Collaborative recall of details of complex emotional material

Ineke Wessel¹, Anna Roos E. Zandstra¹, Hester H. Hengeveld¹, Michelle L. Moulds²

¹ University of Groningen, The Netherlands, ² University of New South Wales, Australia

Short report, word count (excl. title page & references) = 3994

Address for correspondence:  Ineke Wessel

Department of Clinical Psychology and Experimental Psychopathology

University of Groningen

Grote Kruisstraat 2-1

9712 TS Groningen

The Netherlands

Phone  (+31) 50 3637617

FAX  (+31) 50 3637602

e-mail J.P.Wessel@rug.nl
Abstract

Collaborative inhibition refers to the phenomenon that when several people work together to produce a single memory report, they typically produce fewer items than when the unique items in the individual reports of the same number of participants are combined (i.e., nominal recall). Yet, apart from this negative effect, collaboration may be beneficial in that group members remove errors from a collaborative report. Collaborative inhibition studies on memory for emotional stimuli are scarce. Therefore the present study examined both collaborative inhibition and collaborative error reduction in the recall of the details of complex, emotional material in a laboratory setting. Female undergraduates ($n = 111$) viewed a film clip of a fatal accident and subsequently engaged in either collaborative ($n = 57$) or individual recall ($n = 54$) in groups of three. The results show that across several detail categories, collaborating groups recalled fewer details than nominal groups. However, overall, nominal recall produced more errors than collaborative recall.

The present results extend earlier findings on both collaborative inhibition and error reduction to the recall of complex, affectively-laden material. These findings may have implications for the field of eyewitness memory.

KEY WORDS: collaborative inhibition, error-pruning, detail memory, emotion, trauma film paradigm
Collaborative recall of details of complex emotional material

In recent years, there has been an increased interest in the way social influences shape memory. Studies on collaborative recall show that remembering with others may come with costs as well as advantages (Rajaram & Pereira-Pasarin, 2010). The typical collaborative recall experiment (e.g., Weldon & Bellinger, 1997) involves an encoding phase in which participants are individually presented with the material (e.g., wordlists) that they will later be asked to recall. Next, groups of three participants retrieve the studied material and produce a single account by working together. Compared to individual recall, working together is advantageous in that more accurate items are produced. However, a different picture emerges if the output of the group work is compared to nominal recall. Nominal recall contains all non-redundant items reported by three participants who were tested individually. As such, it reflects the potential of items that could have been recalled if these individuals had formed a group. Collaborative recall tends to be inferior to nominal remembering. That is, when study participants recall together, they are likely to produce an account that contains fewer items than when the items recalled by an equal number of separate participants are combined. This poorer memory following collaboration is referred to as collaborative inhibition.

Research shows that collaborative inhibition occurs with various types of materials, including simple stimuli (words, pictures; e.g., Weldon & Bellinger, 1997) and complex, more ecologically relevant stimuli (e.g., stories containing socially relevant information, Reysen, Talbert, Dominko, Jones, & Kelley, 2011; grocery lists, Ross, Spencer, Linardatos, Lam, & Perunovic, 2004; see Rajaram & Pereira-Pasarin, 2010, for an overview). Yet, there seems to be a paucity of literature that has investigated memory for emotionally-laden material (Wessel & Moulds, 2008). To the best of our knowledge, the only study addressing emotional material examined collaborative memory for details of the assassination of the Israeli Prime Minister Rabin more than six years earlier (Yaron-Antar & Nachson, 2006).
Reminiscent of findings with neutral stimuli, nominal recall of such an emotional event contained more details than collaborative remembering. However, nominal recall also contained more errors. In terms of the percentage of correct information (i.e., the percentages correct recall of the total number of information units reported, including errors), collaborative recall was superior to both nominal and individual recall. Thus, although these findings regarding the number of correct items suggest that the costs of collaboration (i.e., inhibition) extend to real-life negative events, they also seem to indicate that collaboration is beneficial in terms of reducing error.

This finding is in contrast with the literature showing that social influence increases false memories (see for overviews, Harris, Paterson, & Kemp, 2008; Rajaram & Pereira-Pasarin, 2010; Wessel & Moulds, 2008). Methods in this literature include introducing false items in collaborative recall (i.e., social contagion, e.g., Roediger, Meade, & Bergman, 2001), manipulating encoding conditions such that participants report different details during a collaborative memory test (i.e., memory conformity, e.g., Gabbert, Memon, & Allan, 2003), and employing lists of words carrying strong semantic associations with a non-presented lure (i.e., DRM paradigm, e.g., Thorley & Dewhurst, 2007). However, these paradigms are designed to capitalize on errors (Rajaram & Pereira-Pasarin, 2010) whereas the standard collaborative recall method used by Yaron-Antar and Nachson (2006) allows for studying spontaneously occurring false memories. A few collaborative recall studies compared such unintended memory errors for neutral stimuli between collaborative and individual conditions. These studies render a mixed pattern of results (Harris, Barnier, & Sutton, 2012; Ross, Spencer, Blatz, & Restorick, 2008). That is, collaboration yielded more (e.g., Basden, Basden, Bryner, & Thomas, 1997), fewer (e.g., Ross et al., 2004; Ross et al., 2008), or comparable levels of intrusion errors (e.g., Weldon & Bellinger, 1997, exp. 1). This mixed pattern may have to do with how group members interact during collaboration. In some studies (e.g.,
Basden et al., 1997) interaction was minimal, because group members took turns recalling words. However, engaging in a free-flowing discussion would increase the probability of error-pruning, that is, the editing out of erroneous items recalled by a minority of the group (Rajaram & Pereira-Pasarín, 2010). In a recent study (Harris et al., 2012) the opportunity for error-pruning was maximized by instructing collaborative groups to reach consensus. Indeed, the consensus groups produced fewer errors than both turn-taking and nominal groups.

Taken together, although collaboration has costs in terms of the number of items recalled (i.e., collaborative inhibition), it may also have benefits for memory accuracy (i.e., fewer errors). This latter effect may be of special relevance for applied areas where the accuracy of event memory is important, such as eyewitness memory. The present study examined whether both collaborative inhibition and collaborative error reduction extend to the recall of details of complex, emotional material in a laboratory setting. Yaron-Antar and Nachson’s study (2006) is unique in that it was concerned with memory for a highly emotional, naturally occurring event. However, the disadvantage of their field approach is that the conditions under which the news of prime-minister Rabin’s assassination was encoded were not standardised. Therefore, in the present study, participants watched an emotional film depicting a fatal accident before they engaged in either individual or group recall. To enhance the probability of error-pruning, groups were instructed to reach consensus. Some types of details may be poorly encoded (e.g., plot-irrelevant or peripheral details such as hair colour, see Christianson, 1992), and may thus be particularly prone to omission or false recall. For this reason we explored both collaborative inhibition and error reduction effects for different categories of details, ranging in their centrality or importance to the general theme of the film-clip.

**Method**

**Participants**
One hundred and eleven female first-year psychology students from the University of Groningen participated in exchange for course credit. They were tested in one of two conditions, collaborative recall \((n = 57)\), or individual recall \((n = 54)\). The study was approved by the ethical committee of the Department of Psychology of the University of Groningen.

**Material**

The emotional film was part of an anti-drink and drive campaign by the Northern Ireland Department of Environment (DOE) running from 2000 to 2005. The clip (“Shame”; [http://www.youtube.com/watch?v=xtJqw--DGI8](http://www.youtube.com/watch?v=xtJqw--DGI8)) showed how a little boy playing soccer in his back garden is killed by an out-of-control car. The length of the film clip was 1:04 min. It was presented on a 21 inch LCD monitor attached to a desktop computer with headphones for sound.

For the purpose of free recall, an opened text document was displayed on the computer monitor. The top of the document contained instructions to participants that they should type everything that they recalled from the film in as much detail as possible. The time that it took each participant to complete the recall protocol was recorded by the researchers with a generic digital kitchen timer.

**Procedure**

Groups of three unacquainted students were tested simultaneously. Triads were randomly assigned to the individual or collaborative recall conditions. Participants were told that the purpose of the experiment was to study the effects of watching an emotional film on group behaviour. The upcoming free recall test was not mentioned. It was stressed to participants that the purpose of the film was to elicit emotion, but they were informed that they could stop watching at any moment they wished. After signing informed consent, participants watched the film individually, each on a separate computer. Subsequently, for the purpose of distraction, the triads in both conditions collaborated on solving a Logic Grid puzzle for five
minutes. Next, the participants received instructions to recall every detail of the film as accurately as possible. They were asked to imagine that they were giving an eye-witness report to the police, and that in this context even seemingly irrelevant details are important. Examples of types of details were given (e.g., clothing, appearance, colours, objects, background). Participants in the collaborative recall condition received additional instructions to work together as a group to produce one recall protocol. Specifically, they were told that every detail that the group agreed on should be typed into the text document on a single computer. One person was to serve as typist (but also should participate in the collaborative recall). In the individual recall condition, instructions were that each participant should work on their separate computer and type in everything that they could recall into the empty text-document.

In both conditions participants were told to take all of the time they needed. Participants were informed that when they reached the point that they were unable to recall any new details, and one minute had elapsed, they could stop the recall test and save their file. At the end of the experiment participants were fully debriefed.

**Coding of Recall Protocols**

The film was transcribed including all visual and auditory details. Based on this transcript, we developed a coding system with a central/peripheral detail distinction in mind (e.g., Burke, Heuer, & Reisberg, 1992). Specifically, there were four broad categories that varied in relevance to the plot of the film in a hierarchical fashion. On the central end, the Actions category reflected the plot of the film. It contained descriptions of the behaviour of or events involving the 4 main characters (i.e., young man, little boy, father and sister; e.g., “The boy is playing soccer”, 1 unit; “The father is crying”, 1 unit). We took Action Details as a less central category. Details in this category were (perceptual descriptions of) objects involved in an action (e.g., “The dark-blue car bursts through the fence”, 3 units). The next, more
peripheral detail category was *Person Details*. This reflected the appearance of the 4 characters in the film, including elements such as age estimations, hair colour, posture, and clothing (e.g., “The *boy* is wearing a yellow shirt”, 3 units). The most peripheral category was *Background Details*, reflecting everything that occurred in the film that was irrelevant to the plot. Examples are locations, surroundings, bystanders, objects (when not part of an action) and music (e.g., “In the background a *woman* is watching”, 2 units). For each recall protocol, the number of correct information units in each category was counted. A nonspecific detail (e.g., “the dark-coloured car”, while it was dark-blue) was counted as partly accurate (½ unit).

In addition, errors were noted. These could reflect distortions (e.g., “The *boy* [instead of girl] was on the swing”) and confabulations (e.g., “The boy called his father”, which did not happen).

The recall stories written by the participants were scored by a rater (ARZ) who was blind to the condition in which the protocol was generated. In order to assess reliability, a second rater (HH) scored the accuracy of twenty-one recall protocols (28.8%). Intraclass Correlation Coefficients (ICC, Shrout & Fleiss, 1979) were calculated using a two-way random model using an absolute agreement definition. Reliabilities were good, with single measures ICC = .89 for actions, ICC = .87 for action details, ICC = .92 for person details and ICC = .87 for background details.

**Pooling of Protocols**

The individual stories were pooled, creating a third condition (Nominal condition, *n* = 18). The pooled protocols contained every unique correct detail in the three individual recall protocols constituting a nominal group. Thus, if participant 1 mentioned the car and the teddy bear and participant 2 recalled the teddy bear and the girl on the swing, the pooled protocol would contain the car, teddy bear and girl on the swing. For errors, we counted each unique detail in the three protocols only once, and we counted only those details as errors that had not
already been counted as a correct detail. For example, when the pooled recall protocol contained one correct index of the colour of the car (dark-blue) and two errors (green, purple), the detail “car-colour” was scored as correct. If however, the protocol contained no correct information but one to three erroneous colours (green, grey, purple) it was counted as one error. Thus, each unique detail was scored once, as either correct or incorrect. This would provide the most optimal comparison with the collaborative condition, in which instructions were to include the details in their report that participants agreed on.

The main dependent variables of interest were net accuracy scores, i.e., the number of correct information units in each category minus the number of errors in that category. In addition, because we were especially interested in the effects of collaboration on error reduction, we analysed the errors separately. In general, the variables were not normally distributed. Therefore, the data were analysed with robust methods (see Wilcox, 2003, for the reasons for preferring these methods over nonparametric methods). Specifically, we used Wilcox’ (2012) Yuen t-test function in R for conducting a priori comparisons between the nominal and the collaborative conditions (collaborative inhibition) and between the individual and collaborative conditions. Robust Cohen’s d effect sizes (\(d_R\); Algina, 2005; Algina, Keselman, & Penfield, 2005) were calculated using \(n = 600\) bootstrap samples.

**Results**

Table 1 presents the 20% trimmed means and square rooted winsorised variances for the net accuracy scores in each detail category. Compared to the nominal groups, the overall net accuracy score was significantly smaller in the collaborative recall condition, \(T_y(20.5) = 8.21, p < .001, d_R = -2.64\). As for the different detail categories, collaborative groups recalled fewer net accurate actions, \(T_y(22.9) = 9.05, p < .001, d_R = -2.86\), action details \(T_y(16.7) = 5.73, p < .001, d_R = -1.86\), person details, \(T_y(17.0) = 4.23, p < .001, d_R = -1.37\), and background details
Collaborative recall of … 10

$T_\gamma(23.0) = 5.80, p < .001, d_R = -1.83$, than the nominal condition. Thus, a collaborative inhibition effect was observed, regardless of detail category.

Compared to the individual recall condition, the total net accuracy score was significantly larger in the collaborative recall condition $T_\gamma(20.5) = 2.47, p = .02, d_R = -0.67$. As for the different detail categories, this difference reached statistical significance for the person detail category, $T_\gamma(22.7) = 2.70, p = .01, d_R = -0.27$, whereas the action detail category showed a trend towards significance, $T_\gamma(26.8) = 1.80, p = .08, d_R = -0.43$. The net accuracy scores in the collaborative and individual recall conditions did not differ significantly with respect to actions, $T_\gamma(19.2) = 0.24, p = .81, d_R = -0.07$, and background details, $T_\gamma(23.9) = 1.05, p = .30, d_R = -0.27$. We repeated all tests for uncorrected scores and found similar results, except that the difference between collaborative and individual groups for action details was also significant, $T_\gamma(27.4) = 2.20, p = .03, d_R = -0.55$.

Table 2 presents the 20% trimmed means and winsorised variances of the numbers of errors in the recall stories. As can be seen in the table, overall, participants in the nominal condition reported significantly more erroneous details than those in the collaborative condition, $T_\gamma(21.5) = 2.97, p = .007, d_R = -0.95$. This difference reached statistical significance for actions, $T_\gamma(21.2) = 2.18, p = .04, d_R = -0.70$, and person details $T_\gamma(17.5) = 2.34, p = .03, d_R = -0.76$, but not for action details, $T_\gamma(18.0) = 0.65, p = .52, d_R = -0.21$, and background details, $T_\gamma(22.0) = 0.24, p = .81, d_R = -0.08$. The collaborative condition did not make significantly more errors than the individual condition in any of the detail categories, all $T_\gamma < 1.6, p > .12$ and $-0.49 > d_R < 0.14$.

**Discussion**

The present study explored the effects of collaboration on the recall of the details of an emotional film clip. In line with expectations, a substantial collaborative inhibition effect was observed. That is, the memory report of three participants working together contained fewer
correct details than was to be expected given the sum of all unique details in the independent reports of three individual participants. By contrast, overall, the collaborative reports contained more details than the recall protocols of single participants. Whereas collaborative inhibition was observed for all detail categories, better performance of the collaborative groups compared to single individuals was most pronounced for details concerning the appearance of the main characters in the film. By and large, the results are in line with the typical finding in this area (see Rajaram & Pereira-Pasarin, 2010, for an overview). Thus, collaborative inhibition seems to be a robust phenomenon that generalizes to the recall of complex emotional material.

It should be noted that we found similar patterns of collaborative inhibition for the absolute number of correct details as well as when we corrected those absolute numbers for errors. In contrast, Yaron-Antar and Nachson (2006) found a reversal of the collaborative inhibition effect when they took errors into account. That is, in this previous study collaborative recall was better than nominal recall, whereas we found that nominal recall was better. These contrasting findings may be due to a different treatment of the errors in the nominal recall condition specifically. Yaron-Antar and Nachson (2006) assigned points for accuracy as well as inaccuracy to a detail such as the name of Rabin’s assassin if a pooled protocol contained both the correct and a false name. However, in the collaborative condition each detail could have been scored only once (i.e., either correct or incorrect) and thus errors in the nominal condition would have received more weight. Instead, to enhance comparability with the collaborative protocols, we counted each detail in the nominal condition only once, as either correct or incorrect. That is, we let correct details prevail over incorrect details, particularly because the collaborative groups had received an instruction to reach consensus. We reasoned that if error-pruning is indeed an important feature of recalling collaboratively
(Rajaram & Pereira-Pasarin, 2010), maximizing the probability of scoring correctly in the nominal protocols would provide an appropriate control.

Despite reducing their importance by taking this approach, analysing false details separately showed that overall, nominal recall still produced more errors than collaborative recall. This was especially the case for details about the appearance of the main characters in the story (person details) and their actions. Our overall pattern of results is in line with Harris et al.’s finding (2012) that consensus collaborative groups made fewer errors than nominal groups. In addition, these authors found that the later individual recall of former consensus group members was more accurate. Interestingly, they observed that consensus group members did not make fewer errors during their actual discussion. This suggests that errors were edited out of the group report and that this improved individual memory accuracy later on. From a forensic psychology point of view, it would be interesting to see whether such error-pruning (see Rajaram & Pereira-Pasarin, 2010) in consensus groups carries over to later individual recall of complex emotional stimulus material. If so, this may imply that under some circumstances, it might be beneficial to have eyewitnesses of the same event work together.

Another interesting question for future research is whether the collaborative inhibition effect in the present study is driven by the same mechanism as in studies that employ the standard paradigm which relies upon word lists. In general, retrieval strategy disruption (Basden et al., 1997) is regarded the best explanation for collaborative inhibition (see Hirst & Echterhoff, 2012; Rajaram & Pereira-Pasarin, 2010). That is, at study individual group members organize the to-be-remembered material in their own fashion to aid later retrieval. During collaborative recall, each group member would follow their own retrieval strategy. Mismatches would undermine each individual’s performance, leading to group output that would be below the potential of all of the group-members combined. Several methodological
features may enhance retrieval strategy disruption (Basden et al., 1997). For example, organization at study would be prompted by instructions that participants’ memory will be tested later on or by the specific nature of the stimulus material (e.g., lists containing exemplars of several categories). The film clip in the present study was encoded incidentally and contained a clear storyline. Thus, participants would be unlikely to organize the material for purpose of generating a retrieval strategy. Moreover, the chronological order of events in the film might impose an obvious retrieval strategy, leaving less room for retrieval disruption due to different strategies to occur in the collaborative groups. Nevertheless, the complex nature of the film (showing many events and details within a limited timeframe) might still have permitted multiple conflicting strategies to emerge during retrieval. Future studies may determine whether retrieval strategy disruption indeed drives collaborative inhibition for rich, affectively laden material. For example, structuring the recall task (e.g., recognition), imposing a single retrieval strategy, should alleviate retrieval strategy disruption and attenuate collaborative inhibition.

In sum, the present study showed that collaborative inhibition generalizes to the recall of complex emotional material. This result adds to the literature on this phenomenon - a literature that to date includes only a minimal number of studies that have employed stimuli that are relevant to real-world emotional events (see Yaron-Antar & Nachson, 2006). The present results also extend earlier findings that collaboration produces fewer errors in the recall of affectively laden material. Future work may determine whether this beneficial effect carries over to individual memory. Answering this question may have implications for the field of eyewitness memory.
Note

Initially, 22 and 19 triads were tested in the collaborative and individual condition, respectively. Screens for depressive (Beck Depression Inventory, second edition; BDI-II, Beck, Steer, Ball, & Ranieri, 1996; van der Does, 2002) and posttraumatic stress symptoms (Trauma Screening Questionnaire, TSQ, Brewin et al., 2002) were administered after the recall test and revealed scores above cut-off for six participants. As high scores may influence recall patterns, protocols of the triads that contained a high scorer ($n = 3$) and protocols from individually tested high scorers ($n = 3$) were not coded. For the nominal condition that constituted three individual recall protocols, the recall data of the remaining six participants in the three affected individual triads were considered as two pooled protocols.
References


doi:10.1177/1745691610388763


Table 1.

20% Trimmed Means (Winsorised Variances) of Net Accuracy Scores per Detail Category for Nominal, Collaborative, and Individual Recall Conditions

<table>
<thead>
<tr>
<th>Recall condition</th>
<th>Nominal</th>
<th>Collaborative</th>
<th>Individual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 18)</td>
<td>(n = 19)</td>
<td>(n = 54)</td>
</tr>
<tr>
<td>Actions</td>
<td>31.7&lt;sub&gt;a&lt;/sub&gt; (85.3)</td>
<td>17.5&lt;sub&gt;b&lt;/sub&gt; (115.9)</td>
<td>17.2&lt;sub&gt;b&lt;/sub&gt; (80.3)</td>
</tr>
<tr>
<td>Action Details</td>
<td>20.3&lt;sub&gt;a&lt;/sub&gt; (148.3)</td>
<td>11.3&lt;sub&gt;b&lt;/sub&gt; (45.5)</td>
<td>9.7&lt;sub&gt;b&lt;/sub&gt; (194.6)</td>
</tr>
<tr>
<td>Person Details</td>
<td>21.5&lt;sub&gt;a&lt;/sub&gt; (189)</td>
<td>14&lt;sub&gt;b&lt;/sub&gt; (61.8)</td>
<td>11.3&lt;sub&gt;c&lt;/sub&gt; (63.1)</td>
</tr>
<tr>
<td>Background</td>
<td>14.4&lt;sub&gt;a&lt;/sub&gt; (57.7)</td>
<td>7&lt;sub&gt;b&lt;/sub&gt; (75.9)</td>
<td>5.8&lt;sub&gt;b&lt;/sub&gt; (86.3)</td>
</tr>
<tr>
<td>Total</td>
<td>87.5&lt;sub&gt;a&lt;/sub&gt; (960.3)</td>
<td>51.7&lt;sub&gt;b&lt;/sub&gt; (567.9)</td>
<td>44.4&lt;sub&gt;c&lt;/sub&gt; (462.3)</td>
</tr>
</tbody>
</table>

20% Trimmed Means with different subscripts differ significantly ($p < .05$).
# Table 2.
20% Trimmed Means (Winsorised Variances) of Number of Errors per Detail Category for Nominal, Collaborative, and Individual Recall Conditions

<table>
<thead>
<tr>
<th>Recall condition</th>
<th>Nominal ((n = 18))</th>
<th>Collaborative ((n = 19))</th>
<th>Individual ((n = 54))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actions</td>
<td>0.58(_a) (2.61)</td>
<td>0.08(_b) (1.75)</td>
<td>0.18(_b) (2.12)</td>
</tr>
<tr>
<td>Action Details</td>
<td>1.17(_a) (13.9)</td>
<td>0.85(_a) (5.44)</td>
<td>0.41(_a) (2.52)</td>
</tr>
<tr>
<td>Person Details</td>
<td>2.83(_a) (23.7)</td>
<td>1.35(_b) (8.36)</td>
<td>1.25(_b) (12.5)</td>
</tr>
<tr>
<td>Background</td>
<td>0.83(_a) (8.11)</td>
<td>0.73(_a) (6.45)</td>
<td>0.47(_a) (2.54)</td>
</tr>
<tr>
<td>Total</td>
<td>5.42(_a) (29.4)</td>
<td>3.08(_b) (21.1)</td>
<td>2.97(_b) (21.4)</td>
</tr>
</tbody>
</table>

20% Trimmed Means with different subscripts differ significantly \((p < .05)\).