Running head: EYEWITNESS MEMORY FOLLOWING DISCUSSION

Eyewitness Memory Distortion Following Co-Witness Discussion: A Replication of Garry, French, Kinzett, and Mori (2008) in Ten Countries

Hiroshi Ito[a], Krystian Barzykowski[b], Magdalena Grzesik[b], Sami Gülgöz[c], Ceren Gürdere[c], Steve M. J. Janssen[d], Jessie Khor[d], Harriet Rowthorn[e], Kimberley A. Wade[e], Karlos Luna[f], Pedro B. Albuquerque[f], Devvarta Kumar[g], Arman Deep Singh[g], William Weber Cecconello[h], Sara Cadavid[i], Nicole Laird[j], Mario J. Baldassari[j], D. Stephen Lindsay[j], & Kazuo Mori[k*]

[a]Department of Psychology, Faculty of Letters, Aichi University, Toyohashi, Japan; [b]Applied Memory Research Laboratory, Institute of Psychology, Jagiellonian University, Kraków, Poland; [c]Department of Psychology, Koç University, Istanbul, Turkey; [d]School of Psychology, University of Nottingham – Malaysia Campus, Semenyih, Malaysia; [e]Department of Psychology, University of Warwick, Coventry, United Kingdom; [f]School of Psychology, University of Minho, Braga, Portugal; [g]National Institute of Mental Health and Neurosciences, Bangalore, India; [h]School of Health Sciences, Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, Brazil; [i]School of Medicine and Health Sciences, Universidad del Rosario, Bogotá, Colombia; [j]Department of Psychology, University of Victoria, Victoria, Canada; [k]Faculty of School Education, Matsumoto University, Matsumoto, Japan.

Abstract

We examined the replicability of the co-witness suggestibility effect originally reported by Garry et al. (2008) by testing participants from 10 countries (Brazil, Canada, Colombia, India, Japan, Malaysia, Poland, Portugal, Turkey, and the United Kingdom, total $N = 486$). Pairs of participants sat beside each other, viewing different versions of the same movie while believing that they viewed the same version. Later, participant pairs answered questions collaboratively, which guided them to discuss conflicting details. Finally, participants took a recognition test individually. Each of the 10 samples replicated the Garry et al. finding: Participants often reported on the final test a non-witnessed answer that their co-witness had stated during the collaboration phase. Such co-witness suggestibility errors were especially likely when the witness had not disputed the co-witness's report during the collaboration phase. The results demonstrate the replicability and generalizability of the co-witness suggestibility effect. (143 words)

Keywords: co-witness suggestibility effect, memory conformity, eyewitness memory, post-event conversation, multi-lab replication project. (5 keywords)

General Audience Summary

When police investigators, journalists, lawyers, and judges aim to determine the facts of a crime, eyewitness testimony can be crucial. Often in criminal cases, multiple witnesses observe the crime. Co-witnesses might remember the details of the event differently, due to differing viewpoints, differences in arousal or attention, or mistakes due to the fallibility of autobiographical memory. Co-witnesses often talk amongst themselves before being interviewed by police. These discussions raise the possibility that some witnesses' subsequent statements to the police, or in court, may be distorted by misinformation they received from other witnesses.

To explore this possibility, applied memory researchers have exposed co-witnesses to subtly different versions of an event and then measured the extent to which those witnesses (who were encouraged to talk to one another about the crime) distorted each other's later reports of that event. Some of these studies used polarized video projectors to present the event: Co-witnesses sat next to each other viewing the same screen while (unbeknownst to them) seeing slightly different versions of a movie (Garry et al., 2008; Kanematsu et al., 1996/2003). After a short delay, co-witness pairs tried to answer questions about the event together, including questions about details that differed between the two movie versions (e.g., one member of the pair might have seen a picture of the *Eiffel Tower* on the wall, whereas the other might have seen a picture of the *Leaning Tower of Pisa*). This collaborative recall created opportunities for co-witnesses to report to their partners their memories of details that differed from what their partners had seen. Finally, participants individually completed a

memory test about what they had observed.

The studies by Kanematsu et al. (1996/2003) and Garry et al. (2008) —conducted in Japan and New Zealand, respectively— found that people sometimes reported seeing details they had only heard about from their co-witness. To assess the replicability of this co-witness suggestibility effect, we collected data using the procedure and materials described by Garry et al. with similar equipment and comparable translated instructions in 10 countries: Brazil, Canada, Colombia, India, Japan, Malaysia, Poland, Portugal, Turkey, and the United Kingdom.

Across the 10 samples, we replicated the effect originally observed by Garry et al. (2008). These results indicate that this co-witness suggestibility effect is robust and common to many cultures. Criminal justice professionals (e.g., investigators, judges, lawyers) and jurors should be aware of the possibility of memory conformity when co-witnesses may have discussed the event in question. We also found that participants conformed to their partner mostly when they had not contradicted their partner's report during the discussion. Further research is needed to reveal the conditions under which discussion among co-witnesses is more versus less likely to lead to false reports and to explore for cultural differences in co-witness dynamics.

(460 words)

Eyewitness Memory Distortion Following Co-Witness Discussion: A Replication of Garry, French, Kinzett, and Mori (2008) in Ten Countries

Crimes are frequently witnessed by two or more people, and co-witnesses to crimes often discuss the event shortly afterward (Paterson & Kemp, 2006; Skagerberg & Wright, 2008). Many studies have shown that such discussion can lead witnesses to integrate elements of a co-witness's report into their own reports of the shared experience (e.g., Gabbert, Memon, & Allan, 2003; Gabbert, Memon, Allan, & Wright, 2004; Gabbert, Memon, & Wright, 2007; Garry, French, Kinzett, & Mori, 2008; Hewitt, Kane, & Garry, 2013; Hope, Ost, Gabbert, Healey, & Lenton, 2008; Kanematsu, Mori, & Mori, 1996/2003; Paterson & Kemp, 2006; Paterson, Kemp, & McIntyre, 2012; Tainaka, Miyoshi, & Mori, 2014; see Condon, Ritchie, & Igou, 2015, for a comprehensive review). This phenomenon, the *co-witness suggestibility effect*, indicates that people's memories can easily be distorted following discussion with a co-witness.

Memory researchers have employed three approaches to examine the co-witness suggestibility effect. In one approach, participants watch a video of a crime with a confederate who subsequently provides the misinformation (Bodner, Musch, & Azad, 2009; Bright-Paul, Jarrold, Wright, & Guillaume, 2012; Jack, Zydervelt, & Zajac, 2014; Paterson & Kemp, 2006; Paterson et al., 2012). In a second approach, participant pairs watch a video on separate screens and (unbeknownst to them) view different versions of the video (Gabbert et al., 2003; Gabbert, Memon, & Wright, 2006; see Wright, Memon, Skagerberg, & Gabbert,

2009, for a review). A third approach, and the focus of this study, uses the *Manipulation of Overlapping Rivalrous Images* technique (hereafter the *MORI technique*; Mori, 2007).

In the MORI technique, two different versions of a video are displayed on the same translucent screen using two rear-projected video projectors that display polarized images. One projector transmits light waves on the vertical plane, whereas the other transmits waves on the horizontal plane. Participants wear polarizing glasses allowing them to see only one of the two versions of the movie. This technique enables researchers to present different versions of the same video to participant pairs without them suspecting the duality. Although they appear to be watching the same video, a set of critical details is manipulated such that each participant sees a different version of each critical detail. In other co-witness techniques, participants may become suspicious about the manipulation when they find differences in their accounts. With the MORI technique, the attenuation of this possibility may enable laboratory approximation of real-life communication of misperceptions between witnesses.

Studies using the MORI technique have demonstrated the co-witness suggestibility effect under various situations and with different ages, genders, and interpersonal relations among the co-witness pairs (French, Garry, & Mori, 2008, 2011; Garry et al., 2008; Hirokawa, Matsuno, Mori, & Ukita, 2006; Kanematsu et al., 1996/2003; Mori & Kitabayashi, 2009; Mori & Mori, 2008; Mori & Takahashi, 2012; Tainaka et al., 2014). Research shows that misinformation gleaned from co-witnesses is particularly powerful when individuals do not have a clear recollection on which to rely, or when the participant views the co-witness as

more confident or as having a more reliable memory (Wright et al., 2009). Co-witness memory conformity is also influenced by participant characteristics. For example, socially avoidant participants tend to conform less whereas socially anxious participants tend to conform more (Wright, Busnello, Buratto, & Stein, 2012; Wright, London, & Waechter, 2010). Partner characteristics may also play a role, as participants tend to conform more towards familiar than unfamiliar partners (French et al., 2008; Hope et al., 2008) and more towards younger than older adults (Davis & Meade, 2013; Meade, McNabb, Lindeman, & Smith, 2017).

The literature may seem to have established the robustness of the co-witness suggestibility effect. However, awareness of threats to replicability has greatly increased among psychologists in recent years (e.g., Simmons, Nelson, & Simonsohn, 2011). The reality of the problem was brought home by Nosek and 269 contributing authors, who reported direct replications of 100 experiments published in three major psychology journals in 2008 (Open Science Collaboration, 2015): Fewer than half of the replication attempts yielded a statistically significant effect. Such findings have led to calls for statistically powerful, preregistered replications to assess the robustness of effects (e.g., Lindsay, 2015; Simons, Holcombe, & Spellman, 2014). “Preregistration” entails specifying in advance the participants, materials, procedures, measures, exclusions, and analyses of a study (Lindsay, Simons, & Lilienfeld, 2016).

To address concerns regarding the replicability of the co-witness suggestibility effect

as found by Garry et al. (2008), we repeated the experiment in 10 countries: Brazil, Canada, Colombia, India, Japan, Malaysia, Poland, Portugal, Turkey, and the United Kingdom. These countries were selected on the basis of interpersonal contacts among researchers interested in co-witness suggestibility. To the best of our knowledge, these are the first tests of the effect conducted in South America, continental Europe, and Asia outside of Japan. We chose the Garry et al. study to replicate with multi-lab samples because it used the MORI technique to create the discrepancies among what co-witnesses observed. As stated above, experiments using the MORI technique have advantages over other procedures, because the discrepancies between the two version of the movies are manipulated without using a confederate (i.e., Paterson & Kemp, 2006) and co-witnesses observe the movies together rather than separately (i.e., Gabbert et al., 2003).

Method

Participants

The study tested 10 samples of university undergraduates collected in two batches, with 486 participants in total (see Table 1). The first batch of samples was collected in Japan, Malaysia, Poland, Turkey, and the United Kingdom. The Japanese team was the first to replicate the experiment, with the same sample size as the original study; $N = 40$. Then, the other four groups followed with a larger sample size of around 60. We had determined the sample sizes at the time of the ethical approval applications before we conducted the experiments. The Japanese sample consisted of 40 students from Aichi University, the

Malaysian sample of 64 Chinese-Malaysian students from the University of Nottingham – Malaysia Campus, the Polish sample of 62 students from Jagiellonian University, the Turkish sample of 60 students from Koç University, and the British sample of 60 students from the University of Warwick.

Table 1. Demographic information of the participants from the 10 samples.

Country	*N*	*Females*	*Males*	*Mean Age*	*SD Age*	*Age Range*
Brazil	40	21	19	21.9	2.58	18-28
Canada[a]	42	30	11	22.6	4.45	19-41
Colombia	40	37	3	19.5	1.10	18-23
India	40	22	18	26.2	4.30	19-37
Japan[b]	40	33	7	20.2	0.58	19-22
Malaysia	64	36	28	20.3	1.80	17-27
Poland	62	41	21	21.8	2.05	18-27
Portugal	40	34	6	19.7	2.44	18-26
Turkey	60	34	26	20.8	1.20	19-24
United Kingdom[a,b]	58	51	3	18.6	1.82	18-31

[a]One Canadian and four UK participants did not indicate a gender.

[b]We removed two Japanese pairs who did not follow the instructions and one UK pair for which there were technical difficulties with the projectors. The Japanese pairs were replaced by two new pairs, but the UK pair was not replaced.

The data from the first batch were analyzed to assess the replicability of the co-witness suggestibility effect and reported at the meeting of the Society for Applied Research in Memory and Cognition in 2017 (see slides [https://osf.io/fsnpx/]) to invite researchers from additional countries. Five more samples were then collected as the second batch[1]: The Brazilian sample consisted of 40 students from the Federal University of Santa Catarina, the Canadian sample of 42 students from the University of Victoria, the Colombian sample of 40 students from Pontifical Xavierian University, the Indian sample of 40 students from the National Institute of Mental Health and Neurosciences, and the Portuguese sample of 40 students from the University of Minho. This second batch of studies was preregistered (https://osf.io/gtwrc/register/564d31db8c5e4a7c9694b2c2)[1]. We set the same sample size of 40 as in the Garry et al. (2008) study for each of the second batch. The results of the two batches were comparable so all are reported together.

All students participated in pairs for course credit or a small financial reward and reported normal or corrected-to-normal vision. We did not control the interpersonal relations of the pairs. Therefore, some pairs knew each other while others did not. They were all naive to the purpose of the study, which had been approved by the local ethics committee of each participating university. We obtained written informed consent from all participants.

[1]The details of the procedure were specified on the Open Science Framework in advance of the second wave of data collection, but we neglected to complete the formal registration process until July 25, 2018. The preregistration ID is http://doi.org/10.17605/OSF.IO/GTWRC.

Materials and Apparatus

We used the video created by Takarangi, Parker, and Garry (2006) and used by Garry et al. (2008). The video depicts simulated thefts committed by an electrician ("Eric") working in someone's house when the homeowner is away. There are two versions of the same movie, which are identical except for eight critical details. Takarangi et al. filmed only one version of the movie and then digitally altered the clip to create the critical details. For example, there is a picture of the *Eiffel Tower* on the wall in one version and a picture of the *Leaning Tower* in the other. The video lasts 6 minutes and 34 seconds and does not contain audio. All materials are available on the OSF page (https://osf.io/j5f82/).

Figure 1 illustrates the equipment set-up for the MORI technique. Both versions of the video were projected onto the back of a translucent projection screen. The screen was made of a 5-mm-thick pane of plain ground glass (45 x 60 cm). We used two LED projectors (TAXAN, KG-PL021X) mounted on a stand, one above the other, with one tilted slightly upward and the other slightly downward, so that their two images overlapped on the screen. The images were about 30 x 40 cm in size on the screen located at a distance of approximately 170 cm from participants. Polarizing filters were attached in front of the lens of each projector, one placed vertically and the other horizontally. For each pair of participants, one set of sunglasses allowed the wearer to view the vertically polarized image while blocking the horizontally polarized one, and the other set of glasses did the opposite.

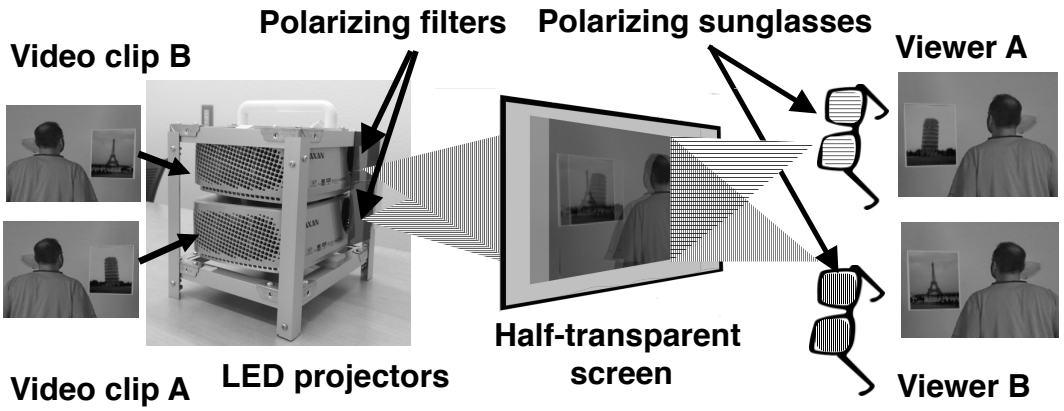

Figure 1. Equipment setting for the MORI technique showing a pair of the differing scenes; the electrician watching a picture on the wall. For the sake of illustration, the polarized lenses in the sunglasses are depicted as having different patterns but in reality the sunglasses looked identical.

Procedure

Following the procedure used by Garry et al. (2008), after providing informed consent, each participant pair was seated side by side in front of the screen. Then the pair was told that the experiment was about people's sensory impressions at different levels of visual acuity. We used the following instructions, or equivalent translated versions (native speakers of Japanese, Polish, Portuguese, Spanish, and Turkish with good knowledge of English and the experimental procedure translated the English instructions into each language):

"We are interested in people's sensory impressions at different levels of visual acuity. Visual acuity basically means how well you can see. So, for example, right now you all should have 100% visual acuity, either because your eyes work properly or because you have correcting glasses on. We want to know what happens to people's

> *sensory impressions when their visual acuity is degraded by different amounts. Today, you will both be in the 95% visual acuity condition. I will give you each a pair of 95% acuity glasses, which will degrade your vision slightly. If you already wear glasses, the acuity glasses should fit over the top."* (Garry et al., 2008, p. 433)

After these instructions, the experimenter handed horizontal or vertical polarizing sunglasses randomly to the two participants. Each pair was taken out of a box labeled "95% acuity sunglasses." To add authenticity to the cover story, other boxes were stacked next to the selected boxes, all labeled with different strengths of acuity. However, those other boxes were empty, and we always told participants that they were in the 95% visual acuity condition. The experimenter then continued as follows:

> *"I am going to show you a short movie of a tradesman called Eric working in a house. Please make sure you watch the movie through the sunglasses (no peeking over the top or around the side) and keep your sunglasses on until I ask you to remove them. We find that people often see best when they keep their head straight rather than tilted."* (Garry et al., 2008, p. 434)

The experimental session consisted of three phases. In Phase 1, participants watched the movie. The two versions of the video were presented on screen, while the participants sitting side-by-side observed the video; one viewer observed one version and the other viewer observed the other version (see the video clip showing how the MORI technique works [https://osf.io/cw9pd/]). Then, the experimenter gave each participant a printed logic puzzle

and participants worked individually on the puzzle for 15 minutes.

In Phase 2, participants took part in a collaborative recognition test. The test was presented with PowerPoint on a laptop or desktop computer. In this test, participants were asked to answer 12 five-alternative questions aloud. On each question, we allowed the pairs to discuss their answers freely whether they matched or not. Each question inquired about a detail of the event and offered five possible responses (e.g., "Eric looked at a picture of ___: The Leaning Tower, The Tower of London, The Eiffel Tower, The Sky Tower, Trump Towers"). Of the 12 questions, four focused on the critical details in the movie, with both correct answers being included among the five response alternatives; the remaining eight questions targeted filler details for which both subjects had seen the same detail.

For each critical question discussed during the collaboration phase, the experimenter recorded the answer that the pair of participants appeared to agree was correct; if each member of the pair maintained a different answer, then the experimenter recorded both answers. The other four conflicting details in the video were not discussed and thus served as controls in the individual recognition test. Which four critical details were discussed alternated across participant pairs. The collaboration phase was audio-video recorded in Canada and Japan, and audio recorded in all other countries except for Brazil, Colombia, and Portugal. After the collaboration task, participants worked individually on multiplication problems (3 digits by 3 digits) for 5 minutes.

In Phase 3, participants completed an individual recognition test comprising 20 two-

alternative forced-choice questions (paper-and-pencil). Of the 20 questions, eight targeted the critical questions (i.e., four discussed and four control questions) and participants had to choose between the alternative that they had seen versus the alternative that their partner had seen. The remaining 12 filler questions concerned details from the movie that had not been tested earlier. The participants indicated their confidence in each answer on a five-point scale, ranging from (1) 'not at all confident' to (5) 'very confident' but we do not report the confidence data here.

Finally, participants were debriefed. The experimenter asked the participants whether they had noticed any anomalies during the video presentation and then explained the experimental manipulation. No participants reported any anomalies to the extent that they were judged to have intuited the critical manipulation. It took about 45-60 minutes to complete the experiment.

Results

First, we examine the co-witness suggestibility effect following the Garry et al. (2008) procedure. Like Garry et al., we analyzed the final-test accuracy scores for the critical questions across the 10 samples using analysis of variance (ANOVA). We then present a meta-analysis of the effect sizes of the suggestibility effects across the 10 samples and the Garry et al. sample. Finally, we compare the answering patterns for discussed items that were disputed versus non-disputed.

Main descriptive statistics and the data for the 10 countries are presented in the Supplemental Materials (downloadable from the OSF site [https://osf.io/qbejt/]). Figure 2 shows average final-test accuracy on the filler, non-discussed, and discussed questions for each country, along with the original NZ results from Garry et al. (2008). Participants' answers on the 12 filler questions on the individual recognition test (which had not been queried during the collaboration phase) provided a baseline measure of memory. Performance on the filler questions was roughly comparable to the 80% accuracy reported by Garry et al. Participants remembered these details above chance level (.50) in each country (all ps < .001). The average proportion of correct responses for the filler questions was entered into a one-way analysis of variance (ANOVA) with country as the between-subjects factor. The effect of country was significant, $F(9, 476) = 11.99$, $p < .001$, $\eta_p^2 = .19$. Pairwise comparisons suggested that this effect was mainly due to lower recognition rates for the filler items in the Brazilian and Turkish samples.

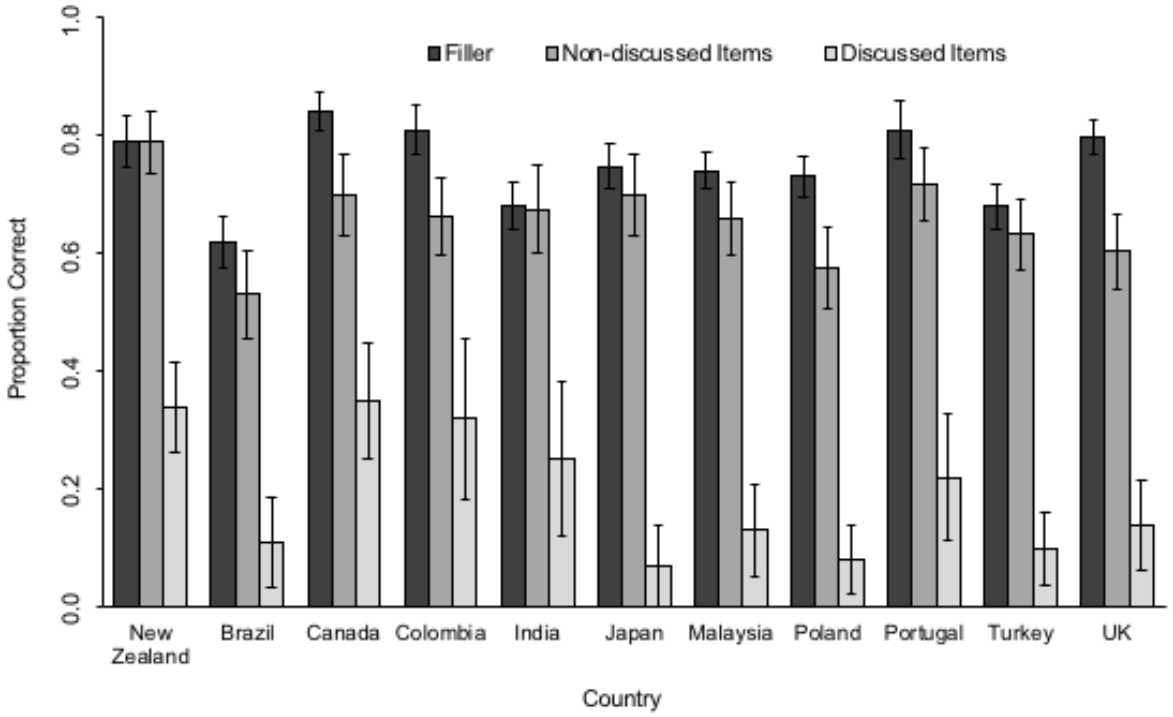

Figure 2. Proportions of correct responses to filler, non-discussed, and discussed items for each country along with the original data from New Zealand. The co-witness effect is the difference between non-discussed and discussed items. Error bars represent 95% CIs.

Following Garry et al. (2008), in scoring final-test accuracy on critical discussed questions, we included only questions for which the experimenter had recorded the co-witness's answer during the collaboration phase (i.e., provided an answer that contradicted what the participant had seen). In some pairs, one member was not exposed to any co-witness details during the collaboration phase (e.g., because they reported the details they had witnessed and their co-witness appeared to go along with those answers). For example, when one participant had seen a red cap and the other participant a black one and the first participant indicated that they had seen a red cap and the second participant did not say anything, then the first participant would not have discussed a co-witness detail whereas the second participant would have. Therefore, "discussed co-witness details" refers to details that

the experimenter recorded as having been reported by the participant's co-witness.

The number of discussed co-witness details for each participant did not significantly vary across the 10 samples, $F(9, 476) = 1.09$, $p = .365$, $\eta_p^2 = .02$. Participants who were not exposed to co-witness details at all were excluded from the subsequent analyses. The number of participants excluded on this basis ranged from a low of 1 participant in the Canadian sample to a high of 13 participants in the Malaysian sample. There were no corresponding cases in the New Zealand sample of Garry et al. (2008).

A repeated-measures ANOVA was conducted on the final-test accuracy score for the critical questions, with discussion (non-discussed vs. discussed) as the within-subjects factor and country as the between-subjects factor. The results showed main effects of country, $F(9, 405) = 4.66$, $p < .001$, $\eta_p^2 = .09$, and discussion, $F(1, 405) = 687.57$, $p < .001$, $\eta_p^2 = .63$, and an interaction, $F(9, 405) = 2.46$, $p = .010$, $\eta_p^2 = .05$. The effect of discussion was significant for each country, all ps $< .001$: Participants were more likely to report the correct answer for non-discussed questions than for discussed ones, thus replicating the original findings of Garry et al. (2008).

To obtain a general estimate of the co-witness suggestibility effect, we conducted a meta-analysis with the effect sizes from Garry et al. (2008) and from our 10 countries (see the Supplemental Materials for technical details on the meta-analysis and formulae). Figure 3 summarizes the main results of the meta-analysis and shows that the effect sizes (Hedges' g) of all the countries showed positive and large values. The meta-analysis revealed a combined

effect size of $g = 1.67$, 95% CI [1.40, 1.94], $Z = 12.19$, $p < .001$. Without the New Zealand sample, the combined effect size was $g = 1.70$, 95% CI [1.40, 1.99], $Z = 11.17$, $p < .001$. These two effect sizes translate to 80.2% (and 81.5%) of the items in which the participant was exposed to misleading information from the co-witness being answered incorrectly. Because participants were on average exposed to misinformation for 1.61 (and 1.58) items, each participant made 1.29 (and 1.29) errors due to the fact that they had been exposed to misinformation from the co-witness.

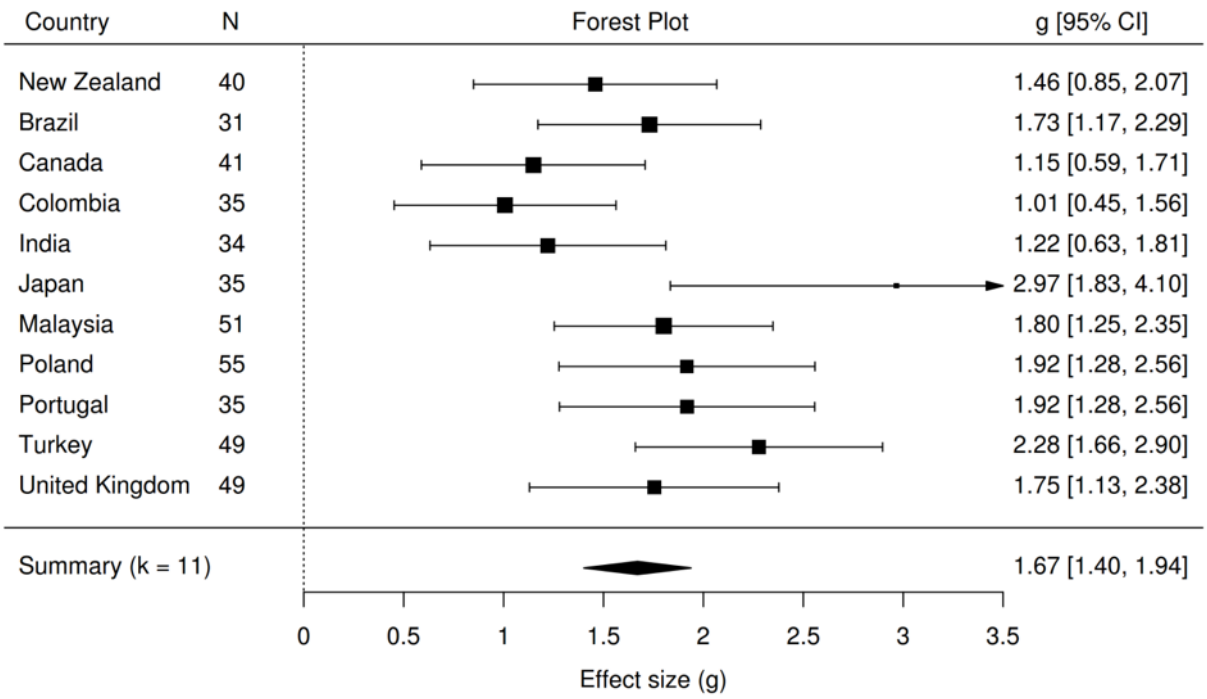

Figure 3. Forest plot with the effect sizes for each country, including the original New Zealand sample. The location of each square on the horizontal axis represents the effect size—the standardized difference between the mean correct responses in the discussed and non-discussed conditions (positive values indicate a higher mean proportion of correct responses to non-discussed items). The lines extending either side of the square represent a 95% confidence interval (CI). The size of each square represents the weight of each study in

the meta-analysis. The diamond shows the result of the meta-analyses, with the center indicating the combined effect size and the spread representing the 95% CI.

The co-witness suggestibility effect varied from 34% in the Colombian sample to 64% in the Japanese sample, which likely caused the significant interaction in the ANOVA reported above. To examine which countries differed, we conducted a series of pairwise comparisons. We computed the proportion of correct responses for non-discussed questions minus the proportion of correct responses for discussed questions per country (i.e., the co-witness suggestibility effect) and compared all samples (Bonferroni correction was applied and alpha was set at .001). The analyses showed a larger effect for the Japanese than the Canadian and Colombian samples ($p = .012$ and $p = .016$). All other pairwise comparisons were not significant, ps > .265.

As noted above, following Garry et al. (2008), final-test accuracy scores on critical discussed questions were contingent on the experimenter having recorded the co-witness's answer during the collaboration phase (which meant that the participant's co-witness had maintained that answer during the collaboration phase). Usually, only one answer was recorded for each critical discussed question; one member of the pair reported the detail that he or she had seen and the other did not dispute that report. However, sometimes each member of the pair stated the answer that he or she had seen. In an exploratory sub-analysis (not reported in Garry et al., 2008), we differentiated between critical discussed questions for which only one co-witness reported the answer he or she had witnessed during the

collaboration phase (i.e., non-disputed co-witness answers) and those for which both co-witnesses reported the answers they had seen (i.e., disputed co-witness answers).

As Table 2 shows, for the vast majority of critical discussed questions, only one answer was recorded during the collaboration phase (i.e., most were non-disputed). Across the 10 samples, there were 708 non-disputed and only 58 disputed co-witness answers. When the co-witness's answer was coded as non-disputed during the collaboration phase, witnesses later erred by reporting that answer on the final test in 626 of the 708 cases (88.4%). In contrast, across the 58 cases of disputed co-witness answers, only in one case (1.8%) did the witness err on the final test by reporting the co-witness's detail.

Table 2. Number of erroneous and correct responses in the non-disputed and disputed cases.

Country	*N of Participants*	*Total cases of "Discussed" items*	*Non-disputed Cases*		*Disputed Cases*	
			Freq. of Errors (%)	*Freq. of Corrects (%)*	*Freq. of Errors (%)*	*Freq. of Corrects (%)*
Brazil	40	64	56 (87.5)	8 (12.5)	0 (0.0)	0 (0.0)
Canada	42	84	52 (61.9)	2 (2.3)	0 (0.0)	30 (35.7)
Colombia	40	63	44 (69.8)	9 (14.3)	0 (0.0)	10 (15.9)
India	40	64	49 (76.6)	9 (14.1)	1 (1.6)	5 (7.8)
Japan	40	62	57 (91.9)	5 (8.1)	0 (0.0)	0 (0.0)
Malaysia	64	98	84 (85.7)	14 (14.3)	0 (0.0)	0 (0.0)
Poland	62	91	83 (91.2)	4 (4.4)	0 (0.0)	4 (4.4)
Portugal	40	63	48 (76.2)	9 (14.3)	0 (0.0)	6 (9.5)
Turkey	60	86	76 (88.4)	8 (9.3)	0 (0.0)	2 (2.3)
UK	58	91	77 (84.6)	14 (15.4)	0 (0.0)	0 (0.0)
Total	486	766	626 (81.7)	82 (10.7)	1 (0.1)	57 (7.4)

There was a striking national difference in the frequency of disputed co-witness answers (i.e., cases in which during collaboration both participants stated the answer that they

had observed in the video). Of the 58 cases of disputed co-witness answers, 30 (51.7%) were in the Canadian sample and 10 (17.2%) were in the Colombian sample, with the remaining 18 cases scattered across the 404 participants from the other eight countries. In four of the samples not a single instance of a disputed answer was recorded. The high accuracy rate on critical discussed items in the Canadian and Colombian samples apparently arose from the high rates of disputed co-witness details in those two samples. When analyses were restricted to critical discussed questions with non-disputed co-witness answers, accuracy did not significantly differ across countries (e.g., for the Japanese sample, 5 out of 62 cases, or 8.1%; for the Canadian sample, 2 out of 54 cases, or 3.7%; for Colombian sample, 9 of 53, or 17.0%), $\chi^2(9) = 13.92$, $p = .125$, Cramer's $V = .140$.

Discussion

The present study aimed to investigate the replicability of the co-witness suggestibility effect across 10 countries. We successfully replicated the findings of Garry et al. (2008) in each sample: Accuracy on the final individual recognition test was lower for critical discussed questions than for non-discussed control questions. The effect size across the 10 samples (Hedges' $g = 1.70$, 95% CI: 1.40–1.94) was comparable to the effect size reported by Garry et al. (Hedges' $g = 1.46$, 95% CI: 0.85–2.07).

The results were for the most part similar across the 10 countries, with the co-witness suggestibility effect statistically significant in each country. That said, the effect was statistically significantly larger in the Japanese sample than in the Canadian and Colombian

samples. An unplanned sub-analysis indicated that this difference was associated with differences between these samples in the frequency with which subjects disputed their co-witness's reports during the collaboration phase. In the Japanese sample, there was not a single case in which the answer to a critical discussed question was coded as disputed, whereas in the Canadian and Colombian samples a nontrivial percentage of co-witness details were coded as disputed.

It is not surprising that subjects were less influenced by co-witness suggestions that they had disputed during the discussion phase than by those they had appeared to accept, but we do not know why disputation was so rare in the Japanese sample, especially compared to the Canadian and Colombian samples. Some readers might speculate that this has to do with Japan being a relatively collectivist culture, but among the countries tested, Colombia is said to be the most collectivistic (Hofstede, 2001), whereas Canada is among the most individualistic. The differences across cultures in the rate of disputing critical answers therefore do not fit an interpretation that the co-witness suggestibility effect would be larger in collectivistic cultures.

Other studies have explored the nature of co-witness discussion and its association with collaborative memory distortions. For instance, Gabbert et al. (2006) reported that, in two studies in which co-witnesses had viewed different versions of the same event and then collaborated before being tested individually, the participant who mentioned a critical detail first was more likely to influence the other participant. Note, however, that Gabbert et al.'s

analyses were collapsed across cases in which both collaborators reported the detail they had witnessed and cases in which only one collaborator reported the detail they had witnessed. The latter cases cannot shed any light on an order effect, per se (see Lindsay, 2007), thus further research is needed to understand the role of order effects, if any, in collaborative memory contamination. The current study, however, reveals a novel finding about the nature of co-witness discussion and its influence on witness memory: Participants were only likely to report erroneous information gleaned from a collaborator if they did not, during the co-witness discussion phase, report the detail they had personally witnessed. That is, when witnesses did not dispute the co-witness report during collaboration, they almost always subsequently erred on the individual recognition test by claiming to have witnessed the detail their co-witness had seen. However, if they had disputed the co-witness's report during collaboration, they almost never later made that error. A related finding comes from Tousignant, Hall, and Loftus (1986). They reported that participants who read slowly (or who were instructed to read slowly) were more likely to detect discrepancies between original and post-event information and were therefore less vulnerable to accepting misinformation. It is possible that the participants in our study who disputed their co-witness' reports detected the difference between what they had seen and what their partner reported during the discussion. As such they were less likely to accept misinformation than the participants who did not dispute their co-witness' reports.

In the context of the co-witness suggestibility effect, this finding is novel, but it fits

with other findings in the literature. For example, in Ackil and Zaragoza (1998), participant-pairs viewed a short video. Immediately after the video, among a series of valid questions, participants were given misleading questions. For half the pairs, one of the two participants was told only to answer the questions about which they were certain of the response. For the other half, one of the two participants was forced to answer the questions and to guess the answer if they were unsure. In both conditions, the partners just listened. One week later, all participants received a source memory test. The participants who were forced to answer the misleading questions often attributed their forced confabulations to the video. Moreover, their partners (who were merely exposed to the forced confabulations), also sometimes misattributed the confabulated details to the video. Similar to the participants in our study who did not dispute their co-witness' reports, the participants who were merely exposed to the forced confabulations seemed to rely on their partners' answers.

A comprehensive theory of collaborative remembering could explain the conditions under which co-witnesses help each other recall details or aid each other in confirming the correctness of details, versus hurt each other's recollections by introducing inaccurate details. That is, co-witness discussion can be a source of memory distortion, but it can also fill memory gaps and enhance accuracy if the co-witnesses' individual reports are accurate (e.g., Vredeveldt, Groen, Ampt, & Van Koppen, 2017). This interpretation is consistent with other evidence that discussants sometimes help each other remember more details and can even correct instances of misleading post-event information (Karns, Irvin, Suranic, & Rivardo,

2009; Ross, Spencer, Blatz, & Restorick, 2008).

The instructions for the final recognition test made explicit that "we are testing your memory for this video." Nonetheless, these instructions do not necessarily mean that when participants selected a co-witness detail on the test that they were experiencing a false memory of having seen that detail in the video. They may instead have knowingly relied on their memories from the discussion phase of what their co-witnesses had said. Further research is required to investigate the subjective phenomenology that accompanies the reports of co-witness details in this procedure.

Our project is an example of an international collaboration focused on direct replication of an established effect. The importance of conducting large-scale replications like this one is particularly significant for issues that have direct implications for real-world practice. It is vitally important that psychologists, both basic and applied, continue to investigate the possibility of file drawer effects leading to effect size overestimation for findings that have influenced practitioners in their relevant fields.

In conclusion, we replicated in 10 countries (i.e., Brazil, Canada, Colombia, India, Japan, Malaysia, Poland, Portugal, Turkey, and the United Kingdom) the co-witness suggestibility effect by using the MORI technique first tested by Garry et al. (2008) in New Zealand. Whether people experience actual memory distortions or simply trust their partner to be correct, our results show that people tend to incorporate elements of the other witness' memory reports into their own memory reports, even when that information contradicts what

they themselves had seen. It is particularly noteworthy that all 10 samples showed similar results with large effect sizes in spite of the cultural differences. The consistency in the size of the effect across countries is good evidence that the co-witness suggestibility effect remains similar in a dyad regardless of the individualistic or collectivistic tendencies of the culture within which the interaction occurs.

Author Contributions

Correspondence concerning this manuscript should be addressed to Kazuo Mori at Matsumoto University. This research was supported by a Grant-in-Aid from the Japan Society for the Promotion of Science (KAKENHI No. 25280050) to Kazuo Mori and a grant from Aichi University (KENKYUJYOSBEI No. C-180) to Hiroshi Ito. KM and HI were also supported by the Joint Research Grant 2017 from the Promotion and Mutual Aid Corporation for Private Schools of Japan to Matsumoto University for a cooperative research with Aichi University.

While conducting the study and writing of the manuscript, Krystian Barzykowski was supported by a grant from the National Science Centre, Poland [no.: 2015/19/D/HS6/00641]. Nicole Laird's work on this project was supported by a Natural Science and Engineering Research Council Discovery Grant to D. Stephen Lindsay.

Hiroshi Ito and Kazuo Mori initiated the project and invited the other authors to take part. The authors conducted the experiments in their respective countries and provided the data for the project. Hiroshi Ito, Karlos Luna, and Steve Janssen analyzed the entire dataset, and all authors discussed the results. Hiroshi Ito drafted the manuscript and all other authors provided critical revisions. All authors approved the manuscript before submission.

We are grateful to Maryanne Garry and Lauren Hewitt (née French) for providing the experimental materials used in their study. We also thank Tahir Chatur for his help setting the equipment in Canada.

References

Ackil, J. K., & Zaragoza, M. S. (1998). Memorial consequences of forced confabulation: Age differences in susceptibility to false memories. *Developmental Psychology, 34,* 1358–1372. doi:10.1037/0012-1649.34.6.1358

Bodner, G. E., Musch, E., & Azad, T. (2009). Reevaluating the potency of the memory conformity effect. *Memory & Cognition, 37*(8), 1069–1076. doi:10.3758/MC.37.8.1069

Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis (2nd Ed.)* (pp. 221–235). New York: Russell Sage Foundation.

Bright-Paul, A., Jarrold, C., Wright, D. B., & Guillaume, S. (2012). Children's memory distortions following social contact with a co-witness: Disentangling social and cognitive mechanisms. *Memory, 20*(6), 580–595. doi:10.1080/09658211.2012.690039

Condon, C. E., Ritchie, T. D., & Igou, E. R. (2015). How dyads reminiscence moderates the relations between familiarity, trust, and memory conformity. *Social Psychology, 46*(2), 65–75. doi:10.1027/1864-9335/a000222

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis.* New York: Routledge.

Davis, S. D., & Meade, M. L. (2013). Both young and older adults discount suggestions from older adults on a social memory test. *Psychonomic Bulletin & Review, 20*(4), 760–765.

doi:10.3758/s13423- 013-0392-5

French, L., Garry, M., & Mori, K. (2008). You say tomato? Collaborative remembering leads to more false memories for intimate couples than for strangers. *Memory, 16*(3), 262–273. doi:10.1080/09658210701801491

French, L., Garry, M., & Mori, K. (2011). Relative—not absolute—judgments of credibility affect susceptibility to misinformation conveyed during discussion. *Acta Psychologica, 136*(1), 119–128. doi:10.1016/j.actpsy.2010.10.009

Gabbert, F., Memon, A., & Allan, K. (2003). Memory conformity: Can eyewitnesses influence each other's memories for an event? *Applied Cognitive Psychology, 17*(5), 533–543. doi:10.1002/acp.885

Gabbert, F., Memon, A., Allan, K., & Wright, D. B. (2004). Say it to my face: Examining the effects of socially encountered misinformation. *Legal and Criminological Psychology, 9*(2), 215–227. doi:10.1348/1355325041719428

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. doi:10.3758/BF03193873

Gabbert, F., Memon, A., & Wright, D. B. (2007). I saw it for longer than you: The relationship between perceived encoding duration and memory conformity. *Acta Psychologica, 124*(3), 319–331. doi:10.1016/j.actpsy.2006.03.009

Garry, M., French, L., Kinzett, T., & Mori, K. (2008). Eyewitness memory following

discussion: Using the MORI technique with a western sample. *Applied Cognitive Psychology, 22*(4), 431–439. doi:10.1002/acp.1376

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107–128. doi:10.3102/10769986006002107

Hewitt, L. Y., Kane, R., & Garry, M. (2013). Speaking order predicts memory conformity after accounting for exposure to misinformation. *Psychonomic Bulletin & Review, 20*(3), 558–565. doi:10.3758/s13423-013-0377-4

Hirokawa, K., Matsuno, E., Mori, K., & Ukita, J. (2006) Relationship between masculinity-femininity and concession in an experimental collaborative eyewitness testimony. *Asian Journal of Social Psychology, 9*(2), 132–139. doi:10.1111/j.1467-839X.2006.00190.x

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations (2nd Edition)*. Thousand Oaks, CA: Sage Publications.

Hope, L., Ost, J., Gabbert, F., Healey, S., & Lenton, E. (2008). "With a little help from my friends…": The role of co-witness relationship in susceptibility to misinformation. *Acta Psychologica, 127*(2), 476–484. doi:10.1016/j.actpsy.2007.08.010

Jack, F., Zydervelt, S., & Zajac, R. (2014). Are co-witnesses special? Comparing the influence of co-witness and interviewer misinformation on eyewitness reports. *Memory, 22*(3), 243–255. doi:10.1080/09658211.2013.778291

Kanematsu, H., Mori, K., & Mori, H. (1996/2003). Kotonaru jitai wo mokugekishita futari no mokugekisha no hanashiai niyoru kioku no hen-you [Memory distortion in eyewitness pairs who observed nonconforming events and discussed them]. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society, 3,* 29–40 [English version was published in *Journal of Faculty of Education* (2003, Shinshu University, Nagano, Japan), *109*, 75–84].

Karns, T. E., Irvin, S. J., Suranic, S. L., & Rivardo, M. G. (2009). Collaborative recall reduces the effect of a misleading post event narrative. *North American Journal of Psychology, 11*(1), 17–28.

Lindsay, D. S. (2007). Order effects in collaborative memory contamination? Comment on Gabbert, Memon, and Wright (2006)? *Psychonomic Bulletin & Review, 14*(5), 1010. doi:10.3758/BF03194137

Lindsay, D. S. (2015). Replication in psychological science. *Psychological Science*, *26*(12), 1827–1832. doi:10.1177/0956797615616374

Lindsay, D. S., Simons, D. J., & Lilienfeld, S. O. (2016, December). Research preregistration 101. *Observer, 29*, 14–16.

Meade, M. L., McNabb, J. C., Lindeman, M. I., & Smith, J. L. (2017). Discounting input from older adults: The role of age salience on partner age effects in the social contagion of memory. *Memory, 25*(5), 704–716. doi:10.1080/09658211.2016.1207783

Mori, K. (2007). A revised method for projecting two different movies to two groups of viewers without their noticing the duality. *Behavior Research Methods, 39*(3), 574–578. doi:10.3758/BF03193028

Mori, K., & Kitabayashi, M. (2009). How child-mother pairs reported what they had witnessed together: An experimental examination using the MORI technique. *Psychology Journal, 6*(2), 60–69.

Mori, K., & Mori, H. (2008). Conformity among cowitnesses sharing same or different information about an event in experimental collaborative eyewitness testimony. *Perceptual and Motor Skills, 106*(1), 275–290. doi:10.2466/pms.106.1.275-290

Mori, K., & Takahashi, R. (2012). Pre-schoolers' reports of conflicting points secretly inserted into a co-witnessed event: An experimental investigation using the MORI technique. *Psychology, 3*(1), 30–35. doi:10.4236/psych.2012.31005

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*, 943. doi:10.1126/science.aac4716

Paterson, H. M., & Kemp, R. I. (2006). Co-witnesses talk: A survey of eyewitness discussion. *Psychology, Crime & Law, 12*(2), 181–191. doi:10.1080/10683160512331316334

Paterson, H. M., Kemp, R., & McIntyre, S. (2012). Can a witness report hearsay evidence unintentionally? The effects of discussion on eyewitness memory. *Psychology, Crime & Law, 18*(6), 505–527. doi:10.1080/1068316X.2010.510117

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. doi:10.1037/0882-7974.23.1.85

Simmons, J. P., Nelson, L. D., Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22,* 1359–1366. doi:10.1177/0956797611417632

Simons, D. J., Holcombe, A. O., & Spellman, B. A. (2014). An introduction to registered replication reports at Perspectives in Psychological Science. *Perspectives in Psychological Science, 6*, 552–555. doi:10.1177/1745691614543974

Skagerberg, E. M., & Wright, D. B. (2008). The prevalence of co-witnesses and co-witness discussions in real eyewitnesses. *Psychology, Crime & Law, 14*(6), 513–521. doi:10.1080/10683160801948980

Tainaka, T., Miyoshi, T., & Mori, K. (2014). Conformity of witnesses with low self-esteem to their co-witnesses. *Psychology, 5*, 1695–1701. doi:10.4236/psych.2014.515177

Takarangi, M. K. T., Parker, S. L., & Garry, M. (2006). Modernizing the misinformation effect: The development of a new stimulus set. *Applied Cognitive Psychology, 20*(5), 583–590. doi:10.1002/acp.1209

Tousignant, J. P., Hall, D., & Loftus, E. F. (1986). Discrepancy detection and vulnerability to misleading postevent information. *Memory & Cognition, 14,* 329–338. doi:10.3758/BF03202511

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of*

Statistical Software, 36(3), 1–48. Retrieved from http://www.jstatsoft.org/v36/i03/

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. doi:10.1111/lcrp.12097

Wright, D. B., Busnello, R. H. D., Buratto, L. G., & Stein, L. M. (2012). Are valence and social avoidance associated with the memory conformity effect? *Acta Psychologica, 141*(1), 78–85. doi:10.1016/j.actpsy.2012.06.008

Wright, D. B., London, K., & Waechter, M. (2010). Social anxiety moderates memory conformity in adolescents. *Applied Cognitive Psychology, 24*(7), 1034–1045. doi:10.1002/acp.1604

Wright, D. B., Memon, A., Skagerberg, E. M., & Gabbert, F. (2009). When eyewitnesses talk. *Current Directions in Psychological Science, 18*(3), 174–178. doi:10.1111/j.1467-8721.2009.01631.x

Supplemental Materials

Table S1. Mean (Standard Error) and [95% Confidence Interval] for each of the main measures per country. For filler, non-discussed and discussed items, proportion of correct recognition is shown.

	Proportion of Correct Recognition: Filler Items	*Proportion of Discussed Items for which Co-Witness Answered*	*Proportion of Correct Recognition: Non-Discussed Items (X)*	*Proportion of Correct Recognition: "Discussed" Items (Y)*	*Co-Witness Effect (X-Y)*
Brazil	.62 (.02) [.58, .66]	.40 (.05) [.30, .50]	.54 (.04) [.45, .63]	.11 (.04) [.02, .19]	43%
Canada	.84 (.02) [.80, .87]	.50 (.04) [.43, .57]	.70 (.04) [.62, .77]	.35 (.05) [.24, .46]	35%
Colombia	.81 (.02) [.77, .85]	.39 (.04) [.32, .47]	.66 (.03) [.58, .73]	.32 (.07) [.18, .45]	34%
India	.68 (.02) [.64, .72]	.40 (.04) [.32, .48]	.65 (.04) [.57, .74]	.25 (.07) [.11, .38]	40%
Japan	.75 (.02) [.71, .79]	.39 (.04) [.32, .46]	.71 (.03) [.64, .79]	.07 (.03) [0, .14]	64%
Malaysia	.74 (.01) [.70, .77]	.38 (.04) [.31, .45]	.61(.03) [.54, .68]	.13 (.04) [.05, .20]	48%
Poland	.73 (.02) [.70, .77]	.37 (.03) [.31, .42]	.58 (.04) [.50, .66]	.08 (.03) [.03, .14]	50%
Portugal	.81 (.03) [.76, .87]	.39 (.03) [.33, .46]	.74 (.03) [.68, .80]	.22 (.05) [.11, .33]	52%
Turkey	.68 (.02) [.64, .72]	.36 (.03) [.33, .46]	.64 (.03) [.57, .70]	.10 (.03) [.04, .17]	54%
United Kingdom	.80 (.01) [.77, .83]	.39 (.03) [.33, .45]	.62 (.03) [.54, .69]	.14 (.04) [.06, .22]	48%

Meta-Analysis Formulae

For the meta-analyses we used the “metafor” package (Viechtbauer, 2010) for R. We ran a random-effects model and the weighting of each study was determined by the inverse of the sampling variance.

As measure for the meta-analysis, we computed Cohen’s *d* for each study following

the formulae in Cumming (2012):

$$\mathrm{d}=\frac{M_{ND}-M_D}{\sqrt{\frac{\mathrm{SD}_{ND}^2+\mathrm{SD}_D^2}{2}}}$$

where M_{ND} and SD_{ND} are the mean and standard deviation for the proportion of correct recognition for non-discussed items, and M_D and SD_D are the mean and standard deviation for discussed items. Then, we applied the bias correction proposed by Hedges (1981):

$$\textit{Hedges' } g = \left(1-\frac{3}{4df-1}\right)d$$

In a within-subjects design *df* is $n - 1$. The sampling variance of g was computed following Equation 12.19 of Borenstein (2009), where n is the number of participants and r is the correlation between memory for items discussed and not discussed:

$$v_g = \left(\frac{1}{n}+\frac{g^2}{2n}\right)2(1-r)$$