




Why collaboration reduces suggestibility: The role of source-monitoring processes and retrieval strategies

Clelia Rossi-Arnaud¹  · Pietro Spataro² · Serena Mastroberardino¹ · Enrica Lucaroni¹ · Anna Maria Giannini¹ · Vincenzo Cestari¹

Accepted: 3 June 2021
© The Author(s) 2021

Abstract

Previous research found that collaboration reduces the tendency to yield to misleading questions. Here, the aim was to determine whether this occurs because collaboration induces a conservative change in response criterion or because it promotes more efficient error-checking strategies. To this purpose, we compared the performance of collaborative and nominal triads in the Gudjonsson Suggestibility Scale. Critically, we recorded conversations during retrieval, allowing us to compute inclusive scores and analyze retrieval strategies. Confirming previous evidence, results showed that collaborative groups yielded less to leading questions; however, the differences between collaborative and nominal groups in yield 1 were canceled when we took into account questions to which at least one participant in the collaborative group gave in but was corrected by collaborators during the discussion (the so-called inclusive scores). This was not the case for yield 2 and total suggestibility scores. Also, the analysis of retrieval strategies indicated that collaborative groups who used process-focused strategies (such as correction and cross-cueing) to a greater extent were less likely to change their responses after receiving the negative feedback and were less suggestible. We conclude that, while the use of error-checking and process-focused strategies played a role in reducing suggestibility in collaborative groups, the administration of negative feedback induced members of collaborative groups to adopt a more conservative response criterion. These results contribute to the understanding of the conditions that maximize the positive effects of collaborative retrieval. They have implications for policymakers and police practitioners, specifying when and how collaboration might be allowed.

Keywords Collaboration · Suggestibility · Misleading questions · Error-checking strategies · Conservative response criterion

Experimental research on memory processes has been typically carried out testing single participants working alone. Recently, however, there has been a surge of interests in the effects of collaboration on memory completeness and accuracy (Harris et al., 2008; Rajaram, 2011; Rajaram & Maswood, 2018). In the collaborative recall paradigm, participants encode to-be-remembered material individually and later collaborate during retrieval to recall as many studied items as possible (Basden et al., 1997; Weldon & Bellinger, 1997). The impact of collaboration on memory is assessed by comparing

the performance of collaborative groups with that of nominal groups (i.e., groups in which the outputs of individual participants are pooled together by counting redundant items only once). The usual outcome of this procedure is that collaborative groups recall a lower number of correct items than nominal groups, a phenomenon known as “*collaborative inhibition*” (Basden et al., 1997; Marion & Thorley, 2016; Weldon & Bellinger, 1997). According to the “*retrieval strategy disruption hypothesis*”, the effect is due to interference caused by exposure to the responses of other members: hearing these responses disrupts the use of idiosyncratic strategies that one adopts to recall studied elements, leading to a less optimal performance (Finlay et al., 2000; Rajaram & Pereira-Pasarin, 2010; Wright & Klumpp, 2004; see Barber et al., 2015, for alternative explanations).

Alongside these negative effects, collaborative remembering also has beneficial effects. When it comes to memory accuracy, for example, considerable evidence indicates that collaborative groups produce fewer errors than nominal

✉ Clelia Rossi-Arnaud
clelia.rossi-arnaud@uniroma1.it

¹ Department of Psychology, Faculty of Medicine and Psychology, University “La Sapienza”, Via dei Marsi 78, 00185 Rome, Italy

² Faculty of Economy, Universitas Mercatorum, Piazza Mattei 10, 00186 Rome, Italy

groups (Henkel & Rajaram, 2011; Ross et al., 2008; Ross et al., 2004; Rossi-Arnaud et al., 2011; Takahashi, 2007; Vredeveldt et al., 2016). Moreover, studies using a multiple recall paradigm, in which a collaborative phase is followed by individual recall, showed that individuals who previously collaborated outperformed individuals who previously worked alone (Basden et al., 2000; Blumen & Rajaram, 2008, 2009; Congleton & Rajaram, 2011; Marion & Thorley, 2016; Rajaram & Pereira-Pasarin, 2007).

Two recent studies expanded this line of research by examining the effects of collaborative retrieval on suggestibility (Rossi-Arnaud et al., 2019; Rossi-Arnaud et al., 2020) – here defined as the degree to which people come to accept misleading information communicated during formal questioning (Gudjonsson & Clark, 1986; Mastroberardino & Marucci, 2013). Specifically, Rossi-Arnaud et al. (2019) used the Gudjonsson Suggestibility Scale (GSS; Gudjonsson, 1984, 1997) to compare the performance of collaborative and nominal dyads. In the GSS, participants listen to a story, provide immediate and delayed (after 50 min) free recalls, and answer a series of misleading questions before and after having received a negative feedback on their performance. Two main results arose from the Rossi-Arnaud et al. (2019) study. First, collaborative dyads remembered the same amount of correct story elements as nominal dyads in the immediate and delayed recall tasks, indicating that the classical effect of collaborative inhibition was not obtained. Second, and most important for the purposes of the present study, collaborative dyads produced a lower number of confabulated elements during the recall phases and were less likely than nominal dyads to give in to misleading questions, both before and after the administration of the negative feedback.

The study by Rossi-Arnaud et al. (2020) used a slightly different paradigm but reached similar conclusions. Collaborative and nominal pairs viewed the videoclip of a bank robbery, provided an immediate free recall and were forced to confabulate answers to a series of false-event questions (i.e., questions referring to details that, although plausible, did not appear in the videoclip). Then, after a short interval (1 h) or a longer delay (1 week), all the pairs were administered a yes/no recognition task in which the misleading statements either matched the questions presented in the confabulation phase or were completely new. As in the Rossi-Arnaud et al. (2019) study, the main outcome was that collaborative pairs were less likely to provide false assents to misleading statements in the final recognition task; furthermore, positive effects of collaboration were shown regardless of whether participants had given confabulated responses to the statements in the previous phase.

Taken together, the results reported by Rossi-Arnaud et al. (2019, 2020) provide solid evidence in support of the conclusion that collaborative remembering reduces suggestibility. However, two different factors might potentially account for

these findings. On the one hand, collaboration might induce a conservative change in response criteria, so that group members were simply less likely to contribute both accurate and inaccurate details (Harris et al., 2012; Ross et al., 2008; Takahashi, 2007; Thorley & Dewhurst, 2007). On the other hand, a different explanation may be that collaboration promotes the use of error-checking strategies, such that group members actively monitor the response accuracy of their collaborators. Harris et al. (2012) tested the validity of this hypothesis in a three-phase experiment in which they compared the performance of consensus, turn-taking, and nominal groups. Results showed that consensus groups (in which members had to reach a collective agreement on each response) produced fewer correct words and fewer intrusions than turn-taking groups (in which members alternated in recalling the studied items) and nominal groups. To examine the role of error-checking strategies, the authors recorded the conversations occurring during recall sessions and computed the so-called “inclusive scores”, which involved considering all the incorrect items that were produced by at least one member but that were subsequently discounted by the group (Harris et al., 2012). When this was done, the advantage of consensus groups on memory accuracy fell below the significance level, suggesting that the members of consensus groups mentioned a number of incorrect items similar to that of turn-taking and nominal groups, but these items were later checked and rejected during group discussions.

Based on this evidence, the present study aimed at determining whether the same mechanism could account for the reduction in the number of false assents to misleading questions observed by Rossi-Arnaud et al. (2019). That is, the primary contribution of our experiment was that, by recording the conversations within collaborative groups and by computing inclusive scores, we aimed at determining whether the positive effects of collaboration could be accounted for by the mutual use of error-checking strategies. We used the same methodology illustrated in our previous study, with two exceptions – namely, participants listened to a forensic story (i.e., the GSS1; Rossi-Arnaud and colleagues used the parallel form GSS2, which involves a non-forensic story) and both the nominal and collaborative groups comprised three members (Rossi-Arnaud et al. examined the performance of dyads). We expected to replicate the main finding reported by Rossi-Arnaud et al. (2019): collaborative triads should be less likely to give in to misleading questions, compared to nominal triads. Most importantly for our purposes, we expected that the latter difference should be eliminated when assessed with inclusive scores. This would suggest that participants in collaborative and nominal triads produced the same number of false assents to misleading questions but that these were later rejected during discussion in collaborative groups.

A second novel contribution of the present study was that we investigated the retrieval strategies used in collaborative

groups during immediate and delayed recall phases and assessed their potential role in predicting the suggestibility of collaborative groups. Vredeveldt et al. (2016) found that participants used two types of strategies during collaboration, namely ‘*process-focused strategies*’ (i.e., strategies focused on the process of remembering together, such as explaining one’s own statements, correcting each other or trying to cue each other) and ‘*content-focused strategies*’ (i.e., strategies requiring participants to elaborate upon their partners’ contributions). Regression analyses indicated that couples who relied primarily on content-focused strategies recalled more information overall, although the model did not predict the accuracy of reported information (see also Vredeveldt et al., 2017). Based on this evidence, we aimed at examining a) whether the same pattern occurred in the present study, and b) whether the use of process-focused or content-focused strategies predicted the suggestibility of collaborative groups. If participants working in collaborative groups are more likely to use error-checking strategies during the retrieval phase, and this factor accounts for their lower tendency to yield to leading questions, then we should expect groups which make a greater use of process-focused to exhibit lower suggestibility.

To summarize, the present study sought to replicate and expand results previously reported by Rossi-Arnaud et al. (2019) in two different ways. First, by determining whether error-checking (i.e., the tendency to check the accuracy of other members’ responses) could explain the reduction in the number of false assents to misleading questions observed in collaborative groups: if this were the case, then the differences in suggestibility between nominal and collaborative groups should be eliminated when scores are computed with the inclusive method (Harris et al., 2012). Second, by determining whether the use of process-focused or content-focused retrieval strategies could predict the suggestibility of collaborative groups. Since these issues were not assessed in the study by Rossi-Arnaud et al. (2019) they are likely to provide novel insights into the mechanisms leading to the reduced suggestibility of collaborative groups.

Method

Participants

Seventy-five graduate and undergraduate students (54 females, 21 males; mean age = 25.13; $SD = 3.21$) from XXX University, volunteered to participate in the experiment. Participants were assigned either to nominal or to collaborative. Groups were either mixed in terms of gender or homogeneous. Nominal and Collaborative triads had comparable proportions of mixed (two females and a male or two males and a female) and same gender (all female or all male) groups.

The distribution of mixed and homogenous triads did not differ between the two conditions: $\chi^2 = 0.02$, $p = 0.87$.

The study was approved by the Institutional Review Board (ethics committee), Department of Psychology, XXX University and an informed consent was signed by each participant before taking part in the study. The procedures used in this study adhered to the principles of the Declaration of Helsinki.

Materials

The Italian version of the Gudjonsson Suggestibility Scale 1 (GSS1; Bianco & Curci, 2015) was used. It consists of a short story describing a robbery, divided into 40 items which are clearly identified and separated by slashes in the written version. The story is read out to the participant who then provides an immediate free recall. After a 50-min interval, the participant recalls the story again (delayed free recall) and is then presented with a set of 20 questions. Five questions concern events that occurred in the story (control questions), whilst 15 concern details and events that were not presented in the original story: the leading questions (e.g., “Was the name of the woman Anna Balducci?” when the correct name was Anna Colucci). After completing the questionnaire, the participant is presented with a negative feedback and asked to answer the 20 questions a second time.

The GSS 1 provides the following measures: a) *Immediate free recall*: the number of correct items recalled immediately after hearing the story (range 0–40); b) *Immediate confabulation*: the number of fabricated (new) or distorted (modified) items reported in the immediate recall; c) *Delayed free recall*: the number of correct items recalled after a 50-min delay (range 0–40); d) *Delayed confabulation*: the number of items fabricated or distorted in the delayed recall; e) *Yield 1*: the number of leading questions to which participants gave in when responding to the questionnaire, before the administration of the negative feedback (range 0–15); f) *Yield 2*: the number of leading questions to which participants gave in after the administration of the negative feedback (range 0–15); g) *Shift*: the number of times participants changed their answers to the control and leading questions after receiving the negative feedback (range 0–20); h) *Total Suggestibility*: the sum of Yield 1 and Shift scores, which reflects participants’ overall suggestibility (range 0–35).

The GSS1 manual provides detailed instructions to score both the free recall reports and the answers to the 15 leading questions (Gudjonsson, 1997). As for the immediate and delayed free recalls, 1 point was assigned whenever the participant accurately reported the general meaning of each idea, even if he/she used a different wording. Since some items of the GSS 1 comprise two elements (i.e., “Anna Colucci”), a partial report (i.e., only “Anna” or “Colucci”) was always assigned 0.5 points. Confabulation scores were computed as

the sum of the total number of distorted and fabricated details provided in the free recall. Distorted elements refer to details mentioned in the story but reported in the wrong way (for instance, if the story reported that the lady was interviewed by “Detective Sergeant Delgado” but the participant recalled “Detective Sergeant Domingo”). Fabricated elements refer to instances in which participants reported a detail that was completely new – i.e., not presented in the original story. Subjects were given 1 point for each distorted or fabricated element reported.

Possible answers to the leading questions are also provided in the manual. Affirmative answers to a leading question were scored 1 point. So, for example, responses like “Yes”, “Maybe Yes”, “I think so”, “One child/Two children” were considered to be yield responses. In contrast, responses such as “I don’t remember”, “this was not mentioned in the story”, “I don’t know”, “I’m not sure” were not considered as yield responses and were therefore assigned 0 points. For shift scores, the changes in the Yield2 responses had to be clear-cut. For example, changes from “No” to “Yes”, from “I don’t know” to “Yes”, or from “Don’t remember” to “Yes” were considered as shifts and scored 1; on the other hand, changes from “Yes” to “Maybe Yes”, from “No” to “Don’t remember”, or from “Yes” to “It’s possible” were not counted as shifts.

Procedure

The GSS1 was administered following instructions in the GSS manual (Gudjonsson, 1997). Participants were read out the story and instructed to listen carefully since they would be later required to recall it as accurately as possible (“I would like you to listen to a short story. Please listen carefully because, when I will have finished, I want you to recall all the information you remember”). They were then provided an immediate free recall (“Now please recall everything you remember from the story”) and, after a 50-min interval filled with unrelated tasks, a delayed free recall. For both the immediate and delayed recall tasks, participants responded at their own pace (i.e., no time limits were specified but most of the triads took between five and ten minutes to complete the task). Then the 20 questions were presented and participants were instructed as follows “Now I’m going to ask you some questions about the story. Try to be as accurate as possible”. At the end of this phase the experimenter provided a negative feedback, irrespective of the participants’ performance, and the questions were presented again (i.e. “You have made a number of errors. It is therefore necessary to go through the questions once more, and this time try to be more accurate.”). Following the GSS instructions, the exact number of errors made by nominal and collaborative groups was not specified and the negative feedback was only related to the questions answered in the last phase – i.e., no feedback was given about the accuracy of the immediate and delayed recalls.

Participants assigned to the nominal triads sat next to each other but did not interact while completing the GSS1 and did not have the possibility to see the responses of the other participants. Participants in the collaborative triads were instructed to collaborate while completing the GSS1 (one of them was randomly chosen to write down the immediate recall, the delayed recall, and the answers to the questions). Following Weldon and Bellinger (1997), participants received no instructions on how to sort out potential disagreements.

Data Coding

The immediate and delayed free recall reports were scored using the GSS1 template. For nominal triads, the total number of correct and confabulated items reported by each participant were calculated and then pooled together, with redundant items counted only once, to obtain the “nominal recall” scores (Basden et al., 1997; Weldon & Bellinger, 1997). For collaborative triads, the total number of correct and confabulated items was calculated and labeled “collaborative recall”. The same procedure was used to compute the nominal and collaborative yield and shift scores from the GSS1 questionnaire. For example, in the case of nominal triads, if participant A gave in to the leading questions 2, 8, and 15, participant B gave in to the leading questions 14, 15, and 20, and participant C gave in to the leading questions 14, 15, and 18, then the nominal yield score of the triad was six (the sum of the responses to questions 2, 8, 14, 15, 18 and 20).

As mentioned above, the conversation occurring between the members of collaborative triads were recorded and coded for computing inclusive scores and assessing the use of different retrieval strategies. Inclusive scores were calculated for those leading questions to which the group did not give in. Following Harris et al. (2012), an inclusive score was assigned if during the verbal interaction one of the participants provided an affirmative response to a leading question that was subsequently discounted by the other members. That is, inclusive scores were calculated for Yield 1 and Yield 2 by adding to the original scores the questions to which one participant gave in during the discussion but was corrected by his/her collaborators.

Retrieval strategies were coded from the interactions occurring in collaborative triads, during immediate and delayed recall, following the scheme proposed by Vredeveltdt et al. (2016). A total of 11 retrieval strategies were examined: *successful cueing* (a cueing attempt followed by retrieval of information by the partners), *failed cueing* (a cueing attempt that was not followed by retrieval by the partners), *acknowledgement* (indicating support for partners’ statements), *correction* (correcting the partners’ statements or questioning their accuracy), *elaboration* (building on the partners’ statements to provide additional information), *explanation* (explaining one’s statement to the partners),

repetition (repeating partners' statements verbatim), *restatement* (rephrasing a partners' statements without changing the content), *renewed remembering* (indicating that the partners' statements triggered memory retrieval), *role division* (statements aimed at dividing or organizing the retrieval task), *relationship positive* (positive comments about the partners' or the triad's ability), *relationship negative* (negative comments about the partners' or the triad's ability). Note that in the present study the last two categories did not occur in participants' interactions and were therefore excluded from the following analyses. Two independent raters coded all the collaborative transcripts: interrater reliability was good ($\kappa = .80$) and disagreements were resolved by discussion.

Results

Immediate and Delayed Recall: Correct Story Elements

The number of story elements correctly recalled by nominal and collaborative triads during immediate and delayed recall tasks (see Fig. 1) were submitted to a mixed 2×2 ANOVA, considering Group (nominal vs. collaborative) and Recall Time (immediate vs. delayed) as the independent factors. Results showed significant main effects of Group, $F(1, 23) = 8.57, p = 0.008, \eta^2_p = 0.27$, and Recall Time, $F(1, 23) = 8.56, p = 0.008, \eta^2_p = 0.27$, indicating that nominal groups ($M = 32.72$) recalled more elements than collaborative groups ($M = 27.33$) and that performance decreased from the immediate ($M = 30.55$) to the delayed recall task ($M = 29.51$). In addition, the two-way interaction was also significant, $F(1, 23) = 6.27, p = 0.020, \eta^2_p = 0.21$. A follow-up analysis of simple effects revealed that recall performance decreased between the immediate and delayed tasks for collaborative groups,

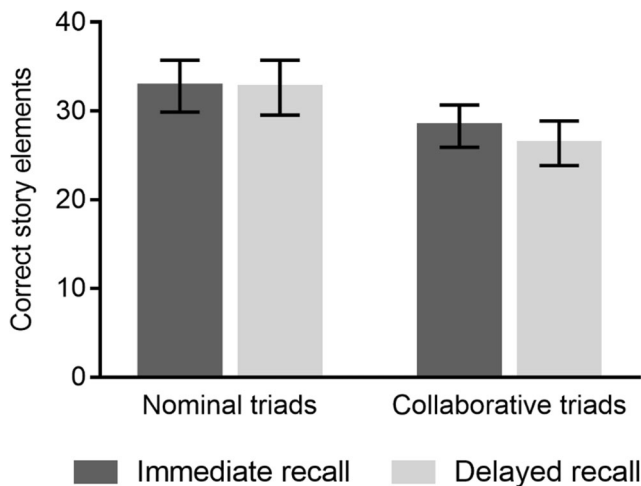


Fig. 1 Mean number of correct story elements remembered by nominal and collaborative triads during the immediate and delayed recall phases. Bars represent 95% confidence intervals

$M = 28.30$ vs. $M = 26.36, F(1, 23) = 18.43, p < 0.001, \eta^2_p = 0.45$, but not for nominal groups, $M = 32.80$ vs. $M = 32.65, F(1, 23) = 0.07, p = 0.78, \eta^2_p = 0.003$. The advantage of nominal groups over collaborative groups was significant in both the immediate and delayed recall tasks, $F(1, 23) = 6.07, p = 0.022, \eta^2_p = 0.21$ and $F(1, 23) = 10.66, p = 0.003, \eta^2_p = 0.32$, respectively.

Immediate and Delayed Recall: Confabulations

The number of confabulated elements (including both distortions and fabrications) reported by nominal and collaborative triads during immediate and delayed recall were also analyzed through a mixed 2×2 ANOVA, considering Group (nominal vs. collaborative) and Recall Time (immediate vs. delayed) as the independent factors. Results showed a significant main effect of Group, $F(1, 23) = 26.63, p < 0.001, \eta^2_p = 0.54$, indicating that collaborative groups ($M = 1.83$) produced less confabulated elements than nominal groups ($M = 4.90$). The main effect of Recall Time and the two-way interaction between Group and Recall Time were not significant, $F(1, 23) = 0.24, p = 0.63, \eta^2_p = 0.01$ and $F(1, 23) = 0.06, p = 0.81, \eta^2_p = 0.003$, respectively.

Yield Responses Before and After the Negative Feedback

Yield scores before and after the negative feedback (see Fig. 2) were submitted to a mixed 2×2 ANOVA, considering Group (nominal vs. collaborative) and Time (before vs. after the negative feedback) as the independent factors. Results revealed a significant main effect of Group, $F(1, 23) = 26.63, p < 0.001, \eta^2_p = 0.66$, indicating that collaborative groups ($M = 2.53$) gave in less to leading questions than

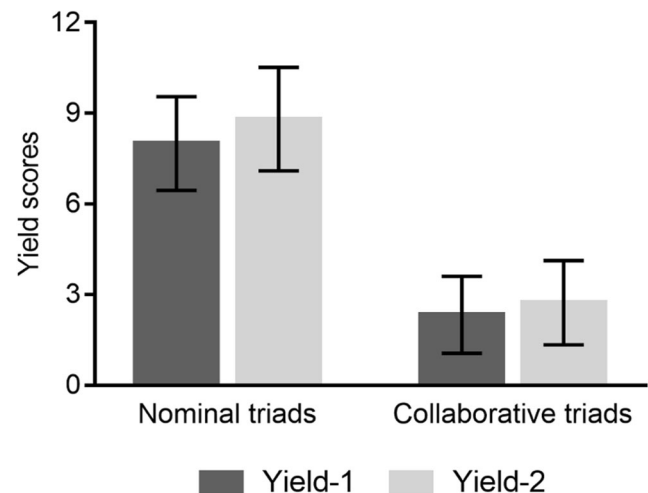


Fig. 2 Mean number of leading questions to which nominal and collaborative groups gave in before (yield-1) and after (yield-2) receiving the negative feedback. Bars represent 95% confidence intervals

nominal ones ($M=8.40$). The main effect of Time and the two-way interaction between Group and Time were not significant, $F(1, 23)=1.42$, $p=0.24$, $\eta^2_p=0.06$ and $F(1, 23)=0.16$, $p=0.69$, $\eta^2_p=0.01$, respectively.

Shift Scores and Total Suggestibility

A couple of t -test for independent samples were computed to determine whether shift scores and total suggestibility differed between nominal and collaborative groups. Collaborative triads ($M=2.80$) were less prone than nominal triads ($M=8.40$) to change their answers after receiving the negative feedback, $t(23)=3.48$, $p=0.002$. Total suggestibility scores were also lower for collaborative than for nominal triads, $M=5.13$ vs. $M=13.60$, $t(23)=6.12$, $p<0.001$.

Inclusive Scores

As previously discussed, inclusive yield scores were calculated for collaborative triads by adding to the original scores the number of questions to which participants gave in during the discussion but were later corrected by other members. The inclusive Yield 1 and Yield 2 scores were again submitted to a mixed 2×2 ANOVA, considering Group (nominal vs. collaborative) and Time (before vs. after the negative feedback) as the independent factors. Results revealed that the main effect of Group was still significant, $F(1, 23)=5.86$, $p=0.024$, $\eta^2_p=0.20$: collaborative groups ($M=5.73$) continued to be less likely to give in to leading questions, compared to nominal groups ($M=8.40$). However, the two-way interaction between Group and Time was also significant, $F(1, 23)=6.56$, $p=0.017$, $\eta^2_p=0.22$. A follow-up analysis of simple effects showed that collaborative groups were less likely than nominal groups to give in to leading questions after receiving the negative feedback (Yield 2), $M=4.53$ vs. $M=8.80$, $F(1, 23)=15.76$, $p=0.001$, $\eta^2_p=0.41$, but not before (Yield 1), $M=6.93$ vs. $M=8.00$, $F(1, 23)=0.55$, $p=0.46$, $\eta^2_p=0.02$. The same analysis indicated that the number of leading questions to which collaborative groups gave in decreased after receiving the negative feedback, $M=6.93$ vs. $M=4.53$, $F(1, 23)=9.23$, $p=0.006$, $\eta^2_p=0.29$, whereas no reduction was apparent for nominal groups, $M=8.00$ vs. $M=8.80$, $F(1, 23)=0.68$, $p=0.41$, $\eta^2_p=0.03$. Lastly, the main effect of Time was not significant, $F(1, 23)=1.64$, $p=0.21$, $\eta^2_p=0.07$.

Total suggestibility scores were also computed a second time by adding the Yield 1 inclusive scores to the shift scores. A t -test for independent samples showed that collaborative triads were still less suggestible than nominal triads, even though the difference was reduced in size, $M=9.73$ vs. $M=13.00$, $t(23)=2.12$, $p=0.045$.

Retrieval Strategies

As illustrated in Table 1, the strategies most often used by members of collaborative groups were ‘elaboration’, ‘acknowledgement’, ‘correction’, ‘repetition’ and ‘successful cueing’. Following Vredeveldt et al. (2016), we performed a principal axis factor analysis (with Oblimin rotation), asking for the extraction of two factors. The solution was quite similar to that reported by Vredeveldt et al. (2016): the first factor explained 27.4% of the variance and was heavily loaded by ‘repetition’, ‘elaboration’, ‘restatement’, ‘renewed remembering’ and ‘acknowledgement’ – it thus identified *content-focused* strategies; the second factor explained 25.5% of the variance and was heavily loaded by ‘correction’, ‘explanation’, ‘successful cueing’ and ‘role division’ – it thus identified *process-focused* strategies (see Table 1). It should be noted that the factor loadings of ‘correction’, ‘explanation’, and ‘successful cueing’ strategies were negative: this means that participants scoring high on this factor were less likely to use these strategies.

We then computed Pearson’s correlations between the factorial scores of the content- and process-focused factors (computed with the regression method) and the variables obtained from the GSS1. To facilitate comprehension, the scores of the process-focused factor were reversed in sign. The results, illustrated in Table 2, showed that the use of content-focused strategies was a) positively and significantly correlated with the number of items retrieved, $r=0.57$, $p=0.026$ for immediate recall and $r=0.57$, $p=0.027$ for delayed recall; and b) negatively correlated with the number of confabulated items produced in the delayed recall, $r=-0.63$, $p=0.012$. Separate analyses showed that the positive effects of content-focused strategies were mostly driven by elaboration ($r=0.55$, $p=0.035$ for immediate recall and $r=0.62$, $p=0.014$ for delayed recall). On the other hand, the use of process-focused strategies showed significant negative correlations with shift scores, $r=0.53$, $p=0.039$, and total suggestibility, $r=0.54$, $p=0.037$, suggesting that participants who used process-focused strategies to a greater extent were less likely to change their answers to leading questions after receiving the negative feedback and were less suggestible. Additional analyses indicated that these negative effects were primarily due to successful cueing ($r=-0.61$, $p=0.015$ for shift scores and $r=-0.52$, $p=0.046$ for total suggestibility) and correction ($r=-0.49$, $p=0.066$ for shift scores and $r=-0.57$, $p=0.026$ for total suggestibility).

Discussion

The present study examined the performance of collaborative and nominal triads in the GSS1, a test specifically aimed at measuring interrogative suggestibility (Gudjonsson, 1984,

Table 1 Descriptive statistics for retrieval strategies and results of the exploratory factor analysis

Retrieval strategies	M (SD)	Factor 1 (content-focused)	Factor 2 (process-focused)
Successful cue	7.26 (2.76)	0.19	-0.84
Failed cue	2.20 (1.47)	0.46	0.21
Acknowledgement	10.93 (3.95)	0.49	0.39
Correction	10.46 (3.60)	0.06	-0.88
Elaboration	34.00 (9.47)	0.77	0.21
Explanation	1.66 (1.34)	-0.26	-0.62
Restatement	3.86 (2.29)	0.65	-0.23
Repetition	9.53 (6.36)	0.90	-0.16
Renewed remembering	0.20 (0.41)	0.58	-0.08
Role division	0.33 (0.48)	0.03	0.63

Note: The numbers in bold represent factor loadings greater than 0.30: the variables associated with these loadings provide a significant contribution to the definition of the corresponding factor

1997). To summarize, we found that collaborative groups recalled a lower number of correct story elements, compared to nominal groups, thereby exhibiting the classical effects of collaborative inhibition in both immediate and delayed recall tests (Basden et al., 1997; Weldon & Bellinger, 1997). On the other hand, collaborative groups were less likely than nominal groups to produce confabulated elements, to give in to leading questions, and to change their answers after the administration of the negative feedback.

When compared with previous results, the current data showed both similarities and differences. The finding that collaboration reduced the number of yield responses, as well as the total suggestibility scores, replicates and extends to a different sample and a different version of the GSS the conclusions reached by Rossi-Arnaud et al. (2019). At the same time, the analyses performed on immediate and delayed recall tests showed robust effects of collaborative inhibition, which were not significant in our previous study. This might be due to the fact that here we tested the performance of triads, whereas Rossi-Arnaud et al. (2019) examined dyads. Larger groups are more sensitive to the negative effects of collaborative inhibition compared to smaller groups. Similarly, Basden et al.

(2000) found a significant collaborative inhibition in four-person groups, but not in two-person groups. This was, in fact, one of the key predictions of the “*retrieval strategy disruption hypothesis*”: participants in larger groups are exposed to a greater number of responses coming from their collaborators and this may interfere to a larger extent with the use of idiosyncratic retrieval strategies (see Marion & Thorley, 2016, for a meta-analysis).

The primary aim of our study was to shed light on the mechanisms underlying the collaborative reduction in yield responses. We showed that the difference between collaborative and nominal groups in yield 1 scores became non-significant when we took into account the questions to which at least one of the participants gave in but was corrected during the discussion by collaborators. This means that members of nominal and collaborative groups provided the same number of affirmative responses to leading questions; however, the latter were more likely to monitor the source of others’ responses and to prune their errors (Harris et al., 2012; Rajaram, 2011). In other words, it was not the case that collaboration decreased the tendency to produce affirmative responses to misleading questions; rather, it enhanced the efficiency of source-monitoring processes (Harris et al., 2012).

While this result corroborates our theoretical predictions, it is equally important to note that the use of the inclusive method did not eliminate the difference between nominal and collaborative groups in yield 2 scores. We believe that this difference might reflect a conservative change in the response criterion used in collaborative groups induced by the administration of negative feedback (Ross et al., 2004, 2008; Rossi-Arnaud et al., 2011; Takahashi, 2007). According to the Gudjonsson and Clark’s model (1986), one factor affecting the probability of giving in to a misleading question is the level of uncertainty about the accuracy of one’s own responses. In agreement with this idea, prior studies showed that members of collaborative groups could inhibit the production of errors if they suspected that their own answers might be

Table 2 Pearson’s correlations between content and process-focused strategies and the GSSI variables

Total sample	Content focused	Process focused
Immediate recall	0.57*	0.00
Delayed recall	0.57*	-0.13
Immediate confabulations	-0.34	0.46
Delayed confabulations	-0.63*	-0.30
Yield 1	-0.37	-0.31
Yield 2	-0.28	-0.23
Shift	-0.26	-0.53*
Total suggestibility	-0.43	-0.54*

wrong (Ross et al., 2004, 2008). What is possible, then, is that the negative feedback increased the participants' uncertainty about the accuracy of their memories. This might have refrained them from contributing low-confidence affirmative responses. The hypothesis is well supported by the finding that the number of leading questions to which collaborative groups gave in decreased after receiving the negative feedback, whereas no such reduction was apparent in the nominal groups.

Another important feature of the present study was that we analysed the use of retrieval strategies during immediate and delayed recall phases. Regarding the immediate and delayed recall performance, our results replicate those reported by Vredeveldt et al. (2016, 2017), showing that couples who spontaneously adopted content-focused strategies (in particular elaboration) recalled more correct story elements and produced less confabulated items. Most importantly, we found that the use of process-focused strategies (in particular cross-cueing and correction) was negatively associated with shift scores and total suggestibility. Thus, triads in which the members cross-cued and corrected each other to a greater extent were less likely to change their responses and exhibited lower levels of suggestibility. These results cannot be explained by the well-known negative association between memory efficiency and suggestibility (Gudjonsson, 1984, 1997; Gudjonsson et al., 2016), because the use of successful cueing and correction strategies did not increase memory performance in the immediate and delayed recall tasks. They instead provide further evidence in support of the view that suggestibility of collaborative groups is critically related to the way in which members interact with each other.

The present results may have relevant consequences for the way in which police investigations are conducted. As suggested by Vredeveldt et al. (2017), around the world, police officers are strongly advised to avoid direct contacts between witnesses, because they might contaminate each other's memories (Gabbert et al., 2006; Wright & Klumpp, 2004). In fact, many previous studies showed that witnesses tend to incorporate into their own memories incorrect details produced by their co-witnesses (Goodwin et al., 2017; Meade & Roediger, 2002). While this is true, the present results, together with those recently reported by Rossi-Arnaud et al. (2020), indicate specific circumstances in which collaboration can have beneficial effects by reducing witnesses' suggestibility. This is especially important in those contexts in which interview procedures might influence eyewitnesses' memories (Gombos et al., 2012; Pezdek et al., 2007). In sum, although we don't want to suggest that witnesses should always be allowed to talk to each other, we believe that future research should try to determine in more details the conditions that maximize the positive effects of collaborative retrieval, by simultaneously limiting its negative consequences (Vredeveldt et al., 2017). This would allow us to formulate

detailed guidelines for policymakers and police practitioners specifying when collaboration should be allowed and when it should be avoided.

Author Contributions CRA and PS developed the research question and CRA, PS and VC designed the experiment. EL collected the data and EL and PS scored the results. PS and SM analyzed the data under the supervision of CRA and VC. CRA, PS and SM drafted the manuscript and all authors participated in the revision process. All authors read and approved the final manuscript.

Funding Open access funding provided by Università degli Studi di Roma La Sapienza within the CRUI-CARE Agreement.

Data Availability The datasets analyzed during the current study are available in the OSF repository: <https://osf.io/ry6ap/>

Declarations

Consent to Participate Informed consent was obtained from all individual participants included in the study.

Ethic Approval The study was approved by the Institutional Review Board (ethics committee), Department of Psychology, Sapienza University. The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

Conflict of Interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Barber, S. J., Harris, C. B., & Rajaram, S. (2015). Why two heads apart are better than two heads together: Multiple mechanisms underlie the collaborative inhibition effect in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(2), 559–566.
- Baden, B. H., Baden, D. R., Bryner, S., & Thomas III, 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.
- Baden, B. H., Baden, D. R., & Henry, S. (2000). Costs and benefits of collaborative remembering. *Applied Cognitive Psychology*, 14(6), 497–507.

- Bianco, A., & Curci, A. (2015). Measuring interrogative suggestibility with the Italian version of the Gudjonsson Suggestibility Scales (GSS): Factor structure and discriminant validity. *Personality and Individual Differences, 82*, 258–265.
- Blumen, H. M., & Rajaram, S. (2008). Influence of re-exposure and retrieval disruption during group collaboration on later individual recall. *Memory, 16*(3), 231–244.
- Blumen, H. M., & Rajaram, S. (2009). Effects of repeated collaborative retrieval on individual memory vary as a function of recall versus recognition tasks. *Memory, 17*(8), 840–846.
- Congleton, A. R., & Rajaram, S. (2011). The influence of learning methods on collaboration: Prior repeated retrieval enhances retrieval organization, abolishes collaborative inhibition, and promotes post-collaborative memory. *Journal of Experimental Psychology: General, 140*(4), 535–551.
- Finlay, F., Hitch, G. J., & Meudell, P. R. (2000). Mutual inhibition in collaborative recall: Evidence for a retrieval-based account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(6), 1556–1567.
- Gabbert, F., Memon, A., & Wright, D. B. (2006). Memory conformity: Disentangling the steps toward influence during a discussion. *Psychonomic Bulletin & Review, 13*(3), 480–485.
- Gombos, V., Pezdek, K., & Haymond, K. (2012). Forced confabulation affects memory sensitivity as well as response bias. *Memory & Cognition, 40*(1), 127–134.
- Goodwin, K. A., Hannah, P. J., Nicholl, M. C., & Ferri, J. M. (2017). The confident co-witness: The effects of misinformation on memory after collaborative discussion. *Applied Cognitive Psychology, 31*(2), 225–235.
- Gudjonsson, G. H. (1984). A new scale of interrogative suggestibility. *Personality and Individual Differences, 5*(3), 303–314.
- Gudjonsson, G. H. (1997). *The Gudjonsson Suggestibility Scales manual*. Psychology Press.
- Gudjonsson, G. H., & Clark, N. K. (1986). Suggestibility in police interrogation: A social psychological model. *Social Behaviour, 1*(2), 83–104.
- Gudjonsson, G., Vagni, M., Maiorano, T., & Pajardi, D. (2016). Age and memory related changes in children's immediate and delayed suggestibility using the Gudjonsson Suggestibility Scale. *Personality and Individual Differences, 102*, 25–29.
- Harris, C. B., Paterson, H. M., & Kemp, R. I. (2008). Collaborative recall and collective memory: What happens when we remember together? *Memory, 16*(3), 213–230.
- Harris, C. B., Bamier, A. J., & Sutton, J. (2012). Consensus collaboration enhances group and individual recall accuracy. *Quarterly Journal of Experimental Psychology, 65*(1), 179–194.
- Henkel, L. A., & Rajaram, S. (2011). Collaborative remembering in older adults: Age-invariant outcomes in the context of episodic recall deficits. *Psychology and Aging, 26*(3), 532–545.
- 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), 1141–1164.
- Mastroberardino, S., & Marucci, F. S. (2013). Interrogative suggestibility: Was it just compliance or a genuine false memory? *Legal and Criminological Psychology, 18*(2), 274–286.
- Meade, M. L., & Roediger, H. L. (2002). Explorations in the social contagion of memory. *Memory & Cognition, 30*(7), 995–1009.
- Pezdek, K., Sperry, K., & Owens, S. M. (2007). Interviewing witnesses: The effect of forced confabulation on event memory. *Law and Human Behavior, 31*(5), 463–478.
- Rajaram, S. (2011). Collaboration both hurts and helps memory: A cognitive perspective. *Current Directions in Psychological Science, 20*(2), 76–81.
- Rajaram S., & Maswood R. (2018). Collaborative memory: A selective review of data and theory. In: John H. Byrne (ed.) *Learning and memory: A comprehensive reference* (2nd edition). Academic Press.
- Rajaram, S., & Pereira-Pasarin, L. P. (2007). Collaboration can improve individual recognition memory: Evidence from immediate and delayed tests. *Psychonomic Bulletin & Review, 14*(1), 95–100.
- Rajaram, S., & Pereira-Pasarin, L. P. (2010). Collaborative memory: Cognitive research and theory. *Perspectives on Psychological Science, 5*(6), 649–663.
- Ross, M., Spencer, S. J., Linardatos, L., Lam, K. C., & Perunovic, M. (2004). Going shopping and identifying landmarks: Does collaboration improve older people's memory? *Applied Cognitive Psychology, 18*(6), 683–696.
- Ross, M., Spencer, S. J., Blatz, C. W., & Restorick, E. (2008). Collaboration reduces the frequency of false memories in older and younger adults. *Psychology and Aging, 23*(1), 85–92.
- Rossi-Arnaud, C., Pieroni, L., Spataro, P., & Cestari, V. (2011). Effects of pair collaboration and word-frequency in recognition memory: A study with the remember-know procedure. *Scandinavian Journal of Psychology, 52*(6), 516–523.
- Rossi-Arnaud, C., Spataro, P., Bhatia, D., & Cestari, V. (2019). Collaborative remembering reduces suggestibility: A study with the Gudjonsson Suggestibility Scale. *Memory, 27*(5), 603–611.
- Rossi-Arnaud, C., Spataro, P., Bhatia, D., Doricchi, F., Mastroberardino, S., & Cestari, V. (2020). Long-lasting positive effects of collaborative remembering on false assents to misleading questions. *Acta Psychologica, 203*, 102986.
- Takahashi, M. (2007). Does collaborative remembering reduce false memories? *British Journal of Psychology, 98*(1), 1–13.
- Thorley, C., & Dewhurst, S. A. (2007). Collaborative false recall in the DRM procedure: Effects of group size and group pressure. *European Journal of Cognitive Psychology, 19*(6), 867–881.
- Vredeveldt, A., Hildebrandt, A., & Van Koppen, P. J. (2016). Acknowledge, repeat, rephrase, elaborate: Witnesses can help each other remember more. *Memory, 24*(5), 669–682.
- Vredeveldt, A., Groen, R. N., Ampt, J. E., & van Koppen, P. J. (2017). When discussion between eyewitnesses helps memory. *Legal and Criminological Psychology, 22*(2), 242–259.
- Weldon, M. S., & Bellinger, K. D. (1997). Collective memory: Collaborative and individual processes in remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(5), 1160–1175.
- 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.