Title

Does collaboration constrain insight problem-solving?

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Pavliuchik, Elena

Knoblich, Günther

Publication Date

2026

Does Collaboration Constrain Insight Problem-Solving?

Elena Pavliuchik (pavliuchik_elena@phd.ceu.edu)[1] & Günther Knoblich[1]
[1]Department of Cognitive Science, Central European University

Abstract

This study examined whether collaboration helps or constrains insight problem solving using the Mutilated Checkerboard Problem, a classic task requiring representational change. One hundred participants solved the problem either individually, using think-aloud protocols (n = 50), or in dyads, using free conversation (n = 25 dyads). Dyads did not outperform individuals in solution rate or solution time, and solved the problem significantly less often than nominal dyads generated from individual performance. In unsolved cases, dyads also produced substantially more covering attempts than individuals, consistent with the possibility that partners became jointly fixated on a misleading strategy. Verbal-protocol analyses showed that successful solving was associated with producing more solution-relevant invariants and referring less often to invariant dimensions that were not relevant to the solution. Overall, collaboration appeared to constrain rather than enhance insight problem solving, suggesting that interaction may sometimes stabilize a shared but unproductive problem representation.

Keywords: problem-solving; collaboration; conversation; mental fixation; representational change

Introduction

Some problems in life call for non-obvious answers. They are difficult because our initial representation of the problem often does not reveal the solution. To reach the goal, we need to change how we represent the problem. This process is commonly referred to as insight (Kaplan & Simon, 1990; Ohlsson, 1992). Many creative ideas and scientific discoveries share this characteristic: pursuing the problem from the same perspective will lead nowhere; restructuring is needed. Importantly, people often face such problems together. Teams in engineering, science, art, and other fields regularly collaborate on challenges that demand representational change. This raises a central question: can people working together solve problems more successfully than individuals working alone when the solution requires insight?

Studying Representational Change with the Mutilated Checkerboard Problem

To study how people collectively solve problems that require representational change, we will use the Mutilated Checkerboard Problem. In this task, participants must determine whether it is possible to cover the remaining 62 squares with 31 dominoes after two diagonal corner squares have been removed (Figure 1).

This problem is well-suited to our investigation because it presents an initially straightforward and misleading problem space. At first glance, it seems possible to cover the 62 squares with 31 dominoes. However, the problem space is vast, with many possible combinations of coverings. Repeated attempts will therefore inevitably lead participants to an impasse: the coverings will always fail. To solve the problem, participants must reject the covering strategy and shift to a new representation that incorporates the alternating colors of the squares. Recognizing color parity means understanding that two white squares have been removed, leaving only 30 white squares and 32 black squares on the board. This immediately leads to the conclusion that the task is impossible: two black squares will always remain uncovered. This shift represents a classic impasse-to-insight sequence (Ohlsson, 1984).

Kaplan and Simon (1990) proposed that, to find a new problem space, people can use helpful constraining heuristics. One such heuristic is identifying the invariants of the problem: properties that remain unchanged during problem-solving. For example, in the case of the Mutilated Checkerboard Problem, one such invariant is that both removed squares are of the same color. Additionally, participants may notice that the squares they fail to cover are always the opposite color of the removed squares. These invariants can help the solver to find a new problem space and lead to a solution. Kaplan and Simon also distinguished between relevant invariants (those that directly lead to the solution) and irrelevant invariants (those that do not contribute to this direct path).

Building on this work, we compared how people solve this problem in dyads versus individually, and also analyzed the number of relevant and irrelevant invariants identified during the problem-solving process.

Why Can Collaboration Foster Insight?

When people solve a problem together, they most often do so through conversation. Conversation can be viewed as a

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

socially distributed system: speakers dynamically coordinate in the specific context and complement each other to accomplish situational goals (Fusaroli et al., 2014). This interaction typically involves alignment, providing a shared structure that supports coordination while also allowing space for task-relevant semantic variability (Mills, 2014; Fusaroli & Tylén, 2015; Tylén et al., 2023). This variability can potentially bring diverse perspectives and interpretations into the problem-solving process.

For example, Schwartz (1995) demonstrated that dyads working on complex problems produced more abstract graphs than individuals. To communicate effectively, dyads developed a shared understanding of the problem that helped them coordinate their different perspectives. Because this representation bridged multiple viewpoints, it tended to be more abstract.

Even more relevant to our study, Voiklis and Corter (2012) demonstrated that dialogue enhances categorization. By comparing silent individual work, think-aloud monologues, and collaborative dialogue, they found that pairs engaged in dialogue outperformed individuals working in silence or thinking aloud. The authors attributed this advantage to the process of negotiating references with another person. In dialogue, participants had to clarify what they meant, which encouraged them to verbalize more relationships between features of the objects being categorized. The think-aloud condition lacked this social dynamic: participants verbalized their thoughts, but did not receive any cues from another person that could prompt further specification. In this way, dialogue helped pairs identify and express more diagnostic features.

Applying this logic to our study, we can predict that collaboration would make it easier to identify the invariant in the Mutilated Checkerboard Problem and, therefore, to solve the problem successfully.

Why Can Collaboration Hinder Insight?

The alternative scenario is derived from the mental fixation literature, which is well-documented in individual problem-solving (Smith & Blankenship, 1991; Smith, 1995). Mental fixation occurs when, while solving a task, a person clings to a strategy or draws upon previous experience, even if it is misleading for the current situation. For example, fixation can lead individuals to repeatedly use the same strategy, even when it makes their performance less effective (Luchins, 1942).

Kohn and Smith (2011) observed this effect in a brainstorming task, where participants, exposed to increasing numbers of ideas, conformed more to the suggested categories, producing fewer novel ideas. Similarly, Bilalić and colleagues (2008) demonstrated how the presence of a familiar but suboptimal solution can interfere with finding the best alternative. In their experiments, experienced chess players were presented with a task that had two possible solutions: one more familiar (five moves) and the other more optimal (three moves). As predicted by the mental fixation hypothesis, players tended to gravitate toward the more familiar, less effective option. However, when the familiar solution was absent, players easily identified the optimal solution. This illustrates the role of mental fixation, where initial familiarity with a suboptimal strategy limits the ability to shift to a more effective one (Bilalic et al., 2014).

Conversation itself can therefore become a source of fixation when dyad members adopt the same strategy and continue to rely on it. To maintain coherence, interlocutors adapt to each other's language and conceptual framing, gradually building a shared understanding of the task (Clark & Schaefer, 1989; Pickering & Garrod, 2009). While this shared understanding can support coordination, it may also constrain exploration if the dyad settles on a misleading interpretation of the problem. In addition, collaboration can be hindered by coordinational costs, such as conflicting goals, turn-taking demands, or difficulties integrating information across partners (Cronin & Weingart, 2007). Applying this logic to our study, dyads may become jointly fixated on an unproductive strategy. Rather than helping them identify the relevant invariant, conversation may lead partners to reinforce each other's misleading assumptions and remain within a shared but ultimately unhelpful problem space.

Experiment

We hypothesized that pairs working together may differ from individuals in terms of success rate, solution time, number of covering attempts, and the number of relevant and irrelevant invariants produced. These differences could go in two opposing directions. On one hand, real-time interaction may promote variability in strategies and the invariants noticed during problem-solving. If this is the case, joint pairs should solve the Mutilated Checkerboard Problem more frequently, solve it faster, make fewer covering attempts, and identify more relevant invariants. On the other hand, from a mental fixation perspective, conversation may cause pairs to fixate on the covering strategy. In this case, we would predict that joint pairs would perform worse than individuals, solving the problem less often, making more covering attempts, and identifying fewer relevant invariants.

Participants A total of 100 participants were recruited through the SONA system at Central European University and received €10 for participation. The study took place in the lab. Participants were randomly assigned to one of two conditions: 50 participants were paired into dyads (i.e., 25 dyads; gender: 34 female, 15 male, 1 declined to answer; age: M = 25.80, SD = 4.20), and 50 participants solved the problem individually (gender: 22 female, 27 male, 1 unspecified; age: M = 27.34, SD = 5.71). Participants did not know each other beforehand and were all fluent English speakers. As the study was exploratory and reliable effect size estimates were unavailable at the time of design, an a priori power analysis was not conducted. No participants were excluded from the analyses.

Procedure Participants first read written instructions for the Mutilated Checkerboard Problem (Figure 1), which were printed on paper together with the checkerboard figure with crossed squares. They were provided with multiple copies of this board and could draw on them freely while working. We collected all sheets to record participants' covering attempts. In the dyad condition, participants were instructed to solve the problem jointly and were told they could talk freely throughout the task. A maximum of 30 minutes was allotted, but participants could submit an answer at any time. If the submitted answer was incorrect, the experimenter asked them to continue. Participants were encouraged to communicate openly during problem solving, and both their conversation and their coverings were video-recorded. In the individual condition, participants solved the problem alone and were instructed to think aloud throughout. Apart from the absence of an interaction partner, the instructions, materials, time limit, and feedback procedure were identical across conditions.

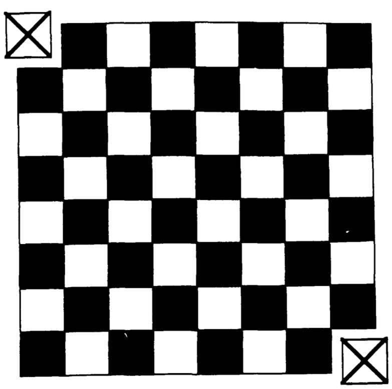

Figure 1: The Mutilated Checkerboard Problem asks whether it's possible to cover the remaining 62 squares (after removing two) with 31 dominoes, where each domino covers two squares. The answer is that it is indeed impossible to cover all the squares, as each domino must cover one black and one white square. With both removed squares being white, 30 white squares and 32 black squares remain, making the task unsolvable.

Results

The data were analyzed using RStudio Version 2025.5.1.513 (RStudio Team, 2025). To transcribe verbal protocols from video into text with timecodes, we used AmberScript.

Solution rates We compared solution accuracy between individuals and interacting dyads (Figure 2). Individuals solved the problem at a rate of 0.20 (binomial SE = 0.06), and interacting dyads solved the problem at a similar rate of 0.24 (binomial SE = 0.09). A logistic regression predicting solution success (0/1) from condition (individual vs. dyad) indicated no reliable difference between conditions, OR = 1.26, 95% CI [0.38, 3.93], $z = 0.40$, $p = .69$.

To benchmark dyad performance against what would be expected if two individuals worked independently, we constructed a nominal-dyad baseline via simulation. In each of 10,000 iterations, the 50 individual participants were randomly shuffled and paired without replacement into 25 nominal dyads. A nominal dyad was coded as successful if at least one of its two members had solved the problem individually. This procedure generated a simulated distribution of expected nominal-dyad solution rates under the assumption that two independent individual solvers contribute their individual probabilities of success. The mean expected nominal-dyad solution rate was 0.36 (binomial SE = 0.10). We then compared the observed real-dyad solution rate of 0.24 to this simulated distribution using a two-sided Monte Carlo randomization test. The p-value was calculated as the proportion of simulated nominal-dyad solution rates that deviated from the simulated mean by at least as much as the observed real-dyad solution rate did. Real dyads solved the problem significantly less often than simulated nominal dyads, indicating that interacting pairs underperformed relative to a baseline constructed from individual performance alone, ($p < .001$).

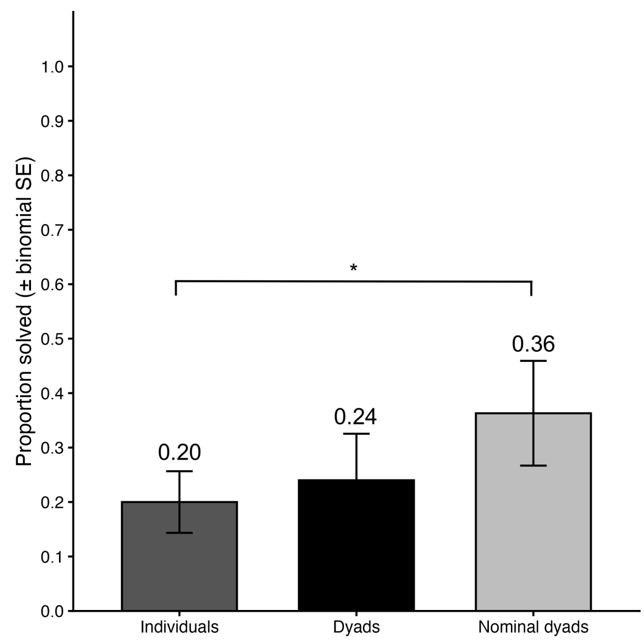

Figure 2: Proportion of solutions for individuals, dyads and nominal dyads. Error bars indicate binomial standard errors of the proportions; for nominal dyads, the mean expected solve rate is shown with binomial SE based on n = 25 dyads per iteration.

Solution times Conditioning on solved trials only, individuals ($n = 10$) solved the problem in a mean of 25.01 min ($SD = 8.20$), whereas interacting dyads ($n = 6$) solved in a mean of 13.04 min (SD = 10.37). A Wilcoxon rank-sum test comparing solution times between individuals and dyads indicated no reliable difference, $W = 45$, $p = .10$ (two-sided), although descriptively dyads showed shorter solution times among solvers.

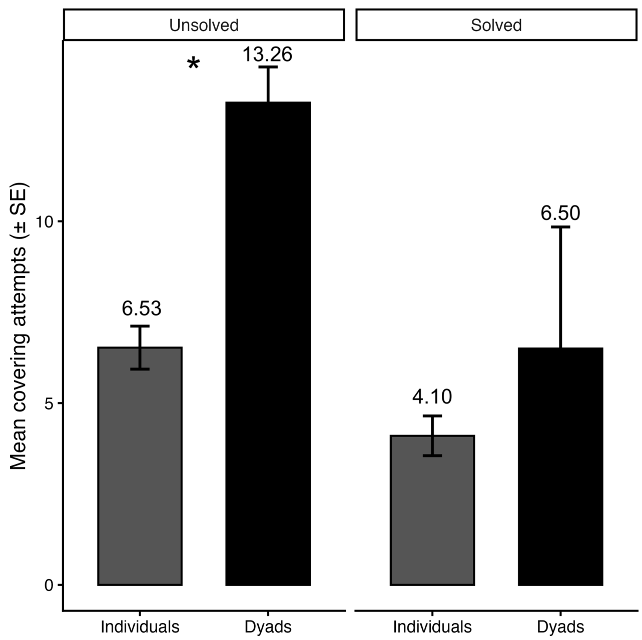

Figure 3: Mean covering attempts for individuals and dyads in solved and unsolved conditions. Error bars indicate standard errors of the mean.

Covering Attempts We analysed the number of covering attempts using a negative binomial GLM with predictors of Group (individuals vs. dyads), Solved (solved vs. unsolved), and their interaction. A covering attempt was operationalised as each instance in which a participant used a new sheet of paper with the printed mutilated board to cover the board, regardless of whether the sheet was marked fully or only partially; participants decided for themselves when to initiate another attempt. Descriptively, mean attempts were 6.53 (SE = 0.59) for unsolved individuals, 4.10 (SE = 0.55) for solved individuals, 13.26 (SE = 0.98) for unsolved dyads, and 6.50 (SE = 3.34) for solved dyads (Figure 3). Dyads produced significantly more attempts than individuals in the unsolved condition (rate ratio = 2.03, 95% CI [1.55, 2.67], $p < .0001$), whereas the dyad-individual difference among solved cases was not significant (rate ratio = 1.58, $p = .122$).

Invariant Analysis Verbal protocols were coded at the level of utterances. Before analysing invariant detection, we first assessed whether the protocols contained sufficient verbal material. Word counts were summarized at the case level, and all cases were retained. Dyads produced on average 1,445 words per case (SD = 1,058), whereas individuals produced on average 1,237 words per case (SD = 599). Within dyads, speech production was comparable across partners: Speaker 1 produced on average 740 words (SD = 534), and Speaker 2 produced on average 706 words (SD = 592). The lowest-production dyad and individual case contained 336 and 309 words, respectively.

We then examined how individuals and dyads detected invariants, that is, problem properties that remain unchanged during problem solving. Following the invariant-analysis procedure introduced by Kaplan and Simon (1990), we coded verbal protocols for mentions of invariant problem properties and classified them as either solution-relevant or non-solution-relevant. Solution-relevant invariants were directly connected to the solution path and included mentions of color parity, same-color removal, color imbalance, and the conclusion that complete covering was impossible. Non-solution-relevant invariants reflected stable or repeated features of the task that did not by themselves imply the solution, such as odd-even counts of squares or board symmetries.

To assess the reliability of the classification scheme, a second coder independently coded a blinded sample of 359 invariant-related utterances, corresponding to approximately half of the full set of 720 invariant-related utterances identified in the verbal protocols. The coding scheme included 13 predefined fine-grained categories, which were also collapsed into the theoretically central solution-relevant versus non-solution-relevant distinction. Agreement on this broader distinction was 86.9%, Cohen's $\kappa = .74$. Agreement on the fine-grained categories was 73.8%, Cohen's $\kappa = .71$. Disagreements were resolved through discussion, and consensus codes were used in the final analyses.

To test whether invariant production varied by session type, relevance, and problem-solving success, we analysed invariant counts using a Poisson generalized linear mixed-effects model with a log link. The model included session type, relevance, problem-solving success, and all interactions as fixed effects, as well as a random intercept for case ID. The analysis revealed a robust interaction between relevance and problem-solving success. Compared with unsolved cases, solved cases showed a stronger distinction between relevant and irrelevant invariants: irrelevant invariants were reduced relative to relevant ones, $b = -0.95$, $SE = 0.26$, $z = -3.67$, $p < .001$. There was no evidence that this relevance-by-success pattern differed reliably between dyads and individuals, $b = -0.41$, $SE = 0.44$, $z = -0.93$, $p = .35$ (Figure 4).

Descriptively, individuals in solved cases produced on average 6.90 relevant invariants (SE = 0.57) and 2.40 irrelevant invariants (SE = 0.65), whereas individuals in unsolved cases produced 4.58 relevant invariants (SE = 0.45) and 4.15 irrelevant invariants (SE = 0.54). A similar pattern was observed for dyads. Solved dyads produced on average 7.33 relevant invariants (SE = 0.72) and 1.83 irrelevant invariants (SE = 0.91), whereas unsolved dyads produced 5.84 relevant invariants (SE = 0.61) and 5.74 irrelevant invariants (SE = 0.89).

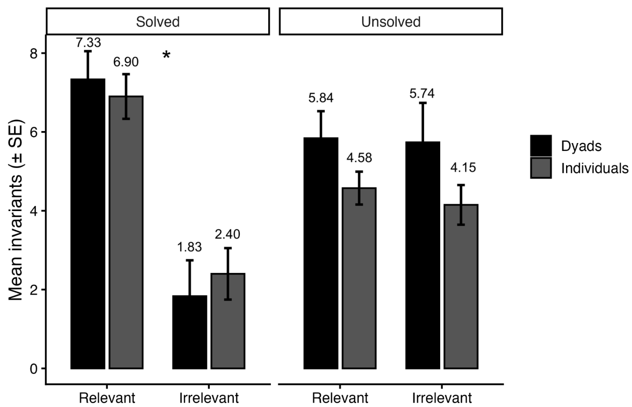

Figure 4: Mean invariant counts for individuals and dyads, separated by relevant versus irrelevant invariants and by solved versus unsolved conditions. Error bars indicate standard errors of the mean.

To complement the count analysis, we also examined the diversity of invariant categories mentioned in each case. Category diversity was defined as the number of distinct invariant types produced within each relevance category. This allowed us to assess not only how often participants mentioned invariants, but also how broadly they explored different invariant dimensions.

We analysed category diversity using a Poisson generalized linear mixed-effects model with session type, problem-solving success, relevance, and their interactions as predictors, as well as a random intercept for case ID. The random intercept variance was estimated as zero, indicating that the model was effectively equivalent to a Poisson GLM. The analysis again revealed an interaction between relevance and problem-solving success. Compared with unsolved cases, solved cases showed a stronger distinction between relevant and irrelevant invariant-type diversity: irrelevant diversity was reduced relative to relevant diversity, $b = -0.72$, $SE = 0.32$, $z = -2.26$, $p = .02$. There was no evidence that this relevance-by-success pattern differed reliably between dyads and individuals, $b = -0.64$, $SE = 0.58$, $z = -1.10$, $p = .27$.

Descriptively, solved individuals mentioned 4.30 distinct relevant invariant types (SE = 0.37) and 1.80 distinct irrelevant types (SE = 0.49), whereas unsolved individuals mentioned 2.75 relevant types (SE = 0.22) and 2.36 irrelevant types (SE = 0.22). A similar pattern was observed for dyads. Solved dyads mentioned 4.33 distinct relevant types (SE = 0.49) and 1.00 irrelevant types (SE = 0.52), whereas unsolved dyads mentioned 2.73 relevant types (SE = 0.22) and 2.45 irrelevant types (SE = 0.30).

Together, the count and category-diversity analyses suggest that successful restructuring was associated with producing more relevant invariants. It was also associated with fewer references to irrelevant invariant dimensions.

Discussion

The goal of this study was to compare individuals and interacting dyads in insight problem solving, a process that requires a qualitative shift in how the problem is represented (Ohlsson, 1992). We assessed performance across solution rate, solution time, covering attempts, and the production of solution-relevant and non-solution-relevant invariants in verbal protocols.

Across measures, dyads did not show an advantage over individuals. Solution rates were comparable for individuals and real dyads, and logistic regression indicated no reliable difference between these conditions. However, real dyads performed markedly worse than nominal dyads constructed from individual performance. The nominal-dyad analysis showed that real dyads solved the problem significantly less often than would be expected from pooling two independent individual solvers. This suggests that simply placing two people in interaction did not improve insight problem solving relative to the baseline expected from independent individual performance.

This pattern aligns most closely with the mental fixation account outlined in the introduction. Interaction may have led partners to remain within a shared problem space rather than increasing the chance that at least one person explored an alternative representation. In other words, conversation may support the construction of common ground, but this shared understanding is not necessarily beneficial if partners converge on an unproductive interpretation of the task.

The covering-attempt data converged with this interpretation. In unsolved cases, dyads produced substantially more covering attempts than individuals. Unsolved individuals produced on average 6.53 attempts, whereas unsolved dyads produced 13.26 attempts. One interpretation is that unsuccessful dyads behaved like two solvers working in parallel, generating more attempts without effectively integrating information in a way that increased the likelihood of insight. A complementary interpretation is that dyads became jointly committed to an unproductive covering strategy. This would be consistent with classic demonstrations of misleading strategy adherence and perseveration in problem solving. In such cases, solvers continue to rely on familiar or initially plausible approaches, even when these approaches prevent discovery of a better solution (Luchins, 1942; Bilalić et al., 2008; Kohn & Smith, 2011).

The invariant analyses provide further insight into the representational processes underlying successful problem solving. Following Kaplan and Simon (1990), we examined whether participants detected invariant properties of the problem and whether these invariants were connected to the solution path. Successful cases were not only associated with the production of relevant invariants, but also with a reduction in irrelevant invariants. This suggests that successful restructuring may involve more than discovering the critical color-parity relation. It may also require disengaging from stable but misleading features of the problem. This extends Kaplan and Simon's emphasis on invariant discovery by suggesting that the reduction of irrelevant invariants may be also informative about successful restructuring. Importantly, however, we did not find evidence that dyads and individuals differed in the overall number of invariants they detected.

An important limitation concerns the use of think-aloud protocols in the individual condition. Because insight problem solving can be affected by verbalization, this design falls within the broader debate on verbal overshadowing. However, this literature has produced conflicting results, with some work suggesting that verbalization can hinder insight problem solving and other work questioning whether verbal protocols necessarily impair performance (e.g., Ball et al., 2015; Fleck & Weisberg, 2004). In the present study, think-aloud protocols were necessary for comparing the content of individual and dyadic problem solving, but they do not provide a silent individual baseline. A strong next methodological step would therefore be to compare silent individuals, think-aloud individuals, and interacting dyads within the same design.

Another limitation is that the proposed alignment-based mechanism remains inferential. The present study did not directly measure lexical convergence and semantic similarity. Therefore, although the results are consistent with the idea that dyads may become jointly fixated on a misleading problem representation, they do not directly show that this occurred through conversational alignment.

Finally, future work should test whether similar collaborative constraints emerge in other insight tasks and in more open-ended creative problems.

The present findings suggest that collaboration does not automatically facilitate insight. In tasks that require representational change, interaction may even constrain problem solving when partners stabilize a shared but misleading interpretation of the problem. Understanding when conversation broadens the search for alternative representations, and when it instead reinforces fixation, remains an important direction for research on collaborative insight.

Acknowledgments

A generative AI tool was used to support language editing, wording alternatives, and clarity revisions. The authors reviewed all suggestions and remain fully responsible for the accuracy, integrity, and originality of the manuscript.

We thank Fanni Takatsy for data curation and project administration. We further thank Nadezhda Moroshkina for her invaluable support.

References

Ball, L. J., Marsh, J. E., Litchfield, D., Cook, R. L., & Booth, N. (2015). When distraction helps: Evidence that concurrent articulation and irrelevant speech can facilitate insight problem solving. Thinking & Reasoning, 21(1), 76–96.

Bilalić, M., McLeod, P., & Gobet, F. (2008). Why good thoughts block better ones: The mechanism of the pernicious Einstellung (set) effect. Cognition, 108(3), 652–661.

Bilalić, M., Graf, M., Vaci, N., & Danek, A. H. (2019). When the solution is on the doorstep: Better solving performance, but diminished Aha! experience for chess experts on the mutilated checkerboard problem. Cognitive Science, 43(8), e12771.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. Cognitive Science, 13(2), 259–294.

Cronin, M. A., & Weingart, L. R. (2007). Representational gaps, information processing, and conflict in functionally diverse teams. Academy of Management Review, 32(3), 761–773.

Fleck, J. I., & Weisberg, R. W. (2004). The use of verbal protocols as data: An analysis of insight in the candle problem. Memory & Cognition, 32(6), 990–1006.

Fusaroli, R., Rączaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. New Ideas in Psychology, 32, 147–157.

Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. Topics in Cognitive Science, 1(2), 292–304.

Kaplan, C. A., & Simon, H. A. (1990). In search of insight. Cognitive Psychology, 22(3), 374–419.

Kohn, N. W., & Smith, S. M. (2011). Collaborative fixation: Effects of others' ideas on brainstorming. Applied Cognitive Psychology, 25(3), 359–371.

Luchins, A. S. (1942). Mechanization in problem solving: The effect of Einstellung. Psychological Monographs, 54(6), i.

Mills, G. J. (2014). Dialogue in joint activity: Complementarity, convergence and conventionalization. New Ideas in Psychology, 32, 158–173.

Ohlsson, S. (1984). Restructuring revisited: II. An information processing theory of restructuring and insight. Scandinavian Journal of Psychology, 25(2), 117–129.

Ohlsson, S. (1992). Information-processing explanations of insight and related phenomena. Advances in the Psychology of Thinking, 1, 1–44.

RStudio Team. (2025). RStudio: Integrated development environment for R. Posit Software, PBC.

Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. The Journal of the Learning Sciences, 4(3), 321–354.

Smith, S. M. (1995). Getting into and out of mental ruts: A theory of fixation, incubation, and insight. In R. J. Sternberg & J. E. Davidson (Eds.), The nature of insight (pp. 229–251). MIT Press.

Smith, S. M., & Blankenship, S. E. (1991). Incubation and the persistence of fixation in problem solving. The American Journal of Psychology, 104(1), 61–87.

Tylén, K., Fusaroli, R., Østergaard, S. M., Smith, P., & Arnoldi, J. (2023). The social route to abstraction: Interaction and diversity enhance performance and transfer in a rule-based categorization task. Cognitive Science, 47(9).

Voiklis, J., & Corter, J. E. (2012). Conventional wisdom: Negotiating conventions of reference enhances category learning. Cognitive Science, 36(4), 607–634.