Co-witness auditory memory conformity following discussion: A misinformation paradigm

Kazuo Mori¹ & Takeshi Kishikawa²
¹: Tokyo University of Agriculture and Technology, Tokyo, Japan
²: Shinshu University, Nagano, Japan

Summary: Twenty-four Japanese undergraduate pairs participated as witnesses to a simulated criminal event. Although the witness pairs watched the same video together, through wireless headphones they experienced two different auditory versions with four differing items without being aware of the discrepancies. After the presentation, the witnesses were led to discuss six items, including two critical ones they had heard differently and another four they had heard in common. Witness memory performance was assessed individually with multiple-choice questionnaires in three sessions: before the discussion, after the discussion, and one week later. The results showed that participants tended to conform to their co-witness more often on the discussed items than on the not-discussed items. Source monitoring analyses on the four critical items revealed that even those participants who conformed were mostly cognizant of the source of their information just after the discussion, but they were prone to source monitoring errors a week later.

Keywords: co-witness effects, auditory memory, conformity, presentation tricks

Skagerberg and Wright (2008) found that in a local trial in Brighton, England, 58% of court witnesses had discussed the crime in question with at least one of their co-witnesses. It has been learned from laboratory studies (e.g., Kanematsu, Mori, & Mori, 1996/2003; Gabbert, Memon, & Allan, 2003) that a crucial source of witness memory distortion can come from co-witnesses. Kanematsu, et al. (1996/2003), utilizing a presentation “trick” (Mori, 2003), simultaneously presented two different versions of a criminal event to co-witness pairs, and then had the witnesses discuss what they had seen. This procedure artificially reproduced situations equivalent to actual ones in which one of the co-witnesses had misperceived the event. Kanematsu, et al. found that the participants often reported what their partners had observed in place of their own memory of the events.

Gabbert, et al. (2003) found a similar co-witness effect. They presented to participant pairs an event video-recorded from two different angles so that one participant could see some details of the event that could not be seen by the other. The Gabbert procedure was a simulation

Acknowledgments. This research was done while the authors were at Shinshu University. It was supported by Grants-in-Aid from the Japanese Ministry of Education, Culture, Sports, Science, and Technology (Grant No. 16653054 and 25280050) to the first author. We are indebted to Rebecca Ann Marck for her superb work in editing the English manuscript. We would also like to express our thanks to Dan Wright and Lauren Hewitt (French) for their helpful comments on the earlier draft.
of the actual situation in which co-witnesses observe an event from different viewpoints. After discussion, the participants “recalled” details that they had not seen themselves, but had acquired from their partners who had actually observed the event. As the experimenters hypothesized, witnesses supplemented their own memories of an event with information gained from a co-witness. Since the original Gabbert, et al. study, co-witness effects have been investigated rigorously by means of the Mori procedure (Hirokawa, Matsuno, Mori, & Ukita, 2006; Garry, French, Kinzett, & Mori, 2008; French, Garry, & Mori, 2008; Mori & Mori, 2008; Mori & Kitabayashi, 2009; French, Garry, & Mori, 2011; Mori & Takahashi, 2012; Hewitt, Kane, & Garry, 2013; Jack, Zydervelt, & Zajac, 2013) and by the Gabbert procedure (Gabbert, Memon, & Wright, 2006; Candel, Memon, & Al-Harazi, 2007; Hope, Ost, Gabbert, Healey, & Lenton, 2008; Paterson, Kemp, & McIntyre, 2012). However, these studies were limited to visual memory of an event. Co-witnesses may observe an event differently, and, after discussing what they witnessed together, their memory may be modified by information shared among them. Similarly, co-witnesses may hear things differently when they witness an event.

There is some research concerning the auditory memory of witnesses, though not as much when compared with eyewitness memory. Yarmey (1992) was a pioneering study in the auditory memory of witnesses. Yarmey investigated the effects of discussion and collaborative remembering by comparing three different conditions: i) collaborative-dyad, in which participant dyads, after having overheard a telephone conversation between two mock criminals, were led to discuss for ten minutes then tested collaboratively for their memory, ii) dyadic-discussion-individual-recall, in which participant dyads were led to discuss for ten minutes, but tested individually, and iii) no-discussion-individual-recall, in which participants were engaged in a 10-minute conversation with the experimenter, then tested individually. Unfortunately, the results were not clear-cut; they showed the discussion would yield statistically more commission errors or fabrications when recalled individually, but would not when recalled collaboratively.

Kerstholt, Jansen, Van Amelsvoort, and Broeders (2006) investigated the effects of accent on speaker identification in line-ups. Three hundred sixty participants heard the target's voice and were asked to identify the person by means of an audio line-up consisting of six voices. Kerstholt, et al. found that speakers with a standard accent were more often correctly identified than speakers with non-standard accents. Philippon, Cherryman, Bull, and Vrij (2007) examined the effects of language familiarity on voice identification performance. Sixty native English speakers were exposed to an auditory event in either a familiar (i.e., English) or unfamiliar language (i.e., French), and were then asked to identify the original speaker's voice. As expected, participants demonstrated significantly better performance for the familiar language condition. Barnecutt, Pfeffer, and Creswell (1999) examined the accuracy of judgments of traffic speed by comparing audio, visual, and audio-visual perception in two experiments with 97 adult participants, and found that traffic speed was over-estimated in the auditory mode and under-estimated in the visual and audio-visual modes. Campos and Alonso-Quecuty (2006) examined the effects of mode of presentation (audio-visual/ auditory-only) on witnesses' free recall for utterances in an experimental criminal conversation. Their results showed that witnesses in the audio-visual condition recalled more correct information than those in the auditory-only condition. Recently, Öhman, Eriksson, and Granhag (2013) compared three types of interviewing techniques to enhance the auditory memory of both child and adult witnesses. The results showed that the interviewing protocol known as the Cognitive Interview was shown
to be beneficial for adults but not effective for enhancing children’s content recall in auditory modality, in contrast to their previous finding in which they found it was effective for child eyewitness testimony (Larsson, Granhag, & Spjut, 2003).

Except for the Yarmey (1992) study, none of these studies focused on the effect of discussion among co-witnesses on their auditory information. However, Yarmey (1992) presented the same telephone dialogue to participants and only had them discuss freely without controlling the discussed items. That might be a crucial reason why the results of Yarmey (1992) were rather inconsistent and not decisive about the effect of discussion. In contrast, the co-eyewitness effect studies using a misinformation paradigm (e.g., Kanematsu, et al., 1996/2003; Gabbert, et al., 2003) deliberately introduced discrepancies into the original source material presented to co-witnesses so that the experimenters were able to control which items were to be discussed by the participants. No co-earwitness studies have been conducted using a misinformation paradigm.

In an infrequently cited article, Asakawa and Imai (2008) investigated the co-witness effect using a modified Mori procedure that focused on discrepancies in auditory information. In their study, one member of each of 20 pairs of undergraduates listened through headphones to one story while the other listened to a different story without either noticing the duality of the stories. After listening to the stories, they were led to discuss what they had heard, and half of the pairs were asked to make a collaborative report, and the other half to make individual reports. A week later, they were asked to recall the stories individually. The results showed that the participants often reported details of the story that they had not actually heard themselves. This tendency was stronger when participants made a collaborative report than when they only discussed and reported individually.

Asakawa and Imai (2008) reproduced the results of Kanematsu, et al. (1996/2003) using a different modality, but unfortunately, the study also included one drawback of the original experimental design. In both of their studies, all the participants were led to discuss all the differing items in the co-witnessed event. Consequently, the effects of discussion were not properly examined in this design. Garry, et al. (2008) improved on the Kanematsu study by introducing a between-items (= within-subjects) experimental design. They presented two versions of a criminal event with eight differing details to participant pairs, and asked them to discuss only half of the differing points. Garry, et al. found that the discussed items were more likely to be reported inaccurately than the not-discussed ones. The present study aimed to replicate the Garry, et al. study in the auditory modality. It was hypothesized that participants would tend to conform to their co-witness more often on the discussed items than on the not-discussed items in the same way as former findings in co-eyewitness studies. It was also hypothesized that the discussion effect would be larger a week later than just after the discussion.

**METHOD**

**Participants**

Twenty-four Japanese undergraduate same-sex pairs (19-to-23-year-olds; 12 male and 12 female) participated as witnesses. All the participant pairs were volunteers from the same college of a Japanese university. It was considered they were a quasi-random sample representing a middle class population of these age groups in rural Japan. They were randomly assigned to the counter-balancing conditions.
**Design**

The study was a two-way 2 (discussed vs. not discussed items) x 3 (before, after-discussion, 1-week delay) repeated measures design. The critical items were counterbalanced so each served equally as a control and discussed item. As in the preceding studies (Kanematsu, et al., 1996/2003 and Asakawa & Imai, 2008), the delayed recall was assessed one week later.

**Presentation Materials**

*Video clip of a simulated criminal event.*-- One version of the Kanematsu, et al. movie was converted into a QuickTime Movie clip and used in the present experiment. It depicted a simulated criminal event lasting about one minute. In the scenario:

A car pulls up in front of a female pedestrian. The driver gets out of the car to ask her for directions. While the pedestrian is explaining the directions to the driver, another passenger sneaks out of the car and steals something from her bag. The pedestrian starts walking away without noticing the theft.

*Two auditory versions of the video clip.*-- Two basically identical auditory versions were prepared to suit the video clip. Then four critical items were inserted into the two versions (cf. Appendix at the end). Of these four, two were mono-polar differences (MD), such as a car horn honking or not, and the other two were bipolar differences (BD), such as two different destinations. The four critical items, in order of their appearance in the event, were these:

1. Car horn: honking vs. no honking (MD)
2. Destination the driver asked of the pedestrian: Fuji Highland vs. Cape Tachibana (BD)
3. Pedestrian’s dialect: Tokyo vs. Osaka dialect (BD)
4. Monologue of the pedestrian: Monologue vs. no monologue (MD)

![Diagram of experimental setting](image)
Co-witness auditory memory conformity

Apparatuses

The experimental apparatuses were prepared according to the following criteria: 1) presenting two different versions of auditory information, one to each of a pair of witnesses, and 2) having the participants think they had witnessed the same event together. In order to achieve the first condition, two channels of wireless headphone sets were used. To fulfill the second condition, the participants observed the same video clip on the same screen simultaneously. The experimental setting is depicted in Figure 1.

Video projection.-- An LCD video projector (EPSON ELP-730) attached to an Apple iBook was placed approximately 60 cm behind a rear screen. The video clip was played with QuickTime Player on the iBook.

Audio-presentation devices.-- Two sets of wireless headphones and transmitters (Sony MDR-IF540RK) were used. The transmitters were placed near the two ends of the rear screen, each being invisible to the participants. In this way, the participant sitting on the right side heard only the sound coming from the transmitter on the right and the other participant sitting on the left heard only the sound from the left transmitter.

Procedure

General instructions.-- Participants entered the psychology laboratory in pairs and sat in chairs placed approximately 1 m apart on opposite sides of a table facing the screen. The participants chose a chair at will. Two sets of wireless headphones were placed on the table. As a cover, the participants were instructed to wear the headsets to eliminate possible noise coming from outside of the laboratory. Then, the experimenter gave general instructions through the headsets, and asked one of the participants to summarize the instruction. This summarization procedure was intentionally conducted to make the participants believe they would each hear the same audio input through the headphones.

Presentation.-- The experimenter turned the room light off. Then, the movie was presented on the rear screen along with the two versions of the auditory information to the participants. After the movie presentation, the light was turned on.

Before-discussion tests.-- Immediately after watching the movie, the participants’ recollection of the event was assessed individually by means of multiple-choice tests. There were 12 questions with four choices each, including four questions on the critical items and eight on the common items. Of the eight common items, four were to be discussed during the ensuing cued discussion session along with two critical items. The remaining four common items only served as fillers. Each question was accompanied by a seven-point confidence rating scale. No time limits were set, but it took about five to ten minutes to complete the tests. The same type of test sheet was used for all three memory-confidence assessments.

Cued discussion.-- After completing the memory test, the participant pairs were cued to discuss with six forced-choice questions presented on the same screen on which they had observed the movie clip. Among them, four (Q1, 2, 4, and 5) were on items they each had witnessed in the same way (i.e., the common items), while two (Q3 and 6) were on the critical items, those that had differed for each participant. The two discussed critical items and the other two not-discussed items were counter-balanced among the witness pairs. The same four questions on the common items were used for all the participant pairs. It took about a minute to answer each question, about eight minutes in total. (See Appendix for the questions used.)
After-discussion tests.-- After the discussion, the participants were asked to report individually on what they had observed and to rate their confidence level on a new test sheet. They were instructed to answer each question individually. In order to assess their source monitoring knowledge, they were instructed to use a red pen when their answer was based on their partner’s answer during the discussion. It took about ten minutes to complete it.

Week-later tests.-- A week later, the participants were again asked to take the same memory test individually. The same instructions as for the After-discussion test were given, including the source monitoring assessment. Following these Week-later tests, another questionnaire was administered inquiring as to whether or not they had noticed any anomaly in the presentation phase. Then, after all the experimental procedures were completed, the participants were debriefed about the experimental purposes and tricks.

RESULTS

Preliminary Analyses

The first issue examined was whether the experimental manipulation had worked as intended and whether there were any fundamental differences between the two groups of participants assigned to each of the two auditory versions. The participant pairs reached agreement on all questions during the discussion. The post-experimental questionnaire revealed that no participant noticed the duality of the auditory sources. Therefore, it was concluded that the experimental manipulation was successful.

Based on the Hirokawa et al. (2006) finding that the masculinity of co-witnesses might play a role in their tendency not to conform to their partners, the effect of gender on memory and co-discussion was assessed before collapsing the data. There were no significant differences between male and female participants in terms of memory accuracy on the common items (3.67 vs. 3.83, 3.71 vs. 3.83, and 3.71 vs. 3.88, average scores for males vs. females at the three assessment times; \( F_{(1,46)} = 1.94, \text{ns} \)). Therefore, in the following analyses, the data from male and female participants were combined.

Memory for Discussed vs. Not-Discussed Items

The number of items correctly recalled on the discussed and not-discussed critical items were counted separately for the 48 participants (= 24 pairs) and averaged. As shown in Table 1, the average number of items correctly recalled on the two discussed items dropped considerably after the discussion. A two-way ANOVA revealed that there was a significant interaction \( (F_{(2,94)} = 27.43, \ p < .01, \ \eta^2 = .10) \): there were significant differences among the average memory scores on the discussed items over the three test sessions \( (F_{(2,94)} = 29.77, \ p < .01, \ \eta^2 = .12) \), while there were no differences for the not-discussed items \( (F_{(2,94)} = 0.16, \ \text{ns}) \). A Tukey post hoc test revealed that there were significant downward gaps between the scores before discussion and those from the After-discussion and Week-later test sessions \( (MSe = 0.21, \ p < .05) \), while there was no significant difference between two post-discussion conditions.
The seven-point confidence ratings were converted into scores from 0 (least confident) to 6 (most confident) and averaged separately for discussed and not-discussed items. There were moderate but significant correlations between the confidence ratings and the accuracy (r = .24, for the discussed items, and r = .28, for the not-discussed items, p < .05). As shown in the bottom half of Table 1, the average confidence scores varied within a small range, but showed a pattern similar to the memory scores. A two-way ANOVA revealed that there was a significant interaction (F(2,94) = 4.33, p < .05, η² = .03), with no differences among the average confidence scores on the not-discussed items (F(2,94) = 0.47, ns), but significant differences among those on the discussed items (F(2,94) = 4.63, p < .05, η² = .03). A Tukey post hoc test revealed a significant decline between the confidence ratings before the discussion and those a week later (MSe = 0.67, p < .05). The confidence rating differences between adjacent rating times (before and after the discussion and after-discussion and a week later) did not reach a significant level.

Source Monitoring Analyses

In the After-discussion and Week-later tests, participants were instructed to use a red pen when their own answer during the discussion was based on their partner’s answer so that we could assess their source monitoring knowledge. Table 2 shows how many “wrong” answers participants made in total on the discussed items while being consciously aware of the actual source of their recollections in the two test periods. In the After-discussion tests, 33 of 39 (= .85) incorrect answers were consciously made, whereas in the Week-later tests, the conscious incorrect answers dropped to only 12 out of 35 (= .34). This can be interpreted to mean that most of the participants who chose “wrong” answers were consciously conforming to their partners in the After-discussion tests. Their source monitoring was highly accurate just after the discussion. However, there were more source monitoring errors a week later, resulting in more erroneous conforming answers (23/35) than conscious answers (12/35). Meanwhile, their source
monitoring was more accurate for the correct answers just after the discussion (54/57 or .95) as well as a week later (48/61 or .79). Statistical analyses with chi-square testing showed a highly significant association of the correctness of memory and the source monitoring in the After-discussion tests, $X^2(1) = 62.212$, $p < .001$, Cohen’s $W = .805$, but not a significant association in the Week-later tests, $X^2(1) = 1.944$, ns, $W = .142$.

**DISCUSSION**

*Co-witness Effects on Memory Distortion in Auditory Modality*

The fundamental finding of the present experiment was that memory conformity effects exist even when the original information was auditory. Ear-witnessed details were distorted when discussed with co-witnesses. The participants reported the correct answers more often for not-discussed critical items than for discussed ones. A similar confidence-rating pattern as that reported in Garry, et al. (2008) was likewise found. As reported above, Asakawa and Imai (2008) also reproduced Kanematsu, et al. (1996/2003), substituting a presentation trick modality, and found basically the same results. Accordingly, one may assume that the other findings of co-witness effects already obtained from eyewitness experiments should be applicable to co-earwitnesses as well.

It should be noted that the present study is not a true earwitness study as compared to Yarmey’s study because both visual and auditory modalities were involved in the present study. Wright and Wareham (2005) investigated whether auditory and visual information interact in the same way as the McGurk effect (McGurk & MacDonald, 1976) even in the perception of a contextually rich scene such as witnessing a criminal event. They found that many participants reported somewhat merged patterns of visual and auditory sources. Therefore, it should be needed to conduct further research on co-witness effects not only focusing in the auditory modality alone but also in various interactions between the visual and auditory modalities.

*Confidence-Accuracy Correlations*

Moderate but significant correlations were found between the confidence ratings and the accuracy in the present study as well as in Garry et al., (2008). It may seem natural to expect a correlation between confidence and accuracy, but it has been known in the eyewitness literature that confidence does not necessarily assure accuracy (See Krug, 2007 for review). Kebbell
(2009) reported that positive confidence-accuracy correlations in eyewitness memory were found when the levels of recalling difficulties of items were varied to include easy questions. This would be one possible explanation of the moderate confidence-accuracy correlation in the present study because the accuracy was very high at first (i.e. before discussion) and declined considerably after the discussion and a week later. It means there was a wide range of difficulty levels in the time course of the present experiment. Whether the confidence-accuracy relations in earwitness memory would be different from those of eyewitness memory is an interesting research topic. Further research in earwitness memory is needed to answer this question.

Recommendations for Forensic Practice

The present finding would support the recommendation for those working in forensics not to let co-witnesses discuss what they have witnessed together, since the present study showed that memory conformity effects exist in auditory information as well as visual. It should be noted that although discussion among co-witnesses may well be one reason for distortion, most of the witnesses recalled the source of their memory just after the discussion. Unfortunately, correct source monitoring deteriorated considerably after one week. As cited before, Skagerberg and Wright (2008) revealed that 58% of court witnesses had discussed the crime in question with at least one of their co-witnesses. It would be difficult to prevent co-witnesses from discussing the event among themselves, but as the present study showed, they may recall the source of their knowledge properly just after the discussion but not later. Therefore, it is recommended that witnesses be interviewed as early as possible. Even if they have already discussed the event with their co-witnesses, they may be able to identify their memory source accurately just after the discussion and less correctly as time goes on.

Limitations of the Present Research and Perspectives for Future Research

The present experiment was designed to investigate the effects of discussion using a within-subjects design. However, the crucial items were not fully counterbalanced owing to practical reasons. There should have been four auditory versions, but there were only two. The results may have been further compromised by the presence of monologue and Tokyo-Osaka dialects.

Since the videotaped event in the present experiment was just one minute in length, only four differing points were inserted into it. These four items were further divided into either discussed or not-discussed items. Thus, the dependent variable of the present study came from memory performance on only two items. It would be desirable to present more critical items as in Garry et al. (2008), in which there were eight differing points between the two versions presented to the co-witnesses. However, there would be a trade-off. The more differing items inserted into the events presented, the longer the presentation phase becomes, and so do the other experimental phases. It took about 45 minutes for each session in the present experiment, including general instructions, video presentation, Before-discussion tests, cued discussions, and After-discussion tests, whereas it took more than an hour for each session in Garry et al. (2008).

The experimental setting used in the present study for presenting two different versions of auditory information to co-witnesses worked perfectly. Therefore, it would be suitable for examining the co-earwitness effects for a variety of factors such as those that have already been investigated in co-eyewitness research using the Mori procedure and more. For example, the effects of the co-witness group size (Mori & Mori, 2008), intimacy among co-witnesses (French,
et al., 2008), mother-child dyads (Mori & Kitabayashi, 2009), credibility among co-witnesses (French, et al., 2011), etc. Recently, Jack, et al. (2013), utilizing the Mori procedure, re-examined the importance of the possible influence of co-witnesses in their testimony and concluded that the co-witness effect was not particularly influential as compared to other social sources of post-event misinformation. It would be interesting to examine whether the conclusion of Jack et al. could be extended to the co-witness effect in auditory information.

REFERENCES


CO-WITNESS AUDITORY MEMORY CONFORMITY


APPENDIX
The Questions Used in the Cued Discussion

Q1: What color was the car? a) Purple, b) Black, c) White, d) Red
Q2: What did the pedestrian do with her bag? a) Put it on the ground in front of her, b) Held it to her chest, c) Hung it on her shoulder, d) Put it on the ground behind her

Q3-1: Where do you guess the pedestrian was from? a) Tohoku area, b) Tokyo area, c) Osaka area, d) Kyushu area

Q3-2: Where did the driver want to go? a) Cape Tachibana, b) Mt. Matsuo, c) Okura Beach, d) Fuji Highland

Q4: What was stolen? a) Wallet, b) Soft toy, c) Mobile phone, d) Documents

Q5: Did she notice the thievery? a) Yes, she did, b) Yes, and she ran after the car, c) Yes, and she was upset, d) No, she didn’t notice it

Q6-1: What did the pedestrian say in the last scene? a) “What strange people!” b) “Did they reach their destination?” c) “Oh, I must go to work.” d) She did not say anything.

Q6-2: How did the car approach her? a) With blinkers and horn, b) With horn but no blinkers, c) With blinkers but no horn, d) Neither blinkers nor horn

---
1) The underlined choices were the correct answers.
2) The question numbers of the critical items are italicized.
3) The critical items had two correct answers depending on the version the participant had heard.