Robust Improvements on Reasoning Performance Following Discussion in Japan

MARIKO BOKU   *Ritsumeikan University*

HIROSHI YAMA   *Osaka City University*

HUGO MERCIER*   *Centre National de la Recherche Scientifique*

Abstract: Group discussion significantly improves performance on intellective problems. However, most experiments have been conducted in Western cultures. Cross-cultural psychology suggests that members of Eastern cultures might be less likely to benefit from group discussion. One experiment in Japan suggested that this was not the case, but this experiment suffered from some limitations. To address these limitations, Japanese participants were asked to solve an intellective task four times: individually (pre-test), in small discussion groups (test), individually again (transfer task, post-test), and individually after a delay (delayed post-test). The results revealed a robust improvement during group discussion. Groups in which at least one member had found the correct answer individually agreed on it during the discussion. Moreover, and in contrast with results obtained in Western cultures, most groups with no such member also found the correct answer. The gains obtained during discussion were maintained in the transfer tasks. This result provides further evidence that the improvement of reasoning performance in group discussion is a universal phenomenon, and provides support for the practice of collaborative learning in Japan.

Key words: group discussion, argumentation, intellective problem, truth wins, Japan.

Small groups of participants engaged in discussion together have been shown to significantly outperform individual participants in the resolution of intellective tasks (Laughlin, 2011). Intellective tasks are tasks for which the correct answer can be conclusively demonstrated using knowledge available to the participants. Most of the tasks used in the psychology of reasoning—such as the Wason (1966) selection task—and many tasks used in judgment and decision making—such as the problems used to demonstrate the conjunction fallacy (Tversky & Kahneman, 1983) and the problems of the Cognitive Reflection Test (Frederick, 2005)—are intellective. On these tasks, groups tend to follow the "truth-wins" scheme: if at least one group member has found the correct answer, she consistently convinces all the group members to accept it (Laughlin, 2011).

This result plays an important theoretical role in debates on the question of the function of human reasoning. According to the argumentative theory of reasoning, the main

function of human reasoning is to exchange arguments with others (Mercier & Sperber, 2011). This theory predicts that reasoning should yield better outcomes when it is used in conversation than in solitary ratiocination. It is thus well supported by the comparison of individual and group performance on intellective tasks (as well as on a variety of others tasks, see Mercier, 2016; Mercier & Sperber, 2011).

However, for these results to fully support the argumentative theory of reasoning, two conditions must hold. First, it is the exchange of arguments that must be chiefly responsible for the improvement in performance during discussion and not, for instance, the forceful imposition of a view by a more confident participant. Second, the improvement in performance during discussion should not depend on a particular cultural setting.

Regarding the first condition, substantial evidence suggests that it is the exchange of arguments that is responsible for the improvement in performance during discussion. Transcripts reveal that participants do exchange arguments at length (Moshman & Geil, 1998; Trognon, 1993).

Moreover, a participant with the correct answer convinces her peers even if she faces group members who all agree on a different answer (the same within these group members) and are more confident than she is (Trouche, Sander, & Mercier, 2014). Finally, participants who change their mind following group discussion tend to perform better on transfer problems than they had before the discussion (Laughlin, 2011; Trouche et al., 2014). Although the argumentative theory of reasoning does not predict that such transfer will always take place, when it does it is a sign that participants have understood why they should change their mind to accept the correct answer.

There is less evidence relevant to the second condition. The vast majorities of experiments demonstrating an improvement in performance during discussion have taken place in Western cultures. It can be argued that Western cultures might be particularly conducive to

obtaining this result. Western cultures tend to value argumentation (Lloyd, 1996; Nakamura, 1964; Peng & Nisbett, 1999), they encourage parents to exchange arguments with their children (Gauvain, Munroe, & Beebe, 2013; Tizard, Hughes, Carmichael, & Pinkerton, 1983), and they teach some forms of reasoning useful in argumentation—such as hypothetical reasoning—in schools from an early age (Luria, 1971; Scribner, 1977; the arguments for and against the peculiarity of Western cultures in this respect are reviewed in Castelain, Girotto, Jamet, & Mercier, 2016; Mercier, 2011; Mercier, Deguchi, Van der Henst, & Yama, 2016).

To the best of our knowledge, only two studies have directly compared the performance of individuals and groups on intellective tasks in non-Western cultures. The first was conducted with participants from a traditional Maya community in Guatemala (Castelain et al., 2016). The results replicated those obtained in the West: poor individual performance, strong improvement during discussion (truth-wins scheme), and good transfer performance. The second study was conducted with Japanese participants (Mercier et al., 2016, Experiment 1). The most basic result—an improvement in group discussion—was also observed. However, there were some limitations to this study. First, there was no transfer task. As a result, it was more difficult to ascertain whether the participants who accepted the correct answer during the discussion had understood the correct reasons for doing so. Second, the truth-wins scheme was only observed in two-thirds of the cases. Although this may seem like a strong result, this means that groups in which a member had the correct answer initially were not significantly more likely to agree on the correct answer than groups in which no participant had the correct answer initially. This result contrasts with the opposite pattern observed in Western cultures (for review, see Mercier, Trouche, Yama, Heintz, & Girotto, 2015). The results of another comparison of individual and group performance in Japan, using a non-intellective task, also suggested potential differences in

the effects of group discussion, compared to Western participants (Mercier et al., 2016, Experiment 2).

Thus, it is still possible that group discussion of intellective tasks operates differently in Japanese participants, compared at least to participants from Western cultures. This would contradict the predictions of the argumentative theory of reasoning, according to which argumentation should yield consistently good results across cultures. This concern is particularly pressing since it has long been argued that Japanese do not fully enjoy the benefits of argumentation (for review, see Mercier, 2011). Some have argued that the Japanese language is inimical to the expression of precise arguments (Nakamura, 1964). Others have claimed that the pressure to preserve social harmony largely precludes Japanese people from expressing disagreements, and thereby from arguing (Becker, 1983, 1986). Finally, some cross-cultural psychologists have suggested that members of East-Asian cultures care little about contradiction, another obstacle to proper argumentation (Peng & Nisbett, 1999). It is thus particularly relevant to soundly test whether Japanese participants see their performance improve following group discussion, and whether this potential improvement is due to argumentation.

The purpose of this study was to evaluate this prediction of the argumentative theory of reasoning by testing whether Japanese participants could derive the same benefits from argumentation as Western participants. We thus replicated with Japanese participants an experiment previously conducted with Swiss participants (Trouche et al., 2014). Participants were asked to complete an intellective task on their own (pre-test). Small groups were then formed, and the participants tried to complete the task together (test). Afterwards, participants were asked to solve a transfer problem, along with a decoy problem added to ensure that a correct answer to the transfer problem would not be an unthinking copying of the answer agreed on during the discussion (post-test). Finally, one transfer question (the same as the question used for the pre-test and the test) was asked again after an interval of 14 weeks (delayed post-test).

Compared to Experiment 1 of Mercier et al. (2016), this experiment was different in two main ways. First, it used a different problem, one for which the classification of answers was easier—they were either completely correct or completely wrong. Second, it introduced transfer tasks—including a delayed post-test delivered 14 weeks after the initial test—to assess the degree of understanding acquired during group discussion. We also introduced a novelty in the analysis by looking at potential gender differences. This analysis seemed particularly relevant for two reasons. One is that gender roles are traditionally strongly marked in Japan, and could thus be more likely to influence the outcome (see, Brinton, 1993; Rosenfeld, Van Buren, & Kalleberg, 1998). The other is that some gender differences in collectivism have been observed (Triandis, Bontempo, Villareal, Asai, & Lucca, 1988). Greater collectivism might reinforce the need to preserve social harmony, which might be detrimental to the exchange of arguments that challenge the beliefs of a majority of group members.

Method

Participants. This experiment involved 95 participants (38 females) from a second-year (19- to 20-year-olds) English class in a Japanese university.

Materials. We used a disjunctive reasoning problem introduced in Levesque (1986):

Paul is looking at Linda and Linda is looking at Patrick. Paul is married but Patrick is not. Is a person who is married looking at a person who is not married? Yes / No / Cannot be determined.

The correct answer is "Yes" (because the statement is true whether Linda is married or not), but most participants attempting to

Table 1   Average group scores at the four phases of the experiment

Phase	Pre-test	Test	Post-test	Delayed post-test
Score (*SD*)	0.07 (0.11)	0.61 (0.47)	0.69 (0.30)	0.74 (0.23)

solve this problem on their own answer "Cannot be determined" (Trouche et al., 2014). As transfer problems, we used an analogue of the problem above with different superficial features (similar to the problem described next), as well as a new problem in which the correct answer is "Cannot be determined":

> Paul is selling bonds to Liz and Liz is selling bonds to Steven. Paul works for a company called GNB and Steven also works for GNB. Is there someone working for GNB selling bonds to someone who does not work for GNB? Yes / No / Cannot be determined.

Procedure. Participants were given a sheet with the first version of the Levesque problem, and they were given 5 min to answer (pretest). Participants were then assigned at random to a group of three to five participants (except for one group of six; a total of 22 groups were formed). The participants in each group were asked to try to reach an agreement on an answer to the same problem. At the end of the 10 min, each participant provided an answer on a separate sheet (test). Because the experiment took place during an English class, the discussions were supposed to be held in English. Finally, participants had 5 min to complete the two transfer tasks individually (post-test). Fourteen weeks after the initial session, participants were again given 5 min to solve the initial task again individually (delayed post-test).

Results

In order to remove any concern with dependencies between data points, most analyses were performed at the level of the group. The actual discussion groups were used for the test,

and the nominal groups composed of the same members for the other phases. Each group was assigned a score between 0 and 1: the proportion of its members that provided the correct answer (for the transfer tasks, which implies providing the correct answer to both problems).

Averaging over group scores, we obtained the following results (see Table 1). Paired *t*-tests revealed a significant improvement between the pre-test and any of the other three phases, all three $t(21) > 5.89$, *p*s < .0001. There were no other significant differences, all three $t(21) < 1.36$, *p*s > .19. To test whether the truth-wins scheme applied, we looked at the influence of the pre-test performance on test performance. In seven (out of 22) groups, one participant had found the correct answer during the pre-test (there were no groups with more than one participant having found the correct answer during the pre-test). All the members of these seven groups provided the correct answer at the end of the discussion.

Some groups in which no members had found the correct solution during the pre-test also found the correct answer during the test. Out of 15 such groups, eight provided the correct answer (with the exception of one of their members, who then provided the correct answer during the post-test).

Given that groups in which at least one member had found the correct solution during the pre-test all ended up accepting the correct answer during the test, to look for potential gender effects on discussion, we focused on the groups in which no member had found the correct solution during the pre-test. Out of these 15 groups, eight had a majority of male participants, one was equally split, and six had a majority of female participants. Of the eight groups having a majority of male participants, six found the correct answer during the discussion, while only one of the groups with a majority of female participants did, which is a significant difference,

$t(12) = 2.77$, $p = .017$. Although the difference between the majority male and the majority female groups narrowed at the post-test and delayed post-tests, it remained marginally significant, both $t(12) > 1.80$, $ps < .1$. Crucially, however, both genders benefited greatly from the discussions, with both post-test and delayed post-tests performance greatly superior to the pre-test performance; for males, both $t(56) > 10.03$, $ps < .0001$; for females, both $t(37) > 6.22$, $ps < .0001$.

Discussion

Japanese participants were asked to solve an intellective task four times: individually (pre-test), in small discussion groups (test), individually (post-test), and individually after a delay (delayed post-test). The two post-tests also included another problem to ensure that participants were not applying indiscriminately the answer acquired during group discussion.

We replicated with Japanese participants an experiment previously conducted with Swiss participants (Trouche et al., 2014) in which we observed large improvement in performance following group discussion, and that the truth-wins scheme applied.[1] Moreover, we have also demonstrated that the participants successfully transferred their group gains to both post-tests. This latter result is particularly significant, since it demonstrates that the performance gains from discussion stem from a genuine understanding of the logic behind the correct answer.

Ideally, the performance in the post-test would have been compared to that of a control condition in which participants would have solved the initial problems twice as well (at the pre-test and test), but always on their

own. However, such tests have never revealed significant improvements in performance (see, Laughlin, 2011). Given the huge improvement in performance observed during the test (from 0.07% correct answer to 0.61%), it is extremely implausible that mere repetition could explain the effect. Moreover, repetition could not explain the application of the truth-wins scheme. As a result, we can be confident that it is the group phase that yielded the improved performance.

Moreover, the good performance at the post-test shows that participants who acceded to the right answer during the group phase understood it, an understanding that is likely to have been gained through discussion. Previous analyses of transcripts in similar tasks (Moshman & Geil, 1998) have shown how participants exchange arguments about the correct answer and how this helps them understand it.

There were two potentially significant differences with previous results obtained in Western cultures. First, a majority of groups in which no member had found the correct answer during the pre-test also found the correct answer following discussion (eight out of 15). For the same task, only three out of 14 Swiss groups found the correct answer (Trouche et al., 2014; and seven out of 27 for another intellective task, see Mercier et al., 2016). This improvement suggests that the improvement observed during the group phase was not entirely due to students who found the right answer during the pre-test explanation to others during the test, but also to a more collaborative form of problem solving.

The second difference is that this improvement among groups in which no member had found the correct answer was nearly completely driven by groups with a majority of male participants. This superior performance for male participants in this case might be explained by the fact that, in Japan, women tend to rate higher on collectivism scales than males (Triandis et al., 1988). As collectivism is related to a desire to preserve social harmony (see, Peng & Nisbett, 1999), more collectivistic participants might have been reluctant to start

[1]In our experiment, the discussion was conducted in English rather than in Japanese. Although having the discussion held in a language that participants were less comfortable with should presumably only have hindered the discussion, it would still be interesting to replicate the current experiments with Japanese participants discussing in Japanese (as was the case in Mercier et al., 2016).

exchanging arguments over a problem on which everybody agreed. We should also note that the gender difference observed across groups in which no member had the correct answer initially might reflect peculiarities of this sample, or of the task used, and that no difference was observed when at least one group member had the correct answer initially. Finally, the main outcome remains that both genders benefitted hugely from the group discussions.

Thus, on the whole, and contrary to what the cross-cultural psychology literature suggests (Becker, 1983, 1986; Nakamura, 1964; Peng & Nisbett, 1999), Japanese participants benefitted more from discussion than Western participants: even groups in which no participant had found the correct answer were often able to discover it during discussion. Although the gender difference observed was significant, it might reflect peculiarities of this sample, or of the task used, and caution should thus be exerted.

These results have theoretical implications. In the cross-cultural psychology literature, the two broad types of cultures who have been thought most likely to differ from Western cultures in terms of reasoning and argumentation are Eastern and traditional cultures (see, Mercier, 2011). Thus, the fact that these improvements in reasoning performance due to argumentation have now been replicated in an Eastern culture (Japan), and that they have been observed in a traditional population (Maya), offers a strong argument in support of their universality. Although more tests in different cultures are needed to establish a firm conclusion, the current results support the thesis that improvements in reasoning performance during group discussion are universal, and that they are universally driven by argumentation, as predicted by the argumentative theory of reasoning (Mercier & Sperber, 2011).

These results also have practical implications. It has been suggested that Eastern cultures might not be able to enjoy the full benefits of collaborative learning (Nguyen, Elliott, Terlouw, & Pilot, 2009)—benefits that have proven to be substantial for a wide variety of problems in Western cultures (Slavin, 1995) and a variety of other cultures (Shwalb & Shwalb, 1995). This result suggests, on the contrary, that members of at least one Eastern culture are perfectly able to reap the learning benefits of group discussion, and thus that there should not be any obstacle to the promotion of collaborative learning in these cultures.

References

Becker, C. B. (1983). The Japanese way of debate. *National Forensic Journal*, *1*, 141–147.

Becker, C. B. (1986). Reasons for the lack of argumentation and debate in the Far East. *International Journal of Intercultural Relations*, *10*, 75–92.

Brinton, M. C. (1993). *Women and the economic miracle: Gender and work in postwar Japan*. Los Angeles: University of California Press.

Castelain, T., Girotto, V., Jamet, F., & Mercier, H. (2016). Evidence for benefits of argumentation in a Mayan indigenous population. *Evolution and Human Behavior*, *37*, 337–342.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*, 25–42.

Gauvain, M., Munroe, R. L., & Beebe, H. (2013). Children's questions in cross-cultural perspective: A four-culture study. *Journal of Cross-Cultural Psychology*, *44*, 1148–1165.

Laughlin, P. R. (2011). *Group problem solving*. Princeton, NJ: Princeton University Press.

Levesque, H. J. (1986). Making believers out of computers. *Artificial Intelligence*, *30*, 81–108.

Lloyd, G. E. R. (1996). *Adversaries and authorities*. Cambridge, England: Cambridge University Press.

Luria, A. R. (1971). Towards the problem of the historical nature of psychological processes. *International Journal of Psychology*, *6*, 259–272.

Mercier, H. (2011). On the universality of argumentative reasoning. *Journal of Cognition and Culture*, *11*, 85–113.

Mercier, H. (2016). The argumentative theory: Predictions and empirical evidence. *Trends in Cognitive Sciences*, *20*, 689–700.

Mercier, H., Deguchi, M., Van der Henst, J.-B., & Yama, H. (2016). The benefits of argumentation are cross-culturally robust: The case of Japan. *Thinking & Reasoning*, *22*, 1–15.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*, 57–74.

Mercier, H., Trouche, E., Yama, H., Heintz, C., & Girotto, V. (2015). Experts and laymen grossly underestimate the benefits of argumentation for reasoning. *Thinking & Reasoning*, *21*, 341–355.

Moshman, D., & Geil, M. (1998). Collaborative reasoning: Evidence for collective rationality. *Thinking & Reasoning*, *4*, 231–248.

Nakamura, H. (1964). *Ways of thinking of Eastern peoples: India, China, Tibet, Japan*. Honolulu: University of Hawaii Press.

Nguyen, P. M., Elliott, J. G., Terlouw, C., & Pilot, A. (2009). Neocolonialism in education: Cooperative learning in an Asian context. *Comparative Education*, *45*, 109–130.

Peng, K., & Nisbett, R. E. (1999). Culture, dialectics and reasoning about contradiction. *American Psychologist*, *54*, 741–754.

Rosenfeld, R. A., Van Buren, M. E., & Kalleberg, A. L. (1998). Gender differences in supervisory authority: Variation among advanced industrialized democracies. *Social Science Research*, *27*, 23–49.

Scribner, S. (1977). Modes of thinking and ways of speaking: Culture and logic reconsidered. In P. N. Johnson-Laird & P. C. Wason (Eds.), *Thinking: Readings in cognitive science* (pp. 483–500). New York, NY: Cambridge University Press.

Shwalb, B. J., & Shwalb, D. W. (1995). Cooperative learning in cultural contexts: An integrative review. *International Journal of Educational Research*, *23*, 293–300.

Slavin, R. E. (1995). *Cooperative learning: Theory, research, and practice* (Vol. 2). London, England: Allyn and Bacon.

Tizard, B., Hughes, M., Carmichael, H., & Pinkerton, G. (1983). Language and social class: Is verbal deprivation a myth? *Journal of Child Psychology and Psychiatry*, *24*, 533–542.

Triandis, H. C., Bontempo, R., Villareal, M. J., Asai, M., & Lucca, N. (1988). Individualism and collectivism: Cross-cultural perspectives on self-ingroup relationships. *Journal of Personality and Social Psychology*, *54*, 323–338.

Trognon, A. (1993). How does the process of interaction work when two interlocutors try to resolve a logical problem? *Cognition and Instruction*, *11*, 325–345.

Trouche, E., Sander, E., & Mercier, H. (2014). Arguments, more than confidence, explain the good performance of reasoning groups. *Journal of Experimental Psychology: General*, *143*, 1958–1971.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*, 293–315.

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology: I* (pp. 106–137). Harmandsworth, England: Penguin.