to interpret old-new recognition data, there are two normal distributions that overlap. These distributions vary along a continuum of memory strength, one being the new-exemplar distribution (reflecting items not previously studied) and the other, shifted to the right, being the old-exemplar distribution (reflecting items that had been previously studied). The response criterion (or evidence threshold) assesses bias favoring old or new judgments and is positioned somewhere between the two distributions. During old/new recognition tests, if the memory strength of an exemplar falls above the criterion, participants would respond old, and if it falls below, participants would respond new.

Experiment 2 will help us understand how this criterion is set for individuals who work alone versus those who work collaboratively. The old/new recognition test used in Experiment 2 required participants to identify whether a tested exemplar was old (had been previously studied) or new (exemplar not previously studied). Specifically, we wanted to examine if collaborating would improve participants' ability to distinguish between studied objects in the same state (old) and unstudied objects (new), as well as the more difficult discrimination involving distinguishing studied objects in the same state from studied objects presented in new states (e.g., studied a whole apple but need to reject the same apple with a bite taken out it). We will elaborate on this further below, but first we discuss how we evaluate recognition memory performance in the current study.

The goal of Experiment 2 is to explore the results from a signal detection theory (SDT; Green & Swets, 1966) approach. SDT is a psychophysical model used to gauge performance, namely how well one can distinguish between signal and noise (Macmillan & Creelman, 2004). It is commonly used to assess variations in human performance in domains like recognition memory (Rajaram & Pereira-Pasarin, 2007; Gronlund & Benjamin, 2018). SDT assesses two indices of performance, sensitivity (*d-prime*) and response bias *c*. Sensitivity captures a participant's ability to successfully discriminate between previously studied stimuli and never before seen stimuli (e.g., old and new responses). Response bias, which is measured as *c,* identifies a participates willingness to make a response. In the current study, we assessed these two measures to conceptualize recognition memory performance and attempt to better understand how collaboration aids exemplar-state binding. More specifically, we were interested in whether collaborating would increase participants' ability to successfully discriminate between old and new exemplars, as well as old exemplars presented in new states, when participants are required to set a criterion. That is, each participant will be setting their own criterion value, which may differ from their collaborating partner's willingness to respond (i.e., *c*), and they will have to decide, as a collaborative pair, whether they will set a more conservative or liberal criterion than they did when they were not collaborating.

To conceptualize the exemplar-state varieties used in Experiment 2, imagine that participants study a Blue Coffee Mug that is full. Participants subsequently were exposed to three possible tests: 1) same exemplar-same state (Blue Coffee Mug that is full), to which they should respond old; 2) same exemplar-new state (Blue Coffee Mug that is empty), which they should identify as new; 3) new exemplar-no state (Toy Hammer that they never studied) that they should identify as new.

Figure 9. On the left is a possible studied exemplar-state stimulus. On the right are the three possible test trials a participant is exposed to. It should be noted that only one exemplar-state test would be displayed at a time, requiring participants to decide whether the tested exemplar is old or new. Below each tested exemplar, we identify the correct decision for each possibility.

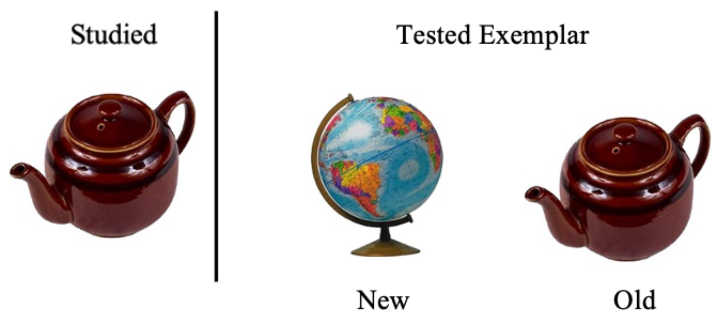

Figure 10. On the left is a possible studied exemplar task stimulus. On the right are the two possible test trials tested a participant is exposed to. It should be noted that only one exemplar test would be displayed at a time, requiring participants to decide whether the tested exemplar is old or new. Below each tested exemplar, we identify the correct decision for each possibility.

It is well known that performance on old/new recognition tests is worse than from 2-AFC tests (Draschkow, Reinecke, Cunningham, & Võ, 2019; Jang, Wixted, & Huber, 2009). Also, due to the challenge of requiring participants to identify previously studied exemplars as new if the state did not match what was studied, we reduced the number of studied and test trials. We wanted to make sure that in this task was not so challenging, or so fatiguing, that it resulted in floor effects. Consequently, we reduced the number of studied trials in half—a decision we discuss further in the Method section.

Experiment 2's change to the old/new recognition task allows us to investigate exemplar-state binding when participants were only provided one exemplar-state test stimulus at a time (see Figure 9, contrast with Figure 1). Given that those who collaborated were able to bind exemplars and state information in Experiment 1B at above chance levels, it could have been because those test stimuli were easier to discuss and compare with their partner. For example, in the 2-AFC exemplar-state task, participants saw the two exemplars from a given category, making more information available to discuss. The presence of both possible options could be used to make an argument regarding which object is correct or which object is wrong. Consider the following: A collaborating participant can reference the exemplar-state to make a claim regarding their decision (e.g., “I know we didn’t study cracked eggs because I paid close attention and know we studied a whole white egg and whole brown egg”). Being able to identify small feature differences like this, a distractor exemplar-state as unfamiliar, and not matching the studied stimuli, is analogous to a recall-to-reject process (see Rotello & Heit, 2000). In Experiment 2, however, participants only have a single test item, which could make buttressing these kinds of arguments more difficult.

If participants can successfully discriminate between old exemplar-state combinations and old exemplars displayed in new states, this would indicate that participants have memory of the exemplar-state as a bound unit. Additionally, by utilizing old/new recognition testing, we can examine how criterion placement differs for individuals working alone compared to when they collaborate. For example, when individuals collaborate, they may be more liberal in their willingness to respond due to the pair's desire to fairly incorporate each other’s contributions.

But individuals may use a more conservative criterion placement when working alone to maximize their performance and limit the number of errors. We will also be able to see how *c* varies for the more difficult tasks that require binding.

In the exemplar-state task, when participants are presented with objects that are presented in a different state than what was studied, we hypothesize that collaborating individuals will show evidence for above chance binding. As for individuals working alone, we hypothesize that they will bind at chance performance due to the increased difficultly of the old/new recognition task. We also hypothesize that the collaborative groups shown exemplar-states they studied will identify them more often as old, and new exemplars more often as new, compared to those who work alone. Lastly, in the exemplar task, we hypothesize that individuals remembering alone, and individuals collaborating, can make these discriminations at above chance levels.

If collaboration enhances the ability to bind exemplars to their correct states, we will evaluate if the benefit arises due to increased hits (due to at least one collaborator making a strong claim that they studied the stimulus), decreased false alarms (due to error pruning), or both. We hypothesize that collaborating will both reduce false alarm rates and increase hit rates. Regarding decreased false alarm rates, our rationale is that the communication and feedback exchange during collaboration will help pairs error prune new exemplars (Rajaram et al., 2020). When only one individual from the collaborative group remembers an exemplar, it is likely that the pair will discuss the disagreement and reject the new exemplar because only one person remembers it (Rajaram et al., 2020). Regarding increased hit rates, this might occur when group members are able to make compelling arguments to explain their memory of an exemplar, which can cue the memory for their partner (Clark et al., 2000). Furthermore, collaborators can provide evidence that they remember an exemplar or the state information using recollection, which can create a convincing argument regarding the acceptance or rejection of an exemplar. We know from prior research that recollection is necessary for the retrieval of bound information (Daselaar & Cabeza, 2014), and therefore, collaboration should facilitate decisions that benefit from recollection. For example, a collaborative group working together to make an old/new recognition decision are more likely to reach a correct decision if they can elaborate on their reasoning for remembering an exemplar-state. They could accomplish this by providing context to their memory of studying an exemplar (e.g., “I remember we saw the striped cup because I have a striped cup just like that in my kitchen.”). Lastly, binding is a demanding task, but collaborating can divide the cognitive load, allowing recollection exchanges to flourish to support a final decision. If support for binding at retrieval is found, this will be another instance in which collaboration is beneficial to memory, and being able to interpret the collaborator's hits and false alarm rates will help detangle the locus of the advantage.

Method

Participants

Relying on the power analysis described in Experiments 1A and 1B, a different group of 38 undergraduate students at the University of Oklahoma participated in exchange for SONA credit or a $10 Amazon gift card. There were 32 females, 5 males, and 1 individual who marked prefer not to say (M = 18.74, SD = 1.14). The alone and noncollaborative conditions contained all 38 individuals. However, when participants made collaborative conditions, they were with a partner, resulting in 19 collaborative pairs in total (19*2 to make a pair).

During our examination of the normality of our data and outliers, we removed five pairs due to their extreme scores. Four of the pairs were removed due to almost perfect *d-prime* performance and one pair was removed due their extreme response bias score.

Materials

Qualtrics was used to present the stimuli to participants and for the completion of all tasks. The same two image sets used in Experiment 1 are used in the exemplar task and the exemplar state task. However, this time, only 240 images were used in each task. We randomly selected 240 images from our previously used images in Experiments 1A and 1B.

Design

The two conditions in this study were the individual condition and the collaborative condition, which follow a between-subjects design. Each condition was exposed to both the exemplar-state task and the exemplar task. The exemplar-state task had three types of tests: 1. same-exemplar; same state, 2. same-exemplar; different state, 3. no-state exemplar, whereas the exemplar task contained only old and new exemplars.

Procedure

The study phase followed the same format as Experiments 1A and 1B. The studied categories were divided into two blocks, just as in Experiment 1B. However, the number of studied trials and tested trials per block was now 60. For example, participants studied 60 categories followed by 60 recognition tests and then studied 60 additional categories followed by the second set of recognition tests. The breakdown of old and new exemplars and exemplar-states is discussed for each task below. The confidence judgment made for each recognition decision used a 5-point Likert scale ranging from (1-not confident at all, 2- a little confident, 3- fairly confident, 4- quite a bit confident, 5- extremely confident). This was the procedure for both the exemplar task and the exemplar state tasks.

Exemplar Task

During each test trial of the *exemplar task*, participants were shown one exemplar and asked to indicate if the exemplar displayed was one that they previously studied (old) or if it is a new exemplar. This task included 30 old exemplars and 30 new exemplars presented in the test phase. The exemplars were counterbalanced between study versions so sometimes exemplars that were presented as old to some participants were presented as new for other participants.

Exemplar-state Task

During the *exemplar-state task*, participants studied exemplars that were either in the same or different states. During the test phase, participants were shown one image at a time and were asked to make an old/new recognition judgment. They needed to identify whether the exemplar and state is the combination they previously studied (old), or if the exemplar and the state are new, or if it is an entirely new exemplar. During the exemplar-state task test phase, there were 20 old exemplar-states, 20 new exemplars, and 20 old exemplars displayed in new states. Again, we counterbalanced the exemplar-state that was presented during the study and test phases.

Results

In order to assess performance, we computed *d-prime*[1] (sensitivity) and *c* (response bias) for the exemplar task (e.g., the globe in the trial of Figure 10), the exemplar-state task, new exemplar (e.g., the toy hammer in the test trial of Figure 9), and the exemplar-state task old exemplar, new state (e.g., the empty blue coffee cup in the test trial of Figure 9). We then compared these measures amongst the alone condition, the noncollaborative condition, and the collaborative condition, by conducting one-way ANOVAs. We conclude by depicting the

[1] A standard correction was used for false alarm rates of 0 and hit rates of 1. The correction for false alarm rates of 0 was 1/(2N) where N represents the maximum number of false alarms. The correction formula for hit rates of 1 was 1-1/(2N) where N represents the number of targets.

confidence-accuracy relationships for each task by creating calibration curves for those tasks completed individually (alone and noncollaborative).

Because were interested in comparing performance amongst all conditions (alone, noncollaborative, and collaborative), we first made sure that the conditions were comparable and, therefore, appropriate to combine in a one-way ANOVA despite stimuli differences. To test this, we conducted a series of independent samples *t*-tests and found no significant difference between the alone and noncollaborative performance.[2] Given this equivalence, we combined the data from those who worked alone with the data from those that made an individual decision before collaborating (noncollaborative) and tested whether this performance differed from those who collaborated.

We present the results by order of importance. First, we present the exemplar-state task for the old exemplar new state performance, followed by the exemplar-state new exemplar performance, and then performance from the exemplar task. In each section below, we compare *d-prime* and *c* for the groups.

Exemplar-state task: same-exemplar; different state performance. In the exemplar-state task, participants had to discriminate between the exemplar-state that they had previously studied and two new tests: 1) that same exemplar but in a different state, 2) new exemplar (presented in no particular state). First, we will analyze participants ability to discriminate same-exemplar,

---

[2] The *t*-tests revealed that there was no significant difference between the group's *d-prime* scores for the exemplar task, $t(74) = 1.35$, $p = .182$, $d = .31$, exemplar-state new exemplar $t(68) = .735$, $p = .465$, $d = .59$, and exemplar-state old exemplar new state $t(74) = .994$, $p = .323$, $d = .64$. We also conducted a series of independent samples *t*-tests comparing *c,* and again, there were no significant differences, exemplar task, $t(74) = 1.70$, $p = .093$, $d = .39$, exemplar-state new exemplar $t(68) = .1.88$, $p = .063$, $d = .30$, and exemplar-state old exemplar new state $t(74) = .122$, $p = .904$, $d = .28$.

different states from old, studied exemplar-states. We conducted a one-way ANOVA to compare whether the three conditions differed in their ability to successfully discriminate between old, studied exemplar-states and new, old exemplars presented in new states. The test on *d-prime* showed that the effect of the condition was significant (alone: $M = 1.77$, $SD = .70$; noncollaborative: $M = 1.63$, $SD = .57$; collaborative: $M = 2.08$, $SD = .53$), $F(2,92) = 3.46$, $p = .036$, $\eta^2 = .07$. Post hoc analyses were conducted with the Bonferroni correction and indicated the collaboration condition was .457 more successful at discriminating between old exemplars presented in new states than was the noncollaborative (**(**p **=** .030, 95% *CI* of the difference = .033 to .881). These findings support the idea that collaborating helped partners correctly identify the bound exemplar-states they had previously studied and reject the old exemplars that were presented in a new state at test. This suggests that the participants remembered the exemplar information and the state information because they successfully bound what they studied, which allowed them to reject these stimuli. We conducted a one-way ANOVA comparing *c* that showed that the effect of condition was not significant (alone: $M = .02$, $SD = .32$; noncollaborative : $M = .03$, $SD = .22$; collaborative: $M = .02$, $SD = .26$), $F(2,92) = .008$, $p = .992$, $\eta^2 = .00$. These findings indicate that all conditions had similar level of response bias.

Exemplar-state task: no-state exemplar performance. In the exemplar-state task, we again asked participants to select whether the presented exemplar or exemplar-state had been previously studied. This time, we tested whether the three conditions varied in their ability to successfully discriminate between old, studied exemplar-states and new exemplars that were displayed with no particular state (e.g., see Figure 9 which displays a toy hammer). The one-way ANOVA comparing *d-prime* showed that the effect of condition was not significant (alone: $M = 2.35$, $SD = .67$; noncollaborative: $M = 2.45$, $SD = .50$; collaborative: $M = 2.70$, $SD = .47$), *F*(2,92) = 2.26, *p* = .111, η^2 = .05. This finding indicated that all of the conditions were able to successfully discriminate between the new exemplars presented to them during the test phase and the previously studied exemplars-state. Similarly, the one-way ANOVA comparing *c* showed that the effect of the condition was not significant (alone: *M* = .31, *SD* = .35; noncollaborative: *M* = .44, *SD* = .25; collaborative: *M* = .33, *SD* = .23), *F*(2,92) = 2.13, *p* = .125, η^2 = .04. These findings indicate that all conditions had similar levels of response bias.

Exemplar task performance: old/new exemplar. In the exemplar task, participants identified whether the exemplar presented at the test was old or new. In order to test whether the three conditions varied in their ability to successfully discriminate between old, studied exemplars and new exemplars, a one-way ANOVA model was performed on *d-prime.* The one-way ANOVA comparing *d-prime* showed that the effect of the condition was not significant (alone: *M* = 2.65, *SD* = .77; noncollaborative: *M* = 2.91, *SD* = .91; collaborative: *M* = 2.97, *SD* = .61), *F*(2,92) = 1.42, *p* = .246, η^2 = .03. We should note that all of the condition performed very well resulting in a ceiling effect. In hindsight, this finding might have been anticipated given that individuals are able to remember large amounts of pictures with good accuracy (Standing, 1973). Nevertheless, by examining the means, we still see a trend that the collaborative condition performed better than both the alone and noncollaborative conditions. This finding again speaks to individuals' ability to recognize previously studied exemplars. Even though the collaborative benefit was lost here, it is possible that, if we avoided a ceiling effect, we may have still seen a benefit. However, it is important to note that in the present experiments, with the focus on binding, the exemplar task serves as a control to show that participants can remember exemplars at above chance levels, which is very clear from all three of our experiments. More importantly, finding this ceiling effect here does not negate our ability to examine our primary questions involving exemplar-state binding.

To determine whether the willingness to make a selection (response bias) differed amongst the three groups, a one-way ANOVA comparing *c* showed that the effect of the condition was significant (alone: $M = .27$, $SD = .27$; noncollaborative: $M = .16$, $SD = .28$; collaborative: $M = .02$, $SD = .28$), $F(2,92) = 5.12$, $p = .008$, $\eta^2 = .10$. Post hoc analyses were conducted using the Bonferroni correction that indicated the alone condition was .246 more conservative in their decision making than the collaborative condition ($p = .006$, 95% *CI* of the difference = .057 to .436). Taken together, these findings indicate that participants in all of the conditions performed very well when it came to their ability to successfully identify previously studied exemplars as old and reject new exemplars. Additionally, we found that the alone condition tended to be more conservative in their decision of reporting an exemplar as old in comparison to the collaborative decision. This supports our hypothesis that when individuals work alone, they employ a more conservative placement to increase their performance and reduce their number of mistakes, however, it seems that when individuals collaborate they want to fairly incorporate each other’s responses which make them more lenient in their criterion placement.

Table 4

Experiment 2 d-prime and *c group comparisons*

	Alone	Noncollaborative	Collaborative
Same-exemplar; different state			
Hit rate	.79(.12)	.77(.11)	.83(.08)
False alarm rate	.21(13)	.22(.09)	.16(.08)
d-prime	1.77(.70)	1.63(.57)	2.08(.53)
c	.02 (.32)	.03(.22)	.02(.26)
No-state Exemplar			
Hit rate	.79(.12)	.77(.11)	.83(.08)
False alarm rate	.09(.10)	.06(.05)	.06(.04)
d-prime	2.35(.67)	2.45(.50)	2.70(.47)
c	.31(.35)	.44(.25)	.33(.23)
Old/New Exemplar			
Hit rate	.83(.11)	.87(.12)	.92(.06)
False alarm rate	.08(.07)	.07(.08)	.09(.07)
d-prime	2.65(.77)	2.91(.91)	2.97(.61)
c	.27(.27)	.16(.28)	.02 (.28)

Note. The standard deviation is provided in parentheses.

Experiment 2

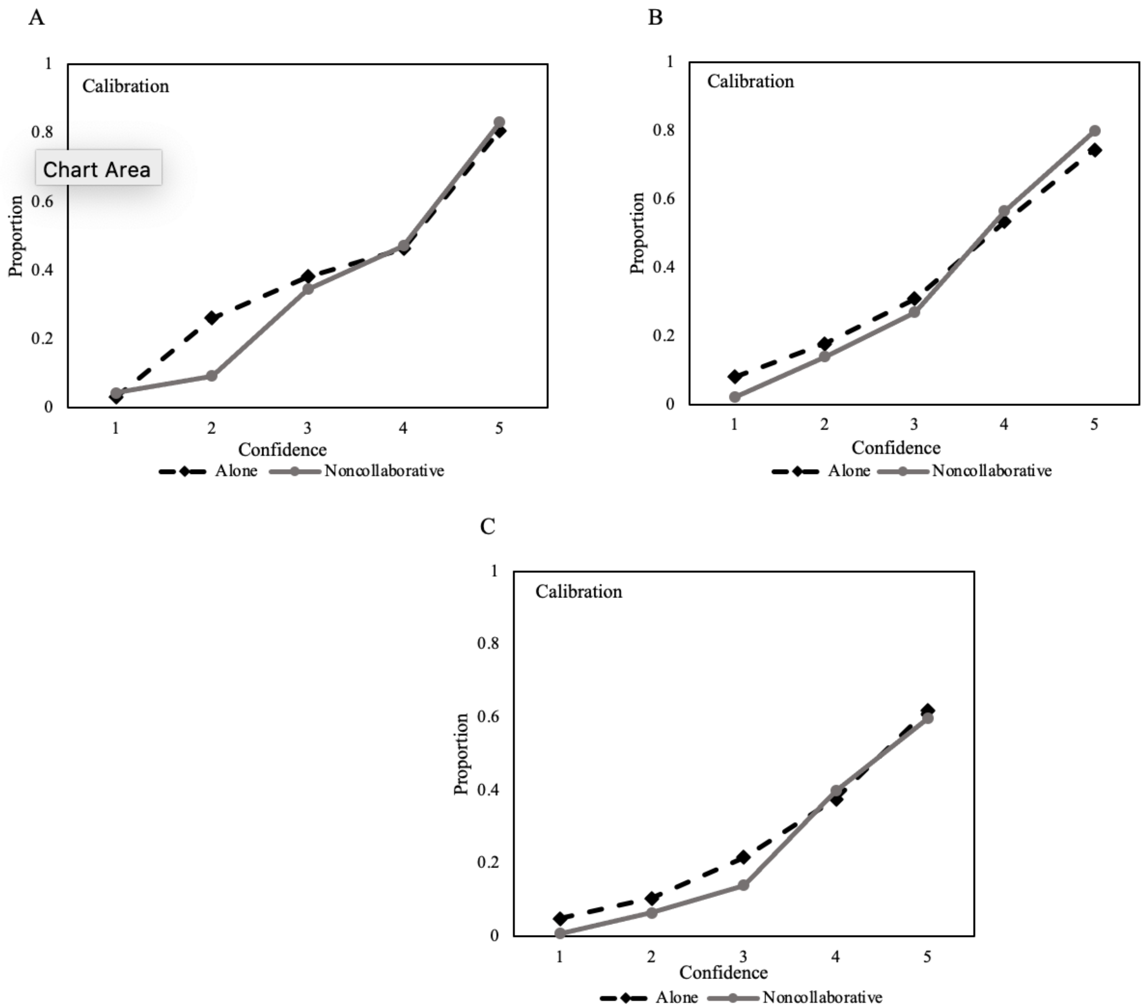

Figure 11. All of the data used to create these calibration curves come from the alone and noncollaborative groups. Each calibration curve plots the proportion correct as a function of confidence. All of the participants made a confidence decision using a 5-point Likert scale (1-not confident at all, 2- a little confident, 3- fairly confident, 4- quite a bit confident, 5- extremely confident). (A) Depicts a calibration curve for the exemplar task; (B) the exemplar-state new exemplar; (C) the exemplar-state old exemplar new state. For both the alone condition and the noncollaborative decision, the likelihood that an identification is accurate rises with the level of reported confidence.

Examining how collaboration helped binding at retrieval.

Lastly, we are interested in understanding how collaboration helped with the binding at retrieval of exemplars and their state information. To be clear, we were only interested when the performance of the collaborative condition was significantly better than the noncollaborative decision, which was when participants made an old/new recognition decision involving the same studied exemplars but presented in different states. First, we averaged the two participants' noncollaborative decisions and created a paired single noncollaborative score for each pair. We then conducted two paired sample *t*-tests between the averaged noncollaborative pair and the pairs collaborative hit rate (HR) and false alarm rate (FAR) for the same-exemplar, different state decisions.

The first paired sample *t*-test compared the HR for the averaged noncollaborative pairs ($M = .77$, $SD = .11$) and the collaborative pairs ($M = .83$, $SD = .08$) for the same-exemplar, different state performance $t(18) = 3.37$, $p = .003$, $d = .77$, 95% CI [.25, 1.28]. We found that the collaborative HR was significantly better than the averaged noncollaborative performance. This indicates that, together, collaborators were able to correctly identify (state "old") which exemplars were studied and their correct state. This supports previous literature by Clark et al. (2000), which also found that the collaborative advantage stemmed from an increase in hit rates.

Next, we conducted a paired sample *t*-test to compare the FAR for the averaged noncollaborative pairs ($M = .22$, $SD = .09$) and the collaborative pairs ($M = .16$, $SD = .08$) for the same-exemplar, different state performance $t(18) = 4.12$, $p < .001$, $d = .95$, 95% CI [.39, 1.48]. The collaborative FAR was significantly lower for the collaborative condition compared to the averaged noncollaborative performance. This finding indicated that when individuals collaborated, they were able to successfully reject exemplar states that they did not study, which led to a reduction of incorrect selections. This decrease in false alarms has also been found by (Rajaram, Maswood, & Pereira-Pasarin, 2020) and supports collaborative pairs' ability to reduce the number of errors by error-pruning.

Another way to understand how collaborative groups made their decisions was to look at how often collaborative groups went with the decision of their best group member. In other words, did the benefit of collaboration arise from one partner's ability to lift the group's performance? We determined which group's member had the better performance (who had the best highest *d-prime* score) when making their noncollaborative decisions. We then used the best group member for each pair to examine the proportion of disagreements that were resolved that were consistent with the best group members' recognition decision. We found that collaborative pairs resolved their disagreements 43% of the time, which indicates that the collaborative advantage was not solely due to the group's best member.

Lastly, we wanted to know if the group's best member from the noncollaborative condition performed better than the collaborative condition. We ran paired samples t-tests on *d-prime* and found that performance was better for the collaborative condition ($M = 2.08$, $SD = .53$) in comparison to the group best member ($M = 1.81$, $SD = .53$) *d-prime*: $t(18) = 2.73$, $p = .014$, $d = .63$, 95% CI [.125, 1.11]. Taken together, it appears that the collaborative advantage was due to factors beyond averaging performance or by simply following the group's best member. Instead, it seems like the communication exchange was a significant contributor driving the group's ability to select correct "old" exemplars and correctly reject "new" exemplars.

In sum, participants performed well on both tasks. Performance for the exemplar task was particularly high for all conditions, which resulted in a ceiling effect. The alone condition was more conservative in their decision-making during the exemplar task compared to the collaborative decision. For the exemplar-state task, when participants were presented with new exemplars during the test, there was no difference in any of the conditions for recognition memory performance. Of primary interest, for the exemplar-state task, when participants were presented with old exemplars with new states, our results indicated that the collaborative condition had a significantly higher *d-prime* compared to the alone condition, indicating that the collaborative condition successfully rejected the old exemplar, new state stimuli. This finding shows that collaboration, once again, has been shown to aid exemplar-state binding. This also provided support for exemplars and their state information being bound because there is evidence that the collaborative groups are able to remember both sets of information well-enough to decipher between rejecting the old exemplar-states that they had studied and old exemplars in new states.

General Discussion

In our Experiment 1A, we sought to replicate Utochkin and Brady 's (2019) Experiment 1A by having participants work alone to complete the exemplar and exemplar-state 2-alternative forced-choice tests. Our findings replicated Utochkin and Brady (2019): we found that individuals had good memory for exemplars they studied and good memory for the state of exemplars being in either the same or differing states. But importantly, when participants needed to identify the exemplar-state information together, they struggled to bind this information and performed near-chance performance.

In Experiment 1B, we tested whether collaboration could enhance memory for exemplar-state binding at retrieval. Participants who remembered individually and then worked collaboratively, demonstrated the ability to remember exemplars, and the states of exemplars they studied. But, surprisingly, they were able to successfully remember this information as a bound unit. In addition, we found that collaborating led to better recognition performance in comparison to noncollaborative decisions for exemplars, correctly choosing the same state of two exemplars when that was what participants initially studied, and with higher proportions for correctly identifying the same state and different state exemplars overall.

Although our Experiment 1A findings contradicted our findings in Experiment 1B, there was one key difference between these experiments; each task (exemplar and exemplar-state) was divided into two blocks to create our within-subjects design. This change in study design cut both the study and test phases in half, which greatly benefited participants' performance. Even individuals working alone were now able to successfully bind exemplars to their studied states. Performance for binding exemplars to their correct state is not perfect, but we found strong support that individuals can successfully bind when the demand for retrieving bound information is decreased.

In Experiment 2, we set out to further understand the benefit that collaboration has in reference to the exemplar-state binding performance. Additionally, we sought to replicate the findings of Experiment 1B, which showed that people working alone could also successfully bind exemplars to their states, despite giving participants a more challenging task. This time, using an old/new recognition test, we found that participants who collaborated were able to discriminate above chance performance for both tasks. These findings indicate that collaborating pairs had the ability to decipher which exemplar state they studied and did not buy into the trap of selecting “old” just because the exemplar was old. An interesting thing to note is that, given our within-subjects design, the same participants who made up the alone condition were unable to make this correct discrimination during the old/new exemplar-state task as often as the collaborative condition did. This is evidence that the ability to collaborate enhances the ability to make these decisions.

We also examined how collaborating helped to bind exemplar-state information for Experiments 1B and 2. In Experiment 1B, we sought to understand how collaborative groups made their decisions when they were required to identify two exemplar-states that were studied in different states (which requires binding). To evaluate this, we looked at how often collaborative groups went with the decision of the best-performing group member. Collaborative pairs went with the decision of their group's best performer 50% of the time, which provides evidence that the collaborative advantage is not due to solely relying on the best person of the group to make the decisions. Interestingly, in Experiment 1B, we found that when we compare the group's best performer to the collaborative condition, the group's best performer did just as well as the collaborative condition at identifying studied exemplars presented in different states. We explored this finding and found that the best performer changed their correct answers in the prior collaborative decision 17% of the time when they collaborated. Said differently, the best group performer changed their correct response 17/50 or approximately 33% of the time. This willingness to change from a correct response could have occurred for multiple reasons. For example, perhaps the best group performer wasn’t 100% sure about their decision and was open to switching if their partner made a strong argument about remembering it. Or maybe the best group member wanted to let their partner provide some input, so they went along with their group members' choice to maintain a cooperative and friendly collaborative exchange. The 2-AFC task in Experiment 1 required participants to get the top and bottom exemplars correct, consequently, even if collaborative groups decided to take turns when they were unsure, this would negatively impact overall performance because both choices needed to be right.

In Experiment 2, we also sought to understand how collaboration was helping to improve exemplar-state binding. Consequently, we examined the HR and FAR of the exemplar-state task when participants were asked to make an old/new recognition decisions involving the same studied exemplars but shown in different states. We found that collaborative pairs' HR was significantly better than the averaged noncollaborative performance when it came to successfully discriminating exemplar-state. This finding replicates previous literature by Clark et al. (2000). Additionally, we found that the FAR was significantly lower for the collaborative pairs compared to the averaged noncollaborative pairs, which also support previous literature (Rajaram, Maswood, & Pereira-Pasarin, 2020), and indicated that collaborative groups were able to error-prune incorrect exemplar-states and limit the number of misidentifications made. We again compared the group's best performer to the collaborative group's performance. However, this time we found that the collaborative group had better *d-prime* scores than the group's best performance. We were able to eliminate the possibility of the best group performer lifting the collaborative group's performance because we found that the collaborative pairs resolved their disagreements by taking the groups best members response 43% of the time. This indicates that the collaborative advantage was not exclusively attributable to the groups' best member. Taken together, it seems that collaborative groups did not rely on the group's best performer for all of the correct answers. Instead, based on the percentages, it seems the collaborative groups had a fairly even exchange of decisions from each member.

We analyzed the group's best performers' performance to understand if this individual was driving the group's decisions. However, given that confidence for both group members was collected individually this could be another approach to analyze the data. For example, the group member with the highest reported confidence could be examined to see if they often led the decision-making in the group. This could be an interesting future direction.

We next want to discuss some additions and changes made in the three experiments that could have played a part in our findings. Experiments 1A and 1B followed the procedure used in Utochkin and Brady 's (2019) Experiment 1. However, after each alone decision, participants provided a confidence judgment. It is possible that the inclusion of confidence judgments could have made participants more mindful of trying to use recollection when making (or justifying) decisions. However, it is important to note that our Experiment 1A included the exact task procedures and the same amount of trials as Utochkin and Brady (2019), and it included collected confidence, *and* we replicated their original findings with comparable means.

For Experiment 2, we decided to cut the amount of studied and tested trials in half because we believed the old/new recognition task was going to be too long and too difficult. In hindsight, it would have been better to include more studied stimuli and test trials so that we could have found greater variability in scores for both tasks, particularly in the exemplar task.

Given that we found that the number of stimuli studied per trial appeared to impact all three experiments, future research should examine at what point binding performance begins to suffer. Because Utochkin and Brady (2019) found that individuals struggled to bind exemplars to their correct state which is also what we found in our Experiment 1A. However, in our Experiments 1B and 2, we found evidence that individuals are able to successfully bind this information when participants study and are tested on fewer stimuli indicating that exemplar-state information is bound however there may be a cap to this ability. To understand more about when binding performance seems to decline researchers should manipulate the amount of studied and tested stimuli.

While our studies were able to provide valuable insights into the advantage that collaborating can have on binding exemplars to their state information at retrieval, there are many potential directions for future research. For example, conducting a microanalyses study could help understand the social exchanges and conversations that help collaborative groups reach better performance. This could be done by recording the collaborative tasks or transcribing audio from the collaborative tasks and then coding how collaborative pairs reach their decisions, solve disagreements, and allow us to gauge individuals' willingness to change their decisions. All of this information would speak to social factors that are hard to estimate when you only have performance variables. Additionally, another avenue for future studies could be to examine bound information by targeting source monitoring, which requires binding. Source monitoring plays an important role in domains such as eyewitness memory and the spreading of misinformation. Experiments focusing on memory for source information might further enhance the collaborative binding advantage because stimuli used in these types of studies tend to be more meaningful to participants.

In summary, we found evidence that exemplar-state binding is possible by individuals who remember individually and that binding performance can be improved when individuals collaborate to remember. However, it seems apparent that the amount of information participants are required to bind impacts this ability. When exemplar-state binding is possible, and when it is not, needs to be further explored to understand better individuals deficiencies and capabilities when binding real-world objects to their correct features.

References

Atkinson, R. C., & Juola, J. F. (1973). Factors influencing speed and accuracy of word recognition. *Fourth International Symposium on Attention and Performance*, (February), 583–611.

Barber, S. J., Rajaram, S., & Aron, A. (2010). When two is too many: Collaborative encoding impairs memory. *Memory and Cognition*, *38*(3), 255–264. https://doi.org/10.3758/MC.38.3.255

Basden, B. H., Basden, D. R., Bryner, S., & Thomas, R. L. (1997). A comparison of group and individual remembering: Does collaboration disrupt retrieval strategies? *Journal of Experimental Psychology: Learning Memory and Cognition*, *23*(5), 1176–1189. https://doi.org/10.1037/0278-7393.23.5.1176

Basden, B. H., Reysen, M. B., & Basden, D. R. (2002). Transmitting false memories in social groups. *American Journal of Psychology*, *115*(2). https://doi.org/10.2307/1423436

Benjamin, A. S., Diaz, M., & Wee, S. (2009). Signal Detection With Criterion Noise: Applications to Recognition Memory. *Psychological Review*, *116*(1). https://doi.org/10.1037/a0014351

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2013). Real-world objects are not represented as bound units: Independent forgetting of different object details from visual memory. *Journal of Experimental Psychology: General*, *142*(3), 791–808. https://doi.org/10.1037/a0029649

Brady, T. F., Konkle, T., Alvarez, G. A., Oliva, A., Boujut, A., Clarys, D., … Yonelinas, A. P. (2020). Yonelinas, 1999.pdf. *Cognition*, *24*(4), 95–100. https://doi.org/10.1016/0749-5978(89)90040-X

Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. A. (2013). Visual Long-Term Memory Has the Same Limit on Fidelity as Visual Working Memory. *Psychological Science*, *24*(6), 981–990. https://doi.org/10.1177/0956797612465439

Chalfonte, B. L., & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory and Cognition*, *24*(4), 403–416. https://doi.org/10.3758/BF03200930

Clark, S. E., Hori, A., Putnam, A., & Martin, T. P. (2000). Group Collaboration in Recognition Memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, *26*(6), 1578–1588. https://doi.org/10.1037/0278-7393.26.6.1578

Daselaar, S., & Cabeza, R. (2014). Age-related decline in working memory and episodic memory: Contributions of the prefrontal cortex and medial temporal lobes. *The Oxford Handbook of Cognitive Neuroscience, Vol. 1: Core Topics.*

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*(1). https://doi.org/10.1037/h0046671

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: a three-component model. *Trends in Cognitive Sciences*, *11*(9), 379–386. https://doi.org/10.1016/j.tics.2007.08.001

Draschkow, D., Reinecke, S., Cunningham, C. A., & Võ, M. L. H. (2019). The lower bounds of massive memory: Investigating memory for object details after incidental encoding. *Quarterly Journal of Experimental Psychology*, *72*(5). https://doi.org/10.1177/1747021818783722

Dunn, J. C. (2004). Remember-Know: A Matter of Confidence. *Psychological Review*, *111*(2). https://doi.org/10.1037/0033-295X.111.2.524

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Gauthier, I., & Tarr, M. J. (2016). Visual Object Recognition: Do We (Finally) Know More Now Than We Did? *Annual Review of Vision Science*, *2*, 377–396. https://doi.org/10.1146/annurev-vision-111815-114621

Green, D. G., & Swets, J. A. (1966). *Signal detection theory and psychophysics. New York, NY: John Wiley and Sons Inc. Wiley & Sons, Inc.*

Gronlund, S. D., & Benjamin, A. S. (2020). *The new science of eyewitness memory*. *Psychology of Learning and Motivation - Advances in Research and Theory* (K. Federme, Vol. 69). https://doi.org/10.1016/B978-0-12-394393-4.00008-X

Hicks, J. L., & Marsh, R. L. (1998). A decrement-to-familiarity interpretation of the revelation effect from forced-choice tests of recognition memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, *24*(5). https://doi.org/10.1037/0278-7393.24.5.1105

Hicks, J. L., & Starns, J. J. (2015). Using multidimensional encoding and retrieval contexts to enhance our understanding ofstochastic dependence in source memory. *Psychology of Learning and Motivation - Advances in Research and Theory*, *62*, 101–140. https://doi.org/10.1016/bs.plm.2014.09.004

Hinsz, V. B. (1990). Cognitive and Consensus Processes in Group Recognition Memory Performance. *Journal of Personality and Social Psychology*, *59*(4), 705–718. https://doi.org/10.1037/0022-3514.59.4.705

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5). https://doi.org/10.1016/0749-

596X(91)90025-F

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, *110*(3). https://doi.org/10.1037/0096-3445.110.3.306

Jang, Y., Wixted, J. T., & Huber, D. E. (2009). Testing Signal-Detection Models of Yes/No and Two-Alternative Forced-Choice Recognition Memory. *Journal of Experimental Psychology: General*, *138*(2). https://doi.org/10.1037/a0015525

Macmillan, N. A., & Creelman, C. D. (2004). *Detection Theory: A User's Guide: 2nd edition*. *Detection Theory: A User's Guide: 2nd edition*. https://doi.org/10.4324/9781410611147

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*(3). https://doi.org/10.1037//0033-295x.87.3.252

Meade, M. L., & Roediger, H. L. (2002). Explorations in the social contagion of memory. *Memory & Cognition*, *30*(7), 995–1009. https://doi.org/10.3758/BF03194318

Nader, K. (2003). Memory traces unbound. *Trends in Neurosciences*, *26*(2). https://doi.org/10.1016/S0166-2236(02)00042-5

Rajaram, S. (2011). Collaboration both hurts and helps memory: A cognitive perspective. *Current Directions in Psychological Science*, *20*(2), 76–81. https://doi.org/10.1177/0963721411403251

Rajaram, S., Maswood, R., & Pereira-Pasarin, L. P. (2020). When social influences reduce false recognition memory: A case of categorically related information. *Cognition*, *202*(August 2019). https://doi.org/10.1016/j.cognition.2020.104279

Rajaram, S., & Pereira-Pasarin, L. P. (2007). Collaboration can improve individual recognition memory: Evidence from immediate and delayed tests. *Psychonomic Bulletin and Review*, *14*(1), 95–100. https://doi.org/10.3758/BF03194034

Rajaram, S., & Pereira-Pasarin, L. P. (2010). Collaborative memory: Cognitive research and theory. *Perspectives on Psychological Science*, *5*(6), 649–663. https://doi.org/10.1177/1745691610388763

Roediger, H L, Meade, M. L., & Bergman, E. T. (2001). Social contagion of memory. *Psychonomic Bulletin & Review*, *8*(2), 365–371. https://doi.org/10.3758/BF03196174

Roediger, Henry L., & McDermott, K. B. (1995). Creating False Memories: Remembering Words Not Presented in Lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory and Cognition*, *28*(6). https://doi.org/10.3758/BF03209339

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Standing, L. (1973). Quarterly Journal of Experimental Psychology Learning 10000 pictures LEARNING 10,000 PICTURES. *Quarterly Journal of Experimental Psychology*, *25*(25).

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*(2). https://doi.org/10.1016/S0959-4388(96)80070-5

Utochkin, I. S., & Brady, T. F. (2019). Independent Storage of Different Features of Real-World Objects in Long-Term Memoryc. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0000664

Vollrath, D. A., Sheppard, B. H., Hinsz, V. B., & Davis, J. H. (1989). Memory performance by decision-making groups and individuals. *Organizational Behavior and Human Decision*

Processes, *43*(3), 289–300. https://doi.org/10.1016/0749-5978(89)90040-X

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*(3), 441–517. https://doi.org/10.1006/jmla.2002.2864

Yonelinas, A. P., Aly, M., Wang, W. C., & Koen, J. D. (2010). Recollection and familiarity: Examining controversial assumptions and new directions. *Hippocampus*, *20*(11), 1178–1194. https://doi.org/10.1002/hipo.20864

Zimmer, H. D., Mecklinger, A., & Lindenberger, U. (2012). *Handbook of Binding and Memory: Perspectives from Cognitive Neuroscience*. *Handbook of Binding and Memory: Perspectives from Cognitive Neuroscience*. https://doi.org/10.1093/acprof:oso/9780198529675.001.0001