

Gadgil, S. M. & Nokes, T.J. (2010). Collaborative facilitation through error-detection: A classroom experiment. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2583-2588). Austin, TX: Cognitive Science Society.

Collaborative Facilitation through Error-Detection: A Classroom Experiment

Soniya Gadgil (smg58@pitt.edu)

Timothy J. Nokes (nokes@pitt.edu)

University of Pittsburgh

Learning Research and Development Center, 3939, O'Hara Street

Pittsburgh, PA, 15260 USA

Abstract

Prior work has shown that individuals working in groups often perform worse than individuals working alone, a finding commonly referred to as collaborative inhibition. In the current work we examine whether engaging in error correction processes can mitigate or eliminate the collaborative inhibition effect and perhaps even facilitate collaborative facilitation. Participants engaged in a writing error-detection and revision task while working either with a partner or individually. On the error-detection task, dyads found more structural flaws in the text, whereas individuals found more surface flaws. Moreover, when comparing dyads nominal groups the dyads did not show the collaborative inhibition effect. A similar pattern of results was found on the revision task. The results are discussed in terms of the underlying cognitive and social processes that support successful collaboration.

Keywords: collaborative learning; error-detection; instruction.

Introduction

When does collaboration lead to robust performance and learning outcomes? A large amount of evidence from past research shows that when individuals collaborate with one or more partners, it leads to better performance outcomes when compared to the average individual (see Hill, 1982; Kerr & Tindale, 2004 for reviews). This result has been found in number different tasks and domains. It is hypothesized that groups are able to “pool” their resources and knowledge to perform better than the average individual whether brainstorming, memorizing lists of words, or solving puzzle problems.

Although groups tend to perform better than the average individual, individuals working in groups often do not perform up to their predicted potential. An extremely robust finding in the collaboration literature is that individuals working in groups actually perform worse than individuals working alone (Andersson & Ronnberg, 1995; Weldon & Bellinger, 1997). This has been referred to as “collaborative inhibition” or “process loss” (Steiner, 1972). It is often measured by comparing the dyad or group performance to nominal group performance. For example, when comparing dyads and individuals, a nominal dyad is formed by randomly pairing two individuals who did not collaborate, and their joint performance is considered, as if they were a collaborating dyad. In a simple list learning task, if a dyad recalled the following letters from a list (a, b, c, d, e, f, g) and two

individuals recalled the following: individual 1 (a, b, e, g, h) and individual 2 (c, d, e, f, i) the dyad performs better than the average individual (7 vs. 5) but worse than the nominal dyad (pooled performance: 7 vs. 9).

Much research has focused on trying to understand the causes of collaborative inhibition. Both cognitive and social factors have been advanced to explain it. Social factors include the free-rider effect or social loafing (Karau & Williams, 1993), evaluation anxiety (Collaros & Anderson, 1969; Mullen, 1983), and diffusion of responsibility. Cognitive factors include cognitive overhead of coordination during collaboration (Steiner, 1972) and disruption of retrieval strategy due to interference caused by the collaborators’ input (Basden, Basden, Bryner, & Thomas, 1997; Finlay, Hitch, & Meudel, 2000; Weldon, Blair, & Huebsch, 2000). Each of these factors has been shown to contribute to the collaborative inhibition effect.

In addition to identifying factors that increase or decrease collaborative inhibition a few studies have shown an elimination of the collaborative inhibition effect or even an advantage for collaborative groups over nominal groups. For example, Wright and Klumpp (2004) compared individuals and two collaborative group conditions during free recall: a “see” condition in which people in the pairs took turns recalling items from a previously studied list and showed each other the words as they were being recalled, and a “no see” condition, in which the participants again took turns recalling the previously seen list, but neither knew which items the other person had recalled. Thus, the “no see” condition was effectively the same as a nominal group, as the participants did not engage in any form of interaction while recalling the items. Not surprisingly, Wright and Klumpp found that the “no see” group performed significantly better than the “see” group, and equal to nominal groups.

In another demonstration of collaborative facilitation, Takahashi and Saito (2004) compared recall of studied story materials by nominal dyads and collaborators. When tested immediately, they found that nominal dyads performed better than collaborators, however, when tested after a one-week delay, collaborators recalled more than nominal dyads.

These findings suggest that there must be some aspect of the task structure in which collaborators engage that play a part in determining collaborative inhibition or

advantage. We propose that if the task structure facilitates the cognitive mechanisms hypothesized to underlie the collaborative advantage, we should be able to overcome the collaborative inhibition effect.

One of the primary mechanisms suggested to underlie successful collaboration is error-detection (Shaw, 1932; Sniezek & Henry, 1989). Groups are hypothesized to engage in a higher degree of error detection and correction compared to individuals. It has been widely documented that detecting your own errors is an important metacognitive skill, however, not many learners have such skills. Moreover, in order to detect an error, it is necessary to have the requisite domain knowledge, which an individual may not possess, but a collaborator may. This results in more errors being detected and corrected. Further, being in an interactive situation, dyads are more likely to engage in constructive processes such as explanation, and therefore more likely to detect errors when things don't compute. There is evidence that by scaffolding learners' interactions to encourage explanation, they were able to form more coherent representations of science concepts (Coleman, 1998).

If error-detection is indeed a mechanism underlying collaborative facilitation, then engaging in an error-detection task collaboratively should help mitigate the collaborative inhibition effect. In other words, the task of error-detection should lead to dyads performing at least as well as nominal dyads. Past studies have proposed error-detection as a mechanism, but not studied it as an aspect of the task structure. In the current experiment, we employed error-detection as the task in which participants would engage either collaboratively or individually.

We decided to test this hypothesis in a college classroom in the context of writing summaries of empirical articles. One reason for choosing this domain was that it provided an ideal open-ended task for students to work on in dyads or individually. Second, research in writing instruction has consistently shown how generating a coherent summary of read material is a challenging task for most students (e.g., Flower, 1979). In any college course with a substantial writing component, especially research reporting, students have the most difficulty summarizing related research succinctly and relating it to their own ideas. The errors that students make are due to imperfect understanding of what constitutes a good summary. It is not intuitive for students to understand the difference between a good summary and a bad one, without engaging in deliberate cognitive processing.

Most of the past work that has found a collaborative advantage has been with simple tasks such as list learning and tested with recall or recognition judgment tasks. We wanted to extend this further to a task involving higher order processing than simple recall.

Finally, testing this paradigm in a real classroom also gave us increased ecological validity. This paradigm has not yet been explored in a controlled experimental way in a real classroom. Investigations of collaborative learning

have been conducted either in a lab setting, where a degree of strict experimental control is possible or in educational settings where factors such as random assignment have been implemented due to various constraints of working in a classroom. Recent endeavors have taken findings from cognitive science and attempted to apply them in authentic learning situations (e.g., Nokes & VanLehn, 2008). We followed in this tradition, and explored this paradigm in a cognitive psychology lab classroom, without sacrificing experimental rigor.

We propose that by collaborating with a peer, students will be more likely to detect flaws in a given summary. Collaborating peers will bring different knowledge to bear on the issue, not all of which will be overlapping. As stated before, we hypothesized that working with a peer will be able to detect a greater number of errors than those working individually. Moreover, we expect that by collaboratively engaging in error-correction, collaborative dyads will outperform nominal dyads, or at least equal them on performance.

We also wanted to see whether the benefits of collaborative error-detection extend to a subsequent on the revision task. In the revision task, students were asked to revise the initial error-ridden summary with the same partner or individually. We hypothesized that because dyads will uncover a greater number of errors to begin with, they will be more likely than individuals to correct those errors, and will perform better than individuals.

Method

Participants

Fifty students from University of Pittsburgh (32 females and 18 males) participated in the study. These students were from three of the lab sections of the course Cognitive Psychology for Majors. Most of the students were upperclassmen (juniors or seniors).

Design

The design was between subjects and students were randomly assigned to either the individual condition or were randomly paired with a partner from the same section without considering gender or ability, and assigned to the collaborative condition. The two main dependent variables of interest were the performance on the error-detection task as measured by number of errors identified and performance on the revision task.

Materials and Procedure

The experiment was conducted over a three-week period, and comprised of homework assignments and in-class activities. The flowchart shown in Figure 1 describes the activities that students performed.

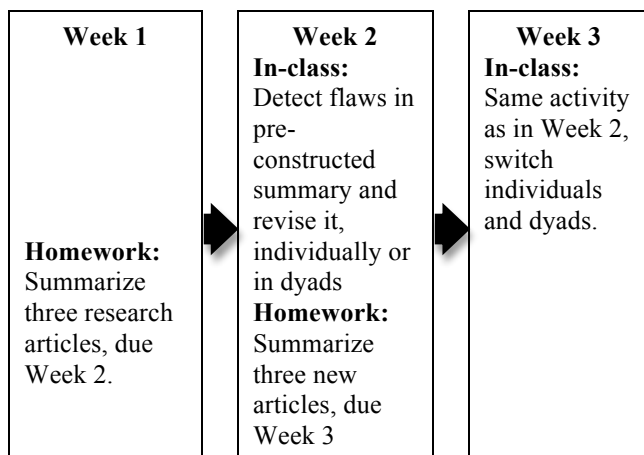


Figure 1: Flowchart of procedure

During week 1, students were asked to summarize three articles on a topic in cognitive psychology. They could choose out of six articles, but one of them was mandatory, because the in-class activity in the following week would be based on that article. The articles were abridged versions of published research articles and consisted of just the Abstract, Method, and Results section. Assigning the summarizing homework prior to the in-class activity ensured that participants were familiar with the task before they worked on it in class.

During the in-class activity in week 2, students were given a pre-constructed summary of the mandatory research article that they had read and summarized in their homework. This summary contained a number of errors. Students were given the same article they had summarized in their homework in order to cut down on time needed to read the article in order to summarize it. During the in-class activity, students were randomly assigned to either the individual or dyadic condition. Each person or dyad got the pre-constructed summary as well as the original abridged article. They were told that this summary was constructed by another student, and their job was to list the flaws in that summary and then rewrite the summary to revise it. At the end of the class, they were assigned a new homework activity in which they summarized three new articles. This homework was due on the class of week 3.

The experiment took place during a regular weekly lab as part of their normal instruction. Students were given 50 minutes to enlist the flaws in the summary and write a revised version. No other scaffolding was provided during the experiment.

A rubric was developed to score students' completed worksheets. First, students' list of flaws was examined to determine how many flaws they could correctly identify. This was compared with a list of all flaws in the document, which could be either structural level flaws or surface level flaws. Structural level flaws included flaws

such as "research question not stated" or "participant characteristics absent". Surface level flaws were stylistic flaws, for example, "summary was not indented" or "italicization in reporting of statistics was incorrect." There were a total of 11 structural level flaws and 6 surface level flaws in each summary. See appendix A for a list of flaws.

Next, a rubric was created to score the revised summaries that students had developed. There were 12 criteria that needed to be fulfilled in order to get full credit. See appendix A for a list of criteria.

Results

We will first describe the performance of dyads and individuals on the error detection task. We will then see whether there is a difference in performance when individuals are randomly paired with another individual to form nominal dyads. This will be followed by an analysis of the scores that dyads or individuals received on the revision task, and subsequently whether dyads and nominal dyads differed on the revision task. Finally, we will see whether the effects of the error-detection activity transferred to a new but related situation, by examining students' performance on the homework assignment immediately after the in-class activity.

As we had hypothesized, dyads performed better than individuals on the error detection task. That is, dyads could detect a higher number of structural-level flaws in the summaries compared to individuals. Dyads could detect 2/3 of the total number of flaws, whereas individuals could detect only 1/2 of them. See Figure 2 for means and standard errors. A 2 (collaboration: dyads versus individuals) x 2 (error: structural versus surface) mixed ANOVA showed no effect of collaboration, $F(1, 33) = 2.33$, ns. However, there was a main effect of error type with structural errors being better identified than surface level errors. In addition, there was a significant interaction of collaboration by error type, such that dyads were better at detecting structural level flaws than individuals whereas there was no difference between them in detecting surface level flaws, $F(1, 34) = 10.83$, $p < .05$.

Next, we looked at dyads versus nominal dyads. We used Kelley and Wright's (2010) procedure to form nominal dyads¹, and looked at the unique number of

Most studies that have compared nominal dyads and collaborators in the past have randomly paired individuals to form nominal dyads. This introduces an unnecessary source of errors, and it is advisable to use all possible pairs of nominal dyads to reduce this error. However, with a sample size of 20, one would need to look at 2×10^{24} pairs of nominal dyads, which is computationally almost intractable. Kelley and Wright (2010) have written a program that randomly selects 10,000 pairs of nominal dyads and then generates a list of nominal dyads with a mean and standard deviation closest to the true mean.

errors identified by each nominal dyads. For example, if one member of the nominal dyad identified errors 1, 2, 3, 4, and 5 and the other identified 4, 5, 6, and 7, their total score was 7. The means and standard errors are shown in Figure 2.

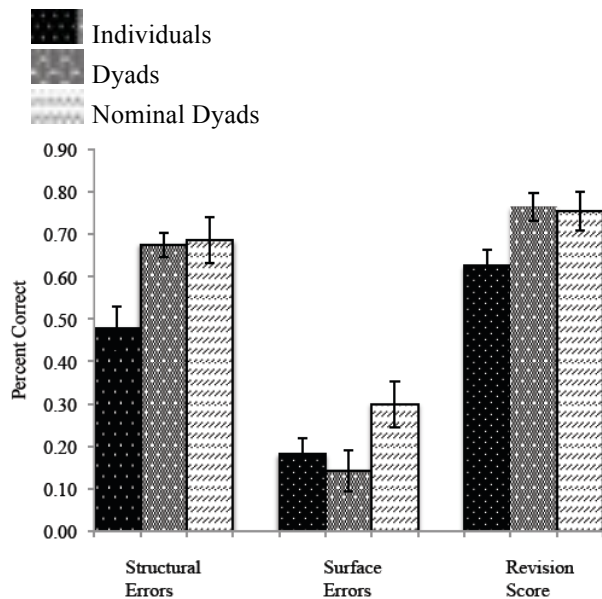


Figure 2: Means and standard errors for individuals, dyads, and nominal groups.

A mixed ANOVA with follow-ups using the LSD procedure ($\alpha = .05$) was performed to examine the effects of collaboration on the number of structural and surface level errors detected. There was a significant main effect of error type such that participants found significantly more structural errors compared to surface errors, $F(1,23) = 83.64, p = 0.00$. There was also a significant main effect of condition such that nominal dyads detected a significantly more number of errors overall, compared to collaborators, $F(1,23) = 17.70, p = 0.048$. There was an marginally significant interaction between condition (nominal vs collaborative) and error-type (structural vs surface), $F(1,20) = 3.45, p = 0.076$. Follow-up tests using Fischer's LSD ($LSD = .15$) showed that the difference between nominal dyads and collaborators was significant only for surface level errors. The two groups were not different on structural level errors. However, both groups found a significantly larger number of structural level errors compared to surface level errors.

Thus, although dyads did not outperform nominal dyads in detecting structural level flaws, they were equally good in terms of their individual performance within the dyad. We found evidence for collaborative inhibition on the surface level flaws. This might be due to social pressure to identify only those flaws that the students' thought would be considered most important, such as structural-level flaws and that perhaps surface-

level flaws were not considered important or so obvious to be easily fixed.

Next, we looked at the performance of dyads and individuals on the revision task. The revised summaries were scored on a rubric where the maximum possible score was 20 points. The means and standard deviations for the revision score are displayed in the last column of Figure 2. A one-way ANOVA revealed that dyads significantly outperformed individuals on revising the summaries, $F(1, 34) = 6.57, p < .05$. Thus, the benefit of error detection activity extended to the actual revision of summaries, and reinforcing the collaborative advantage. We then compared scores on the revision task for nominal dyads and dyads. Similar to the error-detection task, we awarded one point for every criterion that either or both of the two partners got correct in a nominal dyad. A one-way ANOVA revealed that the dyads and nominal dyads were not significantly different from one another, $F(1, 23) = .03, ns$.

To understand how the in-class error-detection activity impacted students' performance on subsequent writing assignments, we looked at their scores on homework assignments immediately following the in-class session. This is a transfer task, because we expected students to apply what they had learned during the in-class activity (error-detection) to generating their own summary of an article.

We expected students who found a greater number of errors to score better on the homework assignment, because they would be less likely to commit the same errors while summarizing an article. We found a marginal correlation on the subsequent homework such that the score on the homework assignment correlated with the number of errors that they detected during the in-class activity $r(44) = .28, p = .058$. There was however no difference by condition, that is the scores of the collaborative participants and individual participants did not differ significantly, after controlling for their performance on the earlier homework, $F(1, 43) = .421, ns$.

Discussion

In the present study, we investigated whether by promoting the mechanisms underlying collaboration, we can overcome the collaborative inhibition effect reported widely in the literature. Our results from this experiment are very encouraging, and provide evidence that by structuring collaborative learning activities according to the cognitive processes underlying it, we can get collaborative learners to perform at least as well as nominal dyads.

We found that engaging in an error-detection task with a partner led to better performance on detecting structural level errors than doing so individually. Even more important is the finding that when the dyads were compared with nominal dyads, they did not do worse, than the nominal dyads unlike many past studies (e.g

Andersson & Ronnberg, 1995). However, individuals found a greater number of surface level errors. One of the possible explanations for this is that dyads focused on the structural level features and ran out of time before getting to the surface level features. Individuals on the other hand, because they could find only a certain number of structural errors, moved on to the surface level errors, and were able to detect more of them. However, as noted before, the overall rate of detection of surface level errors was low, indicating that both individuals and dyads focused more on the structural features.

The other important finding from this study was that dyads performed significantly better than individuals when they revised the flawed summaries. When comparing revision scores of collaborators and nominal dyads, we found no difference between the two. Thus, we have evidence that benefit of error-detection extended to the revision task as well.

We also tested the effects of collaborative error-correction on a measure of transfer when we looked at whether the students' performance on the error-detection task affected their performance on a subsequent homework, which involved generating their own summaries. We found that the number of errors detected during the in-class activity was correlated with their score on the homework assignment. Although we did not find a significant difference between scores of individuals and collaborators, the correlation indicates that students who detected more errors were more likely to perform better on the summarizing task, regardless of condition. There are some caveats to our findings. The first is that since this experiment was conducted in a classroom setting, we could not control all variables as strictly as we would have liked to, in a laboratory setting. We therefore aim to replicate this in a more stringently controlled environment, and understand collaborative error-correction at a more fine-grained level.

Next, we need to replicate this finding in a different domain, and find out whether the effects of collaborative error-detection are robust enough to be found across various domains, such as conceptual physics or mathematics problem solving.

Several issues still need to be addressed in understanding why error-detection leads to better collaborative outcomes. It is clear that error-detection encourages some kind of constructive activity in collaborators that causes them to perform better than individuals. Process data such as verbal protocols can help us better understand what these constructive activities are.

For example, the study by Okada and Simon (1997) found that dyads were more likely than individuals to generate explanations. It would be helpful to analyze process data from collaborative error-detection and understand whether collaborators are more likely to generate explanations for the errors they detected, which

in turn leads to benefits in learning and transfer, and not remain confined to performance alone.

In recent years, scripting of collaborative interaction had been found to be beneficial especially in computer-mediated settings. Understanding how to encourage constructive processes like explanation through collaboration can help create better scripts for collaborative learning.

It is also important to better understand the social dynamics of collaborative learning. For example, what is the role of grounding in collaboration? In our present study, the participants had the required in the task. However, will we find the same effects if less skilled participants are given the same task? What amount of shared knowledge is necessary for successful collaborative learning? All these are open questions that future work needs to address.

In conclusion, we found a robust effect of collaborative error-correction such that collaborators showed better performance compared on a subsequent revision task, and performed as well as nominal dyads. This can have strong educational implications, ranging from applications to classrooms to computer-mediated learning environments.

Acknowledgements

We would like to thank the lab teaching fellows—Daniel Belenky and Jooyoung Jang for their assistance in collecting data for this experiment. Also, thanks to Melissa Patchan for her insightful suggestions during conceptualizing this study and Linnea Warren for help scoring data.

References

- Andersson, J., & Rönnerberg, J. (1995). Recall suffers from collaboration: Joint recall effects of friendship and task complexity. *Applied Cognitive Psychology, 9*, 199-211.
- Basden, B. H., Basden, D. R., Bryner, S., & Thomas, R. L., III (1997). A comparison of group and individual remembering: Does collaboration disrupt retrieval strategies? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 23*(5), 1176-1189.
- Coleman, E. B. (1998). Using explanatory knowledge during collaborative problem solving in science. *Journal of the Learning Sciences, 7*(3/4), 387-427.
- Collaros, P. A., & Anderson, L. R. (1969). Effect of perceived expertness upon creativity of members of brainstorming groups. *Journal of Applied Psychology, 53*, 159-163.
- 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.
- Flower, L. (1979). Writer-based prose: A cognitive basis for problems in writing. *College English, 19*-37.

- Hill, G. W. (1982). Group versus individual performance: Are n+ 1 heads better than one. *Psychological Bulletin*, 91(3), 517-539.
- Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, 65, 681-706.
- Kelley, M. R., & Wright, D. B. (2010). Obtaining representative nominal groups. *Behavior Research Methods*, 42(1), 36-41.
- Kerr, N. L., & Tindale, R. S. (2004). Group performance and decision-making. *Annual Review of Psychology*, 55(1), 623-655.
- Mullen, B. (1983). Operationalizing the effect of the group on the individual: A self-attention perspective. *Journal of Experimental Social Psychology*, 19, 295-322.
- Nokes, T. J., & VanLehn, K. (2008). Bridging principles and examples through analogy and explanation. In the *Proceedings of the 8th International Conference of the Learning Sciences*. Mahwah, NJ: Erlbaum.
- Rajaram, S., & Pereira-Pasarin, L. P. (2007). Collaboration can improve individual recognition memory: Evidence from immediate and delayed tests. *Psychonomic Bulletin and Review*, 14(1), 95.
- Shaw, M. E. (1932). A comparison of individuals and small groups in the rational solution of complex problems. *American Journal of Psychology*, 44, 491-504.
- Sniezek, J. A., & Henry, R. A. (1989). Accuracy and confidence in group judgment. *Organizational Behavior and Human Decision Processes*, 43, 1-28.
- Steiner, I. D. (1972). *Group processes and productivity*. New York: Academic Press.
- Weldon, M. S., & Bellinger, K. D. (1997). Collective memory: Collaborative and individual processes in remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 1160-1175.
- Weldon, M. S., Blair, C., & Huebsch, P. D. (2000). Group remembering: Does social loafing underlie collaborative inhibition? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 26(6), 1568-1577.
- 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.

Appendix A

- List of flaws in summary:
- Structural level:
- H1. Directly copied from text (plagiarized)
 - H2. Details of procedure not clear
 - H3. Gives actual statistics
 - H4. Does not explain results in plain language/ does not define terms
 - H5. References table that is absent in summary
 - H6. Hypothesis not stated
 - H7: Subject characteristics not present
 - H8: IV & DV not clear
 - H9: Experiment design (Between or within not clear)
 - H10: Limitations/ confounds not mentioned
 - H11: Does not interpret results/mention implications for further study
- Surface level:
- L1. Statistics not formatted correctly
 - L2. Reference absent
 - L3. Mentions five conditions instead of six
 - L4. Not indented
 - L5. Does not separate paragraphs
 - L6. APA formatting issues

Appendix B

- Criteria for scoring revised summaries:
1. What is the research question?
 2. What is the hypothesis being tested?
 3. Were participant characteristics (number, age, gender, education etc.) correctly stated
 4. Was the experimental task clear?
 5. Was the experimental design (between or within subjects) correctly stated
 6. Are the dependent variables correctly stated?
 7. Are the independent variables correctly stated?
 8. What were the important points of procedure
 9. What were the major finding/s?
 10. Are confounds/limitations pointed out?
 11. Are findings interpreted in own language and a conclusion stated?
 12. Mechanics (spelling, grammar) and Conciseness/ No unnecessary detail