Abstraction and Cognitive Flexibility in Collective Problem Solving: The Role of Diversity

Tatiana Goregliad Fjaellingsdal (tatiana.goregliad1@uol.de)
Department of Neurology, University of Lübeck, 23562 Lübeck, Germany

Cordula Vesper (cvesper@cc.au.dk)
Department of Linguistics, Cognitive Science and Semiotics and the Interacting Minds Centre, Aarhus University, 8000 Aarhus C, Denmark

Christoffer L. Olesen (clolesen@post.au.dk)
Department of Linguistics, Cognitive Science and Semiotics, Aarhus University, 8000 Aarhus C, Denmark

Kristian Tylén (kristian@cc.au.dk)
Department of Linguistics, Cognitive Science and Semiotics and the Interacting Minds Centre, Aarhus University, 8000 Aarhus C, Denmark

Abstract

Groups of interacting individuals are often found to have an advantage over individuals in contexts of complex problem solving. We suggest that social interaction allows group members to share diverse introspections, perspectives and strategies, promoting the formation of more abstract problem representations, which – in turn – apply more flexibly to new problem contexts. In a reinforcement learning task inspired by the Wisconsin Card Sorting Test (WCST), participants categorized aliens as friendly or dangerous based on an underlying rule specifying feature combinations. After a number of correctly categorized trials, the rule would change (without explicit notification). Participants could solve the task by learning every new rule, but could also discover an underlying abstract rule, which would facilitate faster recovery from local rule changes. We compared pairs of participants individually trained on different rules (diversity pairs), with pairs trained on the same rule (non-diversity pairs), and individuals. We found that diversity pairs outperformed non- diverse pairs and individuals. Our findings suggest that diversity in prior experience benefits groups, likely due to processes of abstraction and cognitive flexibility.

Keywords: problem solving; diversity; social interaction; cognitive flexibility; learning

Introduction

A defining trait of the human species is our ability to solve complex problems (Newell & Simon, 1972). Problem solving is often portrayed as search in a multidimensional search space, and success is contingent on the problem solver's ability to overcome cognitive fixedness, backup-avoidance and other biases to locate an optimal solution (Condell et al., 2010; Duncker, 1945). A number of core cognitive processes have been suggested to support this ability including abstraction, analogy, cognitive flexibility, and conceptual transfer (Condell et al., 2010). Typically, these are considered aspects of the cognitive system of the individual problem- solver; however, a number of studies suggest that groups of interacting problem-solvers often outperform individuals (Bahrami et al., 2010; Dowsett & Burton, 2015; Heller, Keith, & Anderson, 1992), indicating that social interaction might play an important role.

One possible explanation for such group benefits is that groups are able to explore a larger area of the solution space – an effect referred to as the "wisdom of the crowd" (Yi, Steyvers, Lee, & Dry, 2012). However, such effects unlikely to depend on social interaction per se and can also be achieved post hoc by compiling individual contributions (Koriat, 2012; Weldon & Bellinger, 1997). In contrast, an- other account profiles the importance of social interaction (Bang et al., 2014; Voiklis & Corter, 2012): The dialogical exchange of different intuitions, perspectives, and strategies can motivate the formation of more abstract problem representations, that can help overcome cognitive fixedness, and transfer more flexibly to new situations (Schwartz, 1995).

Central to this account is the notion of cognitive diversity (Aggarwal, Williams Woolley, Chabris, & Malone, 2015; Hong & Page, 2004): If interacting individuals share redundant information, the cognitive implication and thus the effect of interaction is minimal. However, if interacting individuals contribute different and complementary information, it might enable them not only to broaden their search in a combinatorial, additive fashion, but also to discover solutions by flexibly combining information to form new and more abstract problem representations. Accordingly, the aim of the current study is to investigate the impact of cognitive diversity on processes of abstraction and flexibility in collective problem solving. Specifically, we test whether such effects can be experimentally induced by manipulating aspects of group members' previous experience with the problem.

Abstraction and Cognitive Flexibility

Individuals differ in their abilities to solve problems, categorize information, and adapt to an unstable environment. These differences become apparent when applying a common neuropsychological test to study cognitive functions, such as learning and flexibility: the Wisconsin Card Sorting Test (WCST; Grant & Berg, 1948; Heaton, Chelune, Talley, Kay, & Curtiss, 1993). In this test, participants are presented with a reinforcement learning task where they have to deduce a categorization rule for cards that differ in a number of features (color, number, and shape of the depicted elements). Whenever a rule is learned, which is indicated by reaching a fixed number of correctly categorized trials, the rule switches (without notification) and a new categorization rule must be learned from positive/negative feedback. In many clinical populations, the ability to deduce a rule, and to flexibly unlearn the rule again to deduce a new rule, is highly impaired (Feldstein et al., 1999).

One mechanism involved in solving the WCST is the ability to form abstract problem representations (Wohlwill, 1957). Abstraction refers to the ability to go beyond observable token phenomena to uncover their underlying organizing principles. This allows individuals to group different experiences as tokens of the same abstract type with the implication that they are better prepared to respond to new tokens even if they differ in surface appearance (Gentner, 1983; Loewenstein, Thompson, & Gentner, 1999).

In the WCST, an individual can learn that the presence of the feature ‘green’ means ‘correct’, but can moreover discover that this is part of a more abstract categorization rule which potentially links other features of the cards to correctness. Furthermore, the WCST is thought to measure cognitive flexibility: the ability to readily update beliefs about the world contingent on environmental evidence and adapt behaviors accordingly (Scott, 1962). This is often measured in so-called *perseveration errors* – the propensity to hinge on to the previous rule even when the feedback states it is now incorrect – indicating problems in detaching from the prior sorting rule.

Diversity in Social Interaction

It is often suggested that the success of collaborative interaction is contingent on the extent to which group members build common ground, i.e. share mutual knowledge and beliefs (Brennan & Hanna, 2009; Clark, 1996). Indeed, inter- acting individuals are observed to adapt to each other across modalities from bodily movement (Louwerse, Dale, Bard, & Jeuniaux, 2012), and visual attention (Richardson, Dale, & Kirkham, 2007), to lexicon, syntax, and situation models (Pickering & Garrod, 2004).

While interactive alignment has been consistently associated with the experience of affiliation and rapport (Marsh, Richardson, & Schmidt, 2009), findings are less clear for collective task performance. Some studies find alignment to be positively correlated with performance (Reitter & Moore, 2014), while in other studies, task performance seems related to the extent to which group members contribute complementary information (Fusaroli & Tylén, 2016). This suggests, that in some contexts of collective problem solving, group members benefit from their differences rather than similarities (Hong & Page, 2004).

Several recent studies support a positive impact of cognitive diversity on collective problem solving (Larson, 2007). The concept of diversity, however, is operationalized in different ways. Group members can differ in basic demo- graphics (age, gender, race, education, etc.), personality traits, cognitive style, skill and expertise, or prior knowledge. Whereas some of these factors are rather stable, inherent traits of the individual (e.g. personality: Asendorpf & Wilpers, 1998), others are more dynamic and continuously shaped by experience (e.g. perspectives and strategies). Since the various dimensions of diversity are often conflated, it can be hard to distinguish whether positive effects on problem solving originate from stable or dynamic factors. In this study, we set out to actively manipulate an element of diversity: participants’ prior experience with the particular problem, in order to assess the impact of this dynamic variable on collective problem solving.

The Present Study

The current pre-registered study aimed at directly testing the hypothesis that diversity in experience influences a group’s problem-solving performance. In particular, we predicted that diversity would promote more abstract problem representations and thus allow the discovery of the under- lying organizing principles governing a problem. To this end, we created a game-like categorization task inspired by the WCST. Participants were presented with a reinforcement learning task where they categorized extraterrestrial aliens as friendly or dangerous based on combinations of different binary features. During the game, the rules changed, making new combinations of features key to correct categorization. While in each individual block of the experiment, participants had to learn the specific combination of features that characterized an alien as friendly or dangerous, they could also discover a more general underlying rule that would facilitate fast recovery from rule changes.

The experimental design thus allowed us to investigate two main parameters reflecting participants’ capacities for abstraction and cognitive flexibility: First, the rate of learning new rules and recovering from rule shifts, indicated by categorization accuracy over the course of trials. Second, the extent to which participants’ errors would be indicative of perseveration, i.e. reproducing the response patterns compatible with a previous rule. Performance in the task was thus contingent on the extent to which participants could abstract from local rule implementations to discover the general underlying rule, and flexibly apply that abstract rule representation to new categorization problems.

We assessed these parameters in three different conditions. Participants performed the task either alone (*individual condition*), or together with a partner who had similar experience with the task (*non-diversity condition*), or together with a partner who had different experience with the task (*diversity condition*). We manipulated task experience in a training session preceding the experimental test session. Non-diversity pair members were trained individually on the same categorization rule before they collaborated in the test session. Diversity pairs were trained individually on different rules before collaborating in the test session.

We hypothesized that diversity pairs would outperform non-diversity pairs who, in turn, would outperform individuals. Specifically, we expected diversity pairs to show higher problem solving accuracy (H1), recover faster from rule changes (H2) and show fewer perseveration errors (H3), indicating increased abilities for abstraction and cognitive flexibility compared to non-diverse pairs and individuals. Moreover, we expected these condition-related effects to emerge from the interaction and thus not be predictable from individual training performance. All predictions and the analysis plan were preregistered at AsPredicted.org (http://aspredicted.org/blind.php?x=277qh8).

Methods

Participants

225 participants (111 females) with mean age 24.1 (SD = 5.16) were recruited from the participant database of Cognition and Behavior Lab, Aarhus University. Most were university students. All gave informed written consent in correspondence with the regulations of the local research ethical committee and received a fixed monetary reward (~$15) for their participation.

Materials and Design

In all parts of the experiment, participants solved the same reinforcement leaning task. They were presented with depictions of extraterrestrial aliens and had to categorize them as friendly or dangerous. The stimulus set consisted of digital images of 32 different aliens that differed on five binary features: color green or blue, arms up or down, eyes on stalks or not, legs slim or fat, spots or no spots (see figure 1a for examples). Each trial of the experiment presented a token alien and two buttons depicting a heart and a skull (see figure 1b). Participants were instructed to make a choice to save friendly aliens (by clicking the heart) and kill dangerous aliens (by clicking the skull). After each trial, they received feedback about the correctness of their choice and a running score that added a hundred points for correct decisions and subtracted hundred points for wrong decisions (with a lower bound of 0 points).

Figure 1: **a**: Stimulus examples. **b**: A token alien with button options as presented in a trial

The categorization rule could only be learned from trial and error over consecutive trials and was characterized by a particular combination of features that decided which aliens would qualify as friendly or as dangerous. For example, in a particular problem, dangerous aliens would have arms up and spots or slim legs. In another problem, dangerous aliens would be green and have arms down or eyes on stalks. Participants' task was to correctly categorize aliens by trial and error.

During the test phase, rule implementations changed several times. This meant that, in the context of the overall experiment, participants would have an advantage if they uncovered the underlying abstract rule behind the different rule implementations: dangerous = feature A and (feature B or feature C), with inclusive 'or'. All problems followed this scheme, and awareness of the abstract rule could thus facilitate faster recovery from rule implementation shifts.

Procedure

Upon entering the lab, participants were randomly assigned to one of three conditions: individuals, non-diversity pairs, and diversity pairs. In all conditions, they were seated at standard pc computers (22" screen) and completed the training session individually. The training session presented participants with the alien categorization game for 14 minutes after which it was automatically terminated. There were five different implementations of the abstract rule (i.e. correct alien feature combinations), counterbalanced between individuals and conditions. Importantly, non-diversity pair members were individually trained on the same rule implementation, while diversity pair members were trained on different problems before they entered the test session (see figure 2).

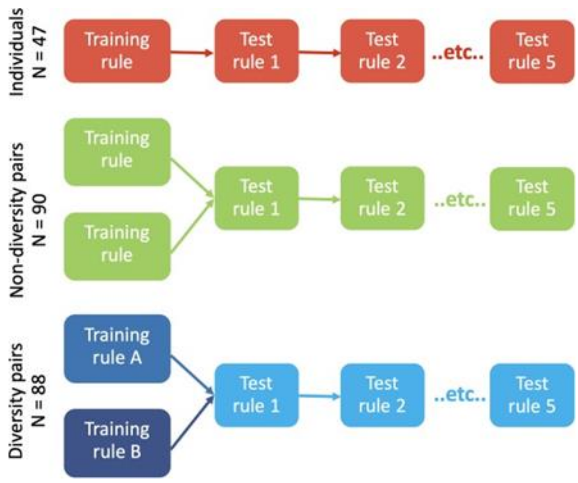

Figure 2: Schematic depiction of the study design.

In the test session, participants of the individual condition continued to work alone, whereas participants in the non- diversity and the diversity conditions relocated to one shared computer and could freely discuss among each other to make joint decisions during test sessions. The task in the test session was the same as during training but with a new rule implementation. As before, the problem proceeded over repeated trials until participants (individuals or pairs) had ten correct trials in a row or reached a timeout of 7 min. Then the problem changed again without any explicit notification. This pattern repeated for 21 minutes and a minimum of two rule shifts (if participants reached the timeout) or more if they learned the new rule implementation and got the necessary ten correct trials in a row.

Analysis

The full dataset consisted of 70084 decisions divided into the training set (40766 decisions) and the test set (29318 decisions). Since only 14 individuals, 2 non-diversity pairs and 9 diversity pairs got beyond rule block 6 within the 21 minutes of the test session, yielding the condition effects very uncertain for later rule blocks, we made a cutoff after rule block 6, leaving 27787 decisions for the analysis of test

performance. For analyses, we used a mixed effects logistic regression approach using the packages lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017) and tidyverse (Wickham et al., 2019) for RStudio (2019). For each analysis, we used model comparisons (relying on AIC, BIC) to identify the fixed effects combination accounting for most variance.

To test H1 (differences in test performance between conditions), we created a model with the binary accuracy score as outcome, and Condition and Trial (centered and scaled trial numbers) as fixed effects. We added random intercepts for participant/pair to account for repeated measures and random intercept and slope by condition for the specific rule implementation to account for item effects. A model including interaction with trial was superior to simpler models (condition only or main effect of condition + trial):

Accuracy ~ Condition * Trial +
(1|participant) + (1 + condition|rule)

To test H2, we made a model including an interaction term with Rule Block. This tests if rule changes had different effects (e.g. more severe disruption) in the three conditions:

Accuracy ~ Condition * Trial * Rule Block +
(1|participant) + (1 + condition|rule)

To test for condition-related effects of perseveration errors (H3), we derived a variable that coded those errors that would have been correct given the previous rule. Since it is very hard to calculate perseveration from training to the first test block in diversity pairs because pair members are trained on different rules, this analysis is only done from second rule change onwards. The model had the binary Perseveration Error variable as outcome, Condition, Trial and Rule Block as predictors. The random effects structure was identical to previous models. Model comparisons revealed the best model to include an interaction term between Condition and Trial, but only a main effect of Rule Block:

Perseveration ~ Condition * Trial + Rule Block +
(1|participant) + (1 + condition|rule)

Lastly, to control for potential confounding differences in training performance (i.e. that diversity pair members by chance performed better in the task from the start), we tested condition-related differences in training performance:

Accuracy ~ Condition * Trial +
(1 + trial|participant) + (1 + condition|rule)

Adding a random slope for trial per participant/pair to this model allowed us to extract individual slope coefficients from the training performance model and assess whether they predicted test performance. For pairs, we averaged the individual slopes. In this case, we report the full model despite the fact that a simpler model without Condition was superior in terms of AIC/BIC:

Accuracy ~ Training Slope * Condition +
(1|participant) + (1 + condition|rule)

Results

Condition significantly predicted test performance (see figure 3a): Diversity pairs performed significantly better than non-diversity pairs, $\beta = 0.13$ (*SE* = 0.05), $z = 2.42$, $p = .01$, *odds ratio* = 1.14, and individuals performed significantly worse than non-diversity pairs, $\beta = -0.17$ (*SE* = 0.04), $z = -4.40$, $p < .001$, *odds ratio* = 0.85. Adding an interaction term for rule block did not improve the model ($\chi(6,19) = 8.59$, $p = .2$), suggesting that the condition effects hold across rule changes (as evident from figure 3b).

Number of perseveration errors decreased at a faster rate over trials in the social than the individual condition, e.g. individuals had a larger number of perseveration errors than non-diversity pairs, $\beta = 0.17$ (*SE* = 0.05), $z = 3.20$, $p < .01$. However, while diversity pairs had fewer perseveration errors than non-diversity pairs, this effect was not significant, $\beta = -0.06$ (*SE* = 0.07), $z = -0.80$, $p = .42$ (see figure 3c).

There were no condition-related differences in training performance, $\chi(4,13) = 4.41$, $p = .35$. However, there were differences in the extent to which training learning slopes were predictive of test performance. This effect was smaller for diversity than for non-diversity pairs, $\beta = -0.34$ (*SE* = 0.13), $z = -2.54$, $p = .01$, *odds ratio* = 0.71. Likewise, the effect was smaller for individuals than for non-diversity pairs, $\beta = -0.19$ (*SE* = 0.10), $z = -1.98$, $p = .047$, *odds ratio* = 0.82 (see figure 3d).

Discussion

As predicted by H1, diverse pairs (trained individually on different rules) outperformed non-diverse pairs (trained on the same rules), and individuals. This advantage of diversity in training persisted over rule changes. While we had predicted the rule blocks to affect the three conditions at different severity (i.e. an interaction between Condition

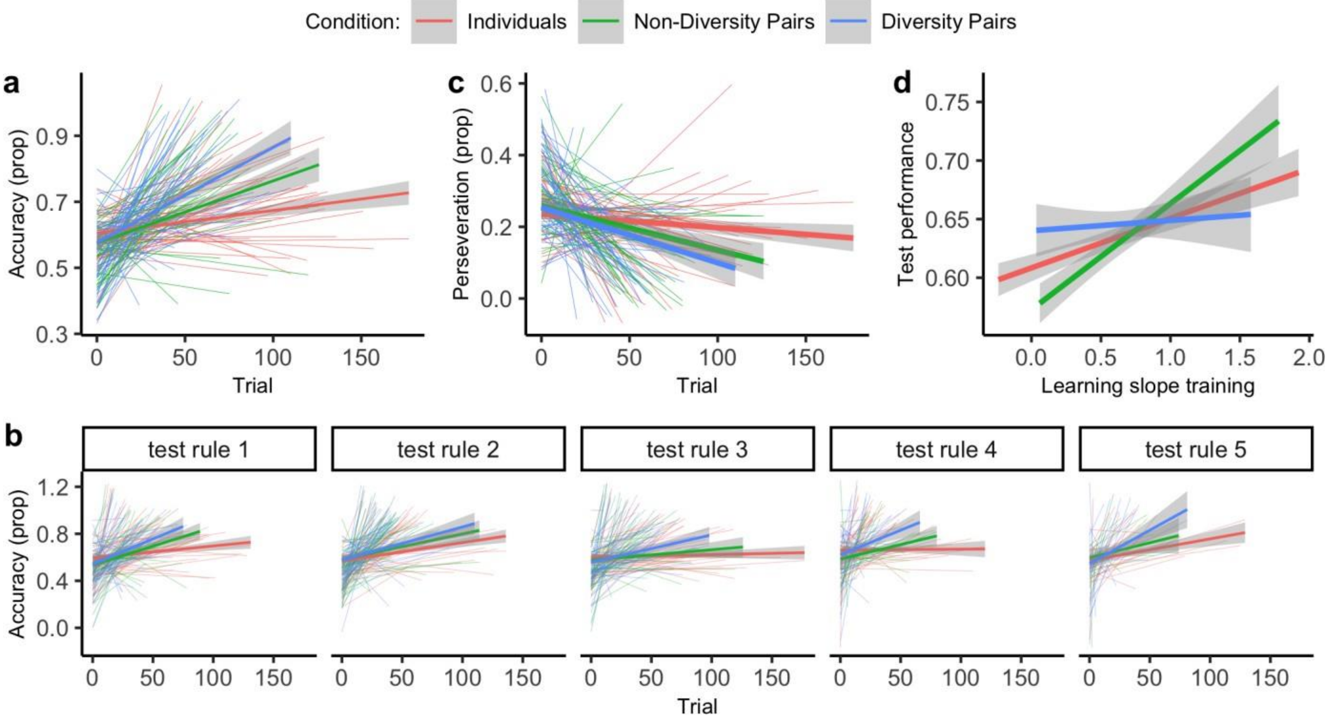

Figure 3: Test session performance. **a**: performance over trials by condition across the test phase (all rule blocks). **b**: performance over trials by condition and rule block. **c**: Perseveration errors over trials by condition. **d**: The extent to which training performance is predictive of test performance in the three conditions. Notice that in the visualizations a-c, the effects are represented as the average probability of a correct answer (binary variable) as a function of trial. Bold lines depict condition level effects, thin lines show individual level effects (individual or pair).

and Rule Block), the differences are stable through the test phase. However, we argue that the fact that diversity pairs sustain their performance benefit relative to the other conditions through repeated rule changes can be regarded support for H2. Moreover, groups were found to recover faster from rule changes than individuals, as indicated by a lower number of perseveration errors. However, while diversity pairs had numerically fewer perseveration errors than non-diversity pairs, this effect was not significant (no support for H3).

What is the basis of this diversity advantage? We suggest that diversity group members benefited from verbal inter- actions (Frith, 2012; Tylén, Weed, Wallentin, Roepstorff, & Frith, 2010; Voiklis & Corter, 2012): By sharing introspections from their training experiences, diversity pairs could come to awareness of higher order similarities between their training rules, which in turn could motivate the intuition that new rules followed the same abstract scheme (Schwartz, 1995). This likely enabled diversity pairs to conduct a more systematic and directed search when encountering new rules, since they could exclude combinations not corresponding to the underlying rule. Interestingly, these effects were obtained with as little as fourteen minutes of individual training, pointing to the flexible and dynamic nature of cognitive diversity.

A potential far-reaching implication is that positive diversity effects can be obtained with quite minimal means (e.g. in terms of education; cf. Canham, Wiley, & Mayer, 2012) and do not depend alone on differences in more stable bio-demographic factors such as personality traits, cognitive style, etc. It is, however, a question for future research whether similar effects are achievable in contexts of more complex problems. Importantly, the performance differences between conditions cannot be reduced to individual pre-test competences, as there were no baseline differences between participants assigned to different conditions in the training phase. Interestingly, while training performance (learning slopes) was generally predictive of test performance, this effect was driven by the non-diversity pairs, and to some extent individuals, while there was no effect for diversity pairs (see figure 3d). This suggests that the benefit of diversity pairs is emergent from the interaction, supporting the idea that individual perspectives are integrated in more abstract problem representations.

A concern could be raised that groups perform better simply because there are working memory advantages of being two individuals to keep track of relevant information. This is indeed a potential explanation for the performance advantage of groups compared to individuals (Andersson & Rönnberg, 1996); however, it cannot account for the diversity effects.

Another concern could be that diversity pairs' performance benefit originates from other factors such as differences in personality or motivation. Although not reported here due to lack of space, these factors were also monitored and controlled for.

In the present study, we have shown that groups can profit from members' diversity in experience to address a problem. We suggest that social interaction among group members with diverse experiences promotes a more abstract problem representation, enabling the group to discover underlying principles of the problem beyond its surface implementation. Moreover, such problem abstraction enhance cognitive flexibility: the ability to transport and apply acquired knowledge across contexts.

Our findings have potential far-reaching implications and applications across a host of contexts, such as work, research, and education, where people can excel by integrating their diverse experiences through dialogical social interaction.

Acknowledgments

This research was funded by an Interacting Minds Centre Seed Grant.

References

Aggarwal, I., Williams Woolley, A., Chabris, C. F., & Malone, T. W. (2015). Cognitive diversity, collective intelligence, and learning in teams. *Proceedings of Collective Intelligence*.

Andersson, J., & Rönnberg, J. (1996). Collaboration and memory: Effects of dyadic retrieval on different memory tasks. *Applied Cognitive Psychology*, *10*(2), 171–181. https://doi.org/10.1002/(SICI)1099-0720(199604)10:2<171::AID-ACP385>3.0.CO;2-D

Asendorpf, J. B., & Wilpers, S. (1998). Personality effects on social relationships. *Journal of Personality and Social Psychology*, *74*(6), 1531–1544. https://doi.org/10.1037/0022-3514.74.6.1531

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally interacting minds. *Science*, *329*(5995), 1081–1085. https://doi.org/10.1126/science.1185718.Optimally

Bang, D., Fusaroli, R., Tylén, K., Olsen, K., Latham, P. E., Lau, J. Y. F., … Bahrami, B. (2014). Does interaction matter? Testing whether a confidence heuristic can replace interaction in collective decision-making. *Consciousness and Cognition*, *26*, 13–23. https://doi.org/10.1016/j.concog.2014.02.002

Brennan, S. E., & Hanna, J. E. (2009). Partner-Specific Adaptation in Dialog. *Topics in Cognitive Science*, *1*(2), 274–291. https://doi.org/10.1111/j.1756-8765.2009.01019.x

Canham, M. S., Wiley, J., & Mayer, R. E. (2012). When Diversity in Training Improves Dyadic Problem Solving. *Applied Cognitive Psychology*, *26*(3), 421–430. https://doi.org/10.1002/acp.1844

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511620539

Condell, J., Wade, J., Galway, L., McBride, M., Gormley, P., Brennan, J., & Somasundram, T. (2010). Problem solving techniques in cognitive science. *Artificial Intelligence Review*, *34*(3), 221–234. https://doi.org/10.1007/s10462-010-9171-0

Dowsett, A. J., & Burton, A. M. (2015). Unfamiliar face matching: Pairs out-perform individuals and provide a route to training. *British Journal of Psychology*, *106*(3), 433–445. https://doi.org/10.1111/bjop.12103

Duncker, K. (1945). The structure and dynamics of porblem-solving processes. *Psychological Monographs*, *58*(5), 1–112.

Feldstein, S. N., Keller, F. R., Portman, R. E., Durham, R. L., Klebe, K. J., & Davis, H. P. (1999). A comparison of computerized and standard versions of the Wisconsin Card Sorting Test. *Clinical Neuropsychologist*, *13*(3), 303–313. https://doi.org/10.1076/clin.13.3.303.1744

Frith, C. D. (2012). The role of metacognition in human social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 2213–2223. https://doi.org/10.1098/rstb.2012.0123

Fusaroli, R., & Tylén, K. (2016). Investigating Conversational Dynamics: Interactive Alignment, Interpersonal Synergy, and Collective Task Performance. *Cognitive Science*, *40*(1), 145–171. https://doi.org/10.1111/cogs.12251

Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science*, *7*(2), 155–170. https://doi.org/10.1016/B978-1-4832-1446-7.50026-1

Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology*, *38*(4), 404–411. https://doi.org/https://doi.org/10.1037/h0059831

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). Wisconsin Card Sorting Test (WCST) manual-revised and extended. *Odessa, FL: Professional Assessment Resources*.

Heller, P., Keith, R., & Anderson, S. (1992). Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving. *American Journal of Physics*, *60*(7), 627–636. https://doi.org/10.1119/1.17117

Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*, *101*(46), 16385–16389. https://doi.org/10.1073/pnas.0403723101

Koriat, A. (2012). When are two heads better than one and why? *Science*, *336*, 360–362. https://doi.org/10.1126/science.1216549

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Larson, J. R. (2007). Deep diversity and strong synergy: Modeling the impact of variability in members' problem-solving strategies on group problem-solving performance. *Small Group Research*, *38*(3), 413–436. https://doi.org/10.1177/1046496407301972

Loewenstein, J., Thompson, L., & Gentner, D. (1999). Analogical encoding facilitates knowledge transfer in negotiation. *Psychonomic Bulletin and Review*, *6*(4), 586–597. https://doi.org/10.3758/BF03212967

Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P.

(2012). Behavior matching in multimodal communication is synchronized. *Cognitive Science*, *36*(8), 1404–1426. https://doi.org/10.1111/j.1551-6709.2012.01269.x

Marsh, K. L., Richardson, M. J., & Schmidt, R. C. (2009). Social connection through joint action and interpersonal coordination. *Topics in Cognitive Science*, *1*(2), 320–339. https://doi.org/10.1111/j.1756-8765.2009.01022.x

Newell, A., & Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice Hall.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190. https://doi.org/10.1017/S0140525X04000056

R Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Reitter, D., & Moore, J. D. (2014). Alignment and task success in spoken dialogue. *Journal of Memory and Language*, *76*, 29–46. https://doi.org/10.1016/j.jml.2014.05.008

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The Art of Conversation Is Coordination. *Psychological Science*, *18*(5), 407–413. https://doi.org/10.1111/j.1467-9280.2007.01914.x

Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. *Journal of the Learning Sciences*, *4*(3), 321–354. https://doi.org/10.1017/CBO9781107415324.004

Scott, W. A. (1962). Cognitive complexity and cognitive Flexibility. *Sociometry*, *25*(4), 405–414. https://doi.org/10.2307/2785779

Tylén, K., Weed, E., Wallentin, M., Roepstorff, A., & Frith, C. D. (2010). Language as a tool for interacting minds. *Mind and Language*, *25*(1), 3–29. https://doi.org/10.1111/j.1468-0017.2009.01379.x

Voiklis, J., & Corter, J. E. (2012). Conventional Wisdom: Negotiating Conventions of Reference Enhances Category Learning. *Cognitive Science*, *36*(4), 607–634. https://doi.org/10.1111/j.1551-6709.2011.01230.x

Weldon, M. S., & Bellinger, K. D. (1997). Collective memory: Collaborative and individual processes in remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(5), 1160–1175. https://doi.org/10.1037/0278-7393.23.5.1160

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wohlwill, J. F. (1957). The abstraction and conceptualization of form, color, and number. *Journal of Experimental Psychology*, *53*(5), 304–309. https://doi.org/10.1037/h0045318

Yi, S. K. M., Steyvers, M., Lee, M. D., & Dry, M. J. (2012). The Wisdom of the Crowd in Combinatorial Problems. *Cognitive Science*, *36*(3), 452–470. https://doi.org/10.1111/j.1551-6709.2011.01223.x