



Memory transmission in small groups and large networks: An empirical study

Vael Gates¹ · Jordan W. Suchow² · Thomas L. Griffiths³

Accepted: 26 September 2021 / Published online: 28 October 2021
© The Psychonomic Society, Inc. 2021

Abstract

When people try to remember information in a group, they often recall less than if they were recalling alone. This finding is called *collaborative inhibition*, and has been studied primarily in small groups because of the difficulty of bringing large groups into the laboratory. To study the dynamics of collaborative inhibition in large groups (Luhmann & Rajaram, *Psychological Science*, 26, 1909–1917, 2015) constructed an agent-based model that extrapolated from previous laboratory experiments with small groups. The model predicts that collaborative inhibition should increase with group size. Here, we evaluate this model by recruiting a large number of participants using crowdsourcing, allowing us to replace the artificial agents in the model with people to study collaborative memory at larger scales. Our empirical results did not match the model predictions: there was no evidence for an increase in collaborative inhibition with group size, despite substantial power to detect such an effect. These findings motivate further empirical work to elucidate the mechanisms of collaborative memory.

Keywords Collaborative memory · Collaborative inhibition · Crowdsourcing · Agent-based modeling

Our memories depend on the people around us and are reshaped and reformed as we interact. Long-term couples, for example, learn how to effectively access each other's memories, remembering details they would not have recalled alone (Wegner, Erber, & Raymond, 1991), and groups collectively form memories that define their values (e.g., Hirst & Echterhoff, 2012). However, sometimes our memories are reshaped in such a way that we remember less together than we would apart, an effect that psychologists have been actively investigating.

Imagine a group of people trying to recall a list of words collaboratively. The group as a whole would recall some number of words. Next suppose that each individual had instead tried to recall the list of words alone. Though the group would likely have recalled a greater number of words than any individual, the group may have performed

even better had they recalled the words individually and then mechanically combined the list, removing duplicates. Indeed, this precise comparison is often made in analyses of the well-established “collaborative memory” task, in which participants listen to a long list of items (often words) and then recall as many items as possible, either as a group or individually. The number of words recalled by the group is compared to the number of words recalled by the “nominal group”: the summed list of an equivalent number of individuals, with duplicate words removed. In the collaborative memory paradigm, nominal groups routinely outperform collaborative groups, a finding called *collaborative inhibition*. This effect has been replicated across many studies and variations on the paradigm (see Marion & Thorley, 2016; Rajaram & Pereira-Pasarin, 2010, for reviews).

The leading theory of collaborative inhibition is the retrieval disruption hypothesis (e.g., Basden, Basden, Bryner, & Thomas, 1997; Rajaram & Pereira-Pasarin, 2010). This hypothesis states that when initially listening to a wordlist, people form idiosyncratic representations of the words, influenced by factors such as the order of the words presented and their semantic relations. When recalling words alone, participants use their idiosyncratic representations to effectively recall the words. However, when placed in groups, other participants, who organized

✉ Vael Gates
vgates@berkeley.edu

¹ Department of Psychology, University of California, Berkeley, Berkeley, CA, USA

² School of Business, Stevens Institute of Technology, Hoboken, NJ, USA

³ Department of Psychology, Princeton University, Princeton, NJ, USA

the words differently, disrupt a participant's recall, leading to reduced performance. This hypothesis predicts that when participants are encouraged to organize information in similar ways, collaborative inhibition will disappear. In fact, when participants are experts in their domain (Meade, Nokes, & Morrow, 2009) or are exposed to similarly ordered information (Finlay, Hitch, & Meudell, 2000), inhibition does not occur.

Psychologists have historically studied collaborative memory in small-scale experiments of dyads or triads in the lab. Bringing groups of participants into the lab to simultaneously recall information—often lists of words—is logistically complex, slow, and expensive. However, in today's world of online communities and internet forums, information is being recalled and exchanged with much larger groups than pairs or triads, and many of the findings from this small-scale work may not be applicable to our daily lives. As people form ever-larger social networks, understanding how information is recalled among large groups of people becomes increasingly important.

To address this lack of understanding of large groups, Luhmann and Rajaram (2015) developed an agent-based model to predict what performance might be like at larger group sizes, based on known factors from empirical small-scale experiments. Computational "agents" follow algorithms for memory storage and retrieval, and can participate in simulated experiments interacting with other agents (details of the model appear in the [Supplementary Material](#)). This agent-based model predicts that collaborative inhibition will increase as group size increases from 1 to 4 participants, as has been suggested by the experimental literature (Basden, Basden, & Henry, 2000; Thorley & Dewhurst, 2007). The model then predicts that collaborative inhibition will continue to increase with group size, until nominal recall reaches ceiling performance as the disruption of idiosyncratic recall strategies is compensated for by sheer group size (see Fig. 1a).¹

Previously, it was a logistical impossibility to experimentally study collaborative memory at this kind of scale. However, using web-based crowdsourcing tools such as Amazon Mechanical Turk it is possible to recruit and organize hundreds of online participants into real-time interactive chatrooms. Here we use this new approach to translate the agent-based model of Luhmann and Rajaram (2015) into an empirical experiment, testing its predictions about the

¹We present model predictions using identical parameter settings to Luhmann and Rajaram (2015) ($\alpha = 0.2$, $\beta = 0.05$ and $\gamma = 0.75$) except we use 60 recalled words rather than 40, because pilot studies indicated that participants were nearing ceiling performance when recalling fewer words (an analysis of the results of this change appears in the [Supplementary Material](#)). Note also that when recalling 60 words, the model prediction of increased collaborative inhibition with group size appears robustly across parameter settings (see [Supplementary Material](#)).

effect of group size on collaborative inhibition. When we simulate the model with parameter settings corresponding to our experimental task, it predicts nominal ceiling performance around a group size of 16. Our experiments thus used group sizes of 2 through 16, meaning that the model predicts collaborative inhibition should increase with group size. More precisely, it predicts an interaction between group size and recall method (collaborative or nominal), with the effect of recall method increasing with group size.

Experiment 1: Collaborative Inhibition in Small and Large Groups

Methods

All experiments were IRB-approved by University of California, Berkeley, Committee for Protection of Human Subjects/Office for Protection of Human Subjects, Protocol ID: 2015-12-8227, Protocol Title: Culture-on-a-chip Computing: Testing Evolutionary Hypotheses through Large-Scale Behavioral Simulations.

Participants A total of 1,138 participants were recruited through Amazon Mechanical Turk. Participants would occasionally repeat the task, as they could choose to complete the task again on Amazon Mechanical Turk despite written advisement against this. A total of 134 participants repeated the experiment more than once (14.6% of participants), and 30.3% of the data was generated by these participants.²

Participants were excluded from the experiment if they did not complete a pre-experiment arithmetic task and they did not contribute words in the main experiment. Participants waited in a virtual waiting room until the group size was met; experiments listed with less than the full group size contained excluded participants. There were two recall conditions in the experiment: collaborative and nominal. Sixteen participants were removed from the collaborative experiments for a total of 561 participants. Nominal groups were matched; thus 561 participants participated in the

²Participants participated an average of 1.22 times. The mean proportion of repeaters across group sizes (both nominal and collaborative) was as follows: group size of 2: 0.33 ± 0.35 (SD), group size of 3: 0.39 ± 0.31 , group size of 4: 0.27 ± 0.21 , group size of 8: 0.27 ± 0.20 , and group size of 16: 0.27 ± 0.09 . The participants who repeated the nominal experiments did not show improvement over time, despite having seen the same wordlists: the correlation between number of repetitions and number of words recalled was $r = -0.03$ (91 data points). Participants did improve across repetitions in the collaborative experiments ($r = 0.47$, 89 data points). However, the proportion of repeaters within experiments was anti-correlated with group size: the correlation between experiments of group sizes 2, 3, 4, 8, and 16 and the proportion of repeaters was $r = -0.086$.

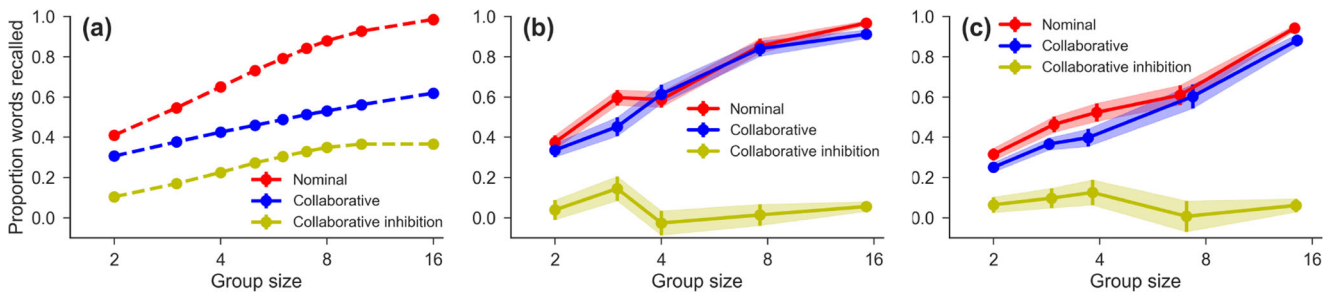


Fig. 1 Effect of group size on collaborative inhibition. (a) Model predictions, produced using code from Luhmann and Rajaram (2015). (b) Experiment 1 behavioral results. The mean proportion of words (\pm SE) recalled in the nominal (red/grey) and collaborative (blue/black)

conditions are shown; collaborative inhibition (yellow/light-grey) is the subtraction of these two values. Note that the horizontal axis is logarithmic. (c) Experiment 2 behavioral results

nominal experiments. The average (\pm SD) number of participants in collaborative and nominal experiments was 15.2 ± 0.7 for groups of size 16, 7.6 ± 0.7 for groups of size 8, 4.0 ± 0.0 for groups of size 4, 3.0 ± 0.0 for groups of size 3, and 2.0 ± 0.0 for groups of size 2. (In group sizes of 4, 3, and 2, groups were only included if they contained the maximum number of participants. In the nominal condition, participants completed the experiments individually and then were only later matched to the appropriate group sizes, so all participants were included.) Participants were paid \$1.00 for completing the task, plus bonuses for time spent waiting for other participants to arrive in a chatroom to begin the experiment (\$5.00/hour).

Procedure Participants were presented with a wordlist: 60 unrelated words each selected from a different category from Overschelde, Rawson, and Dunlosky (2004). (Example words include “diamond,” “hour,” and “uncle.” The 60 words were presented in a random order for each participant.) Each word was presented for two seconds. After seeing the list, participants completed an arithmetic filler task for 30 seconds before advancing to the recall task. Participants were placed in chatrooms alone or with other participants, and were encouraged to type as many of the words they had seen as possible. Participants were free to submit words at any time. Participants were not told how many other participants were in the chatroom: their responses appeared in blue font, and responses from all others appeared in black. They saw all previous words entered and were not permitted to submit any word that had already been submitted (by anyone in the group). This choice—that any words already present on the group list were not redisplayed—was made to encourage participants to read others’ submitted words, and because it more closely matched the lab-based version of the collaborative memory paradigm, in which verbal recall creates social pressure to not repeat words. There was no time limit for the recall task.

Groups contained 2, 3, 4, 8, or 16 participants. For each recall method (nominal or collaborative), 48 groups of 2

were analyzed; 32 groups of 3 were analyzed; 24 groups of 4 were analyzed, and 12 groups of 8 and 16 were analyzed. The sample sizes for group sizes 2, 3, and 4 were consistent with and generally larger than those in previous experiments (Marion & Thorley, 2016). An unbalanced design was used to ensure a similar number of participants for each group size. This was a 2×5 design, crossing recall method by group size. In the nominal recall condition, participants recalled words alone (and were not informed their recall lists would later be pooled with other participants’). In the collaborative recall condition, groups of participants were placed in chatrooms and recalled together. Recalled words that had not been on the original lists were marked as incorrect and not included.

Participants completed a post-experiment questionnaire, in which we asked about participants’ engagement in the task (1–10), the perceived difficulty of the task (1–10), and had a textbox for any other comments.

Results

The dependent variable in these analyses was the number of words recalled for each group; nominal participants’ recall lists (with redundant words removed) were added together according to the appropriate group size. The collaborative inhibition effect is most reliably observed in triads (Rajaram & Pereira-Pasarin, 2010), and we replicated this effect in our behavioral data at group size 3: $t(62) = 2.34$, $p = 0.022$, effect size $d = 0.60$, independent two-sample t -test (Fig. 1b, Table 1).³ Collaborative inhibition in the literature is frequently but not always observed in pairs (Rajaram & Pereira-Pasarin, 2010), but we did not observe this effect in this group size ($p = 0.44$, $d = 0.16$). There are only two studies known to the authors examining tetrads (Thorley

³ $\alpha = 0.05$ for all planned comparisons to follow. Unlike previous studies, we investigate the collaborative inhibition effect at multiple group sizes, while previous studies used $\alpha = 0.05$ at a single group size, justifying its use here.

Table 1 Two-sample *t*-test results for each group size in Experiment 1

	<i>t</i>	<i>p</i>	<i>d</i>	<i>M</i>	95% <i>CI</i>
Group size 2	$t(94) = 0.78$	0.44	0.16	2.33	[-3.53, 8.19]
Group size 3	$t(62) = 2.34$	0.022	0.60	8.69	[1.40, 15.98]
Group size 4	$t(46) = -0.42$	0.67	0.13	-1.58	[-8.94, 5.77]
Group size 8	$t(22) = 0.25$	0.80	0.11	0.83	[-5.74, 7.41]
Group size 16	$t(22) = 2.17$	0.041	0.93	3.33	[0.29, 6.38]

Alongside *t*-test and effect size (Cohen's *d*) statistics we report the mean difference (*M*) in words recalled between the nominal and collaborative groups and associated confidence intervals (*CI*)

& Dewhurst, 2007; Basden et al., 2000). Collaborative inhibition effects were observed in these studies, but we did not observe this effect at group size of 4 ($p = 0.67$, $d = 0.13$).

Given previous observations of the collaborative inhibition effect for small group sizes, Luhmann and Rajaram (2015) extrapolated and hypothesized that the effect of collaborative inhibition would increase with group size (Fig. 1a). There was no significant collaborative inhibition effect at group size of 8 ($p = 0.80$, $d = 0.11$). However, we did find a collaborative inhibition effect at group size 16 ($p = 0.041$, $d = 0.93$, note that effect size is especially large due to small variance in word recall because of ceiling effects).

Overall, using a between-participants two-way unbalanced ANOVA, we did not observe a statistically significant main effect of recall method, $F(1,246) = 2.03$, $p = 0.16$, $\eta^2 = 0.0045$: nominal and collaborative groups did not recall significantly different numbers of words when results from all groups were combined (Fig. 1b). We did observe an expected main effect of group size, $F(4,246) = 49.08$, $p < 0.0001$, $\eta^2 = 0.44$, in that larger group sizes increased word recall. Critically, there was no interaction effect between recall method and group size, $F(4,246) = 1.15$, $p = 0.33$, $\eta^2 = 0.010$. This was the key prediction of the model proposed by Luhmann and Rajaram (2015).

We conducted a power analysis to determine whether our sample sizes were sufficient to detect effects as large as those predicted by the model. We generated 1000 simulations of the entire experiment using the parameter settings from Luhmann and Rajaram (2015) with 60 words. In each simulation, the model was run a number of times matching the sample sizes used in the experiment,⁴ and we conducted an ANOVA testing for a main effect of recall, a main effect of group size, and an interaction effect. In 1000/1000 simulations, the model had $p < 1e-40$ for all these effects. We conclude that if our participants had

⁴Sample sizes: 48 groups for group size 2, 36 groups for group size 3, 24 groups for group size 4, and 12 groups for group sizes 8 and 16.

behaved like the model predicted, we would have had enough power to detect these effects.

Experiment 2: Preregistered Replication with a Modified Task

Because the main finding from Experiment 1 was a null effect for the predicted interaction, and there were methodological concerns (e.g. repeat participants), we conducted a preregistered replication (Experiment 2). In Experiment 1, we used a “free-flowing” collaborative memory procedure, but another common collaborative memory procedure is “turn-taking” (see Marion & Thorley, 2016; Rajaram & Pereira-Pasarin, 2010, for reviews); Luhmann and Rajaram (2015) used a turn-taking procedure. In Experiment 2, we sought to increase our study's similarity to Luhmann and Rajaram (2015) and to lab-based turn-taking paradigms, while reducing methodological concerns that may have suppressed memory interference. To this end, the two largest changes in our replication were using a turn-taking rather than a free-flowing procedure, and having submitted words be read aloud to the other participants to increase similitude to the lab-based experiment.

In lab-based free-flowing experiments, participants recall as many words as they'd like, whenever they would like to, and interact organically with each other. However, in lab-based turn-taking experiments, participants do not interact with each other in a naturalistic way; on each turn, participants say a single word, then wait silently for the next participant to say a single word, while the experimenter records words and monitors turn-taking (Basden et al., 1997; Basden et al., 2000; Thorley & Dewhurst, 2007; for a writing version of this procedure, see Wright & Klumpp, 2004). We sought to replicate this lab-based turn-taking paradigm in our online paradigm, not only to better approximate Luhmann and Rajaram (2015), but also to facilitate comparison of our online collaborative memory study to a traditional lab-based study with similarly reduced naturalistic interaction.

This experiment was preregistered (planned methods and analyses) with the Open Science Framework (<https://osf.io/7ht3g>).

Methods

Participants A total of 1,135 participants were recruited through Amazon Mechanical Turk. Participants were not permitted to repeat the task. As in Experiment 1, participants were excluded from the experiment if they did not complete the pre-experiment arithmetic task and they did not contribute words in the main experiment. Again as in Experiment 1, participants waited in a virtual waiting

room until the group size was met; experiments listed with less than the full group size contained excluded participants. 575 participants were recruited and 36 participants were excluded from the collaborative conditions, for a total of 539 participants. The average (\pm SD) number of participants in the collaborative conditions was 2.0 ± 0.0 for groups of size 2, 2.9 ± 0.3 for groups of size 3, 3.7 ± 0.5 for groups of size 4, 7.3 ± 0.6 for groups of size 8, and 14.5 ± 1.4 for groups of size 16. 560 participants were recruited and 23 participants were excluded from the nominal conditions, for a total of 537 participants. The average (\pm SD) number of participants in the nominal conditions was 2.0 ± 0.0 for groups of size 2, 3.0 ± 0.2 for groups of size 3, 3.9 ± 0.3 for groups of size 4, 6.8 ± 0.8 for groups of size 8, and 14.3 ± 1.2 for groups of size 16. (Across both the collaborative and nominal conditions, groups with less than two participants were re-run.) Participants were paid \$1.00 for completing the task, plus bonuses for time spent waiting for other participants to arrive in a chatroom to begin the experiment (\$5.00/hour) and bonuses for the number of correct words they recalled and participants who transmitted words to them recalled (\$.015/word).

Procedure Participants observed 60 words, randomly selected for each group of participants across all conditions, drawn from a list of 200 words from Overschelde et al. (2004). (Example words include “adverb,” “pine,” and “dollar.” The 60 words were presented in a random order for each participant.) Each group had a different wordlist presented, and within each group, the order of presented words in the wordlist was selected randomly for each participant. Each word was presented for two seconds. After seeing the list, participants completed a 30-second-long arithmetic filler task before advancing to the recall task. Participants were placed in chatrooms alone (the nominal recall condition) or with other participants (the collaborative recall condition), and were encouraged to type as many of the words they had seen as possible.

As in Experiment 1, participants were not permitted to submit any word that had previously been submitted, to increase similarity to the lab-based version of the collaborative memory paradigm in which verbal recall creates social pressure to not repeat words. Moreover, words that were submitted by others in the chatroom, and not by the participant, were read aloud to the participant, again to increase similarity to the lab-based task. Words were read aloud by a text-to-speech computer algorithm that automatically converted participant text during the experiment. To ensure that participants could understand the computerized speech and that their audio was working, before the task began participants had to correctly type in an isolated phrase read aloud by this computerized voice. In the nominal conditions, no words were read aloud, since no

other participants were present in the chatroom. There was no time limit for the recall task.

To increase similarity to the specific methodology of Luhmann and Rajaram (2015), participants in the collaborative conditions were not permitted to submit words whenever they wished (as was true in the nominal condition). Rather, participants took turns to submit words and did so in random order, following Luhmann and Rajaram (2015). (Both turn-based and free-flowing procedures are commonly used in the collaborative memory paradigm, as described in Marion & Thorley, 2016; Rajaram & Pereira-Pasarin, 2010). Participants took part in “rounds,” in which each participant had the option to submit one word during their five-second turn, and turn order was determined randomly within each round. Participants could submit a word that had not already been submitted, press a “Pass” button, wait for their 5-second turn to elapse on each turn, or press “I can’t recall any more.” When participants exited the study by clicking the latter button, they were no longer included in the turn order for the remaining participants.

As in Experiment 1, groups contained 2, 3, 4, 8, or 16 participants. For each recall method (nominal or collaborative), 48 groups of 2 were analyzed; 32 groups of 3 were analyzed; 24 groups of 4 were analyzed, and 12 groups of 8 and 16 were analyzed. The sample sizes analyzed were on par with and generally larger than those in other collaborative memory experiments (Marion & Thorley, 2016).

Participants completed a post-experiment questionnaire, in which we asked about participants’ engagement in the task (1–10), the perceived difficulty of the task (1–10), and had a textbox for any other comments.

Results

Recall that in this experiment, we would observe a collaborative inhibition effect if the nominal groups recalled more words than the collaborative groups. For all group sizes in Experiment 2, we observed that the nominal groups did indeed recall more words than the collaborative groups in absolute number, but this effect was small (Fig. 1c). When results from all groups were combined, nominal and collaborative groups recalled significantly different numbers of words: in a between-participants two-way unbalanced ANOVA, we observed a main effect of recall method, $F(1,246) = 6.47$, $p = 0.012$, $\eta^2 = 0.013$. We ran independent 2-sample t -tests for group sizes 2, 3, 4, 8, and 16, with $\alpha = 0.05$ for all planned comparisons shown (Table 2).

To summarize, we did not observe a statistically significant collaborative inhibition effect for group size 2 ($p = 0.10$, effect size $d = 0.34$). We saw marginally significant effects at group size 3 ($p = 0.056$, $d = 0.50$) and group size 4 ($p = 0.056$, $d = 0.58$) (Fig. 1c). Consistent with previous work showing collaborative inhibition at

Table 2 Two-sample *t*-test results for each group size in Experiment 2

	<i>t</i>	<i>p</i>	<i>d</i>	<i>M</i>	95% <i>CI</i>
Group size 2	$t(94) = 1.64$	0.10	0.34	3.85	[-0.76, 8.47]
Group size 3	$t(62) = 1.95$	0.056	0.50	5.84	[-0.054, 11.74]
Group size 4	$t(46) = 1.96$	0.056	0.58	7.54	[-0.035, 15.12]
Group size 8	$t(22) = 0.086$	0.93	0.037	0.42	[-9.17, 10.00]
Group size 16	$t(22) = 1.76$	0.093	0.75	3.67	[-0.48, 7.81]

Alongside *t*-test and effect size (Cohen's *d*) statistics we report the mean difference (*M*) in words recalled between the nominal and collaborative groups and associated confidence intervals (*CI*)

small group sizes, the effect sizes for group sizes of 3 and 4 were similar to the adjusted estimate of overall mean effect size ($d = 0.56$) for collaborative inhibition (calculated in a meta-analysis by Marion and Thorley (2016), using studies of group sizes 2–4 and compensating for publication bias) and the effect size for group size 3 was similar to that in Experiment 1 ($d = 0.60$). There were no statistically significant differences at group size 8 ($p = 0.93$, $d = 0.037$), nor at group size 16 ($p = 0.093$, but note the large effect size of $d = 0.75$ due to small variance in the proportion of words recalled because of ceiling effects). The number of words recalled did increase with group size, and in a between-participants two-way unbalanced ANOVA, we observed the expected main effect of group size, $F(4,246) = 56.38$, $p < 0.0001$, $\eta^2 = 0.47$. However, once again we found no interaction effect between recall method and group size, $F(4,246) = 0.47$, $p = 0.76$, $\eta^2 = 0.0039$, contrary to the prediction of an increase in collaborative inhibition with group size from the model of Luhmann and Rajaram (2015).

Analyzing the model

The results of our experiments are not consistent with the increase in collaborative inhibition with group size

predicted by Luhmann and Rajaram (2015). However, one could argue that if we had altered the parameters of the model, we would have seen patterns that were more compatible with our empirical results. To address this, we conducted two additional analyses showing that the empirical results do not match model predictions across a range of parameter settings.

Our first analysis asked what predictions would be possible with the model, given the constraint of trying to match the empirical results as closely as possible. We thus did a grid search over the model's parameters, minimizing the squared error between the mean nominal and collaborative results for each group size.

Even for the best-fitting model, the model predictions do not qualitatively match the empirical results for Experiment 1 (Fig. 2) or Experiment 2 (Fig. 3). Specifically, the model is unable to predict a collaborative inhibition effect at small group sizes (as has been shown in the literature and our experiments) even with parameters optimized for reducing the discrepancy between the model and empirical results.

Our second analysis asked a different question: how often would we see model trends that fit the empirical data if we used a broad array of parameter settings? One way to operationalize the differences between the model predictions and empirical results is to check whether there is a smaller or larger collaborative inhibition effect for groups sizes 2, 3, and 4 compared to group sizes 8 and 16. The model predicts a smaller collaborative inhibition effect for group sizes 2, 3, and 4. In our empirical data, we observe a larger collaborative inhibition effect for group sizes 2, 3, and 4. This is one of the most qualitatively salient differences between the model and our results. In the [Supplementary Material](#) we present a detailed analysis showing that the model rarely predicts such an effect, and only does so in the presence of unrealistically strong ceiling effects.

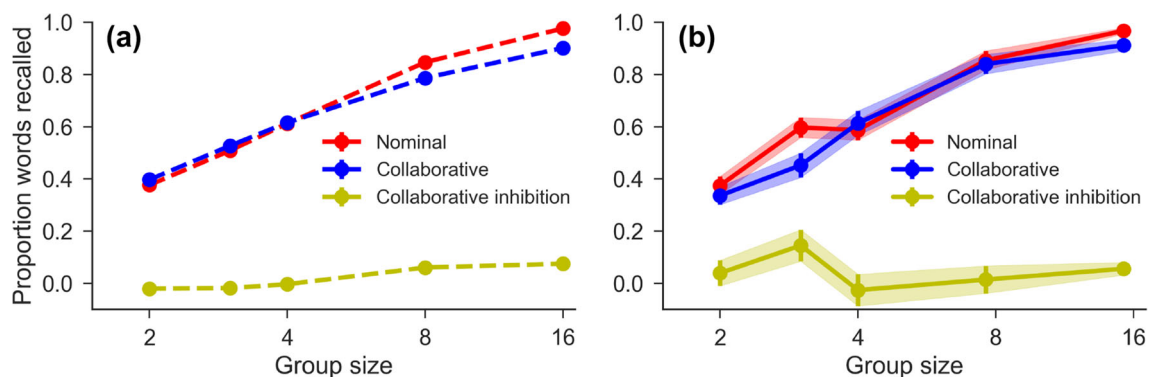


Fig. 2 Best-fit model predictions and behavioral results for Experiment 1. (a) Model predictions with the best-fitting parameters ($\alpha = 0.10$, $\beta = 0.07$, $\gamma = 0.55$, number of rounds = 23), 1000 simulations.

(Compare to the Luhmann and Rajaram (2015) model parameters of $\alpha = 0.20$, $\beta = 0.05$, $\gamma = 0.75$, number of rounds = 20.) (b) Behavioral results for Experiment 1

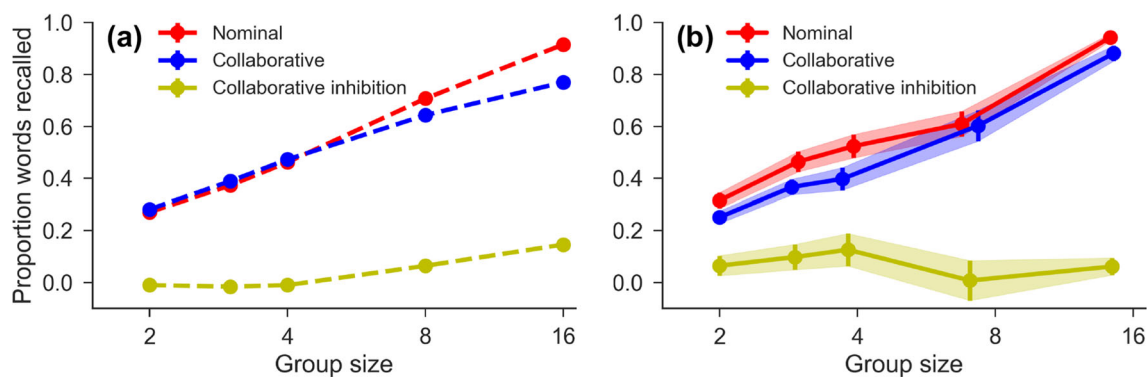


Fig. 3 Best-fit model predictions and behavioral results for Experiment 2. (a) Model predictions with the best-fitting parameters ($\alpha = 0.12$, $\beta = 0.03$, $\gamma = 0.39$, number of rounds = 22), 1000 simulations.

(Compare to the Luhmann and Rajaram (2015) model parameters of $\alpha = 0.20$, $\beta = 0.05$, $\gamma = 0.75$, number of rounds = 20.) (b) Behavioral results for Experiment 2

General discussion

The critical prediction that Luhmann and Rajaram (2015) made is that collaborative inhibition—an effect found robustly at smaller group sizes—would be even greater with group sizes 8 and 16. Under their model, researchers would expect to see an interaction effect between recall method (collaborative or nominal) and group size, reflecting increased collaborative inhibition as group size increases. We did not observe this interaction effect in our data, and believe we had enough power to detect this effect had participants behaved as the model predicted. We now consider possible interpretations of these results.

One possibility is that our online task differed from laboratory studies in ways that had a meaningful impact on the results. First, we note that the online task produced results consistent with those from the laboratory: we observed collaborative inhibition at group size 3 in Experiment 1 and had marginally significant evidence for an effect at group sizes 3 and 4 ($p = 0.56$) as well as a main effect of collaborative inhibition across groups in Experiment 2, and the size of these effects corresponds to those suggested by meta-analysis of laboratory experiments (Marion & Thorley, 2016). Second, we note that all of the factors relevant to the agent-based model—and to the broader theory of retrieval disruption—are present in our online tasks, which closely mimic the structure of the model. If there are critical features of the laboratory setting that are missing from our task, they are equally absent from the model. Third, we note that though our online experiments are not as naturalistic as lab-based free-flowing experiments, in our preregistered replication (Experiment 2), we used a turn-taking procedure which was very similar to the lab-based turn-taking methodology. In lab-based turn-taking procedures, interactions are restricted: participants do not experience interactions like multiple voices and interruptions, irregular overlapping timings, or

normal social cues and organic camaraderie, just as in our Experiment 2. We thus believe that Experiment 2 especially can be fairly compared to lab-based collaborative experiments. Finally, our online tasks—in which people interacted in chatrooms—are increasingly representative of the contexts in which people interact with large groups, making them a reasonable object of study in themselves.

Another possibility is that while retrieval disruption is the dominant force in smaller groups, additional factors come into play at larger group sizes. For example, it might be the case that intense competition as group size increases drives greater recall. Alternatively, at larger group sizes, groups may need to adopt different representational strategies that are less prone to disruption. To determine what these different mechanisms would be, we will need controlled experiments aimed at developing additional cognitive, social, or motivational hypotheses explaining the unexpected lack of consistent collaborative inhibition at large group sizes.

Participant reports hint at what might make large groups distinct. In post-Experiment 1 questionnaires, for group size 16, participants reported time pressure and competitive stress from many people racing to submit responses at once. In smaller groups, the number of responses was less overwhelming and more cooperative. In Experiment 2, participants took timed turns rather than experiencing free-flowing recall, but in larger groups people had to comparatively wait much longer for their turns compared to in smaller groups. These are examples of factors that are comparatively minor in small group collaborative recall, but were qualitatively salient in large group collaborative recall.

In an increasingly interconnected world, it is important to understand how information is shared and remembered by interacting groups of people. Methods like agent-based modeling provide a starting point for developing hypotheses about groups that are larger than we can normally study in the laboratory. However, crowdsourcing and computer-

controlled recruitment offer new tools that we can use to evaluate these hypotheses and extend the scope of our empirical results. Our findings raise the possibility that there may be additional factors that influence collaborative recall in large group sizes, illustrating the importance of translating agent-based models based on lab studies into online studies that can improve our understanding of the interactive behavior of large groups of people.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13423-021-02021-9>.

Acknowledgments Preliminary results from Experiment 1 were presented at the Annual Conference of the Cognitive Science Society and appear in the Proceedings of that conference (Gates, Suchow, & Griffiths, 2017). This work was funded in part by NSF Grant 1456709 to T.L.G., DARPA Cooperative Agreement D17AC00004 to T.L.G and J.W.S., and N.I.H. U.C. Berkeley Neuroscience Training Program Grant T32NS095939 to V.G.

Author Note Materials and data on Open Science Framework: <https://osf.io/txe9a>.

Open Practices Statement The data and materials for all experiments are available at <https://osf.io/txe9a> and Experiment 2 was preregistered.

References

- Basden, B., Basden, D., Bryner, S., & Thomas, R. (1997). A comparison of group and individual remembering: Does collaboration disrupt retrieval strategies? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(5), 1176. <https://doi.org/10.1037/0278-7393.23.5.1176>.
- Basden, B., Basden, D., & Henry, S. (2000). Costs and benefits of collaborative remembering. *Applied Cognitive Psychology*, 14(6), 497–507. [https://doi.org/10.1002/1099-0720\(200011/12\)14:6<497::AID-ACP665>3.0.CO;2-4](https://doi.org/10.1002/1099-0720(200011/12)14:6<497::AID-ACP665>3.0.CO;2-4).
- Finlay, F., Hitch, G., & Meudell, P. (2000). Mutual inhibition in collaborative recall: Evidence for a retrieval-based account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(6), 1556. <https://doi.org/10.1037//0278-7393.26.6.1556>.
- Gates, M. A., Suchow, J. W., & Griffiths, T. L. (2017). Empirical tests of large-scale collaborative recall. In *Proceedings of the 39th Annual Conference of the Cognitive Science Society*.
- Hirst, W., & Echterhoff, G. (2012). Remembering in conversations: The social sharing and reshaping of memories. *Annual Review of Psychology*, 63(1), 55. <https://doi.org/10.1146/annurev-psych-120710-100340>.
- Luhmann, C., & Rajaram, S. (2015). Memory transmission in small groups and large networks: An agent-based model. *Psychological Science*, 26, 1909–1917. <https://doi.org/10.1177/0956797615605798>.
- Marion, S. B., & Thorley, C. (2016). A meta-analytic review of collaborative inhibition and postcollaborative memory: Testing the predictions of the retrieval strategy disruption hypothesis. *Psychological Bulletin*, 142(11), 1141. <https://doi.org/10.1037/bul0000071>.
- Meade, M., Nokes, T., & Morrow, D. (2009). Expertise promotes facilitation on a collaborative memory task. *Memory*, 17(1), 39–48. <https://doi.org/10.1080/09658210802524240>.
- Overschelde, J. V., Rawson, K., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the norms. *Journal of Memory and Language*, 50(3), 289–335. <https://doi.org/10.1016/j.jml.2003.10.003>.
- Rajaram, S., & Pereira-Pasarin, L. (2010). Collaborative memory: Cognitive research and theory. *Perspectives on Psychological Science*, 5(6), 649–663. <https://doi.org/10.1177/1745691610388763>.
- Thorley, C., & Dewhurst, S. (2007). Collaborative false recall in the DRM procedure: Effects of group size and group pressure. *European Journal of Cognitive Psychology*, 19(6), 867–881. <https://doi.org/10.1080/09541440600872068>.
- Wegner, D., Erber, R., & Raymond, P. (1991). Transactive memory in close relationships. *Journal of Personality and Social Psychology*, 61(6), 923. <https://doi.org/10.1037/0022-3514.61.6.923>.
- Wright, D. B., & Klumpp, A. (2004). Collaborative inhibition is due to the product, not the process, of recalling in groups. *Psychonomic Bulletin & Review*, 11(6), 1080–1083. <https://doi.org/10.3758/BF03196740>.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.