Towards a Cognitive Model of Collaborative Memory

Willa M. Mannering
Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405 USA

Suparna Rajaram
Department of Psychology, Stony Brook University, Stony Brook, NY 11794 USA

Michael N. Jones
Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405 USA

Abstract

While humans routinely encode and retrieve memories in groups, the bulk of our knowledge of human memory comes from paradigms with individuals in isolation. The primary phenomenon of interest within the relatively new field of collaborative memory is collaborative inhibition: the tendency for collaborative groups to underperform in free recall tasks compared to nominal groups of the same size. This effect has been found in a variety of materials and group compositions (Rajaram & Pereira-Pasarin, 2010). However, the majority of research in this field is guided by verbal theories without formal computational models. In this paper we adapt the Search of Associative Memory (SAM; Raaijmakers & Shiffrin, 1981) model to collaborative free recall. We present a framework to scale SAM to collaborative paradigms with multiple SAM models working together. Our simulation results with the collaborative SAM model suggest that retrieval disruption, responsible for the part-set cuing effect in individuals, is also the cause of collaborative inhibition when multiple models are working together. Our work provides an existence proof that SAM can act as a unified theory to explain both individual and collaborative memory effects, and offers a framework for future predictions of scaling to increased group sizes, shared knowledge, and spread of false memories.

1. Introduction

Outside of the laboratory we regularly encode and retrieve memories in collaboration with others, but almost all empirical research in human memory has historically involved participants performing tasks in isolation. The experimental study of collaborative memory is a comparatively young field focused on revealing the cognitive mechanisms involved in group interaction in memory tasks. The field primarily scales up experimental paradigms and theories originating in individual memory to the collaborative group level.

The primary focus within the field is collaborative inhibition—the tendency for collaborative groups to underperform in free recall tasks compared to nominal groups of the same size. This effect is robust and has been found in a variety of materials and group compositions (Marion & Thorley, 2016; Rajaram & Pereira-Pasarin, 2010). Currently, there are several competing explanations for the recall deficit, but the explanation with the most empirical support is the
retrieval disruption hypothesis. This hypothesis posits that the deficit from collaboration occurs because individual retrieval strategies are disrupted during group activities (B. H. Basden, Basden, Bryner, & Thomas, 1997). While a large body of experimental research and verbal conceptual frameworks exist in collaborative memory, there are currently no formal computational models to guide the field. The goal of the current paper is a first attempt at modifying a prominent model of individual memory (the SAM model of Raaijmakers & Shiffrin, 1981), well validated at explaining the mechanisms responsible for phenomena in individual memory experiments, to the level of collaborative groups. We develop a collaborative framework and then test the ability of multiple SAM models performing tasks together to produce the patterns of collaborative inhibition seen in human data. Our simulations support the retrieval disruption hypothesis, and provide a formal framework going forward to unify individual and collaborative memory research.

2. Collaborative Inhibition

The experimental paradigm typically used within the collaborative memory field is an extension of classic paradigms previously used and validated in the field of individual memory. This paradigm involves participants learning a list of words, performing a distractor task individually, and then performing a recall task (typically free recall or cued recall) together in small groups (Harris, Paterson, & Kemp, 2008). As expected, collaborative groups perform better in the recall task than individuals. However, to compare group performance, collaborative group recall must be compared to nominal group performance, not individual performance. In both collaborative and nominal group conditions, subjects learn a list of items individually in the study phase. Then, in the collaborative group condition, subjects are asked to work together with other group members to recall items on the list. The collaborative group response is calculated by counting all non-overlapping responses produced by the group. In the nominal group condition, subjects are asked to recall items on the list individually and do not recall together. The nominal group response is calculated by counting the total, non-overlapping responses produced by individual group members. When collaborative group recall performance is compared to nominal group recall performance in this way, there is a significant deficit in recall in the collaborative group (B. H. Basden et al., 1997; Weldon & Bellinger, 1997)—called collaborative inhibition.

2.1 Mechanistic Hypothesis of Collaborative Inhibition

There are three viable theories explaining the collaborative inhibition effect: social factors, the retrieval disruption hypothesis, and the production blocking hypothesis. The social factors hypothesis posits that factors, such as social loafing, are the primary cause of collaborative inhibition. Social loafing as a possible mechanistic explanation for collaborative inhibition is implied by previous group research in a wide variety of fields that show a similar loss of individual productivity (Diehl & Stroebe, 1987; Ingham, Levinger, Graves, & Peckham, 1974; Latane & Nida, 1981). However, while this hypothesis seems intuitive given the social nature of
collaboration, little experimental evidence has been found to support it (Andersson, Hitch, & Meudell, 2006; Weldon, Blair, & Huebsch, 2000).

The most popular mechanistic hypothesis is the retrieval disruption hypothesis which posits that the deficit from collaboration occurs because individual retrieval strategies are disrupted during group recall (B. H. Basden et al., 1997). According to this hypothesis, each group member develops a unique strategy of organizing information in memory during the study phase of a recall task which is then disrupted by mismatched cues from other group members when asked to recall in a group. This hypothesis originates from a mechanistic explanation for the part-set cuing effect found in the individual memory literature.

The individual memory analogue to collaborative inhibition is commonly believed to be the part-set cuing effect (Andersson et al., 2006; B. H. Basden et al., 1997; B. H. Basden, Basden, & Henry, 2000). Typically, when an individual is asked to use cues to aid recall, their recall performance increases (Tulving, 1974). However, the part-set cuing effect produces the opposite. When an individual is presented with a random selection of a memorized list as cues, their recall for the remaining words on the list is inhibited (Nickerson, 1984; Slamecka, 1968). Crucially, the part-set cues must be a random subset of the study list for the effect to occur. It is hypothesized that when randomly chosen, the part-set cues interfere with the subject’s internal organization of the study list, thus interrupting their idiosyncratic retrieval strategy (D. R. Basden & Basden, 1995). It is theorized that in a collaborative setting, group members provide part-set cues for others in the group—causing collaborative inhibition. B. H. Basden et al. (1997) were the first to provide experimental evidence tying collaborative inhibition to the part-set cuing effect and supporting retrieval disruption as a mechanistic explanation for collaborative inhibition. Additionally, there is some evidence to show that retrieval inhibition occurs during collaborative recall (Barber, Harris, & Rajaram, 2015). Retrieval inhibition occurs when unrecalled items are not just disrupted during recall but also inhibited such that they are less likely to be recalled later when the influence of collaboration is removed. Nonetheless, a majority of evidence favors the retrieval disruption hypothesis.

While the retrieval disruption hypothesis has the majority of supporting evidence in the literature, there are some cases where the production blocking hypothesis cannot be ruled out as a mechanistic explanation for collaborative inhibition. The production blocking hypothesis posits that the process of waiting to respond while other group members produce responses inhibits, or blocks, the ability to produce information (Diehl & Stroebe, 1987). This hypothesis has comparatively more empirical support than the social factors hypothesis but less than the retrieval disruption hypothesis. While there is some evidence of dual-processing accounts involving both production blocking and retrieval disruption, most studies involving production blocking during collaborative recall have concluded that while the production blocking hypothesis and retrieval inhibition cannot be ruled out, these mechanisms are insufficient to fully account for collaborative inhibition (Andersson et al., 2006; Finlay, Hitch, & Meudell, 2000).
3. Modeling Collaborative Memory

Currently, research that takes advantage of social media information such as community detection, “fake news” detection, topic modeling, misinformation detection and prevention dominates the group behavior literature. This research stems from the fields of network science and linguistics and tends not to incorporate or consider cognitive concerns in their models. Until now, the only attempt at modeling collaborative memory was made by Luhmann and Rajaram (2015) whose main goal was to model information transmission at network-scale by taking an agent-based modeling approach. Though their main goal was not to model collaborative inhibition, during the verification phase of their model, the authors were able to produce collaborative inhibition when groups of 3 agents were tasked with performing collaborative recall. Additionally, they were able to model some predictions of the collaborative memory field, namely the effect of group size on collaborative memory. However, while this model included psychologically based agents that were able to encode and retrieve memories, the main goal of the study was to examine the effect of information transmission on network behaviors. This model also did not aim to synthesize both individual and collaborative memory outcomes which is the broader goal of the current enterprise.

Because modeling collaborative inhibition was not the main focus of the Luhmann and Rajaram (2015) study, there were a few simplifications and implementation choices which are inconsistent with the individual or collaborative memory literature. First, the agents performed the collaborative recall task in a turn-taking manner to avoid complications of free-for-all recall. While this style of collaboration has some experimental precedent (B. H. Basden et al., 1997), the majority of collaborative memory studies use the free-for-all method because the turn-taking method tends to increase memory intrusions (Meade & Roediger, 2009; Rajaram & Pereira-Pasarin, 2010). Second, while the authors found evidence of collaborative inhibition and emergence of memory similarity during group recall, their model does not explicitly address how the collaborative agents’ memories become more similar. In their model, memory similarity could be detected during the collaboration phase. In behavioral experiments memory similarity is detected in post-collaborative recall, where the cascading effects of retrieval disruption - primarily responsible for collaborative inhibition - as well as re-exposure to others’ recall during collaboration are posited to homogenize memories as detected in the recall performance of former group members (e.g., in the shared/collective memory measures, e.g., Congleton & Rajaram, 2014). The SAM model aims to explicate the mechanism primarily responsible for collaborative inhibition and that also contributes to memory homogenization, i.e., retrieval disruption. Thus, SAM offers a more in-depth and established cognitive model and would be more useful for studying the inhibitory effects of retrieval disruption on collaborative recall.

3.1 Search of Associative Memory (SAM)

An ideal model to study the cognitive mechanisms at play during collaborative memory is the Search of Associative Memory (SAM) model (Raaijmakers & Shiffrin, 1981). The motivation for using SAM over other possible cognitive models is as follows. First, SAM is well-studied and is the most widely used model in episodic memory research (Wilson & Criss, 2017; Wilson, Kellen, & Criss, 2020). Second, SAM is one of the only cognitive models that has previously been shown
to successfully model the part-set cuing effect in individual memory (Raaijmakers & Shiffrin, 1981). Finally, the architecture of the model affords a coherent framework to extend to multiple models working collaboratively. If we can modify SAM to explain collaborative phenomena without changing the fundamental architecture, any of the SAM models in isolation would still retain the explanatory power for the range of behavioral phenomena in individual memory paradigms thus producing a unified account of both individual and collaborative memory phenomena.

SAM is a cue-dependent probabilistic search theory of retrieval and is typically applied to simulations of free recall and free recall with cues. The model makes use of a two-stage memory system: short-term memory and long-term memory. The short-term memory system is where processes such as encoding and rehearsal are carried out. This system is limited in capacity and uses a buffer rehearsal system so that items that co-occur in the short-term buffer for longer tend to have higher associations with each other when committed to long-term memory. The long-term memory system is where information is transferred from short term memory and stored permanently. The long-term memory storage consists of an association matrix of study items and environmental context (item-context information) and item to item-plus-context information (item-item information). The strength of the association between item-context pairs is proportionate to the amount of time an item remained in the short-term memory buffer. The strength of association between item-item pairs is proportionate to the amount of time two items co-occurred in the short-term memory buffer. In this model, item information represents information that would allow a subject to recall the name of an item while context information represents any information available during encoding that’s not directly related to recalling an item name, such as emotions, sensations, or environmental details. Learning associations can occur during the retrieval process in addition to the initial encoding during study.

Retrieval from long-term memory in SAM is a probabilistic, cue-dependent process. When searching through long-term memory during recall, the model uses cues assembled in short-term memory as probes for long-term memory. These cues include context cues, C_T, and words from the study list, W_{1T}, W_{2T}, ... W_{nT}. The T subscript is used to indicate a cue at test. During retrieval, a probe set consists of only C_T or C_T with a word cue. The probability that an item will be sampled from any given probe is dependent on the associations stored between the memory probe and the items stored in memory (see Equation 1a and 1b). Equation 1a gives the probability of sampling a word, W_{iS}, using only context, C_T, as a memory probe. Equation 1b gives the probability of sampling a word, W_{iS}, given both context, C_T, and a word cue, W_{kT}, as a memory probe. The S subscript indicates the item as it is stored in memory. The S() function represents the strength that a given cue will sample an item from memory.

\[
P_S(W_{iS}|C_T) = \frac{S(C_T,W_{iS})}{\sum_{j=1}^{n} S(C_T,W_{jS})} \quad \text{(1a)}
\]

\[
P_S(W_{iS}|C_T,W_{kT}) = \frac{S(C_T,W_{iS})S(W_{kT},W_{iS})}{\sum_{j=1}^{n} S(C_T,W_{jS})S(W_{kT},W_{jS})} \quad \text{(1b)}
\]
Once an item is sampled from memory, the recovery process begins. The probability of successfully recovering sufficient information to name the memory item is based on the association strength between the memory probe and the sampled memory item (see Equation 2a and 2b). Equation 2a shows the probability of successfully recovering a word image, $W_i$, given only context as a memory probe and Equation 2b shows the probability of successfully recovering a word image given both context and a word cue as a memory probe.

$$P_R(W_i|C_T) = 1 - \exp\{-S(C_T, W_{IS})\} \quad (2a)$$

$$P_R(W_i|C_T, W_{kT}) = 1 - \exp\{-S(C_T, W_{IS}) - S(W_{kT}, W_{IS})\} \quad (2b)$$

The original SAM model is able to account for various free recall effects established by behavioral research including list length effects, presentation duration effects, serial position curves, extended recall, and repeated recall. Additionally, the model is easily adapted to simulate categorized free recall and cued recall (Raaijmakers & Shiffrin, 1980). While these feats make SAM a viable cognitive model of recall, the most useful ability for modeling collaborative memory is the previous success in modeling the part-set cuing effect.

### 3.1.1 Part-set Cuing Effect in SAM

In the original formulation of the SAM model, Raaijmakers and Shiffrin (1981) were able to successfully model the part-set cuing effect. Typically, the SAM model uses internal cues to perform free recall. In the case of part-set cuing, the model must first use external cues, provided by the experimenter, to perform recall before transitioning to internal cues once the external cues have been used. The first simulation compared recall performance of a control model, which was given no external cues, and a cued model, which was given external cues. Both models were trained on lists of size 30 and with a presentation time of 2 seconds per item. The cued model was given 15 cues during recall. Additionally, the interitem strength parameter of the models was varied during this experiment to better adhere to the original experiment detailing the part-set cuing effect (Slamecka, 1968). When the control model and the cued model performances were compared, Raaijmakers and Shiffrin (1981) found that the cued models recalled fewer critical items (non-cue items) across all values of the interitem strength parameter, as was predicted by the part-set cuing effect.

This result was interesting not only because the cued group performed worse than the control group but because the performance hit occurred in spite of a factor that aids the cued group. A recovery rule of the SAM model makes it so the probability of recovery after a memory item is sampled using an item-plus-context cue is greater than when a memory item is sampled using a context only cue. That is, during the memory search process, the model can either use context cues or previously recalled item cues to further aid in memory search. In the original model implementation, the model would typically begin free recall by using a context cue (stored during
the study phase) to probe its long-term memory storage. Then, once the model successfully recovers an item from memory, that item is then used in conjunction with context as the next memory probe. So, it is somewhat surprising, given the fact that the recovery rules of the model favor item-plus-context cues, that the cued models would consistently perform worse than the control models.

3.1.2 Adapting SAM to Collaborative Free Recall

Given its success at modeling the part-set cuing effect in individual memory, adapting the SAM model to collaborative recall could provide valuable insights into the cognitive mechanisms behind collaborative inhibition. There were two ways to adapt the SAM model to perform a collaborative recall task. First, a turn-taking procedure could have been implemented that’s similar to the experimental design used by B. H. Basden et al. (1997) and the modeling approach taken by Luhmann and Rajaram (2015). However, as mentioned previously, this method is less common in the literature because the turn-taking method has been shown to increase memory intrusions (Meade & Roediger, 2009; Rajaram & Pereira-Pasarin, 2010). The second method, which we implemented, is to allow a free-for-all recall method. Because of technology limitations, it is difficult to instantiate multiple SAM models and have them interact in real time like humans do. However, there are ways to work around this problem and effectively simulate the free-for-all response method. Additionally, while this method is comparatively more difficult than the turn-taking method, it is more similar to the method used in the majority of the collaborative recall behavioral experiments.

Figure 1 is a flowchart of how we simulated the free-for-all method. First, we created a shared memory buffer, called the group response, between two or more models which represents words “spoken” aloud by the models. To begin, the models perform context recall separately. The first response produced by any of the models is added to the shared buffer and the other models in the collaborative group are able to access this response. Then, the models all use the new response in the shared buffer as a cue for recall. Similar to the context recall phase, the models all perform cued recall separately and the response produced first is added to the shared buffer. The models continue using new responses as cues for recall until all models reach $L_{\text{max}}$ at which point all models return to context recall. This continues until all models reach $K_{\text{max}}$ and the memory search ends. In this way, we are able to simulate the free-for-all method of collaborative recall with the SAM model.
3.1.3 Fitting SAM to Collaborative Recall Data

After implementing a version of SAM able to perform collaborative recall, we fit the model to averages of experimental data of uncategorized list recall (Weldon & Bellinger, 1997) and categorized list recall (B. H. Basden et al., 1997). While fitting the model, we allowed 5 parameters to vary between the collaborative and nominal groups and between categorized and uncategorized conditions. Those 5 parameters were $sam_e$ (the incrementing parameter for context-to-word association), $sam_f$ (the incrementing parameter for word-to-word association), $sam_g$ (the incrementing parameter for word-to-self association), $K_{max}$ (the maximum number of retrieval failures before searching stops), and $L_{max}$ (the maximum number of retrieval attempts).

Figure 2 shows the results of fitting the model to individual, nominal, and collaborative recall on an uncategorized list. Figures 3 and 4 show the results of fitting the model to individual, nominal, and collaborative recall on categorized lists: one list with 6 categories that had 15 words in each (Figure 3) and another list with 15 categories that had 6 words each (Figure 4).
Figure 2. SAM model fit to uncategorized list data taken from the original Weldon and Bellinger (1997) paper detailing collaborative inhibition. Subjects were tested in groups of 3 on a list of 40 unrelated words.

Figure 3. SAM model fit to categorized list data from B. H. Basden et al. (1997). Subjects in groups of 3 were asked to recall from a list of 90 words grouped into 6 total categories with 15 items in each category. The larger category size results in a more prominent collaborative inhibition effect than in Figure 4.
Figures 2-4 show that not only is the SAM model able to accurately reproduce the collaborative inhibition effect found in experimental data, but that it also supports the retrieval disruption hypothesis. The experimental data from B. H. Basden et al. (1997) used to fit the model (Figures 3 and 4) supports this as it shows that collaborative inhibition is stronger when study materials are less organized. In the first condition (Figure 3) study materials are less organized because the group sizes are larger—allowing room for more idiosyncratic organization within categories. In the second condition (Figure 4) the study materials are more organized because the group sizes are smaller—allowing less room for idiosyncratic organization within categories. When the internal organization of study items is dissimilar between group members, collaborative inhibition increases because the cues from other group members are more likely to disrupt individual search strategies. Because the SAM model is able to reproduce this effect, it supports the retrieval disruption hypothesis. Future research will be dedicated towards teasing apart the production blocking and retrieval disruption hypotheses and investigating retrieval inhibition.

In addition to fitting the SAM model to aggregate experimental data, we briefly looked into the effect collaboration had on the similarity of model memory. Luhmann and Rajaram (2015) found that their model was able to produce collaborative inhibition, but they suggested that this was due to the agents’ memories becoming more similar as they collaborated. There is support for this in the experimental literature (Congleton & Rajaram, 2014), suggesting as people collaborate together, their memories become more similar—giving rise to shared memory. Given the increase in memory similarity due to collaborative recall, it seemed pertinent to determine whether SAM exhibited such properties as well.
Figure 5 is the result of calculating the average correlation between two model’s word association matrices every time a new item was added to the group recall buffer. That is, each time a group member successfully “said” a word (and each group member’s internal associations were updated), we calculated the correlation between 2 out of 3 group member’s association matrices. As shown in Figure 5, the average word association correlation between two models in a collaborative group does increase during recall. Additionally, the average increase in word association correlation across 200 collaborative recall groups was $r = 0.09$.

4. Discussion

The implications of collaborative memory research are much larger than participants recalling lists of words in experimental settings. The cognitive mechanisms being studied by this basic science are the same that play a role in crucial applied phenomena such as the spread of misinformation, memory contagion, fake news, eyewitness testimony, and even conspiracy theories. To date, there are no formal computational frameworks within which to understand how the memory mechanisms of individuals interact to produce emergent phenomena when collaborating. In this paper we took a first step towards this goal by modifying the well-validated SAM model of Raaijmakers and Shiffrin (1981), and providing an existence proof that multiple SAM models working together can produce the basic patterns of collaborative inhibition seen in experimental data. The results of the simulation suggest support for the retrieval disruption hypothesis in the
collaborative memory literature. In addition to basic uncategorized lists, collaborative SAM naturally produces the patterns seen in categorized lists, namely greater collaborative inhibition when study materials are less organized. Importantly, each SAM model in isolation would still retain the explanatory power for the range of behavioral phenomena in individual memory paradigms, providing a unified model to understand both individual and collaborative memory.

With our SAM-based framework validated on the standard patterns of collaborative inhibition seen in the literature, we can now go forward fitting individual data from specific experimental manipulations and using the optimal parameters to better understand the internal mechanisms producing the behavioral phenomena. Having a formal computational framework allows the field to generate new predictions and experiments to help differentiate between theories of group memory that are currently unresolvable using only experimental data. In addition, the number of SAM models interacting can be scaled up significantly, a feat not possible in experimental studies, allowing us to better understand how individual cognitive mechanisms give rise to group memory inhibition at scales closer to what we see in shared learning on massive social media discussions.
References


